

ScholarWorks@GSU

Novel Tools to Measure Latent Tuberculosis Infection Among Populations at Higher Risk in the United States

Authors	Vonnahme, Laura A
Citation	Vonnahme, Laura A. "Novel Tools to Measure Latent Tuberculosis Infection Among Populations at Higher Risk in the United States." Dissertation, Georgia State University, 2024. https://doi.org/10.57709/36275860
DOI	https://doi.org/10.57709/36275860
Download date	2026-05-08 12:00:25
Link to Item	https://hdl.handle.net/20.500.14694/14306

**Novel Tools to Measure Latent Tuberculosis Infection Among Populations at
Higher Risk in the United States**

by

Laura Ann Vonnahme
B.S., The University of Iowa
B.A., The University of Iowa
M.P.H., The University of Iowa

A Dissertation Submitted to the Graduate Faculty of Georgia State University in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY IN PUBLIC HEALTH (EPIDEMIOLOGY)

ATLANTA, GEORGIA

30303

Abstract

Novel Tools to Measure Latent Tuberculosis Infection Among Populations at Higher Risk in the United States

The current rate of decline in the tuberculosis (TB) disease case incidence rate will not achieve the United States (U.S.) TB elimination goals. Over 80% of reported U.S. TB cases are attributed to reactivation of latent TB infection (LTBI); to achieve TB elimination LTBI prevalence must be reduced. This doctoral dissertation utilized novel methodological approaches combined with an electronic health record (EHR) data cohort from the largest network of safety net clinics in the U.S. to measure LTBI among populations in the U.S.

In study 1 a systematic review and meta-analysis determined that mean LTBI prevalence estimate in the U.S. general population was 4.04% (95% CI: 3.35%, 4.87%) and 16.49% (95% CI: 14.70%, 18.50%) in the non-U.S.-born population. Current LTBI prevalence estimates in the U.S. were found to have methodological limitations.

Study 2 utilized the EHR data cohort to measure the LTBI care cascade. While 43.5% of the cohort met LTBI screening criteria, only 21.4% were tested, less than half with the recommended test. Among patients diagnosed with LTBI, 29.1% were prescribed LTBI treatment; only 33.6% were prescribed a recommended rifamycin-based regimen.

Study 3 utilized the EHR data cohort to create a novel algorithm to classify individuals into hierarchical definitions for LTBI, TB disease, patients with no TB infection, and patients not evaluated for TB. Using the novel algorithm definitions, the TB incidence rate was 11.8/100,000 persons in the patient population; LTBI prevalence was 14.2% among patients evaluated for TB infection.

This dissertation provides a mean estimate of LTBI prevalence for the U.S. general population and the non-U.S.-born population that can be used to determine resources needed to improve targeted testing and treatment to reduce the burden of TB infection in the U.S. However, updated LTBI prevalence estimates using robust methodology for determining infection are needed. EHR data is a novel data source that can be used to estimate LTBI prevalence in clinical networks among patients at higher risk, as well as to identify gaps in LTBI testing and treatment. Updating national LTBI prevalence estimates, especially for subpopulations at higher risk, and addressing identified gaps in testing and treatment may have a direct impact on improving TB prevention and accelerate progress towards elimination.

APPROVAL PAGE

Novel tools to measure latent tuberculosis infection among populations at higher risk in the
United States

By

Laura Ann Vonnahme

Approved: November 02, 2023

Dr. Christine Stauber, PhD, School of Public Health, Georgia State University
Committee Chair

Dr. Laura Salazar, PhD, School of Public Health, Georgia State University
Committee Member

Dr. Kathryn Winglee, PhD, National Center for HIV, Viral Hepatitis, STD and TB Prevention,
Centers for Disease Control and Prevention
Committee Member

Defense date: November 02, 2023

DEDICATION

This dissertation is dedicated to and was made possible by my greatest loves and biggest cheerleaders: Eric, Lulu, and Abe. To my husband Eric, the sacrifices you have made so I could pursue my dream will never be lost on me. You never doubted me, and I never would have done this without your encouragement and support. To my beautiful daughter, Lulu, who was born just 36 hours after I submitted my comprehensive exams, you are the reason for everything I do. And to my sweet dog Abe, I will be eternally grateful for the long walks to help clear my head, and for your faithful companionship during the long nights working on this dissertation, when you never left my side.

ACKNOWLEDGEMENTS

I want to acknowledge several of my colleagues at the Centers for Disease Control and Prevention for their expertise, guidance, and support: Tracy Ayers, Maryam Haddad, Adam Langer, Amy Krueger, Tom Navin, Julia Raykin, Noah Schwartz, Rebekah Stewart, Rachel Woodruff, and Jonathan Wortham. As well as my collaborators at OCHIN: Matthew Jones, Shane Morgan, Jonathan Puro, Jee Oakley, and Jonathan Todd. While they weren't all a part of the dissertation committee their contributions were immeasurable. I also want to acknowledge the support and encouragement I have received from many Georgia State University faculty, staff, and fellow students over the course of my education. Dr. Terri Pigott who generously offered her expertise in systemic reviews and meta-analysis for the first paper. Dr. Rich Rothenberg who acted as my advisor and provided immeasurable support and encouragement. In addition, many thanks to my fellow classmates Shannon Casillas and Meagan Davis who constantly commiserated with me and propped me up even when I wanted to throw in the towel.

Author's Statement Page

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this dissertation may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

Laura Ann Vonnahme

A handwritten signature in black ink, appearing to read "Laura Ann Vonnahme". The signature is written in a cursive style with a large initial "L" and "A".

Table of Contents

Abbreviations	8
Chapter 1: Introduction, Literature Review and Statement of Purpose.....	9
Introduction and literature review	9
Tuberculosis (TB).....	9
TB and LTBI in the United States.....	10
Estimating LTBI prevalence in the United States.....	12
The LTBI care cascade	14
Using electronic health record (EHR) data to describe LTBI in the U.S.....	18
OCHIN electronic health record data.....	23
Statement of purpose and study aims.....	24
Ethical considerations	25
Appendix A.....	27
Chapter 2: Estimating the prevalence of latent tuberculosis infection in the U.S. general population and non-U.S.–born subpopulation: A systematic review and meta-analysis	28
Introduction	28
Methods.....	29
Search strategy and selection criteria.....	29
Data Analysis.....	30
Results.....	31
Discussion.....	33
Tables and Figures	37
Appendix B.....	41
Chapter 3: Using electronic health record data to measure the latent tuberculosis infection care cascade in safety net primary care clinics	45
Introduction	45
Methods.....	46
Results.....	48
Discussion.....	49
Acknowledgments.....	51
Tables and Figures	53
Appendix C.....	57
Chapter 4: A novel algorithm to identify individuals with tuberculosis infection using electronic health record data.....	61

Introduction	61
Methods.....	63
Assessing diagnostic test results, ICD codes and treatment regimens.....	63
Developing the classification algorithm.....	65
Results.....	66
Defining the classification algorithm	66
Classifying the patient cohort according to the algorithm	68
Discussion.....	69
Tables and Figures	73
Appendix D.....	78
Appendix E	79
Chapter 5. Dissertation summary and future directions in research	82
Summary	82
Future directions in research	84
References	89

Abbreviations

AIDS	Acquired immunodeficiency syndrome
BCG	Bacillus Calmette-Guérin
CDC	Centers for Disease Control and Prevention
CRN	Clinical Research Network
CDM	Common data model
CHC	Community Health Center
DOT	Directly observed therapy
EHR	Electronic health record
FQHC	Federally qualified healthcare center
HIV	Human Immunodeficiency Virus
IRB	Institutional Review Board
IGRA	Interferon-gamma release assay
ICD	International Classification of Diseases
INH	Isoniazid
LTBI	Latent tuberculosis infection
MDR-TB	Multidrug resistant TB
<i>M.tb</i>	<i>Mycobacterium tuberculosis</i>
NHANES	National Health and Nutrition Examination Survey
NTCA	National TB Controllers' Association
PCORI	Patient-Centered Outcomes Research Institute
PCORnet	Patient-Centered Outcomes Research Network
QFT	any generation of QuantiFERON TB test
QFT-GIT	QuantiFERON-TB Gold In-Tube
T-SPOT	T-SPOT®.TB test
TST	Tuberculin skin test
TB	Tuberculosis
USPSTF	U.S. Preventive Services Task Force
XDR-TB	Extensively drug-resistant TB
3HP	3 months of once-weekly isoniazid plus rifapentine
3HR	4 months of daily rifampin
4R	3 months of daily isoniazid plus rifampin

Chapter 1: Introduction, Literature Review and Statement of Purpose

Introduction and literature review

Tuberculosis (TB)

Tuberculosis (TB) is a communicable disease caused by *Mycobacterium tuberculosis (M.tb)*. TB typically affects the lungs but can affect other sites in the body (i.e., extrapulmonary TB), such as the kidney, spine, and brain. TB is spread by individuals with TB disease of the lungs expelling bacteria into the air, for example by coughing, speaking, or singing. Symptoms of TB disease are cough, chest pain, weight loss, fever, night sweats, chills, loss of appetite, and weakness or fatigue¹. TB disease is treatable to cure and the current recommended 6-month, four drug regimen (isoniazid, rifampin, ethambutol, and pyrazinamide) for drug susceptible TB has a reported 85% treatment success rate². However, left untreated, the mortality rate from TB disease is high. A 2021 report by the World Health Organization reported that until the coronavirus disease 19 (COVID-19) pandemic, TB was the leading cause of death globally by infectious disease, even surpassing AIDS/HIV, with an estimated 1.6 million deaths².

Multidrug resistant TB (MDR-TB) is caused by strains of *M.tb* that are resistant to at least two first-line drugs. While MDR-TB is also treatable to cure with second-line drugs, therapy is costly and often twice as long as treatment regimens for drug-susceptible TB. Its estimated that fewer than 20% of individuals with MDR-TB access treatment, and fewer than half of those individuals are treated to cure². Extensively drug-resistant TB (XDR-TB) is caused by *M.tb* strains resistant to isoniazid and rifampin, a fluoroquinolone, and a second-line injectable (amikacin, capreomycin, and kanamycin) or strains that are resistant to isoniazid (INH), rifampin, a fluoroquinolone, and bedaquiline or linezolid. Individuals infected with XDR *M.tb* strains are more likely to progress to TB disease and have higher mortality rates overall³. Treatment options for XDR-TB are very limited and costly, less effective, and may cause severe side effects. Its estimated that only 30-50% of individuals affected with XDR-TB are treated to cure³.

Most individuals infected with *M.tb* have no clinical symptoms of TB disease, representing a latent TB infection (LTBI). It is estimated that over one-fourth of the global population, approximately 1.7 billion individuals, is infected with LTBI^{4,5}. However, LTBI can be treated to cure, and completing treatment for LTBI can reduce the risk of progression to TB disease by 90%⁶. Historically, LTBI treatment regimens have been limited, with a 6- or 9-month regimen of INH as the preferred regimen for treatment of LTBI⁷. However, in 2011 CDC recommended once-weekly isoniazid and rifapentine for 12 weeks by directly observed therapy (DOT) with limitations for use in children and persons living with HIV⁸; the recommendations were further expanded in 2018 to include children, persons living with HIV and for administration by DOT or self-administered therapy⁹. In 2020, the U.S. Centers for Disease

Control and Prevention (CDC) and the National TB Controllers Association (NTCA) published updated comprehensive guidelines for the treatment of LTBI, the first comprehensive update since 2000¹⁰. The guidelines preferentially recommend short-course, rifamycin-based 3- or 4-month treatment regimens over the 6- or 9-month INH monotherapy previously recommended. These short-course rifamycin-based regimens include: 3 months of once-weekly INH plus rifapentine (3HP), 4 months of daily rifampin (4R), and 3 months of daily INH plus rifampin (3HR). Clinical trials have shown that short-course LTBI treatment regimens are safe and effective and have higher completion rates than the longer course INH monotherapy regimens⁹⁻¹².

While TB reactivation rates globally are difficult to determine, it is estimated that without preventative treatment, 1 in 10 individuals with LTBI will develop TB disease^{13,14}. However, individuals with untreated LTBI and opportunistic infections, such as HIV, are estimated to have a 20 times higher risk of developing TB disease in their lifetime. Globally, TB is one of the leading causes of death among people living with HIV². Thus, the current estimate of the global LTBI burden represents a large reservoir of individuals at risk of morbidity and mortality due to TB disease. While global incidence and mortality due to TB disease has been declining since the early 1990s, without addressing the global burden of LTBI, including increasing testing for individuals at higher risk of infection or progression to disease and preventative treatment, goals put forth in the End TB Strategy 2050 will not be achieved^{15,16}.

TB and LTBI in the United States

TB disease incidence in the U.S. has decreased steadily over the past three decades since CDC resolved to work towards TB elimination with a strategy of rapidly identifying and treating cases, and evaluating exposed contacts to limit secondary cases due to recent TB transmission¹⁷. In 2019 (the end date for the cohort used in this analysis) the U.S. reported the lowest number of TB cases and the lowest incidence rate to date, with 8,916 cases and an incidence rate of 2.7 cases per 100,000 persons, achieving an overall 66.6% decrease in case count and 73.9% decrease in incidence rate¹⁸. However, the annual rate of decline in case count and incidence rate remains far below the average of the previous two decades and will not achieve the U.S. TB elimination goal of <1 case per 100,000, as the current incidence rate is 27 times the elimination threshold¹⁹.

Non-U.S.–born persons are disproportionately affected by TB infection and account for most reported TB disease cases in the U.S. In 2019, 71% of reported TB disease cases were among non-US.–born persons, with their case incidence rate 15.5 times that of US.–born persons²⁰. Among non–US.–born persons residing in the United States, TB rates were highest among Asians (25.7 per 100,000), followed by Native Hawaiians/Pacific Islanders (25.1), blacks/African Americans (19.5), Hispanics/Latinos

(10.2), and American Indians/Alaska Natives (5.3) and were lowest among whites¹⁸. The top five countries of birth among non-US.-born persons with incident TB in 2019 were Mexico (1,165 cases; 18.4% of non-U.S. -born cases), the Philippines (790; 12.5%), India (573; 9.1%), Vietnam (503; 8.0%), and China (387; 6.1%)¹⁸.

Since 2014, over 80% of reported TB cases in the U.S. have been attributed to reactivation of LTBI, rather than to secondary contact as a result of recent transmission^{21,22}. Based on the 2011–2012 National Health and Nutrition Examination Survey (NHANES), it is estimated that 4.7% of the U.S. population, or 13 million individuals, have TB infection, with the large majority being classified as having LTBI. Again, non-US.-born persons are disproportionately affected, as an estimated 16.9% of non-U.S.-born persons living in the United States have LTBI, compared to 2.8% in the U.S.-born population^{23–25}. Therefore, the majority of the burden of TB disease in the U.S. can be attributed to reactivation of LTBI among non-U.S.-born persons likely infected with *M.tb* in their countries of birth²⁶. Thus, increasing testing and treatment for LTBI among non-U.S.-born persons in the U.S. may be one of the most effective interventions in moving closer to TB elimination. Additionally, increasing testing and LTBI treatment may decrease morbidity and mortality among individuals at higher risk of progressing to TB disease, such as persons living with HIV.

National surveillance for TB disease was established in 1953 to collect information on new cases of active TB disease from reporting jurisdictions who submitted documented cases and operational data in aggregate²⁷. Since 1985, jurisdictions have been reporting case level data in a standardized format²⁷. While TB disease is nationally notifiable and reporting is mandated for all states, there is no established national surveillance for LTBI, making estimating LTBI prevalence in the U.S. difficult. This is partially due to the historical prioritization of identifying and treating TB cases, and evaluating exposed contacts to reduce the risk of progression to TB disease²⁸. However, as the annual rate of decline in TB cases and the incidence rate remain stagnant, it has been determined that the current rate of decline will not achieve U.S. TB elimination goals¹⁸. Recent statistically modeling has shown that LTBI prevalence in the U.S. must be reduced in order to achieve TB elimination^{29–31}. Accordingly, the national strategy for TB elimination is shifting to emphasizing LTBI detection and treatment³². However, in the absence of national surveillance, up to date, nationally representative data sources for estimating LTBI prevalence have been limited. As such, it has been difficult to ascertain the resources necessary to reduce the burden of LTBI among populations at higher risk in the U.S.

Estimating LTBI prevalence in the United States

As previously mentioned, 2011–2012 NHANES data estimate 4.7% [95% confidence interval: 3.4-6.3%] of the U.S. population, or 13,276,000 [9.6-17.8] million people, has TB infection, with the large majority being classified as having LTBI. NHANES is the only population-based national survey of LTBI prevalence. NHANES is a large, representative, population-based survey that can provide disease prevalence estimates for the U.S. population. The survey is a series of cross-sectional, nationally representative surveys and uses a complex, stratified, multistage probability cluster sampling design to select a nationally representative sample of the U.S. noninstitutionalized population³³. The 2011–2012 NHANES contained a TB infection testing component that included testing using tuberculin skin test (TST) and interferon gamma release assay (IGRA) tests to estimate TB infection prevalence in the U.S.

These prevalence estimates, in addition to now being over a decade old, have several limitations. First, while NHANES is meant to be nationally representative and can produce stable estimates for non-Hispanic black, non-Hispanic white, Hispanic and Asian persons, it is not designed to produce stable estimates for other race/ethnicity groups, such as American Indians, Native Hawaiians, Pacific Islanders, or Alaska Natives; race/ethnicity groups at higher risk of TB infection¹⁹. In addition, a single 2-year NHANES cycle is designed to provide stable estimates for conditions of 10% prevalence or greater, and with little geographic variability – neither apply to LTBI^{34,35}.

Second, NHANES used TB diagnostic tests to determine prevalence, and although a diagnosis of TB disease can be confirmed by the detection of *M.tb* in a patient sputum sample, there is no diagnostic gold standard for diagnosis of LTBI³⁶. Further, neither TST nor IGRA diagnostic tests can distinguish between latent infection and active disease. Additional medical evaluation, including chest radiograph, are required to distinguish LTBI from TB disease^{37,38}. The TST diagnostic test also has several limitations³⁹. It requires two clinical visits, one for administration and one for reading, and clinicians must be trained to read TST results accurately. Of particular concern, false positives can occur as a result of nontuberculous mycobacterial infection or previous Bacille Calmette-Guerin (BCG) vaccination^{39,40}. The BCG is a vaccine for TB disease and is used in many countries with high prevalence of TB disease to prevent childhood TB meningitis and miliary disease⁴¹. However, BCG is not recommended for use in the U.S. due to its variable efficacy against adult pulmonary TB and its interference with TST reactivity^{41–43}. Thus, the limitations of the TST are especially relevant to populations at higher risk of infection, such as persons experiencing homeless, who have poor return rates for TST readings, and non-U.S. – born individuals, who have high rates of BCG vaccination⁴⁴.

The IGRA blood test was introduced in 2001 and is more specific and equally as sensitive as the TST; additionally, it only requires a one-time blood draw. IGRA blood tests are also not affected by prior BCG vaccination and are less likely to give a false-positive result⁴¹. CDC guidelines recommend the IGRA test over the TST in individuals ages 2 or older, especially for individuals traditionally at higher risk for TB infection^{39,45}. If an individual reports having received effective treatment for active TB or LTBI, interpretation of positive diagnostic TB test results can be more difficult as the TST is believed to remain positive for life, and it is still undetermined whether an IGRA blood test remains positive after treatment to cure⁴⁶⁻⁴⁸. Initial 2011–2012 NHANES-based estimates of LTBI prevalence were refined to exclude persons who reported prior treatment for TB disease or LTBI, and it was subsequently estimated that 12.6 [10.5-14.7] million individuals, or 4.5% of the national population had TB infection²⁵. However, these estimates relied on individuals self-reporting prior history. Recall bias may have resulted in misreporting of prior treatment, and it cannot be assumed that all individuals did complete treatment to cure.

Additional methodologies that used mathematical models to back-calculate the prevalence of LTBI in the U.S. have also produced national, as well as state and local, LTBI prevalence estimates^{49,50}. Haddad et al. designed a straightforward method using genotyping results from active TB cases to derive an estimate of untreated LTBI nationally and for local jurisdictions⁴⁹. The mathematical model used annual TB incidence averaged during 2008–2015, an estimate of ≈ 0.084 cases of reactivation TB/100 person-years among US residents with LTBI, and a uniform population-level 0.10% annual risk for progression to active disease to estimate local and national LTBI prevalence. They estimated that 3.1% [2.2-5.2% based on higher or lower risk progression assumptions] of the U.S. population, or 8.9 million [6.3-14.8 million based on higher or lower risk progression assumptions] have TB infection. These estimates are lower than the 2011–2012 NHANES estimates, but ultimately the NHANES estimates are within the stated range of 2.2-5.2% based on higher or lower risk progression assumptions²³. This method had limitations related to relying on genotyped TB cases. In particular, in jurisdictions with longstanding genotyping clusters, the method may have overestimated recent TB infection, resulting in an underestimation of LTBI prevalence.

Mirzazadeh et al. employed a mathematical model using all incident TB disease cases during 2013–2017 not attributed to recent TB transmission, annual average number of TB cases, and reactivation rates to back-calculate estimated prevalence of untreated LTBI nationally and at the state level, both in the total population, as well as in subpopulations⁵⁰. They estimated 2.7% [2.6-2.8%] in the U.S. general population, or 8.6 million [8.3-8.8 million], had untreated LTBI in 2015. These estimates are similar to

estimates by Haddad et al. but lower than estimates from 2011–2012 NHANES data. Both Mirzazadeh et al. and Haddad et al. suggest that 2011–2012 NHANES estimates based on TST may have overestimated LTBI prevalence due to cross reaction with prior BCG vaccine especially, since prior BCG vaccine is more prevalent among non-U.S.-born. Haddad et al. and Mirzazadeh et al. report the most recent LTBI prevalence estimates for the U.S. general population; both used national TB surveillance system data that is over 5 years old, and a LTBI reactivation rate based on surveillance data from 2006-2008²¹.

Regardless of data source or methodology, all estimates of LTBI prevalence in the U.S. underline the disparity in TB infection based on country of birth^{23–25,50}. Using IGRA testing data from the 2011–2012 NHANES data, Yelk Woodruff et al. estimated that 16.9% of the noninstitutionalized, non-U.S.-born population ≥ 6 years old in the United States had LTBI²⁴; this is compared to 2.8% prevalence in the equivalent U.S.-born population as reported by Miramontes et al²³. The mathematical model employed by Mirzazadeh et al. estimated that 13.9% of the non-US.-born population had LTBI, respectively, compared to 1.0% in the U.S.-born population. However, LTBI prevalence among non-US.-born varies widely based on country of birth²⁶. In addition, TB incidence rates for a country as reported by WHO, did not consistently equate to the TB incidence in the U.S. for persons born in the corresponding country⁵¹.

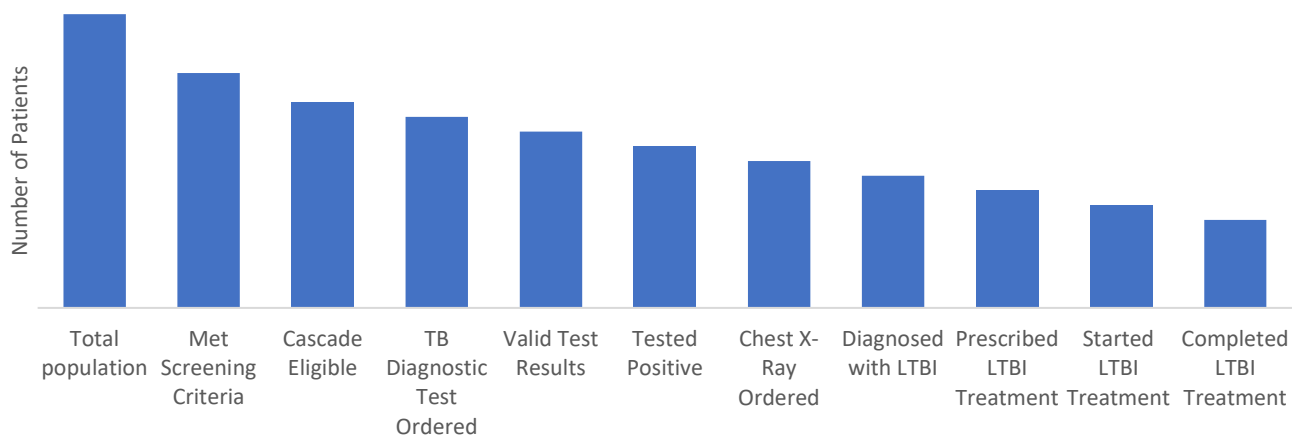
Thus, it is important to attain updated and specific LTBI prevalence estimates in the U.S., especially for specific subpopulations at higher risk. It is difficult to understand resource needs, and direct them towards populations at highest risk, without knowing an accurate, up to date estimate of the LTBI burden in the U.S. A thorough systematic review and meta-analysis assessing LTBI prevalence estimates among the U.S. general population and subpopulations, including an assessment of data sources and methodologies employed for determining prevalence, would be critical to determining resources needed to address the burden of LTBI in the U.S. In addition, alternate data sources should be explored to determine if they may provide more updated LTBI prevalence estimates and be a viable data source for producing consistent and flexible national estimates.

The LTBI care cascade

To go beyond LTBI prevalence estimates, the LTBI care cascade is a tool that can be used to assess TB infection diagnostic and LTBI treatment outcomes along the LTBI care continuum among individuals at higher risk for TB infection. Identifying and successfully managing patients with LTBI is a multistep process that links efforts in appropriate testing for patients at higher risk, notification of results, access to care to ensure diagnosis, and the offering, initiation, and completion of treatment⁵². Points along the care cascade, as well as the quantification of the cascade, vary depending on the setting, available data, and the aims of the analysis^{53,54}. However, a complete LTBI care cascade typically includes steps from

identifying people at risk for TB infection based on CDC and USPSTF recommendations, to diagnostic evaluation and diagnosis, to the offering, initiation and completion of treatment for LTBI^{53,54}. A hypothetical LTBI care cascade is described in Figure 1.1.

Figure 1.1. Hypothetical LTBI care cascade



A critical element of the LTBI care cascade is to identify individuals who meet screening criteria for TB infection. This step is typically defined based on current TB screening guidelines^{36,55}, but the at risk population can be defined to include all persons meeting any of the criteria or focus on a specific subpopulation of individuals at risk (e.g., non-US.-born) depending on the setting and available data elements^{36,55}. Current CDC recommendations stratify individuals at higher risk for TB infection into two groups: 1) individuals at higher risk for exposure to or infection with TB and 2) individuals at higher risk for developing TB disease once infected^{36,56}. Individuals in the former group are recommended for TB screening and individuals in the latter category should be screened for TB as part of regular medical care. Table 1.1 below details screening criteria for both groups as reported in Latent Tuberculosis Infection: A Guide for Primary Health Care Providers⁵⁶.

Table 1.1 Groups at High Risk for TB Infection and TB Disease

Individuals at higher risk for exposure to or infection with TB	Individuals at higher risk for developing TB disease once infected
<ul style="list-style-type: none"> • Contacts of people known or presumed to have infectious TB disease • People who were born in or who frequently travel to countries where TB disease is common, including Mexico, the Philippines, Vietnam, India, China, Haiti, and Guatemala (in general, people born in Canada, Australia, New Zealand, western European countries, or northern European countries are not 	<ul style="list-style-type: none"> • People living with HIV • Children younger than 5 years of age • People recently infected with M. tuberculosis (within the last 2 years) • People with a history of untreated or inadequately treated TB disease • People who are receiving immunosuppressive therapy such as tumor necrosis factor (TNF)-alpha antagonists, systemic corticosteroids

<p>considered at high risk for TB infection unless they spent time in a country with a high rate of TB).</p> <ul style="list-style-type: none"> • People who currently live or used to live in large group settings where TB is more common, such as homeless shelters, prisons, jails, or nursing homes • Employees of high-risk congregate settings • Health care workers who serve patients with TB disease • Populations defined locally as having an increased incidence of LTBI or TB disease, including medically underserved populations, low-income populations, or people who abuse drugs or alcohol • Infants, children, and adolescents exposed to adults who are at increased risk for LTBI or TB disease 	<p>equivalent to/greater than 15 mg of prednisone per day, or immunosuppressive drug therapy following organ transplantation</p> <ul style="list-style-type: none"> • People with silicosis; chronic renal failure; leukemia; or cancer of the head, neck, or lung • People with diabetes mellitus • People who have had a gastrectomy or jejunoileal bypass • People who have low body weight • People who use substances (such as injection drug use) • Populations defined locally as having an increased incidence of disease due to <i>M. tuberculosis</i>, including medically underserved and low-income populations
---	---

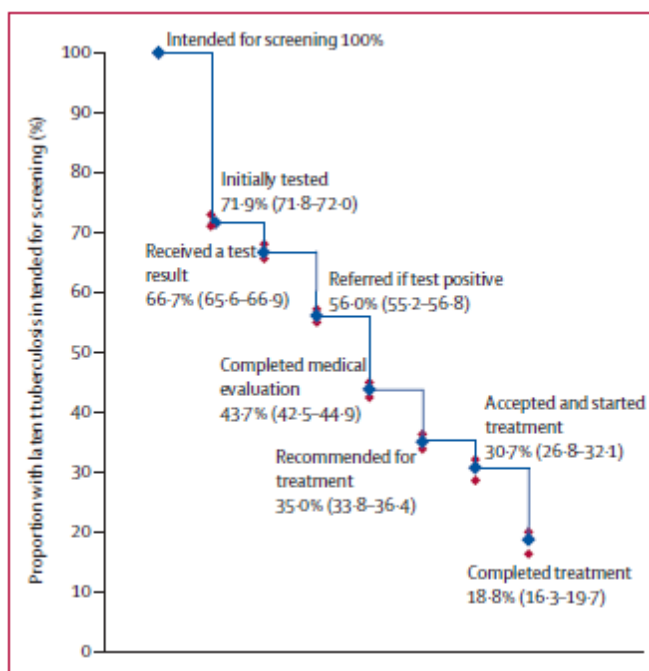
In 2016, the U.S. Preventative Services Task Force (USPSTF) issued recommendations for LTBI screening and testing⁵⁵. Based on a systematic review, synthesis of evidence, and expert review USPSTF recommended screening for 1) persons born in, or are former resident of, countries with increased TB prevalence and 2) persons who live in, or have lived in, high-risk congregate settings (e.g., homeless shelters, and correctional facilities)^{55,57}. The Task Force recognized that contacts of individuals with active TB, health care workers, and workers in high-risk congregate settings may be at increased risk of exposure, but because screening would be conducted as part of public health or employee health surveillance, they did not review evidence for these populations. In addition, they recognized other populations are at increased risk for LTBI or progression to active disease including persons who are immunosuppressed (e.g., persons living with HIV, patients receiving immunosuppressive medications such as chemotherapy or TNF–alpha inhibitors or received an organ transplant) and patients with silicosis (a lung disease)^{58,59}. However, because screening in these populations is part of disease management or indicated prior to the use of certain medications, they did not review evidence on screening in these populations.

TB screening recommendations of the complexity recommended by CDC and USPSTF can be difficult to implement in a clinical setting. Especially in a primary care setting where clinicians have limited time to address multiple screenings, and for other diseases that traditionally have higher rates of morbidity and mortality in the U.S. general population than TB infection (e.g., colon cancer, heart disease, etc.). To simplify and improve implementation of TB screening guidelines in a clinical setting, validated risk

assessment tools and recommendations focused on specific populations (e.g., non-U.S.–born) have been developed and modeling has shown them to be cost effective, as well as likely to reduce the burden of TB infection among populations at higher risk to the same extent as following USPSTF or CDC guidelines exactly^{60,61}. However, it is unknown to what extent TB screening guidelines, from simplified risk assessments to the more detailed CDC or USPSTF recommendations, are being implemented in primary care settings.

The LTBI care cascade can indicate whether TB screening recommendations or risk assessment tools related to screening populations at high risk and LTBI treatment are being implemented in specific settings. Subsequently, it can also be used to identify points for potential interventions to improve diagnostic and preventive treatment outcomes. A systematic review and meta-analysis of studies focusing on the LTBI care cascade explored losses along each step of the care cascade. The LTBI care cascade described in the analysis was modeled after the HIV care cascade and is shown in Figure 1.2.

Figure 1.2. Losses and drop-outs at each stage of the cascade of care in LTBI as reported by Alsdurf et al.⁵³



The analysis showed higher losses of patient retention in the steps of initial screening, completing medical evaluation (e.g. chest x-ray performed), and starting and completing treatment⁵³. Additional research, conducted in more specific settings looking at losses along the LTBI care cascade have shown similar gaps in care^{52,54,62,63}.

The literature assessing the LTBI care cascade is lacking in detailed methodology and a thorough description of a fully defined cascade of care. Some studies focus on the earlier steps in the cascade of care, while others focus on the latter steps related to treatment initiation and completion⁵³. Studies include or exclude points along the cascade depending on available data. Almost all studies focus on a specific population (e.g., persons experiencing homelessness, migrants), or a population that has already been identified for TB screening, limiting the ability to determine if TB screening guidelines are being implemented in a broader setting, such as in primary care clinics. While the LTBI care cascade has been described in public health departments or specialty clinics (e.g., AIDS/HIV clinics)^{53,54}, they are not primary care settings and thus not where the majority of the population at highest risk in the U.S. would be seeking routine medical care. It is important to describe the LTBI care continuum in settings where individuals at higher risk of TB infection seek care to understand and subsequently address gaps in screening, diagnostics, and treatment.

Using electronic health record (EHR) data to describe LTBI in the U.S.

The Center of Medicaid and Medicare Services defines EHR data as the digital form of patients' records that include patient demographics, medical history, allergies, test results, and treatment plan⁶⁴. While EHRs provide the opportunity to analyze a large amount of patient information across a large population, utilizing EHR data for analytic purposes presents unique challenges as the data was not originally collected for analytic purposes^{65,66}. However, EHR data has been used to support public health, and a systematic review in 2018 showed that there were more positive than negative factors in the use of EHR for analysis related to public health purposes and disease surveillance⁶⁴. In particular, public health surveillance can utilize EHRs to track occurrence of infectious diseases, as well as to make population and public health predictions⁶⁴. EHRs also have the potential to initiate collaboration and improve integration of clinical care and public health practices and recommendations⁶⁷.

EHR data are an untapped data source for determining LTBI prevalence estimates in the U.S., as well as for describing the LTBI care cascade including screening, diagnosis and treatment initiation and completion among individuals recommended for TB infection screening. EHR data can also be used to determine if and to what extent TB infection diagnostic testing and LTBI treatment recommendations are being implemented in clinical practice, especially in clinical networks that serve a high proportion of patients at higher risk for TB infection. EHR data is particularly appropriate for LTBI prevalence estimates for several reasons. First, with an estimated 13 million individuals in the U.S. with LTBI, extracting EHR data for LTBI prevalence estimates, rather than implementing nationwide surveillance, would relieve a large reporting burden on local and state health departments. Additionally, diagnosing LTBI is complex

and rarely based only on a positive screening test. Rather, LTBI is often diagnosed based on a combination of test results, the absence of symptoms, and patient history considering exposure risk factors. EHR data contains many variables considered when diagnosing LTBI and all can be incorporated into a robust and standardized definition of LTBI. This definition would be flexible and could be altered if new science or screening diagnostics are introduced. In addition, EHR data can capture longitudinal data, and subsequently determine diagnostic testing, LTBI prevalence and treatment rates over time, or after the implementation of an intervention or new screening diagnostic. EHR data may also be able to bypass some of the previously discussed limitations associated with using NHANES, or a standardized LTBI reactivation rates (e.g., mathematical models).

Initial complexities of estimating LTBI prevalence using EHR data are related to the iterative process of defining LTBI diagnosis based on multiple EHR data elements available, including, but not limited to a patient's TB diagnostic results, ICD codes and treatment regimens prescribed. Additionally, free text fields in EHR data are generally inaccessible due to concerns related to patient identifying information, though they may contain information relevant to LTBI diagnostic and treatment outcomes (i.e., chest x-ray results). Jenks et al. has previously determined the proportion of patients with LTBI using EHR data in a single center setting by analyzing prescription drug patterns for isoniazid, rifampin and rifapentine among patients determined to be at higher risk due to country of birth. Results were validated with a chart review of all records⁶⁸. Among all those prescribed isoniazid or rifampin (none were prescribed rifapentine), 73% were confirmed to having been treated for LBTI and 15% for pulmonary TB disease. While this methodology highlights the sensitivity of utilizing prescription drug records to determine LTBI diagnosis, it may underestimate LTBI prevalence, since not all individuals diagnosed with LTBI are recommended for or agree to treatment^{53,69,70}. An additional study of administrative claims based EHR data also noted the limitations of relying on prescription drug records only for determining LTBI diagnosis⁷¹. Specifically, caution must be exercised when assessing rifampin prescriptions as a proxy for LTBI diagnosis since rifampin is used to treat health conditions other than TB disease or LTBI⁷².

There is no prior research that incorporates multiple EHR elements (e.g., prescription drug records, TB diagnostic test results and ICD codes) into a LTBI diagnosis definition in an analysis of EHR data. A systematic review summarized validated methods for identifying active TB cases⁷³ in health administrative data using solely or in combination: prescription drug records, ICD diagnostic codes or laboratory data⁷⁴. Only 14 studies met inclusion criteria, and the positive predictive value (PPV) of each study's validated algorithms varied widely. However, the review found that the lowest PPVs were found

in algorithms relying solely on ICD codes to identify TB in EHR records. Several studies included also noted the miscoding of LTBI as active TB, especially since (at the time of the study in 2016) there was no ICD code to indicate LTBI diagnosis. Algorithms that used TB drug prescriptions in combination with ICD codes or diagnostic results^{67,73}, provided improved diagnostic accuracy. Thus, a specific LTBI definition that incorporates additional EHR data elements, including ICD codes or prescription drug records, may allow for better identification of LTBI diagnosis within EHR data, as well as improved discrimination between TB disease cases versus latent infection cases.

Incorporating multiple EHR data elements into a definition of LTBI is critical as there are limitations with defining LTBI diagnosis based solely on diagnostic test result or ICD code, or treatment regimen prescribed. As previously described, TB infection diagnostic tests cannot distinguish between latent infection and disease. In addition, a prior history of TB or BCG vaccine complicates the results. Furthermore, there is no prior research describing how LTBI ICD codes are utilized in a clinical setting to indicate diagnosis since the current ICD-10 code for LTBI diagnosis (Z22.7) was implemented in the 2020 release of the ICD-10 clinical modification on October 1, 2019⁷⁵. Prior to the introduction of this code, there were only codes to indicate a positive diagnostic test result in the absence of disease (ICD-9: 795.51, 795.52; ICD-10: R76.11, R76.12). For example, R76.11 is a code indicating “a nonspecific reaction to cell mediated immunity measurement of gamma interferon antigen response without active tuberculosis”. Anecdotally, clinicians have indicated that these codes were used to initiate different clinic specific protocols or further diagnostic testing and may not be reliable as a definite LTBI diagnosis. Finally, as previously mentioned, defining LTBI diagnosis only on the presence of a treatment regimen prescription would underestimate prevalence since not all individuals diagnosed with LTBI are recommended for or initiate treatment^{53,69,70}.

In addition, a vetted algorithm could be applied to estimate LTBI prevalence in any EHR cohort that utilizes data elements common in all EHR systems (e.g., lab results, ICD codes, prescription drug records). Ideally, a robust methodology would allow for the definition to be applied broadly, without requiring significant work upfront and without having to perform a resource-intensive chart review such as the one implemented by Jenks, et al. While across all EHRs, data related to lab tests and results are limited since lab coding is not standardized across different labs used by clinics, in the context of LTBI diagnostic testing, TST and IGRA tests and results have proven to be extractable^{68,76}.

Since EHR data is generally unstandardized across different healthcare systems it would be difficult to generate an overall estimate of LTBI prevalence in the U.S. However, EHR data can be used to estimate the burden of LTBI among a higher risk population by directly capturing data from

organizations where these individuals seek care, such as safety net clinics. Safety net clinic is a broad term to define clinics that provide health care to patient who experience barriers to accessing care due to lack of coverage, geographic isolation, language and culture, mental illness and homelessness⁷⁷. While safety net clinics include Indian/tribal health clinics, rural health clinics and others, most safety clinics are community health clinics (CHC). Many CHCs are designated as federally qualified health centers (FQHC) or FQHC look-alikes and receive grants and/or benefits under the Health Center Program as authorized under section 330 of the Public Health Service Act⁷⁸. The National Association of Community Health Centers defines CHCs as nonprofit, patient-governed organizations that provide high-quality, comprehensive primary health care to America's medically underserved communities, serving all patients regardless of income or insurance status⁷⁹. In 2021, CHCs served over 30 million patients across all 50 states, with 1 in 11 Americans seeking care at a CHC. The population seeking care at CHCs are populations likely at higher risk for TB infection. CHCs are the primary care home of historically underserved communities in the U.S.; 20% are uninsured, 60% are publicly insured, 90% are low-income and 65% are members of racial and/or ethnic minority groups^{78,79}. In the U.S., TB rates are highest among racial or ethnic minorities with rates being highest among Asians, Native Hawaiians/Pacific Islanders and blacks/African Americans, and Hispanics/Latinos; TB rates are lowest among those of white race^{18,20}. In addition, evidence suggests that non-U.S.-born individuals are often uninsured⁸⁰⁻⁸³, and as such, likely seek care at CHCs or safety net clinics. Non-U.S.-born, regardless of their immigration status, and undocumented migrants can also access no cost or low-cost health care services at CHCs⁷⁹. CHCs also served over 1.3 million persons experiencing homelessness and almost 7 million of their patients are best served in a language other than English⁷⁸; both populations at higher risk for TB infection in the U.S.^{18,51,84-86}.

While LTBI estimates generated using an EHR data cohort from a network of safety network clinics or CHCs may not be representative of the general U.S. population, they may be representative of a subpopulation at higher risk for TB infection. As such, successful public health interventions to increase LTBI screening and treatment implemented in these clinical networks would likely be impactful, and their effect could be analyzed. The ability to generate overall and subpopulation estimates of LTBI prevalence at sufficiently short intervals is useful for planning and evaluating public health interventions. Additionally, secondary analysis of EHR data already collected by health networks is a feasible and immediate option. Mechanisms for accessing EHR data for secondary analysis are in the early stages of development and, while complex and costly upfront, may prove to be a consistent and affordable resource in the future.

EHR data also allows for the description of an expansive set of demographic and clinical characteristics of LTBI patients. Key demographic variables, such as age, race, ethnicity, and gender, are readily available in all EHRs. A limitation of EHR data across most EHR systems is that country of birth is frequently incomplete. This is especially limiting in the context of TB infection since non-U.S.–born persons are recommended for screening based on CDC and USPSTF screening recommendations^{36,37,55}. However, previous research has shown that preferred language is an appropriate proxy for non-U.S.–born persons who meet screening criteria for LTBI^{68,87,88}.

Additionally, EHR data contains a broad set of variables that can identify other risk factors for acquiring TB infection or progressing to TB disease based on U.S. screening recommendations^{36,37,55}. Depending on the EHR network or data repository, EHR data collects many demographic, social and clinical variables to identify individuals with risk factors for TB or progressing to TB disease. An important risk factor for TB infection is close contact with a person with infectious TB disease. EHR data can identify these individuals by an ICD diagnosis code that indicates the patient had contact with a suspected exposure to TB disease. This diagnostic code is available in all EHRs; however, it is unclear if this ICD code is entered consistently by clinicians. Many medical related risk factors for TB infection or progressing to TB disease can also be identified in EHR data. Data pertaining to medical risk factors are robust and can be determined by ICD diagnostic codes, diagnostic test results or prescribed treatment. With these multiple variables, a medical risk factor for TB can be validated; for example, an HIV diagnosis defined by ICD diagnostic code can be confirmed by the presence of a relevant prescription or a positive diagnostic test. Additionally, EHR data allows for increased validation when identifying medical risks. For example, individuals at risk for progressing to TB disease due to treatment with an TNF-alpha inhibitor can be identified by a relevant prescription in the medical record, rather than just relying on ICD diagnostic codes for conditions where a TNF-alpha inhibitor is likely to be prescribed.

EHR data is limited in being able to determine an individuals' risk based on living in a congregate setting such as homelessness or prior incarceration, both of which are risk factors for TB infection, and are not frequently collected in an EHR. In addition, EHRs do not typically contain data elements related to patient occupation and as such, EHR data cannot identify an occupational risk related to being a healthcare worker, or working in a high-risk setting, such as a homeless shelter or correctional facility.

Finally, EHR databases may have data going back many years or even several decades. Thus, EHR data can report trends over time on diagnostic testing and treatment regimens. Data can be used to report trends on diagnostic testing occurring in clinical practice, the type of diagnostic tests used (TST vs. IGRA), and the use of LTBI treatment regimens (short course vs. traditional regimens) over the last

decades. EHR data can also determine if and how well screening and treatment recommendations are being implemented across clinical networks where individuals at increased risk for TB infection seek care. Additionally, with longitudinal data, the estimated effect of any future interventions to increase screening or encourage short-term treatment regimens can be determined. EHR data is a viable and novel option for exploring LTBI prevalence estimates in the U.S., and more specifically, diagnostic screening and treatment initiation and completion outcomes among populations at higher risk for TB infection.

OCHIN electronic health record data

OCHIN is a nonprofit health information network serving the largest network of community-based and safety net providers in the U.S., with approximately 500 organizations, 21,000 providers and over 6 million patients across 30 states⁸⁹. They provide clinical support services, research and analytics, and technological innovations that include deploying and hosting a full suite of Epic and NextGen EHRs and practice management solutions for their members^{90,91}. The mission of OCHIN is to reach underserved communities receiving care at safety net clinics, including but not limited to: CHCs, FQHCs, FQHC look-alikes, public health departments, rural health clinics, correctional facilities, and HIV/AIDS care organizations. They are the largest network of safety clinics in the U.S., with the majority of the clinics being CHCs. As a nationwide network their mission is to remove barriers to health care delivery by addressing health disparities that are systemic, avoidable, and unjust. The OCHIN patient population is one that is traditionally underserved. Among their over 6 million active patients, 1 out of 2 patients are on Medicaid and one fourth are uninsured. Nearly 1 out of 3 prefer care in a language other than English, and 2 out of 5 patients identify as people of color; nearly one third are of Hispanic race⁸⁹⁻⁹¹.

OCHIN maintains the ADVANCE Clinical Research Network (CRN), the largest safety net clinic research database in the U.S. in partnership with Fenway Health, Health Choice Network and Oregon Health and Science University; the research database contains EHR data from the former two organizations and the OCHIN EHR research database, with the OCHIN EHR research database contributing the large majority of patient data. A common data model (CDM) is used to collate EHR data from all clinics into this research-ready database that contains aggregated, longitudinal data going back to 2004⁹². The CDM ensures that EHR data elements across multiple clinics and health networks are standardized to allow for data and information exchange between different health care and EHR applications. ADVANCE, funded by the Patient-Centered Outcomes Research Institute (PCORI), contains over 8 million patients records from 182 health systems and 1,688 clinic sites and is growing rapidly^{89,93}. ADVANCE CRN is one of 12 CRNs in the PCORI National Patient-Centered Clinical Research Network

(PCORnet). PCORnet was created in 2013 to provide a resource for conducting patient-centered clinical research to meet the needs of the healthcare community. To improve the nation's capacity to conduct comparative effectiveness research, PCORnet created a large, highly representative electronic data infrastructure for conducting clinical outcomes research. All data from participating CRNs in PCORnet is translated into the CDM, ensuring research can be replicated effectively across all network levels. Therefore, analyses completed on an OCHIN EHR dataset using variables from the CDM can be applied to EHR datasets from the larger PCORnet. Detailed information and a data dictionary pertaining to the ADVANCE implementation of the PCORnet CDM can be found at their website⁹³.

The OCHIN EHR research database and the ADVANCE CRN contains EHR data on a population with a higher proportion of patients at higher risk of TB infection than the general population. As such, the OCHIN EHR research database and the ADVANCE CRN is a relevant cohort of patients for estimating LTBI prevalence using EHR data elements, as well as for generating an LTBI care cascade to determine gaps in TB infection diagnostic testing and LTBI treatment outcomes.

[Statement of purpose and study aims](#)

This three-manuscript doctoral dissertation proposes to utilize a novel data source and other methodologies to measure LTBI among populations at higher risk in the U.S. This will be done by conducting a systematic review and meta-analysis to estimate LTBI prevalence in the U.S. and by using EHR data elements to define LTBI diagnosis, as well as to describe and identify gaps in the LTBI care cascade among a population with a large proportion of patients at higher risk for TB in the U.S.; specifically, those seeking care at a large network of safety net clinics. Two of the manuscripts will utilize a data cohort from the OCHIN EHR research database to achieve research aims. A data cohort from the OCHIN EHR research database was preferred over one from the ADVANCE CRN because the OCHIN EHR research database contained an expanded set of variables in the CDM relevant to defining LTBI diagnosis and the LTBI care cascade.

The first study will perform a systematic review and meta-analysis to determine an overall mean LTBI prevalence estimate for the U.S. general population and the non-U.S.-born subpopulation. The systematic review will identify data sources and methodologies used to calculate the available U.S. LTBI prevalence estimates in the literature. The strengths and limitations of the data sources and methodologies will be described and will be used to determine if updated LTBI prevalence estimates for the U.S. are needed. In addition, the meta-analysis will assess heterogeneity among reported prevalence estimates. This systematic review and meta-analysis will address specific gaps in our knowledge of LTBI prevalence in the U.S. by finding all estimates, examining heterogeneity among estimates, data sources,

and methodologies, and by concluding whether updated estimates, based on updated population denominators or new data sources, should be pursued.

The second and third studies will utilize a data cohort from the OCHIN EHR research database to describe LTBI among the population seeking care at clinics within the OCHIN safety net clinical network. The second study will measure the LTBI care cascade among this population from identifying people at risk for TB infection, to diagnostic evaluation and diagnosis, to initiation of treatment for LTBI. This analysis will create a novel and thorough methodology for describing the LTBI care cascade using EHR data elements. In addition, this analysis will identify gaps in testing for infection, diagnosis of LTBI or LTBI treatment in the LTBI care cascade among a population at higher risk. Finally, it will be determined if recommendations related to TB diagnostic testing and LTBI treatment are being implemented in this network of safety net clinics that provides care to a large proportion of patients at higher risk for TB infection.

The third paper will use OCHIN EHR data to create a novel algorithm for identifying LTBI among patients using a combination of EHR data elements. This algorithm will utilize diagnostic laboratory test results, diagnostic ICD codes and prescription drugs records to determine patients with a LTBI diagnosis. This definition of LTBI will ultimately provide an alternative methodology for estimating LTBI prevalence within a clinical network using EHR data elements. This algorithm will provide a detailed methodology that can be applied to any EHR cohort that utilizes data elements common in all EHR systems (e.g., lab results, ICD codes, prescription drug records) to estimate LTBI prevalence. The algorithm will then be utilized to determine TB incidence and LTBI prevalence in the OCHIN network of safety net clinics.

[Ethical considerations](#)

Both studies 2 and 3 using the OCHIN EHR data cohort were reviewed by and conducted under the authority of the Centers for Disease Control and Prevention and were determined to not be human subjects research as the primary intent of the studies is routine disease surveillance and data will be used only for disease control programs or policy purposes. The studies were conducted consistent with applicable federal law and CDC policy (see e.g., 45 C.F.R. part 46, 21 C.F.R. part 56; 42 U.S.C. §241(d); 5 U.S.C. §552a; 44 U.S.C. §3501 et seq). All studies in this dissertation received review from the Georgia State University Institutional Review Board and were determined to be non-human subjects' research.

As previously described, studies 2 and 3 will utilize a data cohort from the OCHIN EHR research database. This dataset includes all patients seen in at least one OCHIN clinic on at least one occasion from 2012 to 2019; data will not include any personally identifying information. The limited dataset contains specific categories of variables and/or variables that have been derived from these variables

(See Appendix A). OCHIN disclosed this limited data set to CDC for the purposes of secondary analysis after a signed Data Use Agreement was in place. The agreement commenced on August 19th, 2020, and shall continue in full force and effect for an initial period of one (1) year (the "Initial Term") and shall automatically renew for consecutive one (1) year periods (each a "Subsequent Term") until August 31st, 2024, provided however that either party may terminate this Agreement by providing the other with not less than ninety (90) days' notice prior to the commencement of any Subsequent Term.

The OCHIN dataset became a Federal record once received by the CDC, and is subject to Federal laws and rules governing data release and Federal records retention laws. Accordingly, one copy of the dataset shall be retained by CDC for archival purposes pursuant to the Federal Records Act, as amended and codified in Title 44 of the United States Code.

Appendix A

Data elements contained in the OCHIN limited dataset

Category	Examples include, but are not limited to:
Clinic-level identifiers	Masked site ID number, facility location (zip code)
Demographics	Gender, date of birth, ethnicity, race, socioeconomic status (federal poverty level), country of origin
Diagnoses	ICD-9 diagnosis codes, ICD-10 diagnosis codes, diagnosis name, diagnosis date
Encounters	Encounter ID, encounter type, encounter date, payor, payor type
Enrollment	When patient was enrolled in health plan or was seen at a clinic
Immunizations	Immunization name, code, dates
Lab orders and Results	Lab test code, lab test name, lab test result, lab order and result dates
Medication orders and Dispensings	Medication ID, medication name, number of months prescribed, number of refills prescribed, fill status, order and dispensing dates
Patient identifier	Masked patient ID number
Patient information	Body weight, last Body Mass Index (BMI), homeless status, tobacco use, prison status, primary language, country of birth, country, migrant/seasonal worker status
Procedures	Procedure code, procedure name, procedure type, procedure dates
Vitals	Blood pressure, weight, tobacco use, dates

Chapter 2: Estimating the prevalence of latent tuberculosis infection in the U.S. general population and non-U.S.–born subpopulation: A systematic review and meta-analysis

Introduction

As the annual rate of decline in TB disease case count and incidence rate remain stagnant, it has been determined that the current rate of decline will not achieve the U.S. TB elimination goals¹⁸. Because over 80% of reported TB cases in the U.S. are attributed to reactivation of LTBI^{21,22}, recent statistical modeling has shown that LTBI prevalence in the United States must be reduced in order to achieve TB elimination^{29–31}. Accordingly, the national strategy for TB elimination is shifting to emphasizing LTBI detection and treatment³².

Targeted testing for TB infection and treatment among populations at higher risk for TB infection and progression to disease are important in reducing the burden of TB in the United States.³² Non-U.S.–born persons are disproportionately affected by TB infection. In 2019, 71% of reported TB disease cases were among non-U.S.–born persons, with their case incidence rate 15.5 times that of U.S.–born persons²⁰. Additionally, one estimate using data from NHANES 2011–2012 estimated 16.9% of non-U.S.–born persons living in the United States have LTBI, compared to 2.8% in the U.S.–born population^{23–25}. Therefore, the majority of the burden of TB disease in the U.S. can be attributed to reactivation of LTBI among non-U.S.–born persons likely infected with *M. tb* in their countries of birth²⁶. Thus, increasing testing and treatment for LTBI among non-U.S.–born persons in the U.S. may be one of the most effective interventions in moving closer to TB elimination for the U.S.

While TB disease is nationally notifiable and reporting is mandated for all states, there is no established national surveillance for LTBI, making estimating LTBI prevalence for the general population and populations at higher risk difficult in the U.S. Widely used estimates for LTBI prevalence in the U.S. come from the NHANES 2011–2012 cycle^{24,25,94}. NHANES uses TB infection testing results in a nationally representative sample of the U.S. noninstitutionalized population³³ to estimate TB infection prevalence in the U.S.; estimates were calculated for the U.S. general population, as well as for the non-U.S.–born population. Other methods have used national TB surveillance data to “back calculate” prevalence based on the assumption that LTBI cases have a uniform risk of progressing to TB disease across geographic, demographic and clinical factors^{49,50}. In addition, novel data sources using specific study cohorts have been explored to estimate LTBI prevalence including administrative claims data⁷¹ and electronic health data^{68,76}.

However, U.S. LTBI prevalence estimates, both in the general population and in populations at higher risk of infection such as the non-U.S.–born population, have not been systematically reviewed,

nor have the data sources and methodologies for determining LTBI infection been compiled and compared. In the absence of national surveillance, up to date, nationally representative data sources for estimating LTBI prevalence have been limited. Estimates of LTBI prevalence in the U.S., especially among the non-U.S.–born who are at higher risk for infection, are critical to knowing the breadth and estimating the cost of expanding targeted testing and treatment.

We aim to systematically review the published data on LTBI prevalence estimates in the United States in the general population and the non-U.S.–born subpopulation. This systematic review and meta-analysis will address specific gaps in our knowledge of LTBI prevalence in the U.S. by finding all estimates, examining heterogeneity among estimates, data sources, and methodologies, and by concluding whether updated estimates, based on updated population denominators or new data sources, should be pursued.

Methods

Search strategy and selection criteria

We conducted a systematic review and meta-analysis of observational data that estimate LTBI prevalence in the U.S. general population and the non-U.S.–born population at higher risk of TB infection. We conducted electronic database searches of MEDLINE, PubMed Central, NCBI Bookshelf via PubMed, and Embase via Ovid. The search was limited to English language studies published between January 1, 2000, and December 31, 2022. The following search strings were used to search in PubMed (MeSH = medical subject headings): (prevalence[MeSH Terms]) AND ((tuberculosis, latent[MeSH Terms]) OR (infection, latent tuberculosis[MeSH Terms])) OR (latent tuberculosis infection[MeSH Terms]) OR (latent tuberculosis[MeSH Terms]) AND (united states[MeSH Terms]); and Ovid: prevalence OR estimates AND latent TB OR latent tuberc* OR tuberc* infection OR inactive tuberc* AND united states. To search grey literature, GreyNet International and Grey Literature Report were searched using the phrase: tuberculosis infection in the United States. Finally, ProQuest Dissertations and Theses were searched to identify relevant dissertations and theses using the following string: diskw(Latent Tuberculosis) AND (prevalence) AND (United States).

The PRISMA 2020 guidelines were used to report on this systematic review and meta-analysis⁹⁵. All relevant articles were screened in two stages using a title and abstract screening and a full-text screening tool (Appendix B). Studies selected for full-text review based on the title and abstract were required to be observational studies reporting on LTBI or TB infection prevalence, with the study population being either the U.S. general population or the non-U.S.–born population. We excluded all randomized controlled trials, case-control studies, reviews, editorials, or letters to the editor; it is

important to note there were no randomized controlled trials or case-control studies reporting LTBI prevalence estimates for the U.S. general population or non-U.S.–born population. Articles selected for full-text review were screened using more specific eligibility criteria including requiring that the article estimate LTBI or TB infection prevalence in either the U.S. general population or non-U.S.–born population as a proportion (%) or population estimate (N). In addition, articles had to directly report the population denominator used, or describe the data source used, to estimate prevalence. If an article reported an LTBI point prevalence (%) but did not directly report the population prevalence estimate (N) or the population denominator (e.g., the U.S. general population) used to calculate the point prevalence, the authors were contacted to ascertain those numbers directly, because both values were required for the meta-analysis. If the data source for the population denominator was cited (e.g., 2010 U.S. census data), but the number not reported and unable to be obtained from the author, the population denominator data source was referenced directly to determine the population denominator. If articles did not meet all the eligibility criteria, or the LTBI population estimate(s) or the overall population denominator(s) was unable to be determined, the article was excluded from the systematic review and subsequent meta-analysis.

Data Analysis

Data were abstracted using a standardized data abstraction tool (Appendix B). Report data were collected including authors, year published, literature type (published, unpublished, peer-reviewed, etc.), and whether the article reported a prevalence estimate for the U.S. general population and/or the non-U.S.–born population. Data were also extracted on the data sources used in the methodology for determining LTBI prevalence, related to both determining the number with LTBI infection (numerator) and the overall population (denominator). Data were extracted on the data source(s) time frame, population restrictions, sampling technique used (if relevant), and sample size (if relevant). Additionally, specific codes were collected related to the denominator to indicate if the denominator was unreported in the study, and subsequently whether an estimated denominator based on census data was used in the meta-analysis.

Data on all reported LTBI prevalence estimates for the U.S. general population and the non-U.S.–born population were extracted from each study. Multiple estimates for each population could be extracted from a single study. In addition, data on the population point prevalence (%), the population estimate (N) and the denominator (U.S. general population and the non-U.S.–born population) were extracted. Data were extracted on how LTBI was defined for estimating prevalence, and whether diagnostic tools or other data were used to determine LTBI prevalence. Finally, data were extracted on

the risk of bias related to representation of the U.S. general population, and the methodology for defining LTBI.

A distribution of the LTBI prevalence estimate in the U.S. general population and in the non-U.S.–born population were estimated independently using a random effects model. The random effects model was chosen over a fixed effects model because the studies included in the meta-analysis were not replications, but rather had known variations in data sources and methodologies for defining LTBI; the random effect model supports the broadest possible inferences. The mean prevalence estimates were transformed using the log method, but final estimates were transformed back to proportions for ease of interpretation and final recommendations. The metafor package⁹⁶ in R v4.1.2 was used to conduct the random-effects analysis and tau-squared was estimated using restricted maximum likelihood methods (REML). Prediction intervals were calculated to indicate the potential range of mean prevalence estimates. The heterogeneity test of variance, Q , was used to determine if tau-squared, the estimate of the variance of the true effect size (i.e., LTBI prevalence) distribution, was significantly different from zero to determine whether to reject the null hypothesis that effect sizes were homogeneous.

In addition to the meta-analysis, the representativeness of the sample of studies gathered and the way they defined LTBI were reported and summarized.

Results

There were 2319 articles identified in the systematic review after de-duplication. After screening for eligibility based on title and abstract, 7 articles were identified for full-text review; 6 were included in the qualitative synthesis and meta-analysis (Figure 2.1). An overview and summary of the articles included in the systematic review and meta-analysis is in Table 2.1. Of the 6 articles included, 5 of the articles reported 7 separate LTBI prevalence estimates for the U.S. general population, and 5 articles reported 7 separate estimates for the non-U.S. population. One article reported an estimate only for the U.S. general population (Haddad et al., 2018) and 1 article reported an estimate only for the non-U.S.–born population (Yelk Woodruff et al., 2021); the 4 other articles reported a prevalence estimate for both the U.S. general population and the non-U.S.–born population. Two articles (Miramontes, et al, 2015 and Mancuso et al, 2016) each reported two estimates for the U.S. general population and two estimates for the non-U.S.–born population.

For the U.S. general population, articles reported a range of LTBI estimates from 2.7% (2.6%–2.8%) to 5.0% (4.2%–5.8%), estimating that between 8.5 million (8.3–8.8 million) to 14.1 million (11.9–16.4 million) persons in the U.S. have TB infection. In the non-U.S.–born population, LTBI prevalence estimates from included articles reported a range from 13.8% (no confidence intervals reported) to

20.5% (16.1%–25.8%), estimating between 5.5 million (4.6–6.3 million) to 8.1 million (6.4–10.2 million) non-U.S.–born persons have TB infection.

Five of the seven (71.4%) reported U.S. general population estimates and six of seven (85.7%) non-U.S.–born estimates used 2011–2012 NHANES data to determine LTBI infection based on either a positive tuberculin skin test (TST) or positive QuantiFERON-TB Gold In-Tube (QFT-GIT). Miramontes et al. and Mancuso et al. reported U.S. general population and non-U.S.–born population estimates based on a positive TST defined as an induration of ≥ 10 mm; each study indicates that they adjusted the weights to account for missing TST readings or nonparticipation, though the methodologies used differ between papers. Both articles also reported LTBI prevalence estimates for both populations based on a positive QFT-GIT result. Vonnahme et al. used 2011–2012 NHANES to report LTBI prevalence estimates for both the U.S. general population and non-U.S.–born population defining LTBI as those with a positive QFT-GIT and no reported prior TB treatment based on the 2011–2012 NHANES questionnaire data. Finally, Yelk Woodruff et al. reported LTBI prevalence estimates for the non-U.S.–born population only using 2011–2012 NHANES data, also defining infection as a positive QFT-GIT, but adjusting for QFT-GIT specificity and sensitivity.

The two remaining articles included in the qualitative synthesis and meta-analysis both used a similar back calculation method to report LTBI prevalence estimate for the U.S. general population and non-U.S.–born population. Haddad et al. used a mathematical formula that uses the annual TB incidence averaged during 2008–2015, the proportion of TB cases attributed to recent transmission, and a uniform population-level 0.1% annual risk for progression to active disease (based on Shea et al.²¹) to derive a U.S. general population LTBI prevalence estimate. The calculation assumes all genotyped TB cases not attributed to recent transmission arose from LTBI and that the same proportion of non-genotyped cases are attributed to recent transmission as genotyped cases. Mirzazadeh et al. used the same mathematical formula and assumptions, but averaged annual TB incidence based on 2013–2017 data, and applied previously derived population-specific annual risk rates for progression to active TB disease from the published literature, including those from Shea et al.^{21,50,97}. These populations specific rates were for the U.S.-born and non-U.S.–born populations iteratively classified into three groupings based on five medical risk factors, five age groups and four race/ethnic categories; specific rates and additional methodology are described in the appendix of the Mirzazadeh, et al. article⁵⁰., Haddad et al. more simply rounded up the Shea et al. overall population TB re-activation rate estimate of 0.084 to 0.1% to use as part of their back-calculation methodology.

There was a total of 7 U.S. general population prevalence estimates included in the meta-analysis using a random effects model. The mean prevalence estimate of LTBI in the U.S. general population was 4.04% (95% CI: 3.35%, 4.87%); the mean effect size was significantly different from zero ($p < 0.0001$). The forest plot shows the mean LTBI prevalence estimates across all studies, the methodology employed by each paper to determine prevalence, and the mean prevalence estimate for the U.S. general population from the random effects model (Figure 2.2); the 95% confidence intervals for the prevalence estimates across studies are very small and not visible on the plot. We reject the null hypothesis of effect size homogeneity across studies as determined by the Q test indicating that the tau-squared value (0.0633, SE=0.0365) is significantly different from zero ($Q[df = 6] = 3869068.0262$, $p < 0.0001$). Therefore, the observed amount of variance across the studies is larger than we would expect for sampling variance, and we reject the null hypothesis of effect size homogeneity across studies. The mean prevalence prediction interval is 2.39%–6.84%.

Seven non-U.S.–born population prevalence estimates were included in the meta-analysis using a separate random effects model to estimate the mean effect size. The mean LTBI prevalence estimate in the non-U.S.–born population was 16.49% (95% CI: 14.70%, 18.50%); the mean effect size was significantly different from zero ($p < 0.0001$). The forest plot in Figure 2.3 shows the mean LTBI prevalence estimates among the non-U.S.–born population, the methodology employed by each author to determine prevalence, and the mean prevalence estimate for the non-U.S.–born from the random effects model. Again, the 95% confidence intervals for the prevalence estimates across studies were very small and are not visible on the plot. Similar to the U.S. general population prevalence estimate random effects model, we reject the null hypothesis of effect size homogeneity across studies as determined by the Q test indicating that the tau-squared value (0.0241, SE=0.0139) is significantly different from zero ($Q[df = 6] = 1239344.2025$, $p < 0.0001$). Therefore, the observed amount of variance across the studies is larger than we would expect for sampling variance, and we reject the null hypothesis of effect size homogeneity across studies. The mean prevalence prediction interval is 11.91%–22.83%.

Discussion

This is the first study to systematically review the published data on LTBI prevalence estimates in the U.S. general population and the non-U.S.–born population. This systematic review and meta-analysis address specific gaps in our knowledge of LTBI prevalence in the U.S. by finding all estimates, examining heterogeneity among estimates, data sources, and methodologies, and by concluding whether updated estimates, based on updated population denominators or new data sources, should be pursued. These

estimates can be used to determine the breadth and resources that need to be dedicated to targeted testing, as well as LTBI treatment completion to reduce the burden of TB infection in the U.S.

This study found the overall mean LTBI prevalence estimate in the U.S. general population to be 4.04% (95% CI: 3.35%, 4.87%) and 16.49% (95% CI: 14.70%, 18.50%) in the non-U.S.–born population based on data published from 2000 to 2022. There was significant heterogeneity among effect sizes in both estimates, as seen in the tests for heterogeneity, as well as the overall mean estimate prediction intervals, which was very wide (U.S. general population: 2.39%–6.84%; non-U.S.–born population: 11.91%–22.83%). Heterogeneity between effect sizes was likely due, in part, to the use of different data sources as four of the studies utilized NHANES data and the other two studies used national TB surveillance data, an LTBI reactivation rate, and an annual risk of progression to TB disease to back-calculate LTBI prevalence. In addition, among studies that used NHANES data, methodologies differed among studies on how TB infection was determined when estimating prevalence. In general, studies using NHANES data used a positive diagnostic test result to determine infection, which had a risk of bias due to the possibility of overestimating LTBI infection prevalence, especially the estimates employing TST positivity to determine prevalence. Previous research has shown that diagnostic tests may remain positive even after successful treatment to cure, especially TST⁴⁴. In addition, prior BCG vaccination causes a reaction to the TST and can subsequently lead to a false positive test³⁹. This is especially relevant in the LTBI prevalence estimate in the non-U.S.–born population as BCG is used in many countries with high prevalence of TB^{39–41}.

In addition, the representative population for the 2011–2012 NHANES is limited to the noninstitutionalized, civilian U.S. population ages 6 years and older. Thus, estimates exclude a portion of the pediatric population, as well as incarcerated individuals. This is particularly relevant to TB infection as those who live or have lived in high-risk congregate settings such as prisons or jails are at higher risk for TB infection and are recommended for TB screening based on CDC and USPSTF guidelines^{36,55,56}. Finally, while NHANES is meant to be nationally representative and can produce stable estimates for non-Hispanic black, non-Hispanic white, Hispanic and Asian persons, it is not designed to produce stable estimates for other race/ethnicity groups, such as American Indians, Native Hawaiians, Pacific Islanders, or Alaska Natives; race/ethnicity groups at higher risk of TB infection¹⁹. In addition, a single 2-year NHANES cycle is designed to provide stable estimates for conditions of 10% prevalence or greater, and with little geographic variability – neither apply to LTBI^{34,35}. The two other studies used national TB surveillance data to back calculate prevalence based on number of TB cases, the proportion of TB cases attributed to recent transmission and an annual risk rate of progression to TB disease. The risk of a

biased LTBI prevalence estimate using this methodology is due to assumptions built into the back calculations^{22,35,98}. The methodology assumes that all genotyped TB cases not attributed to recent transmission arose from LTBI. This approach may overestimate the prevalence of LTBI in localities with many pediatric TB cases, which tend to be difficult to confirm by culture and subsequently receive genotyping. Conversely in localities with longstanding genotyping clusters, this methodology could have underestimated LTBI prevalence^{35,49}.

Finally, none of these estimates can be considered recent. NHANES estimates are utilizing data collected from over a decade ago (i.e., 2011–2012 NHANES). While the back-calculation methods can be updated using more recent estimates, in the published literature the methods use TB surveillance data from 2008–2015 (Haddad, et al.) and 2013–2017 (Mirzazadeh et al.); thus, most recent LTBI prevalence estimates are at least 5 years outdated. In addition, none of the prevalence estimates for either the U.S. general population or the non-U.S.–born population used updated and available 2020 U.S. census denominators. As such, the mean prevalence estimates calculated in the meta-analysis may still not reflect current LTBI prevalence rates in the United States.

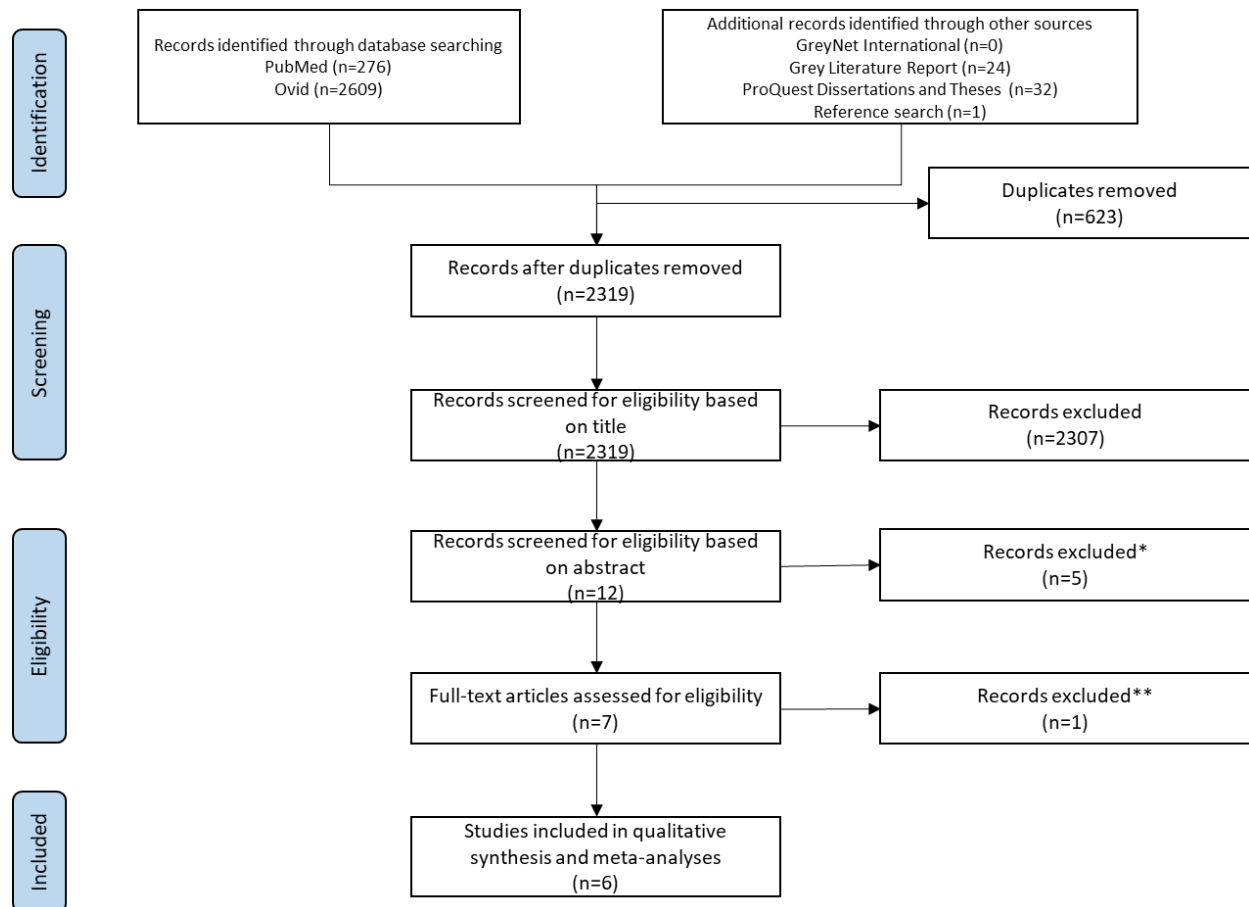
Overall, it is difficult to ascertain from this study whether NHANES data or back-calculation methodologies using TB surveillance data, and their subsequent methodologies for determining LTBI, provided a more accurate estimate LTBI prevalence in the U.S. general population and non-U.S.–born population. The differences between estimates overall may not warrant a difference in determining public health resources to dedicate to TB infection testing and treatment; however, the difference in resources to estimate LTBI prevalence using NHANES data or back calculation methodologies is significant. NHANES is a resource intensive nationally representative survey which consists of an interview and examination component for 5,000 persons^{35,99}. In addition, the survey does not consistently assess TB infection and has not included a TB infection testing component since the 2011–2012 survey³⁵. The NHANES 2019–2020 intended to include a TB infection test for a limited population, however, due to NHANES suspending operations in March 2020 due to the COVID-19 pandemic, the data collection for NHANES 2019–2020 was not completed and the collected data are not nationally representative¹⁰⁰. NHANES data may not be a reliable data source for consistently determining future LTBI prevalence estimates in the U.S. Back calculated LTBI prevalence can be readily and consistently reproduced using national and local TB surveillance data using straightforward calculations⁴⁹. In addition, these methods for estimating prevalence could be applied in any jurisdiction with an established tuberculosis surveillance system.

A limitation of this meta-analysis is that several studies use the same data sources, such as 2011–2012 NHANES data, but report different prevalence estimates due to differences in methodology for determining LTBI infection. A correlated hierarchical model would have been the more robust methodology for accounting for multiple studies using the same data source. However, due to the low number of both U.S. general population ($n=7$) and non-U.S.–born population ($n=7$) prevalence estimates, the more robust correlated hierarchical model was not feasible, and we employed a random effects model which could not account for effect size dependency within studies (e.g., a study reporting multiple prevalence estimates using NHANES data).

Currently cited LTBI prevalence estimates for the U.S. general population and the non-U.S.–born population have several limitations and reflect outdated numerator and denominator data that range from 5 to 12 years old. In addition, there is heterogeneity amongst currently published estimates due to differences in data sources and methodologies for determining LTBI prevalence. This analysis provides a mean estimate of LTBI prevalence for the U.S. general population and the non-U.S.–born population that can be used to determine the breadth and resources that need to be dedicated to targeted testing, as well as LTBI treatment completion, to reduce the burden of TB infection in the U.S. However, due to the ongoing lack of LTBI surveillance data, updated LTBI prevalence estimates using robust methodology for determining infection are needed to determine updated LTBI prevalence estimates in the U.S. for both the general population and the at higher risk non-U.S.–born population. Currently, there are limited number of data sources being utilized to determine LTBI prevalence estimates in the U.S. This conclusion suggests that novel data sources, such as electronic health record data, and methods to analyze these data, should be explored further as possible resources for determining and updating U.S. LTBI prevalence estimates.

Tables and Figures

Figure 2.1. PRISMA flow diagram outlining literature search



*Includes an abstract that was removed due to it representing the same analysis and results in an included peer reviewed journal article (Yelk, 2021)

**This article reported point estimates for LTBI prevalence in the U.S. general population as a proportion (percentages), but without estimating the underlying number of persons with LTBI. Ultimately, this article was excluded from the meta-analysis after a discussion with the authors who indicated that due to the methodology used to estimate those percentages, they were not able to produce an estimated number of persons with LTBI¹⁰¹.

Table 2.1. Summary of studies included in qualitative synthesis and meta-analysis

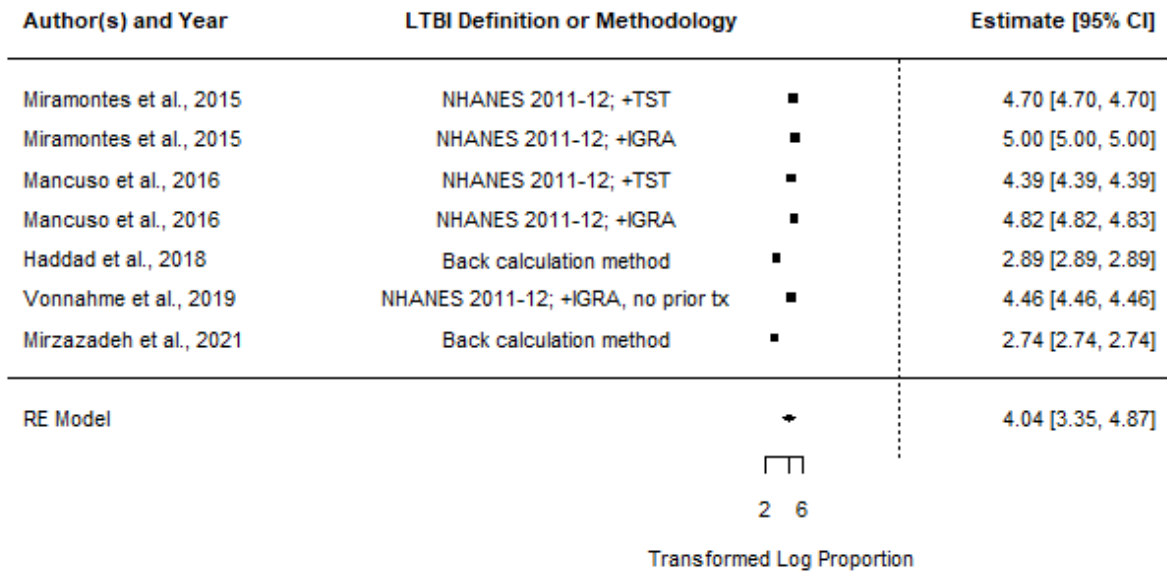
First Author (Year)	U.S. general population			Non-U.S.–born population			LTBI definition or methodology	LTBI infection data source	Population denominator data source
	LTBI prevalence estimate %(95%CI)	LTBI Population estimate N(95%CI)	Total population estimate (N)	LTBI prevalence estimate %(95%CI)	LTBI population estimate N(95%CI)	Total population estimate (N)			
Miramontes R et. al. ²³ (2015)	4.7 (3.4-6.3)	13,276,000 (9,604,000- 17,795,000)	282,460,000	20.5 (16.1-25.8)	8,139,000 (6,392,000- 10,243,000)	39,700,801	Positive TST defined as induration ≥ 10 mm; NHANES- provided 2-year examination weights were adjusted to account for missing TST reading. Also employed smoothing techniques to address digit preferences for 8 and 9 mm TST measurements.	2011–2012 NHANES (noninstitutionalized civilian U.S. population ages 6 years and older)	2011 American Community Survey data
	5.0 (4.2-5.8)	14,123,000 (11,863,000- 16,383,000)	282,460,000	15.9 (13.5-18.7)	6,312,000 (5,360,000- 7,424,000)	39,700,801	Positive QFT-GIT	2011–2012 NHANES (noninstitutionalized civilian U.S. population ages 6 years and older)	2011 American Community Survey data
Mancuso JD et al. ⁶² (2016)	4.4 (3.1-6.1)	12,398,000 (8,869,000- 17,230,000)	282,460,000	19.8 (15.2–25.4)	9,018,000 (6,917,000- 11,570,000)	45,624,000	Positive TST defined as induration ≥ 10 mm; NHANES- provided 2-year examination weights were further adjusted for nonparticipation in TB testing so that it would represent the applicable study population (adjustments not described in detail)	2011–2012 NHANES (noninstitutionalized civilian U.S. population ages 6 years and older)	2011 American Community Survey data
	4.8 (4.0-5.8)	13,628,000 (11,411,000- 16,241,000)	282,460,000	15.9 (13.3-19.0)	7,264,000 (6,045,000- 8,678,000)	45,624,000	Positive QFT-GIT	2011–2012 NHANES (noninstitutionalized civilian U.S. population ages 6 years and older)	2011 American Community Survey data
Haddad MB et al. ⁴⁹ (2018)	3.1 (2.2-5.2)	8,932,528	309,349,689*				A mathematical formula that uses the annual TB incidence averaged during 2008–2015, the proportion of TB cases attributed to recent transmission, and a uniform population-level 0.1% annual risk for progression to active TB disease to derive LTBI prevalence estimates. Assumes all genotyped TB cases not	<ul style="list-style-type: none"> National TB Surveillance System reported cases averaged during 2008-2015 Proportion of genotyped TB cases not attributed to recent transmission^{22,98} 	U.S. Census 2010 population data

							attributed to recent transmission arose from LTBI and that the same proportion of non-genotyped cases are attributed to recent transmission as genotyped cases.	<ul style="list-style-type: none"> • 0.1% annual risk for progression to TB disease²¹ 	
Vonnahme LA et al. ²⁵ (2019)	4.5 (not reported)**	12,604,138 (10,500,000-14,700,000)	282,460,101	13.8 (not reported)**	5,488,989 (4,600,000-6,300,000)	39,700,801	Positive QFT-GIT and no self-reported prior TB treatment	2011–2012 NHANES (noninstitutionalized civilian U.S. population ages 6 years and older)	2011 American Community Survey data
Mirzazadeh A, et al. ⁵⁰ (2021)	2.7 (2.6-2.8)	8,561,899 (8,307,006-8,844,338)	312,295,448	13.9 (13.5-14.3)	5,845,369 (5,681,850-6,015,484)	42,167,973	A mathematical formula that uses the annual TB incidence averaged during 2013–2017, the proportion of TB cases attributed to recent transmission, and previously derived population-specific annual risk rates for progression to active TB disease from the published literature to derive LTBI prevalence estimates. Assumes all genotyped TB cases not attributed to recent transmission arose from LTBI and that the same proportion of non-genotyped cases are attributed to recent transmission as genotyped cases.	<ul style="list-style-type: none"> • National TB Surveillance System reported cases averaged during 2013-2017 • Proportion of genotyped TB cases not attributed to recent transmission^{22,98} • Previously derived population-specific annual risk rates for progression to active TB disease from the published literature^{21,97} 	2015 American Community Survey midpoint estimates
Yelk Woodruff R, et al. ²⁴ (2021)				16.9% (13.1-21.5)	6,709,000 (5,201,000-8,536,000)*	39,700,801	Individuals with positive IGRA test adjusted for IGRA sensitivity and specificity	2011–2012 NHANES (noninstitutionalized civilian U.S. population ages 6 years and older)	2011 American Community Survey data

*Population estimate was not reported in article; referenced U.S. Census 2010 population data directly to obtain population estimate

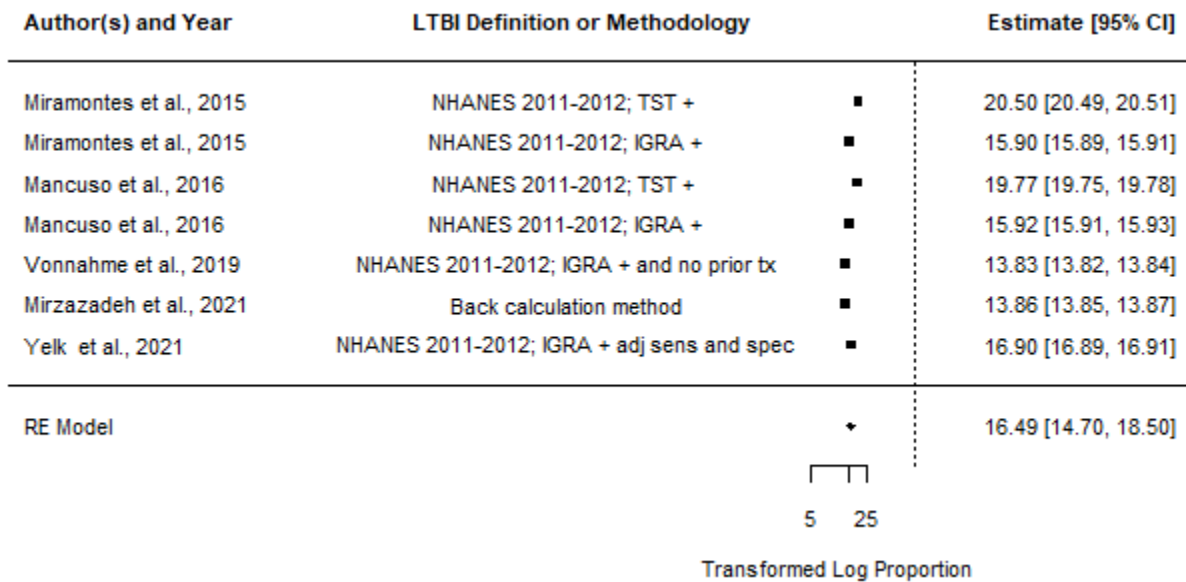
**Estimate reported by author

Figure 2.2. Forest plot of mean prevalence estimate of LTBI infection in the U.S. general population



RE Model ($Q=3869068.0262$, $df=6$, $p < .0001$; $I^2=100\%$)

Figure 2.3. Forest plot of mean prevalence estimate of LTBI infection in the non-U.S.-born population



RE Model ($Q=1239344.2025$, $df=6$, $p < .0001$; $I^2=100\%$)

Appendix B

Title and abstract screener and full-text study eligibility and data extraction tool

Data question	Code
Title/Abstract screener	
1. Is this document about latent TB infection (LTBI) or TB infection prevalence?	1 Yes 0 No [STOP] 999 unsure
2. Is this document an observational study of LTBI or TB infection?	1 Yes 0 No [STOP] 999 unsure
3. Is the study population in this document either the U.S. general population OR the non-U.S.-born population in the U.S.?	1 Yes 0 No [STOP] 999 unsure
Full-text review	
Study eligibility	
1.1 Does this article report on LTBI or TB infection prevalence?	1 Yes 0 No [STOP] 999 unsure
1.2 Is the study cohort from the United States?	1 Yes 0 No [STOP] 999 unsure
1.3 Is the study population either the U.S. general population OR the non-U.S.-born population, or is the sample meant to represent either of these populations?	1 Yes 0 No [STOP] 999 unsure
1.4 Does the article report LTBI prevalence in either the U.S. general population OR the non-U.S.-born population as a population estimate [N] or proportion of the total population [%]?	1 Yes 0 No [STOP] 999 unsure
1.5 Does the article report the denominator used to determine LTBI prevalence (e.g., the U.S. general population)?	1 Yes 0 No [STOP] 999 unsure
Data Extraction	
Report	
2.1 First author (last name, first name)	Free text
2.2 Year published	Year
2.3 Publication type (e.g., peer-reviewed, dissertation, report)	Free text
2.4 Does this publication report LTBI prevalence estimates for the U.S. general population?	1 Yes 0 No [skip 2.9-2.15, 2.33-2.39] 999 unsure
2.5 Does this publication report LTBI prevalence estimates for the non-U.S.-born population?	1 Yes 0 No [skip 2.16-2.22, 2.40-2.46] 999 unsure
Data sources	
2.6 How many total data sources are used in the overall methodology for determining LTBI prevalence?	Number
2.7 Name the data source(s) used	Free text
2.8 Time frame for each data source	Dates
U.S. GEN POP: Data source - LTBI infection in the U.S. general population (i.e., prevalence numerator)	
2.9 Name the data source used to determine LTBI prevalence for the U.S. general population (i.e., prevalence numerator)	Free text
2.10 Time frame of the data source	Dates
2.11 Was prevalence restricted to a subpopulation? (e.g., NHANES represents nonincarcerated individuals ages 6 and up)	1 Yes 0 No 999 unsure
2.12 If yes, describe the restrictions or the population that the numerator represents	Free text
2.13 What type of sampling technique was used for the data source?	1 probability 2 non-probability 999 unsure 888 none
2.14 Code the sampling technique for the data source (if relevant)	<ul style="list-style-type: none"> • 1 simple random sampling • 2 systematic random sampling • 3 stratified random sampling • 4 cluster sampling • 5 multistage cluster sampling • 6 stratified multistage cluster sampling

	<ul style="list-style-type: none"> • 7 convenience sampling • 8 venue-based time sampling • 9 purposive sampling • 10 quota sampling • 11 chain referral or snowball sampling • 12 respondent-driven sampling • 999 unsure • 888 none
2.15 Sample size for the data source (if relevant)	Number
NON-USB: Data Source - Defining LTBI infection in the non-U.S.–born population (i.e., prevalence numerator)	
2.16 Name the data source used to determine LTBI prevalence for the non-U.S. general population (i.e., prevalence numerator)	Free text
2.17 Time frame of the data source	Dates
2.18 Was prevalence restricted to a subpopulation? (e.g., NHANES represents nonincarcerated individuals ages 6 and up)	1 Yes 0 No 999 unsure
2.19 If yes, describe the restrictions or the population that the numerator represents	Free text
2.20 What type of sampling technique was used for the data source?	1 probability 2 non-probability 999 unsure 888 none
2.21 Code the sampling technique for the data source (if relevant)	<ul style="list-style-type: none"> • 1 simple random sampling • 2 systematic random sampling • 3 stratified random sampling • 4 cluster sampling • 5 multistage cluster sampling • 6 stratified multistage cluster sampling • 7 convenience sampling • 8 venue-based time sampling • 9 purposive sampling • 10 quota sampling • 11 chain referral or snowball sampling • 12 respondent-driven sampling • 999 unsure • 888 none
2.22 Sample size for the data source (if relevant)	Number
DENOMINATOR: Data source and definition of the overall population (i.e., prevalence denominator)	
2.23 Was the denominator (N) unreported in this analysis? (i.e., answered NO to Q1.5)	1 Yes [Skip 2.24-2.25] 0 No
2.24 Did the author provide the denominator for this analysis?	1 Yes [Skip 2.25] 0 No
2.25 Will the denominator be estimated from census data for the meta-analysis?	1 Yes [continue coding based on census data being the data source] 0 No
2.26 Name the data source(s) used to determine the overall U.S. general population and/or the non-U.S.–born population (i.e., denominator)?	Free text
2.27 Time frame of the data source	Dates
2.28 Was the denominator restricted to a subpopulation (other than the non-U.S.–born population)? (e.g., NHANES represents nonincarcerated individuals ages 6 and up)	1 Yes 0 No 999 unsure

2.29 If yes, describe the restrictions or the population that the denominator represents	Free text
2.30 What type of sampling technique was used for the data source?	1 probability 2 non-probability 999 unsure 888 none
2.31 Code the sampling technique for the data source (if relevant)	<ul style="list-style-type: none"> • 1 simple random sampling • 2 systematic random sampling • 3 stratified random sampling • 4 cluster sampling • 5 multistage cluster sampling • 6 stratified multistage cluster sampling • 7 convenience sampling • 8 venue-based time sampling • 9 purposive sampling • 10 quota sampling • 11 chain referral or snowball sampling • 12 respondent-driven sampling • 999 unsure • 888 none
2.32 Sample size for the data source (if relevant)	Number
Outcome measures – LTBI prevalence in the U.S. general population	
2.33 How many prevalence estimates for the U.S. general population are reported in this publication?	Number
<i>Instructions: for EACH prevalence estimate reported complete Q2.34-2.39</i>	
2.34 National LTBI prevalence estimate	% [confidence limits]
2.35 National LTBI population estimate	N [confidence limits]
2.36 Report the total U.S. general population used to calculate prevalence	Number
2.37 Were prevalence estimates based on diagnostic test results?	1 Yes 2 No [Skip 2.38] 999 unsure
2.38 If yes, what diagnostic test results were used to calculate estimates? (Select all that apply)	<ul style="list-style-type: none"> • 1 tuberculin skin test (TST) • 2 interferon gamma-ray assay (IGRA) • 3 Both TST and IGRA • 4 TST or IGRA • 5 Chest radiograph • 6 Self-reported diagnosis • 7 Other, please describe: • 8 None • 999 unsure
2.39 Describe the methodology for determining LTBI infection	Free text
Outcome measures – LTBI prevalence in the non-U.S.-born population	
2.40 How many prevalence estimates for the non-U.S.-born population are reported in this publication?	Number
<i>Instructions: for EACH prevalence estimate report complete Q2.41-2.46</i>	
2.41 non-U.S.-born LTBI prevalence estimate	% [confidence limits]

2.42 non-U.S.–born LTBI population estimate	N [confidence limits]
2.43 Report the total non-U.S.–born population used to calculate prevalence (total number, if relevant)	Number
2.44 Were prevalence estimates based on diagnostic test results?	1 Yes 2 No [Skip 2.44] 999 unsure
2.45 If yes, what diagnostic test results were used to calculate estimates? (Select all that apply)	<ul style="list-style-type: none"> • 1 tuberculin skin test (TST) • 2 interferon gamma-ray assay (IGRA) • 3 Both TST and IGRA • 4 Chest radiograph • 5 Self-reported diagnosis • 4 Other, please describe: • 5 None • 999 unsure
2.46 Describe the methodology for determining LTBI infection	<ul style="list-style-type: none"> • Free text
Risk of bias	
3.1 Adequate representation of U.S. general population (selection bias)	<ul style="list-style-type: none"> • 1 yes = low risk of selection bias • 2 unclear; insufficient information (inclusion or exclusion criteria unclear; methodology does not clearly weight populations in analysis) • 3 no = high risk (specific populations or groups are not represented in the sample population; convenience sample is used)
3.2 Adequate definition of LTBI?	<ul style="list-style-type: none"> • 1 yes = low risk of over or underestimation of LTBI • 2 = unclear; insufficient information to determine definition of LTBI (methodology for defining LTBI is unclear) • 3 no = high risk (methodology uses single diagnostic test to define LTBI, or a diagnostic test with many limitations [e.g., TST])

Chapter 3: Using electronic health record data to measure the latent tuberculosis infection care cascade in safety net primary care clinics

This manuscript was published in the American Journal of Preventive Medicine (AJPM) Focus, Volume 2 Issue 4, December 2023: <https://doi.org/10.1016/j.focus.2023.100148>.

Introduction

Tuberculosis (TB) elimination in the United States is defined as a national incidence rate of <1 case per 1,000,000 population. Currently the incidence rate is 27 times the elimination threshold¹⁰². Over 80% of TB disease cases in the United States are attributed to latent TB infection (LTBI) reactivation despite progression to TB disease being preventable with treatment^{21,22}. Appropriate screening and treatment of LTBI are critical steps for achieving TB elimination in the United States. The Centers for Disease Control and Prevention (CDC) and the U.S. Preventive Services Task Force (USPSTF) recommend TB screening for populations at increased risk for infection, progression to TB disease, or transmission to others. These include people who were born or lived outside the United States in countries where TB is more common, individuals who are immunocompromised such as persons living with HIV, and those who live or have lived in high-risk congregate settings such as homeless shelters, prisons or jails^{36,55,56}.

Primary care clinics are important settings for LTBI screening and treatment. Safety net clinics deliver primary care health care services to patients who experience barriers to accessing care due to insufficient insurance coverage, geographic isolation, language and culture, mental illness, or homelessness^{77,103}. Most safety net clinics are community health centers (CHC), which serve over 30 million persons in the U.S.; 65% belong to racial or ethnic minority groups, 1.3 million persons experiencing homelessness, and 7 million persons are best served in a language other than English^{78,79}. Safety net primary care clinics are especially relevant for LTBI since they serve populations at higher risk for TB infection, such as non-U.S.-born persons and persons experiencing homelessness^{18,51,78-86}. They are key settings to engage in efforts to expand TB screening, testing, and treatment.

Electronic health record (EHR) data have the potential to identify persons who should be tested for LTBI and their outcomes through the cascade of care in safety net clinics. OCHIN is a nonprofit health information technology collaborative that provides linked EHR services to safety net clinics across 30 states, making it the largest network of safety net clinics in the United States⁹³. OCHIN maintains an EHR research database of primary care delivery data from safety net clinics and includes information from over 6 million patient records that are demographically similar to the national profile of patients in CHCs^{78,89}. OCHIN utilizes a common data model to combine EHR data from clinics into this research-ready database that contains aggregated, longitudinal data starting in 2004⁹².

The LTBI care cascade (LCC) is a tool that measures the progression of patients across the care continuum from identifying people at risk for TB infection through treatment completion, with several steps in between⁵³. The LCC is a useful tool to improve TB prevention and identify steps where patients are lost, which can be used to improve TB prevention in clinics and healthcare systems^{53,104,105}. However, measuring the LCC can be difficult and has not been well characterized outside of public health clinics or specific populations (e.g., migrants, persons experiencing homelessness, contacts to TB cases), nor has EHR data been used previously to characterize a complete LCC^{53,54}. We used a cohort from the OCHIN EHR research data warehouse to characterize the LCC among their network of safety net clinics and to identify opportunities for potential future interventions to improve TB infection screening, diagnosis, and treatment.

Methods

EHR data was extracted for patients with any encounter at an OCHIN member clinic between January 1, 2012, and December 31, 2019. The analytic cohort was limited to patients considered established in primary care, defined as at least two ambulatory visits (in-person clinic visits) to the same OCHIN member clinic before or during the study period. Data extracted included demographics, medical risk factors, lab results related to LTBI diagnostic tests, diagnostic codes (ICD-9 and ICD-10) associated with the patients' visit history or active problem list, and prescription drug records. Data elements contained in the dataset were not restricted based on encounter date and all patient history up until September 2020 (when data was extracted) was included in the data cohort; patient history from prior to 2012 was limited and was not considered inclusive of all patient history.

The first step in the LCC was to identify patients that should be tested for TB infection. We applied screening criteria adapted from the California Adult Tuberculosis Risk Assessment¹⁰⁶, a simplified risk assessment tool based on CDC and USPSTF TB screening guidelines^{36,55,56} that recommends screening for patients who meet one or more of the following criteria: 1) birth, travel or residence in any country other than the United States, Canada, Australia, New Zealand, or a country in western or northern Europe for at least one month; 2) immunosuppression, current or planned, because of HIV infection, organ transplant, or treatment with TNF-alpha antagonists, steroids or other immunosuppressive medication; or 3) known close contact to someone with infectious TB disease during their lifetime. To align with the USPSTF recommendations⁴, we also included individuals who were likely staying in congregate settings.

EHR variables were not always available to match exact risk criteria. Table 3.1 describes sources referenced, screening criteria, and equivalent definitions and variables in the EHR. For the first criterion,

country of birth was the preferred variable as a surrogate to identify people who had lived outside the U.S. due to birth, travel or residence; among patients without country of birth records, non-English language preference was used as a proxy. Preferred language has been found to be a reasonable proxy for non-U.S.–born persons who meet screening criteria for LTBI^{68,87}. Since we relied on preferred language, we generalized all non-U.S.–born patients with a recorded country of birth to be at risk.

For the second criterion, immunosuppression was defined as patients who ever had or currently have a HIV diagnosis, organ transplant, or aftercare following an organ transplant based on the presence of specific ICD-9 and ICD-10 diagnostic codes recorded in patients' EHR. Also included were individuals ever prescribed immunosuppressive medications (Table 3.1, footnote c) that place individuals at higher risk for TB reactivation^{58,59}. Both generic and brand name drugs were searched for in patient prescription drug records. Finally, close contacts to someone with TB disease were identified by a specific diagnostic code recorded in the patients' EHR.

We also included persons who live in, or have lived in, high-risk congregate settings by including individuals who reported ever experiencing homelessness, and individuals who ever had an encounter at an OCHIN-member clinic that is associated with a correctional facility. Homelessness status was determined by a registration variable and social history variable in the EHR; these variables are updated based on clinic-specific protocols. We also searched for homelessness ICD codes recorded in patients' EHRs. When a criterion was defined by the presence of an ICD code or a drug prescription, patients were considered to have met specific criterion if the ICD code or drug prescribed was present in their record at any point up to the end of the study period.

After determining the proportion of the cohort that met screening criteria, we removed patients from the LCC whom had an ICD diagnostic code for 1) TB disease or 2) a prior history of TB because they would not have been eligible for LTBI testing or treatment. Patients with these codes were removed regardless of whether they also had an ICD diagnostic code for LTBI. Next in the LCC, we determined whether patients had received a TB infection diagnostic test ever until the end of the study period: an interferon-gamma release assay (IGRA) test, including any generation of QuantiFERON TB test (QFT), or T-SPOT.TB test (T-SPOT), or a tuberculin skin test (TST). Among those tested, we determined the proportion that had a valid diagnostic result, defined as a positive or negative result. We defined a positive result as at least one positive IGRA test result (if multiple were performed) or at least one positive TST when only TST results were available. A detailed diagram describing how diagnostic test results were defined is available in the online-only content (Appendix C, Figure S1).

We defined patients diagnosed with LTBI in the LCC as those with an LTBI diagnostic code in their visit diagnosis history or active at any time on the problem list following a positive diagnostic test in their EHR. A positive diagnostic test is not sufficient for diagnosing LTBI; a diagnosis requires excluding TB disease through a symptom review, focused physical exam, and a chest radiograph³⁸. However, EHR variables that would allow identification of patient symptom reviews as having occurred or results of a chest radiograph, without requiring individual chart review, do not exist.

Among those diagnosed with LTBI, we determined the proportion that were prescribed an LTBI treatment regimen. We identified recommended LTBI treatment regimens including isoniazid and rifapentine, isoniazid alone, rifampin or rifabutin alone, or isoniazid and rifampin or rifabutin^{8–10}. For a drug to be considered part of an LTBI treatment regimen it had to be prescribed after a positive TB diagnostic test. To reduce potential misclassification of TB disease treatment regimens as LTBI regimens the aforementioned regimens had to be prescribed without the following drugs: pyrazinamide, ethambutol, cycloserine, ethionamide, clofazimine, or linezolid. For regimens where multiple drugs were prescribed, drugs had to be prescribed on the same day. If a patient was prescribed more than one LTBI regimen, we deferred to the most recently prescribed regimen. Since rifampin and rifabutin can be used to treat other conditions^{72,107}, an algorithm was created to determine whether these prescriptions were part of a treatment regimen for TB disease or LTBI and is available in the online-only content (Appendix C, Supplemental Methods, Table S1, Table S2). Treatment initiation and completion could not be assessed using standard EHR data elements captured.

This activity was reviewed by CDC and was conducted consistent with applicable federal law and CDC policy.[§]

Results

1,973,017 patients were included in the analytic cohort due to having had at least two ambulatory visits to the same OCHIN primary care facility before or during the study. The majority of the analytic cohort were aged 25–44 years (57.9%) and 56.2% (n=1,109,252) were female. 32% identified as Hispanic; 41.3% were non-Hispanic white, 15.7% were non-Hispanic black, 4.8% were Asian; less than 1% were Native Hawaiian or other Pacific Islander, or American Indian or Alaskan Native. Nativity, or country of birth, was frequently missing (87.3%). Preferred language was captured for 98.2% of the analytic cohort and 26.8% preferred a non-English language. The demographics for the analytic cohort are described in Table 2.

[§] See e.g., 45 C.F.R. part 46.102(l)(2), 21 C.F.R. part 56; 42 U.S.C. §241(d); 5 U.S.C. §552a; 44 U.S.C. §3501 et seq.

Among the analytic cohort, 43.5% (n=858,767) met screening criteria (Figure 3.1). 13.7% (n=117,468) reported a non-U.S. country of birth, 59.1% (507,519) reported a non-English language preference; 2.7% (n=22,788) had an HIV diagnosis, 23.2% (n=199,578) were prescribed an immunosuppressive medication, 22.3% (n=191,693) had experienced homelessness and 4.0% (n=34,221) had a clinical encounter in a correctional facility (Table 3.2). Among those who met our screening criteria, 0.8% (n=7,198) were removed from the LCC because of an ICD code for either TB disease (n=6,083) or a personal history of TB disease in their medical record (n=1,610); 495 patients had an ICD code for both.

Among those patients who met screening criteria and were eligible to continue through the LCC, 21.4% (n=182,454) were tested for TB infection. Over half of total tests performed were TSTs (59.6%, n=115,867), with 40.4% (n=78,665) of testing being performed by IGRA. Among those tested, 85.6% (n=156,159) had a valid test result. Approximately one fourth (23.8%) of TSTs performed did not have a result recorded that could be readily identified; 96.6% (n=75,989) of IGRAs performed had valid results. Among those with a valid diagnostic result, 12.9% (n=20,088) had a positive result, and 82.0% (n=16,465) of those also had an ICD code for LTBI. Thus, 10.5% (n=16,465) of patients who met screening criteria and had a valid test result met our LTBI diagnosis criteria. Country of birth was unknown for the majority of those diagnosed with LTBI (60.3%, n=9,929); 80.8% (n=13,009) reported a non-English language preference.

Among those diagnosed with LTBI, 29.1% (n=4,791) were prescribed an LTBI regimen. The majority (66.4%, n=3,180) were prescribed isoniazid alone, 25.3% (n=1,214) were prescribed rifampin or rifabutin alone, 7.4% (n=356) were prescribed isoniazid and rifapentine, and 0.9% (n=41) were prescribed isoniazid and rifampin or rifabutin. Among those prescribed an LTBI regimen, approximately half (50.6%) had a country of birth recorded, with 95% being non-U.S. born; 80% reported non-English language preference.

Discussion

We created a novel methodology for constructing an LCC using EHR data. Previous studies have used EHR data to determine the proportion who met TB screening criteria, were tested, diagnosed with LTBI, or prescribed treatment, but none have reported on the LCC in totality^{68,108–110}. In addition, previous screening criteria definitions were limited by the study population or setting, or based solely on preferred language and self-reported nationality^{53,54,68}. Our definition was applied to a broad cross-sectional cohort that also included individuals who met criteria due to immunosuppression, close contact to a TB case, or history of homelessness or incarceration. This is the first study to characterize

the LCC across a large network of primary care safety net clinics; thus, setting a baseline for the at-risk population seeking care in this setting and how TB screening and diagnostic recommendations are being implemented.

Based on CDC and USPSTF screening criteria, a large proportion of this clinical network patient population is at high risk for TB infection, as 44% of the cohort met the recommended TB screening criteria based on available EHR data. The OCHIN cohort was reflective of a high-risk population; 10% of patients with a valid test result being diagnosed with LTBI is slightly lower than national estimates of TB infection prevalence among non-U.S.–born persons and higher than the estimate of 5% among the U.S. general population²³. Another indicator that this population is at high risk for TB infection is that approximately 0.7% of those who met screening criteria were diagnosed with TB disease at some point during the study period. Assuming the population that met screening criteria remained stable throughout this period, that represents a TB incidence rate of 88.5/100,000 annually among those who met screening criteria, higher than the reported incidence rate of 34.0/100,000 among non-U.S.–born individuals reported in 2019. Thus, the OCHIN clinical network represents an appropriate study population for understanding LTBI testing and diagnosis in patients at high risk of TB infection, and for potential interventions to increase screening, improve diagnostic testing, and increase treatment for LTBI.

Our findings demonstrate that there is potential for improvement in TB screening, diagnostic testing, and treatment in this network of clinics. Over 78% of patients who met our screening criteria were not tested for infection. The majority of testing was performed by TST, which presents an opportunity for increased use of the IGRA diagnostic test; the preferred test in a population who met screening criteria based on non-U.S.–birth and other risk factors such as experiencing homelessness^{39,111}. Treatment prescription was also low in the cohort, and only 33% were prescribed a rifamycin-based treatment regimen. This may reflect that U.S. guidelines for preference of shorter, rifamycin-based therapy over isoniazid monotherapy were not published until 2020¹⁰. However, in 2011 CDC recommended once-weekly isoniazid and rifapentine for 12 weeks (3HP) by directly observed therapy (DOT) as an equal alternative to other regimens⁸; the recommendations were further expanded in 2018 to include additional populations and types of therapy administration⁹. Thus, the large proportion of OCHIN patients being prescribed longer course treatment with isoniazid monotherapy may also represent an implementation gap to be further explored between CDC recommendations and a change in clinical care to using shorter-course therapy.

There were limitations to our study. While this was a large sample, results are not generalizable to all safety net clinics in the United States, and the population is not representative of the overall U.S. population at risk for TB infection. In addition, it is possible some patients did not go through steps in the LCC successively, e.g., a diagnostic ICD code for LTBI might have been entered into a patient EHR in the absence of a positive test, causing us to underestimate some steps. Another limitation is that our evaluation of the LCC was limited to routine EHR data that was captured in structured fields. We also could not determine if there were specific reasons that steps were not completed. For example, we could not determine if testing or treatment were offered and declined, or if treatment that was prescribed was initiated or completed by the patient. In addition, how and what data is routinely captured in an EHR can vary across health care settings and across providers within a practice.

Our approach for determining patients' risk for infection was an innovative methodology, but one that likely underestimated the proportion in the cohort who met screening criteria. Country of birth was missing for most patients and EHR data does not capture information on travel or prior residence outside the U.S. Also, while non-English language preference is a reasonable proxy for non-U.S.-born persons who may be at increased risk, many non-U.S.-born persons are proficient English speakers and may not have an alternate preferred language in the EHR^{68,87,88,112}. Our methodology was also not able to capture a complete social history of homelessness or incarceration. Finally, we may have underestimated the number of people diagnosed with LTBI by requiring it be listed as an ICD code.

Our results show that an LCC can be constructed using available EHR data from a large network of primary care safety net clinics. We identified opportunities for improvements in EHR data, such as routinely capturing country of birth as a surrogate measure of risk for TB exposure. We also identified several opportunities in the OCHIN clinical network for improvements, including increased testing among patients with potential risk and increased use of the current, recommended diagnostic tests and treatment regimens. Developing standardized approaches to using EHR data could help advance TB elimination in the U.S. by facilitating partnerships between primary care providers and public health.

[Acknowledgments](#)

Oscar Espinosa, Rose Gunn, Kuan-Chieh Huang, Amy Krueger, Shane Morgan, Thomas Navin, Theresa Schmidt, Noah Schwartz, Matt Whipple, Jonathan Wortham, Amy Zlot. This analysis was fully funded by the Centers for Disease Control and Prevention (CDC) of the U.S. Department of Health and Human Services (HHS). The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by CDC/HHS, or the U.S. Government. This work was conducted with the Accelerating Data Value Across a National Community Health Center Network (ADVANCE) Clinical

Research Network (CRN). The ADVANCE network is led by OCHIN in partnership with Health Choice Network, Fenway Health, and Oregon Health and Science University. ADVANCE is funded through the Patient-Centered Outcomes Research Institute (PCORI), contract number RI-CRN-2020-001.

Tables and Figures

Table 3.1. Tuberculosis screening recommendation criteria definitions used in the OCHIN electronic health record data cohort

Source	Screening criterion	Screening criteria equivalent in EHR	EHR variable(s) used
California Adult Tuberculosis Risk Assessment^a	<i>Birth, travel, or residence in any country other than the United States, Canada, Australia, New Zealand, or a country in western or northern Europe for at least one month</i>	Birth in any country other than the United States OR If country of birth missing, non-English language preference	Country of birth, preferred language
	<i>Immunosuppression, current or planned, due to HIV infection, organ transplant, or treatment with TNF-alpha antagonists, steroids or other immunosuppressive medications</i>	HIV	HIV ICD codes: 042, 079.53, 795.71, V08 (ICD-9) or B20, B97.35, Z21, 098.7x (ICD-10) ^b
		Organ transplant	Organ transplant ICD codes: V42.x (ICD-9) or Z94 (ICD-10) ^b
		Aftercare following organ transplant	Aftercare following organ transplant ICD codes: V58.44x (ICD-9) or Z48.298 (ICD-10) ^b
		Individuals prescribed immunosuppressive medications	Immunosuppressive drugs prescribed ever before the end of the study period ^c
<i>Known close contact to an infectious TB case during lifetime</i>	Close contact with a person with infectious TB disease	TB close contact ICD codes: V01.1 (ICD-9), Z20.1 (ICD-10) ^b	
U.S. Preventive Services Task Force recommendations	<i>Persons who live in, or have lived in, high-risk congregate settings (e.g., homeless shelters and correctional facilities)</i>	Homelessness status	Ever homeless on or before the end of the study period as indicated by registration variable or in social history. Homelessness ICD codes: V60.0, V60.1 (ICD-9) or Z59.0x (ICD-10) ^b
		Encounter in correctional facility	Ever had an encounter on or before the end of the study period in a facility in the OCHIN collaborative that is listed as correctional

EHR=electronic health record

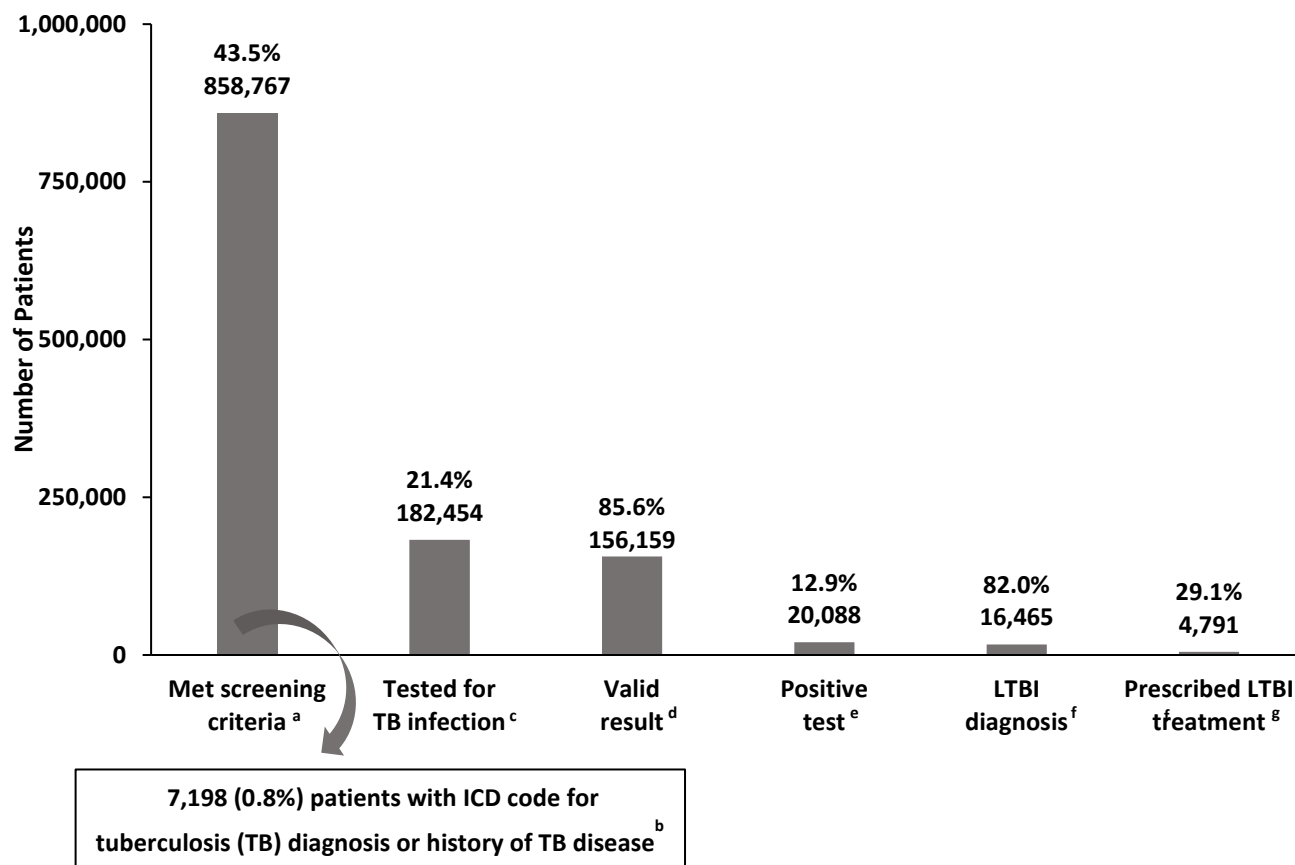
ICD=International Classification of Diseases

^aThis assessment follows U.S. Preventive Services Task Force and Centers for Disease Control and Prevention recommendations^{36,55,56}

^bDiagnostic codes present in visit history and/or active at any time on patient problem list entries ever before the end of the study period

^cAbatacept, adalimumab, anakinra, baricitinib, brodamulab, canakinumab, certolizumab pegol, etanercept, golimumab, guselkumab, infliximab, ixekimumab, methylprednisolone, prednisolone, prednisone (nontopical), risankizumab, rituximab, sariliumab, secukinumab, tildrakizumab, tocilizumab, tofacitinib, upadacitinib, ustekinumab; drug and brand names were searched for in patient prescription records.

Figure 3.1. Latent tuberculosis infection (LTBI) care cascade among patients with two or more ambulatory visits to same OCHIN primary care clinic, 2012–2019



^aPatients that met screening criteria described in Table 1.

^bTuberculosis (TB) disease diagnostic codes including 010, 011, 012, 013, 014, 015, 016, 017, 018 (International Classification of Diseases [ICD]-9) or A15, A16, A17, A18, A19 (ICD-10). Personal history of TB codes including V12.01 (ICD-9) or Z86.11 (ICD-10).

^cInterferon-gamma release assay (IGRA) test or a tuberculin skin test (TST).

^dValid result defined as positive or negative.

^eIf patient had any IGRA result that was valid (positive or negative) the result of that test was used to determine patient test result. Regardless of multiple IGRA tests and results, if patient had any positive IGRA test they were considered latent TB infection (LTBI) test positive. If patient had no valid IGRA result, then a patient with any positive TST was considered to have tested positive.

^fLTBI diagnostic codes including 795.51, 795.52 (ICD-9) or R76.11, R76.12, Z22.7 (ICD-10).

^gIsoniazid and rifapentine, isoniazid alone, rifampin or rifabutin alone, or isoniazid and rifampin or rifabutin; where rifampin and rifabutin dosages consistent with a TB disease or LTBI treatment regimen.

Table 3.2. Characteristics of patients with two or more ambulatory visits to same OCHIN primary care clinic, 2012–2019

Patient Characteristic	Total analytic cohort N=1,973,017		Met screening criteria ^a N=858,767		LTBI diagnosis ^b N=16,465		Prescribed LTBI treatment ^c N=4,791	
	N	%	N	%	N	%	N	%
Age (years) ^d								
0–4	83,412	4.2	30,342	3.5	94	0.6	37	0.8
5–14	239,937	12.2	97,298	11.3	798	4.8	291	6.1
15–24	267,341	13.5	91,027	10.6	1,564	9.5	479	10.0
25–44	605,815	30.7	257,577	30.0	6,198	37.6	1,782	37.2
45–64	536,536	27.2	274,239	31.9	5,586	33.9	1,597	33.3
≥65	239,974	12.2	108,284	12.6	2,225	13.5	605	12.6
Unknown/missing	2	0.0	0	0.0	0	0.0	0	0.0
Sex								
Male	863,493	43.8	385,017	44.8	7,113	43.2	2,103	43.9
Female	1,109,252	56.2	473,654	55.2	9,351	56.8	2,688	56.1
Other	192	0.0	68	0.0	1	0.0	0	0.0
Unknown	80	0.0	28	0.0	0	0.0	0	0.0
Race/Ethnicity ^e								
Hispanic	631,726	32.0	428,042	49.8	7,430	45.1	2,653	55.4
White, not of Hispanic origin	815,365	41.3	225,665	26.3	1,904	11.6	408	8.5
Black, not of Hispanic origin	310,337	15.7	110,227	12.8	3,402	20.7	818	17.1
Asian	94,655	4.8	59,989	7.0	3,157	19.2	762	15.9
Native Hawaiian or other Pacific Islander	8,332	0.4	2,771	0.3	66	0.4	20	0.4
American Indian or Alaskan Native	10,958	0.6	4,025	0.5	27	0.2	9	0.2
Multiple Race	16,148	0.8	4,639	0.5	40	0.2	7	0.1
Unknown/missing race	85,496	4.3	23,409	2.7	439	2.7	114	2.4
Positive TB diagnostic test	26,087	1.3	21,448	2.5	16,465	100.0	4,791	100.0
Tuberculin skin test (TST) ^f	11,348	43.5	8,830	41.2	6,120	37.2	1,181	24.7
Interferon-gamma release assay (IGRA) ^f	14,739	56.5	12,618	58.8	10,345	62.8	3,610	75.3
Diabetes ^{g,h}	236,635	12.0	128,909	15.0	2,778	16.9	809	16.9
Screening Recommendation Criterion								
Nativity ⁱ								
U.S.-born	132,484	6.7	31,566	3.7	390	2.4	157	3.3
Non-U.S.-born	117,468	6.0	117,468	13.7	6,146	37.3	2,267	47.3
Unknown	1,723,065	87.3	709,733	82.6	9,929	60.3	2,367	49.4
Language preference								
English	1,409,652	71.4	339,411	39.5	3,175	19.3	975	20.4
Non-English	528,022	26.8	507,519	59.1	13,209	80.2	3,804	79.4
Unknown	35,343	1.8	11,837	1.4	81	0.5	12	0.3
Immunosuppression								
HIV ^{h,j}	22,788	1.2	22,788	2.7	637	3.9	254	5.3
Prescribed immunosuppressive medication ^k	199,578	10.1	199,578	23.2	1,653	10.0	511	10.7
Organ transplant/aftercare following organ transplant ^{h,l}	2,900	0.1	2,900	0.3	17	0.1	4	0.1
Close contact with a person with infectious TB disease ^{h,m}	5,536	0.3	5,536	0.6	521	3.2	210	4.4
Ever experienced homelessness ^{h,n}	191,693	9.7	191,693	22.3	2,801	17.0	821	17.1

Ever had an encounter in a correctional facility ^o	34,221	1.7	34,221	4.0	336	2.0	48	1.0
---	--------	-----	--------	-----	-----	-----	----	-----

LTBI=Latent tuberculosis (TB) infection

^aPatients that met screening criteria described in Table 1.

^bPatients with an LTBI diagnosis as described in the latent TB infection care cascade in Figure 1.

^cPatients who were offered LTBI treatment as described in the LTBI care cascade in Figure 1.

^dAs of 1 January 2020

^ePersons who identified as Hispanic were categorized as Hispanic regardless of race. Persons who did not identify as Hispanic or had an unknown ethnicity were categorized by race.

^fDenominator is total positive TB diagnostic tests.

^gDiabetes international classification of diseases (ICD) codes: 249.0x, 250.0x (ICD-9), E08.x, E09.x, E10.x, E11.x, E13.x (ICD-10)

^hDiagnostic codes present in visit history and/or active at any time on patient problem list entries ever before the end of the study period.

ⁱBased on recorded country of birth

^jHuman immunodeficiency virus (HIV) ICD codes: 042, 079.53, 795.71, V08 (ICD-9) or B20, B97.35, Z21, 098.7x (ICD-10)

^kPrescribed one of the following ever before the end of the study period: Abatacept, adalimumab, anakinra, baricitinib, brodamulab, canakinumab, certolizumab pegol, etanercept, golimumab, guselkumab, infliximab, ixekinumab, methylprednisolone, prednisolone, prednisone (nontopical), risankizumab, rituximab, sariliumab, secukinumab, tildrakizumab, tocilizumab, tofacitinib, upadacitinib, ustekinumab; drug and brand names were searched for in patient prescription records.

^lOrgan transplant ICD codes: V42.x (ICD-9) or Z94.x (ICD-10); aftercare following organ transplant codes: V58.44x (ICD-9) or Z48.2xx (ICD-10)

^mTuberculosis close contact ICD codes: V01.1 (ICD-9), Z20.1 (ICD-10)

ⁿPatient ever homeless on or before the end of the study period as indicated by registration variable or in social history or homelessness ICD code (V60.0, V60.1 [ICD-9] or Z59.0x [ICD-10]) present in visit history or active at any time on patient problem list entries ever before the end of the study period.

^oPatient ever had an encounter on or before the end of the study period in a facility in the OCHIN collaborative that is listed as correctional.

Appendix C

Supplemental materials

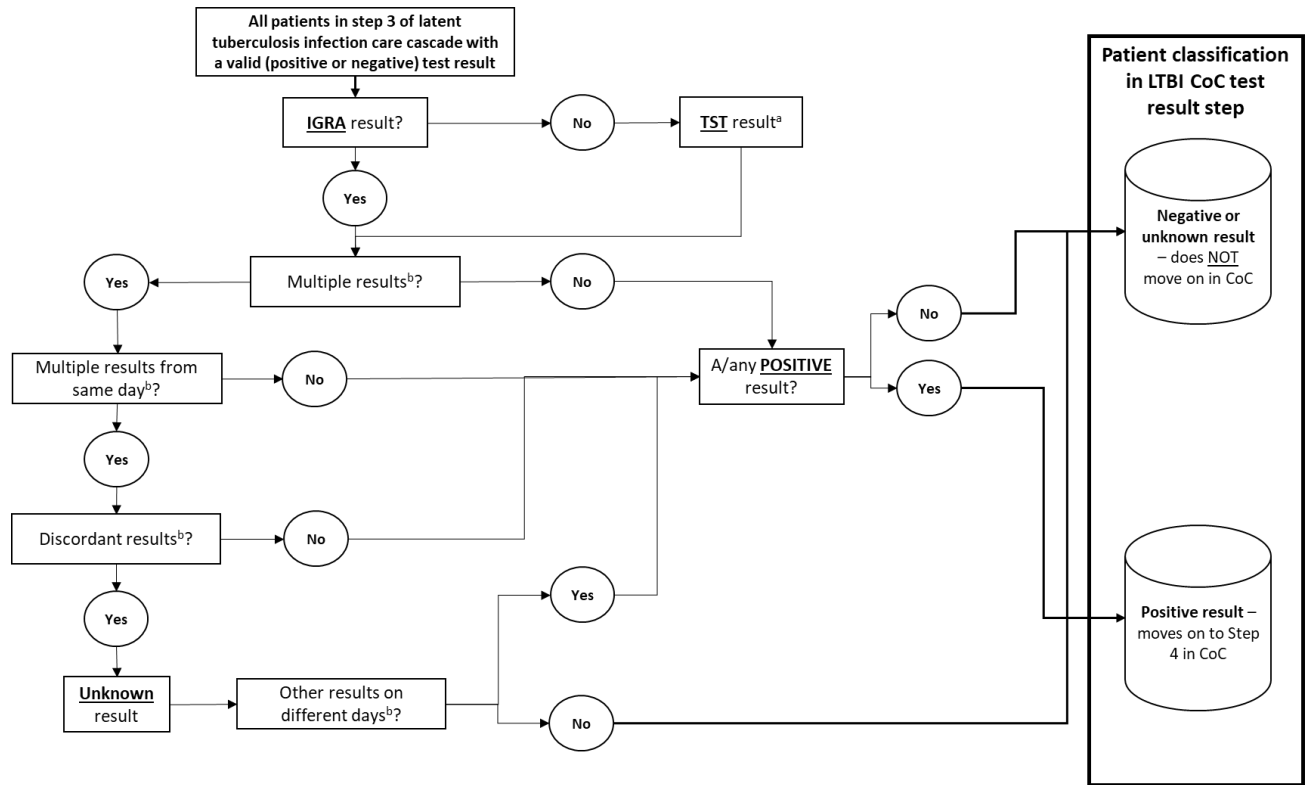
Figure S1. Defining diagnostic test results for patients in the latent tuberculosis infection care cascade using electronic health record data variables

Supplemental Methods. Algorithm for determining rifampin or rifabutin prescriptions as part of latent tuberculosis infection or disease treatment regimens

Table S1. Algorithm for determining rifampin or rifabutin prescriptions as part of latent TB infection or disease treatment regimens

Table S2. Non-TB or LTBI conditions treated with rifampin or rifabutin

Figure S1. Defining diagnostic test results for patients in the latent tuberculosis infection care cascade using electronic health record data variables



IGRA=interferon-gamma release assay; TST=tuberculin skin test

^aSince this population has a known valid result (positive or negative) and individuals do not have a valid IGRA result at this point in logic, must have valid TST result

^bSame test type (i.e., IGRA, TST)

Supplemental Methods. Algorithm for determining rifampin or rifabutin prescriptions as part of latent tuberculosis infection or disease treatment regimens

Rifampin is used frequently to treat other infectious diseases⁷², thus a prescription for rifampin alone is not necessarily indicative that a patient is being treated for latent tuberculosis (TB) infection (LTBI) or TB disease. We created an algorithm for determining whether a rifampin or rifabutin prescription present in an EHR record was indicative of a LTBI or TB disease treatment regimen based on other drugs prescribed with rifampin or rifabutin, frequency and duration of rifampin or rifabutin prescriptions, and presence of other ICD-10 codes indicating conditions that are treated with rifampin or rifabutin (Table S1). This algorithm does not try to distinguish between rifampin or rifabutin prescribed for treatment of TB disease versus LTBI.

If rifampin or rifabutin is prescribed with isoniazid or pyrazinamide it is considered to be indicative of a TB or LTBI treatment regimen. If prescribed in the absence of isoniazid or pyrazinamide, with a dose frequency greater than once per day, the rifampin or rifabutin prescription is classified as not indicative of a TB or LTBI treatment regimen; if the frequency is once per day or unknown, but the duration is less than or equal to 30 days or over 365 days it is also considered to not be indicative of a TB or LTBI regimen. If rifampin or rifabutin is prescribed in the absence of isoniazid or pyrazinamide, with a frequency of once per day or unknown and a duration greater than 30 days but less than or equal to 365 days, then ICD-10 codes for other conditions treated with rifampin or rifabutin are considered. If there is an ICD-10 code for a condition typically treated with rifampin or rifabutin (Table S2) then the rifampin or rifabutin prescription is classified as not prescribed as part of a TB or LTBI regimen. However, if there is no evidence of an ICD-10 code for those conditions in the patient's record, then it is assumed that the rifampin or rifabutin prescription is being prescribed as part of a TB or LTBI treatment regimen. This algorithm was created in close collaboration with TB clinicians and many resources related to LTBI and TB treatment regimens, drug dosing, and ICD coding were utilized^{10,107,113-120}.

Table S1. Algorithm for determining rifampin or rifabutin prescriptions as part of latent TB infection or disease treatment regimens

Prescribed with isoniazid or pyrazinamide ^a	Frequency	Duration ^b	ICD-10 code for condition treated with rifampin ^c	Classification
Yes	NA	NA	NA	TB or LTBI
No	>1/day	NA	NA	not TB or LTBI
No	1/day or unknown	duration <= 30 days	NA	not TB or LTBI
No	1/day or unknown	duration > 365 days	NA	not TB or LTBI
No	1/day or unknown	365 days >= duration > 30 days or unknown	Yes	not TB or LTBI
No	1/day or unknown	365 days >= duration > 30 days or unknown	No	TB or LTBI

TB=Tuberculosis

LTBI=Latent TB infection

^aPrescribed the same day

^bDuration calculated using prescription quantity*(refills+1); if refills were missing assumed there were zero refills; if there was another prescription within 45 days then they were added together; if frequency is unknown, assumed frequency was 1 to calculate duration.

^cAnaplasmosis, bartonella infections, brucellosis, cholestatic pruritis, staph endocarditis, hidradenitis suppurativa, leprosy, meningococcal prophylaxis, pulmonary NTM severe and non-severe, staphylococcus infections, group A strep carriage (eTable 2).

Table S2. Non-TB or LTBI conditions treated with rifampin or rifabutin

Indication	International Classification of Diseases (ICD)-10 codes
Anaplasmosis	A79.82
Bartonella infections	A44.x, A79.0
Brucellosis	A23.x
Cholestatic pruritis	L29.8, L29.9
Staph endocarditis	I33.0, T82.6x, T82.7x
Hidradenitis suppurativa	L73.2
Leprosy	A30.x
Meningococcal prophylaxis	Z20.811
Pulmonary nontuberculosis mycobacterium, non-severe	A31.x
Pulmonary nontuberculosis mycobacterium, severe	A31.x
Staphylococcus infections	B95.61, B95.62, B95.7, B95.8; M86.x; M00.0x, T84.5x, T84.6x, T84.7x; G00.3, G04.2, G06.x, G07, G08; R78.81; T85.7x; T81.4x; S91.x; L97.x; L89.x
Group A Strep carriage	Z22.338

Chapter 4: A novel algorithm to identify individuals with tuberculosis infection using electronic health record data

Introduction

In 2019, 8,920 cases of tuberculosis (TB) disease were reported in the United States, with an incidence rate of 2.7/100,000 persons¹⁸. While this total case count and incidence rate are the lowest ever reported, the average annual decline in incidence has slowed over the last several years and it has been determined that the current rate of decline will not achieve the U.S. TB elimination goals¹⁸. Recent statistical modeling has shown that LTBI prevalence in the U.S. must be reduced in order to achieve TB elimination²⁹⁻³¹. Accordingly, the national strategy for TB elimination is shifting to emphasize LTBI detection and treatment³². However, in the absence of national surveillance, up to date, nationally representative data sources for estimating LTBI prevalence have been limited. Frequently reported U.S. LTBI prevalence estimates use national survey data from NHANES or mathematical modeling, but both have methodological limitations and are not consistently updated to reflect available data^{24,35,49,50,94}. As such, it has been difficult to consistently ascertain the resources necessary to reduce the burden of LTBI among populations at higher risk in the U.S. Alternate data sources should be explored to determine if they may provide more updated LTBI prevalence estimates and be a viable data source for producing consistent and flexible national estimates.

EHR data are an untapped data source for determining LTBI prevalence estimates in the U.S. With an estimated 13 million individuals in the U.S. with LTBI, extracting EHR data for LTBI prevalence estimates, rather than implementing nationwide surveillance, would relieve a large reporting burden on local and state health departments. In addition, diagnosing LTBI is complex and rarely based only on a positive screening test, but rather based on a combination of test results, the absence of symptoms, and patient history considering exposure risk factors. EHR data contains the majority of variables considered when diagnosing LTBI and may be incorporated into a robust and standardized definition of LTBI. This definition would be flexible and could be altered if new diagnostic tools were introduced.

Currently, there is no published methodology describing a LTBI diagnosis definition incorporating multiple EHR data elements. A systematic review summarized validated methods for identifying active TB cases in health administrative data using solely or in combination: prescription drug records, ICD diagnostic codes or laboratory data⁷⁴. The review found that the lowest positive predictive values were found in algorithms relying solely on ICD codes to identify TB in EHR records, whereas those that used TB drug prescriptions in combination with ICD codes or diagnostic results^{67,73} provided improved diagnostic accuracy. To detect patients with LTBI using EHR data, a prior study analyzed

prescription drug patterns for isoniazid, rifampin and rifapentine and validated the LTBI diagnosis with a chart review⁶⁸. This methodology highlighted the sensitivity of utilizing prescription drug records to determine LTBI diagnosis, but concluded that LTBI prevalence may have been underestimated since not all individuals diagnosed with LTBI are recommended for treatment and only 30-60% agree to or initiate treatment^{53,69,70}. An additional study of administrative claims-based EHR data also noted the limitations of relying on prescription data; specifically when relying on rifampin prescriptions as a proxy for LTBI diagnosis since rifampin is used to treat health conditions other than LTBI⁷².

A specific LTBI definition that incorporates additional EHR data elements, other than prescription drug records, may allow for better identification of LTBI diagnosis within EHR data, as well as improved discrimination between TB disease cases versus latent infection cases. In addition, the definition could be applied to estimate LTBI prevalence in any EHR cohort that utilizes data elements common in all EHR systems (e.g., lab results, ICD codes, prescription drug records). While there are likely system-specific complexities, data related to lab tests and results, ICD codes and prescription drug records have proven to be extractable^{68,71,73,76}. Ideally, a robust definition would allow for broad application, without requiring significant work upfront and without having to conduct a resource-intensive chart review as in previous studies^{68,73}.

The burden of LTBI among a higher risk population could be estimated by directly applying a definition of LTBI diagnosis to EHR data from organizations where these individuals seek care, such as safety net clinics. Safety net primary care clinics are especially relevant when trying to target populations at high risk for LTBI since they serve populations at higher risk for TB infection^{18,51,78,79,84-86}. While an LTBI diagnosis definition and subsequent LTBI estimates generated using an EHR data cohort from a network of safety net clinics may not be representative of the general US population, the estimates may be representative of a subpopulation at higher risk for TB infection. As such, successful public health interventions to increase LTBI screening and treatment implemented in these clinical networks would likely be impactful, and their effect could be analyzed. In addition, estimating LTBI prevalence among subpopulations can help identify and subsequently prioritize subpopulations with higher rates of infection. The ability to generate overall and subpopulation estimates of LTBI prevalence at sufficiently short intervals is useful for planning and evaluating public health interventions. Estimates could also be generated pre and post implementation of interventions within a clinical network to improve our ability to detect infection, and thus determine an intervention's impact on LTBI prevalence. For example, for an intervention focused on increasing testing among higher risk populations, the LTBI definition using EHR data elements could be used to detect an increase in LTBI cases diagnosed.

We developed a novel algorithm using EHR data elements to classify individuals in an EHR data cohort from the OCHIN safety net clinical network into hierarchical definitions for LTBI, as well as TB disease, or determine if patients have no TB infection or were not evaluated for TB. Subsequently, we used the algorithm to determine LTBI prevalence estimates in the OCHIN clinical network. We evaluated multiple EHR variables related to diagnostic tests, diagnostic codes, and treatment to define each classification in the algorithm.

Methods

EHR data was extracted for patients with any encounter at an OCHIN member clinic between January 1, 2012, and December 31, 2019. Data extracted included demographics, medical risk factors, lab results related to LTBI diagnostic tests, diagnostic codes (ICD-9 and ICD-10) associated with the patients' visit history or active problem list, and prescription drug records. Data elements contained in the dataset were not restricted based on encounter date and all patient history up until September 2020 (when data was extracted) was included in the data cohort; patient history from prior to 2012 was limited and was not considered inclusive of all patient history.

We used this EHR data cohort from the OCHIN safety net clinical network to develop a novel algorithm aimed at identifying LTBI cases with the highest possible accuracy using EHR data elements. As a consequence of determining LTBI cases, we also defined TB cases within the algorithm, as well as those patients who have no TB infection, or were not evaluated for TB infection. The algorithm assesses TB diagnostic test results, ICD-9 and -10 diagnostic codes and TB or LTBI treatment regimens when determining how to classify a patient within the algorithm. We classified all individuals in the cohort into a hierarchical algorithm with classifications for definite, probable, or possible TB disease or LTBI, no TB infection, or not evaluated for TB infection. The algorithm classifies every possible combination of TB diagnostic test result, TB or LTBI ICD diagnostic code, and TB or LTBI treatment regimen.

Assessing diagnostic test results, ICD codes and treatment regimens

In assessing diagnostic test results, a patient was determined to have a positive result, negative result, no valid result, or the test was not performed. To determine a patient's diagnostic test result, we searched for the results of all IGRA and TST tests performed; the type of test performed, and the longitudinal nature of multiple tests if relevant. For a patient with multiple results, the most recent test with valid results was used to determine an individual's diagnostic result for algorithm classification. Discordant valid results (two tests, one positive and the other negative) from the same day was an unknown result since neither test could be determined to be the most recent. If the patient has no

further valid results on a different day, they were determined to have no valid result. A detailed diagram describing how diagnostic test results was defined using EHR data variables is depicted in Figure 4.1.

A positive diagnostic test alone is not synonymous with an LTBI diagnosis. Rather, it requires excluding TB disease through a symptom review, focused physical exam, and a chest radiograph³⁸. Thus, it is important to identify whether a patient has either a TB or LTBI diagnostic code, or both in their visit diagnosis history or active on the problem list. ICD-9 or ICD-10 codes for LTBI (ICD-9: 795.51, 795.52; ICD-10: R76.11, R76.12, Z22.7) and TB (ICD-9: 010, 011, 012, 013, 014, 015, 016, 017, 018; ICD-10: A15, A16, A17, A18, A19) were identified in patient visit diagnosis history or were determined to be active on the patient problem list at any time in their EHR.

Finally, for each patient it was determined if a TB regimen or LTBI regimen was prescribed. It is important to note that treatment regimens were assessed at the drug level. In looking for regimens prescribed, individual drug prescriptions were identified, as EHR data does not identify treatment regimens in totality. In searching for TB or LTBI regimens, we identified recommended TB and LTBI treatment regimens in patients' records. TB treatment regimens vary in the duration of the regimen, the types of anti-TB drugs prescribed, and the dose and frequency of the drugs¹¹⁴. Regimens also vary based on whether the patient has drug-susceptible or drug resistant TB disease^{114,115}. In searching for drug susceptible TB regimens, we specifically looked for the only recommended regimen at the time of the study for treating adults with drug susceptible TB consisting of isoniazid (INH), rifampin (RIF), pyrazinamide (PZA), and ethambutol (EMB) (referred to hereafter as HRZE)¹¹⁴. Based on the recommendations of TB clinicians, we also searched for six other TB regimens and a single MDR-TB treatment regimen. Finally, we also searched for the following LTBI regimens in the patients records: INH and rifapentine, INH alone, RIF or rifabutin alone, or INH and RIF or rifabutin^{9,10} (Table 4.1).

For the aforementioned regimens where multiple drugs are prescribed, drugs had to be prescribed on the same day. Because treatment regimens for drug susceptible, drug resistant and multidrug resistant TB, and for LTBI vary widely^{10,114,115}, we also searched for all combinations of anti-TB drugs that have been determined to be prescribed on the same day to identify other possible TB or LTBI treatment regimens prescribed among the cohort. In consultation with TB clinicians, it was determined if any of these drug combinations represented a TB or LTBI regimen. If a patient had multiple regimens prescribed on different days, the most recent regimen was used to determine a patient's classification in the algorithm. A list of the anti-TB drugs that we searched for is contained in Appendix D.

A complication in assessing treatment regimens is that rifamycin drugs, such as rifampin and rifabutin, can be used to treat other conditions⁷². Rifampin is used frequently to treat other infectious

diseases, thus a prescription for rifampin alone is not necessarily indicative that a patient is being treated for TB infection or disease. We previously created an algorithm for determining whether a rifampin or rifabutin prescription present in an EHR record was indicative of a LTBI or TB disease treatment regimen based on other drugs prescribed with rifampin or rifabutin, frequency and duration of rifampin or rifabutin prescriptions, and presence of other ICD-10 codes indicating conditions that are treated with rifampin (Chapter 3, Appendix C, Supplemental Methods, Table S1, Table S2). This algorithm does not distinguish between rifampin or rifabutin prescribed for treatment of TB disease versus LTBI. Since rifampin and rifabutin can be used to treat other conditions^{72,107}, this analysis used this previously created algorithm to determine whether these prescriptions were part of a treatment regimen for TB disease or LTBI.

Developing the classification algorithm

In assessing diagnostic test results, ICD codes and treatment regimens for each patient, we defined gold standard definitions to include in the definitions, at minimum, for definite TB disease, definite LTBI, no TB infection and not evaluated for TB infection. If a patient record contained a positive diagnostic test, an ICD code for TB disease only and had been prescribed a TB regimen, they would be classified as definite TB disease, as all EHR data elements indicate TB disease. If a patient record contained a positive diagnostic test, an ICD code for LTBI only, and had been prescribed a LTBI regimen, they would be classified as having definite LTBI. For a patient to be considered to have no TB infection they must have a negative TB diagnostic test, no ICD code for TB nor LTBI, nor any indication of a prescription for a TB or LTBI regimen. A patient with no indication of a diagnostic test being performed, no TB or LTBI ICD, nor a TB or LTBI regimen was classified as having not been evaluated for TB infection.

In classifying patients, we first determined if a patient had a TB or LTBI regimen, and then looked to their ICD codes, and subsequently any diagnostic test results. This methodology placed higher weight to prescribed treatment regimens, rather than ICD codes or diagnostic tests, as recommended in the current literature^{53,67–69,73}. A prescription of an LTBI treatment regimen is a clear indication of diagnosis, whereas there are limitations to relying on diagnostic tests, as well as ICD diagnostic codes for LTBI, as previously described. There is no prior research describing how LTBI ICD codes are utilized in a clinical setting to indicate diagnosis since the current ICD-10 code for LTBI diagnosis (Z22.7) was implemented in the 2020 release of the ICD-10 clinical modification on October 1, 2019⁷⁵. Prior to the introduction of this code, there were only codes to indicate a positive diagnostic test result in the absence of disease (ICD-9: 795.51, 795.52; ICD-10: R76.11, R76.12). For example, R76.11 is a code indicating “a nonspecific reaction to cell mediated immunity measurement of gamma interferon antigen response without active

tuberculosis". Anecdotally, clinicians have indicated that these codes were used to initiate different clinic specific protocols or further diagnostic testing and may not be reliable as a definite LTBI diagnosis. In addition, clinicians and clinics have indicated that code usage may depend on billing and insurance. A secondary outcome of this analysis was to describe the utilization of LTBI ICD codes in the EHR in relation to a patient's diagnostic test results, and LTBI treatment. We determined the number of patients with a positive diagnostic test that do and do not have an LTBI ICD code and determined the number of patients prescribed LTBI treatment that have an ICD code for LTBI.

In determining how to classify each combination of test result, ICD code and recorded treatment regimen, there was significant input from physicians regularly treating patients for LTBI and TB disease. A group of five clinicians who regularly treat patients with TB and LTBI considered each unique combination of treatment regimen, ICD code and diagnostic result in the context of the current literature previously described, as well as their own clinical practice. Once the algorithm was defined, the entire patient cohort was classified into algorithm categories, and we determined the incidence rate of TB disease in the patient cohort and a prevalence estimate of LTBI in the patient cohort, as well as a LTBI prevalence estimate among only those patients with any indication of being evaluated for TB infection in their record (any TB diagnostic test, TB or LTBI ICD code, or TB or LTBI treatment regimen).

Results

Defining the classification algorithm

To define the classification algorithm, we first defined regimen categories made up of the specific regimens in Table 4.1 and TB and LTBI regimens found in the dataset when searching for all drug regimens combinations of anti-TB drugs prescribed on the same day. We defined 7 treatment regimen categories to indicate if a patient had been prescribed 1) HRZE (this includes patients who were prescribed HRZE followed by INH + RIF as this represents the initiation and continuation phase of this TB regimen¹¹⁴), 2) other TB regimen, 3) an MDR-TB regimen, 4) either a TB or a LTBI regimen, 5) INH + rifapentine, 6) other LTBI regimen, or 7) no treatment regimen. TB clinicians used categories 1-6 to define all unique anti-TB drug combinations prescribed on the same day; they were also allowed to define a drug combination as not a TB nor a LTBI regimen. Discordant classifications of regimens among clinicians were discussed, and a final classification was determined. After clinician input, and after review of all the other TB and LTBI regimens present in the data set it was determined that HRZE and INH + rifapentine would be separate treatment categories on their own since HRZE and INH + rifapentine are used exclusively to treat TB disease and LTBI, respectively^{10,114}.

In total, there were 89 other TB regimens in addition to HRZE, and 12 MDR-TB regimens; there were no instances of the single MDR-TB regimen that was specifically searched for in the data cohort (Bedaquiline, Pretomanid, Linezolid). However, this is not unexpected as this regimen was only approved by FDA in August 2019^{121,122} and was unlikely to be represented in the data. There were 8 regimens identified as either a TB or LTBI regimen, and 15 other LTBI regimens, in addition to INH and rifapentine. Regimens defined as other TB regimens, MDR-TB regimens, either TB or LTBI regimens, or other LTBI regimens are listed in Appendix E. Of note, while Levofloxacin only and Moxifloxacin only are anti-TB drugs that are used to treat LTBI in MDR-TB contacts, they were excluded as part of the other LTBI regimens since they are primarily used to treat non-TB infections¹⁰⁷ and MDR-TB rates in the U.S. are extremely low¹⁹.

To classify a patient in the algorithm, a flow chart was followed that looks first for a TB or LTBI treatment regimen, then TB or LTBI ICD code or codes, then TB diagnostic test results in a patient record (Figure 4.2). It was first determined if the patient was prescribed HRZE, if yes it was then determined if they had a TB or LTBI code, or both codes, and then a classification was made. In change to the pre-determined gold standard definitions to indicate TB disease, those with HRZE TB treatment regimen and a TB code were considered a gold standard indication of TB disease and defined as definite TB. Diagnostic test result was not considered for those prescribed HRZE since prescription of that regimen is specific to treatment of TB diagnosis, and a TST or IGRA diagnostic test is not relevant in determining classification. If a patient did not have HRZE prescribed, but was prescribed another TB regimen, then classifications were subsequently determined based on ICD codes present, and then diagnostic test results, or lack thereof. If the patient did not have HRZE or other TB regimen, but an MDR regimen, then classification was made based on that regimen, then ICD codes and then diagnostic test results.

From there if patients had not been classified, but had a LTBI regimen, and specifically INH + rifapentine, they were classified as having definite LTBI, since this regimen is only used to treat LTBI and is a clear indication of diagnosis. Those with another LTBI regimen were then classified based on ICD codes, and subsequently diagnostic test results. Patients prescribed INH + rifapentine (independent of ICD codes and diagnostic test results), or patients prescribed another LTBI regimen and having an LTBI code only and a positive diagnostic test were considered to have clear indication of LTBI and were classified as definite LTBI and the gold standard definition for LTBI diagnosis. This was a slight deviation from the pre-determined gold standard definition for LTBI, as those patients prescribed INH + rifapentine were considered definite LTBI, regardless of ICD codes and diagnostic results. This decision

was made due to limitations of using ICD codes to determine LTBI diagnosis^{67,68,73-75} or diagnostic test results^{36,39,41-43,46-48}

If patients had no TB regimen, then ICD codes and subsequently diagnostic test results were used to make a classification. Finally, patients with no TB or LTBI regimen, no TB or LTBI codes, but a negative diagnostic test result were determined to have had TB infection ruled out. Patients with no TB or LTBI regimen, no TB or LTBI codes, and no diagnostic test were determined to have not been evaluated for TB infection. Ultimately, all combinations TB or LTBI treatment regimen, TB or LTBI ICD codes, and TB infection diagnostic test result were classified in the algorithm (Table 4.2).

In summary, in using the algorithm to determine if patients have LTBI, the following general definitions were used to determine definite, probable, possible LTBI based on treatment regimen, ICD codes and diagnostic test results:

- *Definite LTBI:*
 - 1) Prescribed INH + rifapentine

OR

 - 2) positive diagnostic test + prescribed LTBI treatment or LTBI ICD code only.
- *Probable LTBI:*
 - 1) Prescribed LTBI treatment (other than INH + rifapentine) or LTBI ICD code only but no positive diagnostic test.
- *Possible LTBI:*
 - 1) LTBI treatment (other than INH + Rifapentine) or LTBI ICD code BUT with conflicting evidence (TB ICD code or negative diagnostic test; treatment is designated as either a TB or LTBI regimen)

OR

 - 2) positive diagnostic test only without any other indication of LTBI (i.e., no LTBI treatment or LTBI ICD code).

Classifying the patient cohort according to the algorithm

There were 3,645,477 patients in the 2012-2019 OCHIN patient cohort, and 88.4% (n=3,221,706), were classified as not being evaluated for TB infection in the algorithm, having no evidence of a TB or LTBI treatment regimen, a TB or LTBI code, or a TB infection diagnostic test (Table 4.3). The remaining 11.6% (n=423,771) of the patient cohort were determined to have been evaluated for TB infection based on the presence of a TB diagnostic test, TB or LTBI ICD code, or TB or LTBI drug regimen prescribed in the patient EHR. Among those evaluated, 63.3% (n=268,327) were determined to have TB ruled out based on the presence of negative diagnostic test only; 15.3% (n=65,019) were evaluated for TB based on the

presence of a diagnostic test only but there was no valid result (i.e., borderline/indeterminate, unknown/unread diagnostic results). Thus, 21.3% (n=90,425) of patients evaluated for TB infection had a classification of definite, probable, or possible TB or LTBI. Among those evaluated for TB infection, 24,169 (5.7%) were classified as definite LTBI, 35,966 (8.5%) were classified as probable LTBI, and 18,731 (4.4%) were classified as possible LTBI. Only 429 (0.1%) were classified as definite TB, 1,101 (0.3%) as probable TB, and 10,029 (2.4%) were defined as possible TB.

Using the condition of “definite TB”, the TB incidence rate was 11.8/100,000 persons in the OCHIN population during this time period. The LTBI prevalence estimate was a 0.7% in the entire patient population during this time period when defined as patients with definite LTBI; expanding the definition to also include patients with probable LTBI, the LTBI prevalence estimate increases to 1.6%. Assessing LTBI prevalence amongst only those patients with an indication of being evaluated for TB infection (n=423,771), prevalence increases to 5.7% when defined as patients with definite LTBI; expanding to include those with probable LTBI increases prevalence to 14.2%.

A secondary outcome of this analysis was to describe the utilization of LTBI ICD codes in the EHR in relation to a patient’s diagnostic test results, and LTBI treatment. Overall, of all patients with an LTBI ICD code, 47.3% (n=34,462) had an LTBI ICD code only in their EHR without an indication of a TB or LTBI treatment regimen and no diagnostic test result; 36.0 of patients with an LTBI ICD code had supporting evidence of an LTBI diagnosis in their EHR, such as a LTBI treatment regimen and/or a positive diagnostic test result. However, 5.3% of patients with a LTBI ICD code had conflicting evidence of diagnosis in the form of a TB ICD code and/or a TB treatment regimen; 11.4% had conflicting evidence in the form of a negative diagnostic test only. Assessing patients prescribed a LTBI regimen specifically, among all patients prescribed INH + rifampentine, 83.4% (n=1,340) had a LTBI ICD (and no TB ICD code) in their EHR. Among those prescribed another LTBI regimen, 79.3% (n=9,057) had a LTBI ICD (and no TB ICD code). There were 31,859 patients with a positive test, and 5.4% (n=1,708) had an indication of TB disease in their EHR (TB ICD code, TB treatment regimen). However, among those with no indication of TB disease in their EHR but with a positive test, 75.9% (n=22,873) had an LTBI ICD code.

Discussion

This analysis creates a novel methodology for estimating LTBI prevalence in an EHR cohort, and possibly for estimating broader prevalence estimates among the U.S. general population or higher-risk populations in the U.S. Using the algorithm classification of definite TB, we found a TB incidence rate of 11.8/100,000 persons in the OCHIN population. This is 4 times higher than the reported U.S. TB incidence rate of 2.7/100,000 persons in 2019¹⁸. This estimate indicates that this population is at higher

risk for TB infection than the U.S. general population, which harmonizes with results reported in Chapter 3, where it was determined that 44% of the EHR cohort met recommended TB screening criteria based on CDC and USPSTF screening criteria^{36,55,56}. This rate is representative of the incidence of TB in the OCHIN cohort, since it's assumed that the majority of incident TB cases during this time frame would present for care at their clinic due to being symptomatic. While the LTBI prevalence rate of 0.7% in the cohort was smaller than reported prevalence estimates, this cannot be considered representative of the prevalence in the total cohort since only 11% had an indication of being evaluated for TB infection. Because LTBI is a latent infection and those infected are asymptomatic, they wouldn't present for care for symptoms related to LTBI. As such patients need to be screened to determine risk, and then subsequently tested. Therefore, a large percent of those patients who were not evaluated for TB infection could have LTBI. The LTBI prevalence estimate defined as definite and probable LTBI among those with an indication of being evaluated for infection (14.2%) aligns more closely with other reported estimates for LTBI among the non-U.S.-born being population^{24,25,49,50}.

This definition of LTBI incorporates multiple EHR data elements and allows for better identification of LTBI diagnosis within EHR data, as well as improved discrimination between TB disease cases versus LTBI cases. While the algorithm defined 24,169 patients as having definite LTBI, if the definition of LTBI had been limited to only those with an indication of having an LTBI treatment regimen, only 13,034 patients would have been identified as having LTBI, indicating a portion of patients with LTBI would have been missed. If LTBI ICD code had been the sole indicator of LTBI diagnosis, 16% with conflicting evidence to indicate TB disease or no TB infection would have been included, and possibly misclassified. In addition, in allowing for hierarchical definitions (definite, probable, possible), we were able to define patients with supporting evidence of LTBI diagnosis (e.g., LTBI treatment regimen and ICD code, or ICD code and positive diagnostic test result) as a more definitive case of LTBI. Whereas patients with singular evidence of LTBI or conflicting evidence (e.g., TB code or negative diagnostic test) lower in the hierarchical structure of the algorithm.

LTBI ICD codes may not be a reliable of LTBI diagnosis when relied on solely to determine diagnosis. Of all patients with an LTBI ICD code, 58.7% had an LTBI ICD code only in their EHR without any indication of a TB or LTBI treatment regimen and no diagnostic test result, or an ICD code and a negative diagnostic test only. This could be an indication that LTBI ICD codes are being used to initiate different clinic specific protocols or further diagnostic testing and may not be reliable as a definitive LTBI diagnosis, as it seems that these patients were not ultimately tested or were tested but had negative diagnostic results. In addition, 5.3% of patients had conflicting information in their EHR that may

indicate a TB diagnosis. Thus, if LTBI ICD codes had been used solely to determine LTBI diagnosis in this analysis, it's possible that up to 64% of individuals with an LTBI ICD code may have been misclassified. However, among those patients with an LTBI treatment regimen, 80% had an LTBI ICD code in their record indicating that LTBI treatment regimen, in combination with an LTBI ICD code, is a more reliable indicator of LTBI diagnosis. In addition, among those with a positive diagnostic test result, and without any indication of TB disease, 76% also had an LTBI ICD code; again, indicating that an LTBI ICD code in addition to other confirmatory evidence of LTBI is a better indicator of diagnosis than a LTBI ICD code alone.

This analysis is the first analysis to describe how LTBI ICD codes are utilized in clinical settings by describing diagnostic test results and treatment regimens prescribed for those patients in the cohort with and without an LTBI ICD code. The analysis showed that future analyses should be cautious when relying on an ICD code solely to determine LTBI diagnosis; however, when combined with another indicator of infection, such as LTBI treatment or a positive diagnostic test, an LTBI ICD code may be a more reliable indicator of infection. Future research should look further at the specific use of the current, and more specific, ICD-10 code for LTBI diagnosis (Z22.7) that was implemented in the release of the ICD-10 clinical modifications on October 1, 2019⁷⁵. Due to the EHR cohort end dates, this analysis could not fully determine whether or how this code is being used in clinical setting, or in comparison to other codes that indicate TB infection in the absence of TB disease. Clinical education and increased awareness of this newly available LTBI diagnosis code, as well as appropriate reimbursement for services when this is coded is submitted, may improve adoption of this code in a clinical setting. Subsequently, this would make the code a more reliable indicator for LTBI diagnosis in the analysis of EHR data in the future.

A significant limitation of this algorithm is that it cannot be validated with surveillance data. The only option for validating the algorithm is to perform a chart review to confirm diagnosis of LTBI or TB disease. A chart review was outside the scope of this analysis since a de-identified data cohort from the OCHIN research database was used. In addition, a chart review would be resource intensive, requiring full review of individual patients' EHR by clinicians. A chart review would require an agreement with a specific clinic or smaller network of clinics in the OCHIN network. A more feasible option, and a future direction in this research, is for this algorithm be applied to another EHR cohort in a smaller setting, where a chart review is feasible and can confirm diagnosis of TB disease or LTBI.

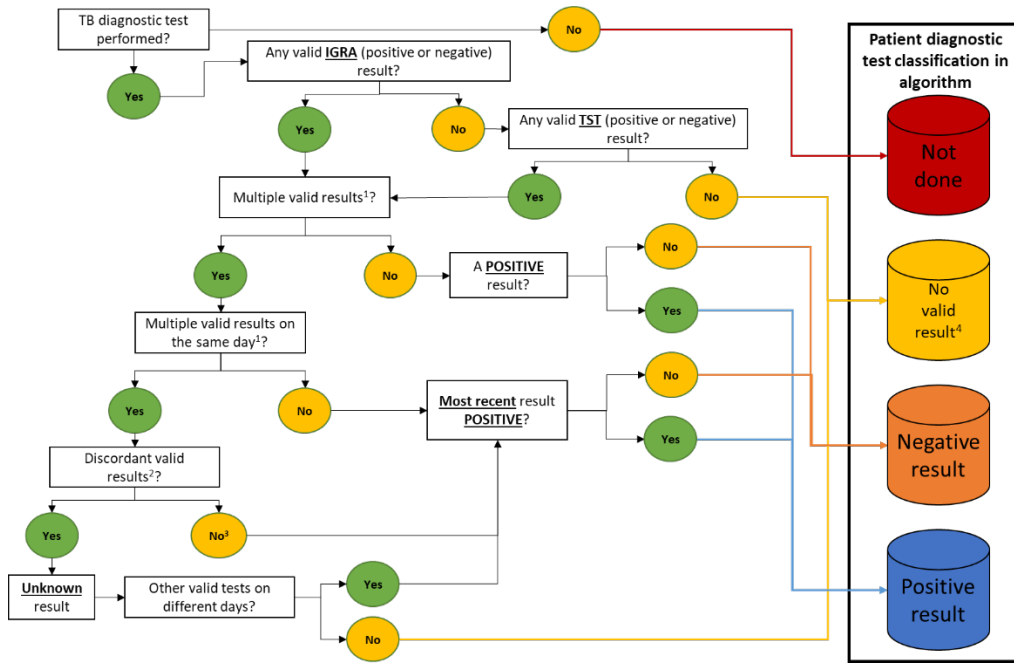
In addition, EHR data contained in unstandardized fields or notes that would be useful for algorithm classification and provide additional specificity in determining TB or LTBI diagnosis, were not

available. Incorporating these EHR data elements into further development of the algorithm would be beneficial and increase specificity and sensitivity of the algorithm. For example, TST results may be indicated outside of lab data and chest radiograph results are typically reported in physician notes. There are also tests specifically for TB diagnosis (e.g., smear, culture) contained in lab results that were not attainable for this analysis. The algorithm also made broad assumptions with regards to how diagnostic testing is performed, ICD codes are entered, and how treatment is prescribed across clinics on the OCHIN clinical network. Clinic specific standard operating procedures may vary widely by clinic. For example, an ICD code for LTBI may be entered for every patient with a positive diagnostic test, regardless of whether they are eventually diagnosed with LTBI or not. The classification of these categories was subjective and based on currently available research and recommendations, as well as input from clinical and epidemiological experts. Additionally, we did not consider the longitudinal nature of the EHR data elements in the algorithm. For example, conflicting evidence in an EHR record related to TB diagnosis and a LTBI diagnosis could represent LTBI that progressed to TB disease over several months or years. Incorporating the longitudinal nature of testing, diagnosis and treatment into this algorithm would also likely improve its performance in classifying patients.

This algorithm lays the foundation for a methodology that can be adapted and subsequently applied to other systems in multiple settings: large network wide EHR cohorts to determine LTBI prevalence, or small clinic populations to better direct resources aimed at increasing LTBI testing and treatment. Separately, this analysis can be utilized, in addition to national cross-sectional surveys and mathematical models based on TB surveillance data, to estimate LTBI prevalence in the U.S. in an effort to try to estimate LTBI prevalence among higher-risk groups. In considering additional analytic research, the TB and LTBI algorithm could be used in machine learning and predictive models applied to large EHR datasets, with the LTBI classification(s) as the outcome in the model. These models could include a myriad of other EHR variables as predictive variables, many not previously researched and frequently unavailable in other data sources, to determine if they are predictive of an LTBI diagnosis. Subsequently leading to the identification of additional, previously unknown, risk factors for TB infection that could help clinicians further improve testing and treatment for LTBI. In addition, it could lead to the development of improved and more robust recommendations for TB screening to target specific populations at higher risk of infection.

Tables and Figures

Figure 4.1. Defining diagnostic test results using electronic health record (EHR) data variables to determine classification in the TB and LTBI algorithm



IGRA=interferon-gamma release assay; TST=tuberculin skin test

¹Same test type

²Defined as a positive and a negative result on the same day

³Must consider multiple results from different days

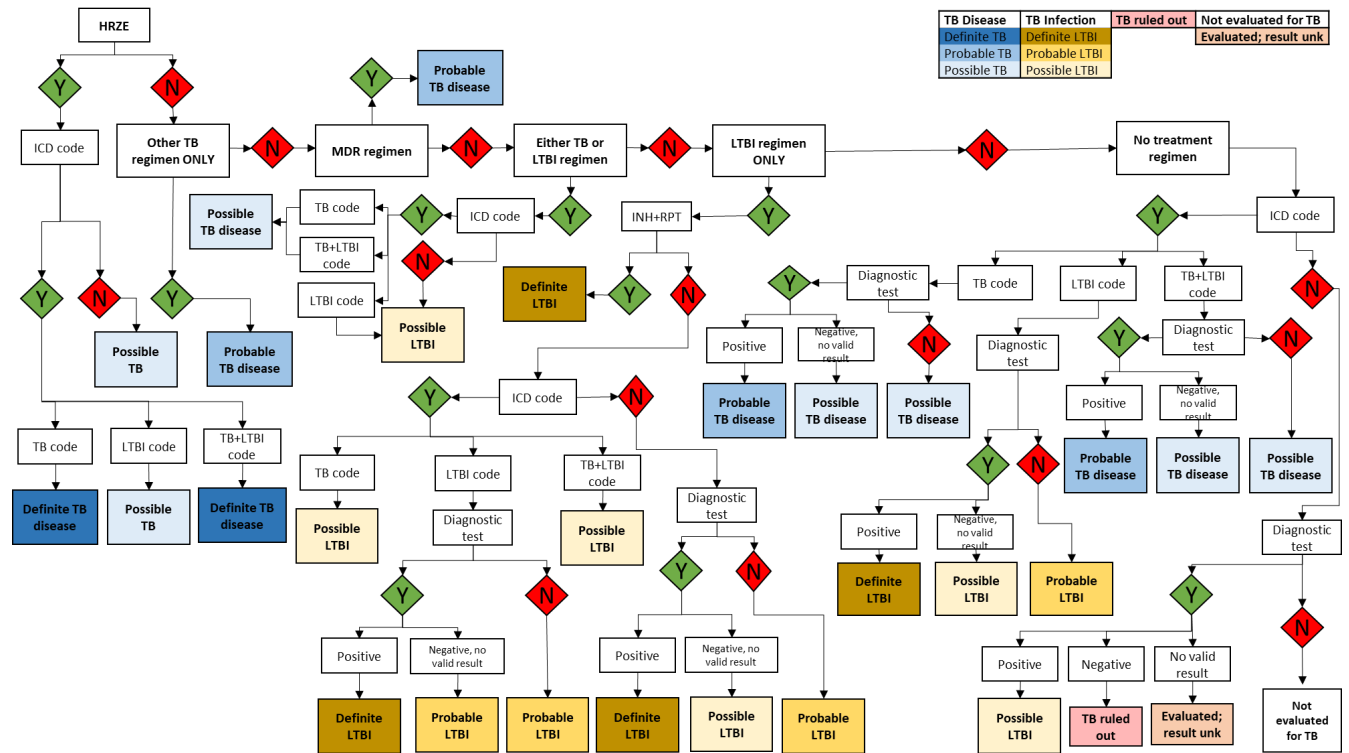
⁴Borderline/indeterminate, unknown/unread TST or IGRA results

Table 4.1. TB disease and LTBI treatment regimens

TB disease treatment regimens	LTBI treatment regimens
Isoniazid, rifampin, pyrazinamide, ethambutol	Isoniazid, rifapentine
Isoniazid, rifampin, pyrazinamide	Isoniazid
Isoniazid, rifampin, pyrazinamide, ethambutol, levofloxacin	Rifampin or rifabutin
Isoniazid, rifabutin, pyrazinamide, ethambutol, levofloxacin	Isoniazid, rifampin or rifabutin
Isoniazid, rifampin, pyrazinamide, ethambutol, rifabutin, levofloxacin	
Isoniazid, rifabutin, pyrazinamide, ethambutol, moxifloxacin	
Rifampin, pyrazinamide, ethambutol, levofloxacin	
Bedaquiline, Pretomanid, Linezolid*	

*MDR-TB treatment regimen; this regimen was approved by FDA in August 2019^{121,122} and is unlikely to be represented in the data.

Figure 4.2. TB and LTBI algorithm flow diagram



TB Disease	TB Infection	TB ruled out	Not evaluated for TB
Definite TB	Definite LTBI		Evaluated; result unk
Probable TB	Probable LTBI		
Possible TB	Possible LTBI		

HRZE= Isoniazid (INH), rifampin or rifabutin, pyrazinamide, ethambutol
 Unk= unknown
 RPT= rifapentine
 LTBI= latent tuberculosis (TB) infection
 ICD= international classification of diseases

Table 4.2. TB and LTBI algorithm classification schema

ICD code	Diagnostic ¹ result	Treatment						
		HRZE ²	Other TB regimen ³	MDR regimen ⁴	Either TB or LTBI regimen ⁵	INH + rifapentine	Other LTBI regimen ⁶	No treatment regimen
TB ⁷	Positive	Definite TB	Probable TB	Probable TB	Possible TB	Definite LTBI	Possible LTBI	Probable TB
	Negative	Definite TB	Probable TB	Probable TB	Possible TB	Definite LTBI	Possible LTBI	Possible TB
	No valid result ⁹	Definite TB	Probable TB	Probable TB	Possible TB	Definite LTBI	Possible LTBI	Possible TB
	Not done	Definite TB	Probable TB	Probable TB	Possible TB	Definite LTBI	Possible LTBI	Possible TB
LTBI ⁸	Positive	Possible TB	Probable TB	Probable TB	Possible LTBI	Definite LTBI	Definite LTBI	Definite LTBI
	Negative	Possible TB	Probable TB	Probable TB	Possible LTBI	Definite LTBI	Probable LTBI	Possible LTBI
	No valid result ⁹	Possible TB	Probable TB	Probable TB	Possible LTBI	Definite LTBI	Probable LTBI	Possible LTBI
	Not done	Possible TB	Probable TB	Probable TB	Possible LTBI	Definite LTBI	Probable LTBI	Probable LTBI
TB ⁷ + LTBI ⁸ code	Positive	Definite TB	Probable TB	Probable TB	Possible TB	Definite LTBI	Possible LTBI	Probable TB
	Negative	Definite TB	Probable TB	Probable TB	Possible TB	Definite LTBI	Possible LTBI	Possible TB
	No valid result ⁹	Definite TB	Probable TB	Probable TB	Possible TB	Definite LTBI	Possible LTBI	Possible TB
	Not done	Definite TB	Probable TB	Probable TB	Possible TB	Definite LTBI	Possible LTBI	Possible TB
No TB or LTBI code	Positive	Possible TB	Probable TB	Probable TB	Possible LTBI	Definite LTBI	Definite LTBI	Possible LTBI
	Negative	Possible TB	Probable TB	Probable TB	Possible LTBI	Definite LTBI	Possible LTBI	TB ruled out
	No valid result ⁹	Possible TB	Probable TB	Probable TB	Possible LTBI	Definite LTBI	Possible LTBI	Evaluated for TB; result unknown
	Not done	Possible TB	Probable TB	Probable TB	Possible LTBI	Definite LTBI	Probable LTBI	Not evaluated for TB

¹QuantIFERON TB test (QFT) or T-SPOT.TB test (T-SPOT), or a tuberculin skin test (TST); diagnostic result determined by using algorithm described in Fig 4.1

²Isoniazid, rifampin or rifabutin, pyrazinamide, ethambutol

³See other TB regimens in Appendix E

⁴See MDR regimens in Appendix E

⁵See either TB or LTBI regimens in Appendix E

⁶See other LTBI regimens in Appendix E

⁷TB Disease codes include ICD-9: 010, 011, 012, 013, 014, 015, 016, 017, 018 and ICD-10: A15, A16, A17, A18, A19

⁸LTBI Infection codes include ICD-9: 795.51, 795.52 and ICD-10: R76.11, R76.12, Z22.7

⁹Includes borderline/indeterminate, unknown/unread diagnostic results

Table 4.3. Classification of OCHIN patient cohort using TB and LTBI algorithm

	Definitions					N	% (of total cohort)	% (of total evaluated for TB ¹⁰ [n=423,771])
	Regimen		ICD Code		Diagnostic result ⁸			
Definite TB	HRZE*		TB ⁶		NA	323	0.01%	0.08%
	HRZE*	AND	TB ⁶ and LTBI ⁷ code		NA	106	0.00%	0.03%
Total						429	0.01%	0.10%
Probable TB	Other TB regimen ¹		NA		NA	122	0.00%	0.03%
	MDR regimen ²		NA		NA	8	0.00%	0.00%
	None	AND	TB ⁶	AND	positive	323	0.01%	0.08%
	None	AND	TB ⁶ and LTBI ⁷ code	AND	positive	648	0.02%	0.15%
Total						1,101	0.03%	0.26%
Possible TB	HRZE*	AND	LTBI ⁷		NA	37	0.00%	0.01%
	HRZE*	AND	None		NA	36	0.00%	0.01%
	Either TB or LTBI regimen ³	AND	TB ⁶		NA	214	0.01%	0.05%
	Either TB or LTBI regimen ³	AND	TB ⁶ and LTBI ⁷ code		NA	108	0.00%	0.03%
	None	AND	TB ⁶		negative, no valid result ⁹ ; not done	7,774	0.21%	1.83%
	None	AND	TB ⁶ and LTBI ⁷ code		negative, no valid result ⁹ ; not done	1,860	0.05%	0.44%
Total						10,029	0.28%	2.37%
Definite LTBI	INH+RPT only		NA		NA	1,606	0.04%	0.38%
	Other LTBI regimen only ⁴	AND	LTBI ⁷	AND	positive	5,891	0.16%	1.39%
	Other LTBI regimen only ⁴	AND	None	AND	positive	168	0.00%	0.04%
	None	AND	LTBI ⁷	AND	positive	16,504	0.45%	3.89%
Total						24,169	0.66%	5.70%
Probable LTBI	Other LTBI regimen only ⁴	AND	LTBI ⁷	AND	negative, no valid result ⁹ ; not done	3,166	0.09%	0.75%
	Other LTBI regimen only ⁴	AND	None	AND	not done	1,004	0.03%	0.24%
	None	AND	LTBI ⁷	AND	not done	31,796	0.87%	7.50%
Total						35,966	0.99%	8.49%
Possible LTBI	Other LTBI regimen only ⁴	AND	TB ⁶		NA	231	0.01%	0.05%
	Other LTBI regimen only ⁴	AND	TB ⁶ and LTBI ⁷ code		NA	717	0.02%	0.17%
	Other LTBI regimen only ⁴	AND	None	AND	negative, no valid result ⁹	251	0.01%	0.06%
	Either TB or LTBI regimen ³	AND	LTBI ⁷ or none		NA	121	0.00%	0.03%
	None	AND	LTBI ⁷	AND	negative, no valid result ⁹	10,306	0.28%	2.43%
	None	AND	None	AND	positive	7,105	0.19%	1.68%
Total						18,731	0.51%	4.42%
TB ruled out	None	AND	None	AND	negative	268,327	7.36%	63.32%
Evaluated for TB; result unknown	None	AND	None	AND	no valid result ⁹	65,019	1.78%	15.34%
Not evaluated for TB	None	AND	None	AND	not done	3,221,706	88.38%	NA
Total						3,645,477		

*Isoniazid, rifampin or rifabutin, pyrazinamide, ethambutol
¹See other TB regimens in Appendix E
²See MDR regimens in Appendix E
³See either TB or LTBI regimens in Appendix E
⁴See other LTBI regimens in Appendix E
⁶TB Disease codes include ICD-9: 010, 011, 012, 013, 014, 015, 016, 017, 018 and ICD-10 A15, A16, A17, A18, A19 (excluding codes for TB Infection)
⁷TB Infection codes include ICD-9: 795.51, 795.52 and ICD-10: R76.11, R76.12, Z22.7 (excluding codes for TB Disease)
⁸QuantiferON TB test (QFT) or T-SPOT.TB test (T-SPOT), or a tuberculin skin test (TST); diagnostic result determined by using algorithm described in Fig 4.1
⁹Includes borderline/indeterminate, unknown/unread diagnostic results
¹⁰Patient with one or more of the following present in patient record: TB diagnostic test, TB or LTBI ICD code, TB or LTBI drug regimen prescribed; i.e. sum of all those classified in the algorithm (excluding those not evaluated for TB)

Appendix D

Anti-tuberculosis drugs used to treat TB disease and LTBI

Isoniazid
Rifampin
Pyrazinamide
Ethambutol
Streptomycin
Rifabutin
Rifapentine
Ethionamide
Amikacin
Capreomycin
Ciprofloxacin
Levofloxacin
Ofloxacin
Moxifloxacin
Cycloserine
Para-Amino Salicylic Acid
Linezolid
Bedaquiline
Delamanid
Clofazimine
Meropenem
Imipenem
Pyridoxine (vitamin B6)
Pretomanid

Appendix E

TB and LTBI treatment regimens identified in the OCHIN EHR cohort, 2012-2019

<i>Other TB regimens</i>	<i>MDR regimens</i>	<i>Either TB or LTBI regimens</i>	<i>Other LTBI regimens</i>
INH, PZA, EMB	Ethionamide Moxifloxacin Linezolid Cycloserine PZA Amikacin	INH, RIF	INH
INH, PZA, RIF	Ethionamide Levofloxacin RIF Moxifloxacin Linezolid PZA	INH, Rifabutin	RIF
INH, PZA, Rifabutin	Aminosallyclic Moxifloxacin Ethionamide Amikacin	INH RPT RIF	Rifabutin
INH, EMB, RIF	Clofazimine Levofloxacin	INH RPT Rifabutin	RIF, PZA
INH, EMB, Rifabutin	Ethionamide Levofloxacin Linezolid Cycloserine Amikacin	Moxifloxacin INH RIF	Rifabutin RIF
Ethionamide Levofloxacin INH RIF Moxifloxacin Linezolid PZA EMB Amikacin	Levofloxacin Linezolid Cycloserine EMB Amikacin	INH Rifabutin RIF	Moxifloxacin EMB
Ethionamide Levofloxacin INH RIF Moxifloxacin Linezolid Cycloserine PZA EMB Amikacin	Levofloxacin Linezolid Cycloserine PZA Amikacin	INH RPT Rifabutin RIF	Levofloxacin EMB
Levofloxacin INH RIF Linezolid PZA EMB Amikacin	Levofloxacin Linezolid EMB Rifabutin	Levofloxacin Moxifloxacin INH	RPT RIF
Levofloxacin INH Rifabutin RIF PZA EMB	Levofloxacin Linezolid Ethionamide		Levofloxacin Moxifloxacin EMB
INH Rifabutin RIF Moxifloxacin PZA EMB	Levofloxacin Linezolid PZA Ethionamide		Ethionamide INH
Levofloxacin INH RIF Linezolid PZA EMB	Levofloxacin Linezolid Meropenem		Levofloxacin, Ethionamide
INH RIF Moxifloxacin Linezolid PZA EMB	Rifabutin Moxifloxacin Linezolid Cycloserine Amikacin		Moxifloxacin, Ethionamide
Levofloxacin INH RIF Moxifloxacin PZA EMB			Levofloxacin, Cycloserine
INH RPT RIF Moxifloxacin PZA EMB			Moxifloxacin, Cycloserine
INH RPT Rifabutin RIF PZA EMB			Levofloxacin, PZA (no longer recommended in 2020 guidelines ¹⁰ due to toxicities)
Levofloxacin INH Rifabutin Moxifloxacin Linezolid EMB			Moxifloxacin, PZA (no longer recommended in 2020 guidelines ¹⁰ due to toxicities)
Ethionamide Levofloxacin INH RIF PZA EMB			PZA, EMB (no longer recommended in 2020 guidelines ¹⁰ due to toxicities)
Levofloxacin INH RIF Moxifloxacin Linezolid PZA EMB Amikacin			
Ethionamide INH RIF Moxifloxacin Linezolid PZA EMB Amikacin			
INH Rifabutin Moxifloxacin Linezolid Cycloserine PZA EMB Amikacin			

Levofloxacin INH RIF Moxifloxacin Linezolid PZA EMB			
INH Rifabutin RIF Moxifloxacin Linezolid PZA EMB			
Levofloxacin INH Rifabutin RIF Linezolid PZA EMB			
Levofloxacin INH RPT RIF Moxifloxacin PZA EMB			
INH RIF Moxifloxacin Linezolid Cycloserine PZA EMB			
Levofloxacin INH Moxifloxacin Linezolid Cycloserine PZA EMB			
Levofloxacin INH RIF			
Levofloxacin EMB Rifabutin			
Levofloxacin INH RPT			
Levofloxacin PZA INH RIF			
Moxifloxacin EMB RIF INH			
PZA INH RPT RIF			
Levofloxacin INH RIF PZA EMB			
INH RPT RIF PZA EMB			
INH RIF Moxifloxacin PZA EMB			
INH Rifabutin RIF PZA EMB			
PZA EMB Rifabutin INH			
Levofloxacin RIF EMB INH			
Levofloxacin PZA EMB RIF			
Levofloxacin INH Rifabutin PZA EMB			
PZA EMB RIF			
INH RIF Linezolid PZA EMB			
Levofloxacin INH Rifabutin RIF EMB			
Ethionamide Levofloxacin RIF PZA EMB			
Levofloxacin INH RIF Linezolid EMB			
Levofloxacin Moxifloxacin EMB RIF			
Levofloxacin RIF Linezolid PZA EMB			
Moxifloxacin PZA EMB RIF			
Moxifloxacin PZA INH RIF			
INH Rifabutin Moxifloxacin PZA EMB			
INH Rifabutin RIF Moxifloxacin EMB			
Levofloxacin EMB Rifabutin INH			
Levofloxacin EMB Rifabutin RIF			
Levofloxacin PZA EMB INH			
Levofloxacin PZA EMB			
Levofloxacin Rifabutin RIF Moxifloxacin EMB			
Levofloxacin Rifabutin RIF PZA EMB			
Linezolid Moxifloxacin EMB Rifabutin			
Moxifloxacin EMB INH			
Levofloxacin Rifabutin			
Levofloxacin EMB RIF			
Levofloxacin INH RPT RIF			
Moxifloxacin EMB RIF			
Moxifloxacin Rifabutin			
Levofloxacin Moxifloxacin RIF			
Moxifloxacin EMB Rifabutin			
Moxifloxacin INH RPT			

Levofloxacin INH Rifabutin			
Levofloxacin Moxifloxacin Rifabutin			
Cycloserine Moxifloxacin RIF			
Levofloxacin Linezolid Rifabutin RIF			
Levofloxacin Rifabutin EMB Amikacin			
Levofloxacin INH			
Moxifloxacin INH			
Levofloxacin Linezolid RIF			
Moxifloxacin PZA EMB INH			
Levofloxacin PZA RIF			
EMB Rifabutin RIF			
Imipenem EMB RIF			
Levofloxacin EMB INH			
Levofloxacin Imipenem RIF			
Levofloxacin Linezolid INH			
Levofloxacin Moxifloxacin INH Rifabutin			
Linezolid INH			
Linezolid Moxifloxacin RIF			
Moxifloxacin INH RPT RIF			
Moxifloxacin PZA RIF			
PZA EMB Rifabutin			
Levofloxacin INH Aminosalicylic			

Chapter 5. Dissertation summary and future directions in research

Summary

This dissertation includes three analyses that address current knowledge gaps in measuring LTBI among populations at higher risk in the U.S. While the total TB case count and TB incidence rate in 2019 were the lowest ever reported, the average annual decline in incidence has slowed over the last several years and it has been determined that the current rate of decline will not achieve the U.S. TB elimination goals¹⁸. Recent statistical modeling has shown that LTBI prevalence in the U.S. must be reduced in order to achieve TB elimination²⁹⁻³¹. Accordingly, the national strategy for TB elimination is shifting to emphasizing LTBI detection and treatment³². However, in the absence of national surveillance, up to date, nationally representative data sources for estimating LTBI prevalence have been limited. As such, it has been difficult to consistently ascertain the resources necessary to reduce the burden of LTBI among populations at higher risk in the U.S. Alternate data sources and novel methodologies must be explored to determine if they may provide updated LTBI prevalence estimates in the U.S., as well as identify gaps in LTBI testing and treatment in settings where populations at higher risk seek care to subsequently develop solutions to overcome barriers.

The first study conducted the first ever systematic review and meta-analysis to determine overall LTBI prevalence estimates for the U.S. general population and non-U.S.-born subpopulation. Currently cited LTBI prevalence estimates in the U.S. for the general population and the non-U.S.-born subpopulation vary and utilize different data sources and methodologies while also representing varying time frames. However, published LTBI prevalence estimates have never been systematically reviewed. A mean estimate of LTBI prevalence in the U.S., especially among populations at higher risk, is critical to know the breadth of infection to subsequently expand targeted testing for TB infection and treatment for LTBI. Using a random effects model, the analysis found an overall mean LTBI prevalence estimate in the U.S. general population of 4.04% (95% CI: 3.35%, 4.87%) and 16.49% (95% CI: 14.70%, 18.50%) in the non-U.S.-born population based on data published from 2000 to 2022. However, significant heterogeneity was found amongst currently published estimates due to differences in data sources, as well as methodologies for determining LTBI prevalence. In addition, this study found that currently cited LTBI prevalence estimates for the U.S. have several limitations and reflect outdated data collected and reported from 5-12 years ago. While this analysis provides overall mean LTBI prevalence estimates for the U.S. using current estimates available in the literature, updated LTBI prevalence estimates using a robust methodology for determining LTBI are needed for the U.S. general population and the non-U.S.-born subpopulation.

The second and third studies utilized a data cohort from the OCHIN EHR research database to describe LTBI among the population seeking care at clinics within the OCHIN safety net clinical network. EHR data are an unexplored data source with the potential to identify gaps in LTBI testing and treatment, as well as for determining LTBI prevalence, among populations at higher risk seeking care at primary care settings. Primary care clinics are important settings for LTBI screening and treatment, and safety net primary care clinics are especially relevant for LTBI since they serve populations known to be at higher risk for TB infection, such as non-U.S.-born persons and persons experiencing homelessness^{18,51,78-86}. They are key settings to engage in efforts to expand TB screening, testing, and treatment. Diagnosis and treatment of LTBI in safety net primary care settings that serve patients at risk for TB may increase uptake of TB prevention efforts and accelerate progress towards TB elimination.

In the second study the OCHIN EHR data cohort was used to measure the LTBI care cascade. Steps measured along the care cascade include patients that met TB screening criteria based on current recommendations, were tested for TB infection, diagnosed with LTBI, and prescribed treatment for LTBI. The analysis also aimed to determine if there were gaps in the LTBI care cascade related to whether testing and treatment recommendations were being implemented in the OCHIN clinical network. The study found that a large proportion of the OCHIN clinical population meets TB screening criteria (43.5%). However, only 21.4% were tested for TB, and only 40.4% were tested using the recommended IGRA test. In addition, among those diagnosed with LTBI, just 29.1% were prescribed LTBI treatment. Among those prescribed treatment only 33.6% were prescribed a recommended rifamycin-based regimen which have been shown to have improved treatment outcomes. This study developed a novel methodology for using EHR data elements to measure the LTBI care cascade. Gaps were identified in LTBI testing and treatment indicating that recommendations, with regards to preferred diagnostic tests and treatment regimens, are not being implemented in this network of safety net clinics. These gaps are in line with previous literature and therefore are not unique to this network. Addressing these gaps may have a direct impact on improving TB prevention in primary care clinics and accelerate progress towards TB elimination in the U.S.

The third study also used the OCHIN EHR data cohort to create a novel algorithm for identifying LTBI among patients using a combination of EHR data elements including diagnostic test results, ICD codes and prescribed treatment regimens. The algorithm classifies the entire OCHIN patient cohort into hierarchical definitions for LTBI and TB disease or defines patients as having no TB infection or having not been evaluated for TB infection. Applying the algorithm to the OCHIN EHR cohort, 11.6% of the patient were determined to have been evaluated for TB infection. The TB incidence rate in the

population was 11.8/100,000 persons during this time period; this is 4 times higher than the reported U.S. TB incidence rate, indicating that this population may be at higher risk for TB infection than the U.S. general population. The LTBI prevalence estimate among those patients evaluated for infection was 14.2%, which aligns closely with reported LTBI prevalence estimates among the non-U.S.-born population^{24,25,49,50}.

This algorithm can be applied in a stepwise process to other EHR data cohorts to determine TB incidence in the population or LTBI prevalence among those evaluated, as well as estimate the proportion of patients evaluated for TB infection. This definition of LTBI incorporates multiple EHR data elements and allows for better identification of LTBI diagnosis within EHR data, as well as improved discrimination between TB disease cases versus LTBI cases. In addition, the algorithm definition of LTBI provides an alternate method for estimating LTBI prevalence within a clinical population or network using readily available EHR data elements.

[Future directions in research](#)

These three studies collectively highlight the gaps in knowledge with regards to estimating the overall burden of LTBI prevalence in the U.S., and in LTBI testing and treatment among populations at higher risk. With an estimated 5 million in the U.S. with LTBI²³, and the majority of TB cases attributed to LTBI reactivation in the U.S.^{21,22}, these gaps inhibit progress towards TB elimination.

The first study highlighted the gaps in currently cited LTBI prevalence estimates in the U.S. Updated and robust estimates are necessary to determine the breadth and resources that need to be dedicated to targeted testing, as well as LTBI treatment completion to reduce the burden of TB infection in the U.S. While this study reports an updated mean prevalence estimate for the U.S. general population and the non-U.S.-born population utilizing current literature, the study found that prevalence estimates cited in current literature have many limitations. Significant heterogeneity was observed between prevalence estimates likely due to the different data sources and methodologies used to calculate prevalence. The majority of the estimates using NHANES data relied on diagnostic test results, that have many known limitations, to determine LTBI. Other estimates back calculated LTBI prevalence based on TB surveillance data; these estimates also apply assumptions with several limitations. Finally, none of these estimates are considered recent, as all NHANES estimates were based on data from 2011-2012, and back-calculated prevalence estimates utilized surveillance data from 2008-2015 and 2013-2017.

For any future iterations of NHANES it is important to consider whether including the TB infection questionnaire and testing component is worthwhile given the resources necessary to include it

in the annual NHANES, as well as the limitations with regards to how sampling is performed for the populations at higher risk for TB infection and the diagnostic test used to determine infection. Given that the population at higher risk for TB infection may not be accurately sampled, nor the diagnostic test used represent true LTBI, maintaining the status quo in the utilization of NHANES for LTBI prevalence estimates should be reconsidered. Resources should be focused on improving sampling techniques and improved testing diagnostics for only populations at higher risk in the U.S. This could include enhancing sampling for certain populations that are at risk or limiting TB related questions and testing to the non-U.S.-born population or other at-risk populations in the U.S., such as those of Asian, American Indian, Native Hawaiian, Pacific Islander, or Alaska Native race. In addition, given its many limitations, the TST should no longer be considered a valid methodology for determining LTBI prevalence estimates, especially among populations at higher risk. The TB testing component of NHANES should be limited to IGRA testing only and possibly be expanded to possibly include other testing to rule out TB disease and confirm LTBI. The questionnaire component of NHANES should consider additional questions; for example, if the individual had received a BCG vaccination which could affect the interpretation of the outcome of the diagnostic test.

Back calculation methods that produce LTBI prevalence estimates are less resource intensive as they don't require complex sampling, nor in person questionnaires and testing. They can be calculated locally, nationally or at the state by any organization that does national TB reporting. Back calculation methods reported in both Haddad et al. and Mirzazadeh et al. apply annual risk rates for progression to active TB disease based on Shea et al. and utilize annual TB incidence based on surveillance data. While back calculated prevalence estimates should obviously utilize updated TB surveillance data, future research should also focus on updating annual risk rates for progression to active disease as the Shea et al. is based on data from over a decade ago. In addition, the accuracy of LTBI prevalence estimates may be improved if annual risk rates for progression to active TB disease are specified by population (e.g., the non-U.S.-born population, individuals with HIV, diabetes, known immunosuppression, etc.) rather than applied uniformly, as the risk of activation to TB disease varies widely based on demographic and medical risk factors⁹⁷.

In the absence of national LTBI surveillance, future research needs to utilize updated data sources that are nationally representative to estimate LTBI prevalence in the U.S. Innovative data sources, as well as novel methodologies for determining infection that do not rely solely on diagnostic testing, are needed for more robust estimates. In addition, future research needs to consider the sustainability of LTBI prevalence estimates and consider if, how frequently, and how easily these

estimates can be updated. In addition, the necessity and burden of calculating prevalence estimates for the U.S. general population should be considered. It may be more feasible and resourceful to focus solely on data sources from populations at higher risk in the U.S. and more accurate methodologies for determining infection. This may mean focusing on more specific data sources, such as data from clinical networks that serve populations at higher risk or focusing on data sources that specifically target populations at higher risk either separately or collectively. Future research should continue to research the feasibility of utilizing the new LTBI ICD-10 code implemented in late 2019 (Z22.7) to estimate LTBI prevalence in the U.S. While the utility of this code is currently unknown and may not be widely used due to its novelty, this also presents an opportunity for implementation of interventions, both clinical and educational based, to improve knowledge and adaption of this specific ICD code. Implementation science strategies can be used to determine which interventions were most effective in increasing utilization of the code and why; these strategies could then be implemented broadly across clinical and/or physician networks as a means to improve utilization of the ICD code. In turn, this code could then be utilized in a proxy method for surveillance of LTBI and determining LTBI prevalence estimates in the U.S.

This dissertation has shown the utility of using EHR data as a novel data source estimating LTBI prevalence estimates in populations at higher risk, and for determining whether and where there are gaps in clinical care for LTBI, specifically related to testing and treatment. EHR data has the potential to modernize surveillance by providing detailed data on large populations in a timely, automated and sustainable manner and has proven successful in other pilot projects for other diseases and chronic conditions^{67,123}. However, exploration of EHR data as a means for determining national LTBI prevalence estimates has not been broached until this dissertation. Future research must continue to utilize EHR data not only as a means for determining LTBI prevalence and identifying gaps in clinical care, but also for determining where and which interventions to improve testing and treatment would be most effective and efficient.

In continuing to utilize EHR data, future research must build on the novel methodology developed in this dissertation for constructing an LTBI care cascade using EHR data elements. Our analysis set a baseline methodology, as well as a baseline for the at-risk population seeking care in safety net clinic settings and how TB screening and diagnostic recommendations are being implemented. Future research should focus on how to further identify high risk populations within an EHR, specifically non-U.S.-born individuals since our methodology was limited by high levels of missingness related to patient country of birth; as such, we likely underestimated the at-risk population

in our LTBI care cascade. In addition, we were unable to determine treatment initiation or treatment completion due to limited prescription refill data. Future research should seek to utilize prescription refill data as a proxy for determining whether patients initiated and subsequently completed treatment.

Research must also take the next step forward to go beyond observational studies and partner with primary care providers, especially those that provide care to populations at higher risk, to develop and implement interventions that effectively address the gaps that this dissertation identified with regards to LTBI testing and treatment. These gaps may be due to lack of physician training or education in TB infection and subsequently not knowing who to screen or treat, or due to limited time during clinical visits. EHR-based interventions should be explored to determine if they can address these gaps. EHR messaging and alerts that fire based on specific screening criteria entered into EHR data fields can be used to notify physicians that patients are considered high risk and a test for infection should be ordered. EHR systems can suggest the test to perform for straightforward ordering or can even bypass the physician and implement standing orders for a diagnostic test that needs only a clinician confirmation. EHR alerts and reminders have led to increased screening for other common conditions, including other infectious conditions such as hepatitis B^{124,125}. Implementation of EHR messaging and alerts may remove barriers to improved screening related to lack of education around screening recommendations, or time limitations of physicians performing clinical duties. EHR-based identification of individuals at higher risk removes the burden of screening patients from the physician and may improve rates of testing. In addition, EHR messaging can also improve reporting, as well as uptake of recommended treatment regimens. For example, an EMR SmartSet can suggest the utilization of the new LTBI ICD code in indicating LTBI infection for relevant patients or suggest recommended drugs to prescribe to initiate a short course regimen for LTBI treatment.

Finally, this dissertation developed a novel algorithm for determining LTBI infection in EHR data using multiple EHR data elements. This algorithm was truly novel in that there are none similar cited in the literature for other diseases (i.e., using multiple EHR data elements), since diagnostic testing is a more reliable marker for other infections or ICD codes for other disease or conditions are more reliably entered. This algorithm can be applied to other EHR data cohorts, and specifically future studies should focus on applying it to an EHR dataset where the algorithm could be validated; specifically, where a chart review could be performed to confirm algorithm classification for patients. In addition, future studies could build on the foundational algorithm by continuing to incorporate other EHR data elements into the algorithm classification. This could include, but is not limited to, additional diagnostic lab results such as culture and smear results that could confirm a TB disease diagnosis or absence of disease, and

chest radiograph results. In addition, incorporating the longitudinal nature of testing, diagnosis and treatment could improve the sensitivity of the algorithm. For example, if a patient had both an LTBI and TB ICD code determining the longitudinal nature of when these codes are entered would improve classification, as a TB code following an LTBI code may be an indication that a latent TB infection may have progressed to TB disease.

The large percent of the OCHIN clinical population that was found to meet screening criteria for TB infection (43%) supports the reported TB incident rate and LTBI prevalence estimates generated using the algorithm definitions. Both estimates were likely to be higher than those cited in the literature for the U.S. general population given the population was at higher risk; this lends credibility to the TB and LTBI classification definitions in the algorithm even given the aforementioned limitations. Expanding upon the current algorithm to include additional data elements would improve up these estimates further. This algorithm can be applied broadly to large clinical network EHR data to determine LTBI prevalence among populations at higher risk, helping to determine where clinical interventions, such as those mentioned previously, could be most effective. In addition, this algorithm could be applied to smaller clinics, such as community health centers, who may need to estimate LTBI prevalence due to a local outbreak among their clinical population or funding deficiencies. The algorithm can be applied easily to an EHR dataset and with far fewer resources required than a chart review. In addition, clinics could tailor the algorithm to fit specific clinical protocols. For example, if they routinely use certain LTBI codes to distinguish between LTBI diagnosis versus an indicator for further testing to rule out TB disease, those nuances could be built into the algorithm to improve accuracy.

Finally, in considering additional research, it is feasible that the TB and LTBI algorithm could be used in machine learning models using large EHR datasets, with the LTBI classification(s) as the outcome in the model. These models could include a myriad of other EHR variables, many not previously researched and frequently unavailable in other data sources, to determine if they are predictive of an LTBI diagnosis. Overall, future research that continues to develop standardized approaches to using EHR data, to both determine LTBI prevalence in specific populations as well as gaps in care, could help advance TB elimination in the U.S. by facilitating partnerships between primary care providers and public health.

References

1. Centers for Disease Control and Prevention (CDC). Basic TB Facts. Published 2016. <https://www.cdc.gov/tb/topic/basics/default.htm>
2. WHO. *Global Tuberculosis Report 2021*.; 2021. <https://www.who.int/publications/i/item/9789240037021>
3. Centers for Disease Control and Prevention (CDC). Extensively Drug-Resistant Tuberculosis (XDR TB). Published 2016. <https://www.cdc.gov/tb/publications/factsheets/drtb/xdrtb.htm>
4. Cohen A, Mathiasen VD, Schön T, Wejse C. The global prevalence of latent tuberculosis: a systematic review and meta-analysis. *Eur Respir J*. 2019;54(3):1900655. doi:10.1183/13993003.00655-2019
5. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med*. Published online 2016. doi:10.1371/journal.pmed.1002152
6. CDC. Latent TB Infection Treatment Maximizing Adherence Fact Sheet. Published 2016. <https://www.cdc.gov/tb/publications/factsheets/treatment/ltbiadherence>
7. Centers for Disease Control (CDC). Targeted tuberculin testing and treatment of latent tuberculosis infection. *MMWR Morb Mortal Wkly Rep*. 2000;49(RR-6). doi:10.1164/ajrccm.161.supplement_3.ats600
8. Centers for Disease Control and Prevention (CDC). Recommendations for use of an isoniazid-rifapentine regimen with direct observation to treat latent Mycobacterium tuberculosis infection. *MMWR Morb Mortal Wkly Rep*. 2011;60(48):1650-1653.
9. Borisov AS, Bamrah Morris S, Njie GJ, et al. Update of Recommendations for Use of Once-Weekly Isoniazid-Rifapentine Regimen to Treat Latent Mycobacterium tuberculosis Infection. *MMWR Morb Mortal Wkly Rep*. Published online 2018. doi:10.15585/mmwr.mm6725a5
10. Sterling TR, Njie G, Zenner D, et al. Guidelines for the Treatment of Latent Tuberculosis Infection: Recommendations from the National Tuberculosis Controllers Association and CDC, 2020. *Am J Transplant*. Published online 2020. doi:10.1111/ajt.15841
11. Haas MK, Aiona K, Erlandson KM, Belknap RW. Higher Completion Rates With Self-administered Once-weekly Isoniazid-rifapentine Versus Daily Rifampin in Adults With Latent Tuberculosis. *Clin Infect Dis*. 2021;73(9):e3459-e3467. doi:10.1093/cid/ciaa1364
12. Njie GJ, Morris SB, Woodruff RY, Moro RN, Vernon AA, Borisov AS. Isoniazid-Rifapentine for Latent Tuberculosis Infection: A Systematic Review and Meta-analysis. *Am J Prev Med*. 2018;55(2):244-252. doi:https://doi.org/10.1016/j.amepre.2018.04.030
13. Centers for Disease Control and Prevention (CDC). Tuberculosis.
14. World Health Organization. *Latent Tuberculosis Infection Updated and Consolidated Guidelines for Programmatic Management*. Vol 38.; 2019.
15. Wejse C. Tuberculosis elimination in the post Millennium Development Goals era. *Int J Infect Dis*. 2015;32:152-155. doi:https://doi.org/10.1016/j.ijid.2014.11.020

16. Churchyard GJ, Swindells S. Controlling latent TB tuberculosis infection in high-burden countries: A neglected strategy to end TB. *PLOS Med*. 2019;16(4):e1002787.
17. Centers for Disease Control and Prevention (CDC). A Strategic Plan for the Elimination of Tuberculosis in the United States. *MMWR Suppl*. 1989;38(S-3):1-25. <https://www.cdc.gov/MMWR/preview/MMWRhtml/00001375.htm>
18. Schwartz NG, Price SF, Pratt RH, Langer AJ. Tuberculosis — United States, 2019. *MMWR Morb Mortal Wkly Rep*. Published online 2020. doi:10.15585/mmwr.mm6911a3
19. Centers for Disease Control and Prevention (CDC). *Reported Tuberculosis in the United States, 2019.*; 2020.
20. Filardo TD, Feng PJ, Pratt RH, Price SF, Self JL. Tuberculosis - United States, 2021. *MMWR Morb Mortal Wkly Rep*. 2022;71(12):441-446. doi:10.15585/mmwr.mm7112a1
21. Shea KM, Kammerer JS, Winston CA, Navin TR, Horsburgh CR. Estimated rate of reactivation of latent tuberculosis infection in the United States, overall and by population subgroup. *Am J Epidemiol*. Published online 2014. doi:10.1093/aje/kwt246
22. Yuen CM, Kammerer JS, Marks K, Navin TR, France AM. Recent Transmission of Tuberculosis — United States, 2011–2014. *PLoS One*. 2016;11(4):e0153728. doi:10.1371/journal.pone.0153728
23. Miramontes R, Hill AN, Woodruff RSY, et al. Tuberculosis Infection in the United States : Prevalence Estimates from the National Health and Nutrition Examination Survey, 2011-2012. 2015;128:2011-2012. doi:10.1371/journal.pone.0140881
24. Yelk Woodruff R, Hill A, Marks S, Navin T, Miramontes R. Estimated Latent Tuberculosis Infection Prevalence and Tuberculosis Reactivation Rates Among Non-U.S.-Born Residents in the United States, from the 2011-2012 National Health and Nutrition Examination Survey. *J Immigr Minor Heal*. 2021;23(4):806-812. doi:10.1007/s10903-020-01065-8
25. Vonnahme LA, Haddad MB, Navin TR. Factoring prior treatment into tuberculosis infection prevalence estimates, United States, 2011-2012. *Emerg Infect Dis*. Published online 2019. doi:10.3201/eid2510.190439
26. Collins JM, Stout JE, Ayers T, et al. Prevalence of Latent Tuberculosis Infection Among Non-US-Born Persons by Country of Birth-United States, 2012-2017. *Clin Infect Dis an Off Publ Infect Dis Soc Am*. 2021;73(9):e3468-e3475. doi:10.1093/cid/ciaa1662
27. Yelk Woodruff RS, Pratt RH, Armstrong LR. The US National Tuberculosis Surveillance System: A Descriptive Assessment of the Completeness and Consistency of Data Reported from 2008 to 2012. *JMIR Public Heal Surveill*. 2015;1(2):e15. doi:10.2196/publichealth.4991
28. Centers for Disease Control and Prevention (CDC). Prevention and control of tuberculosis in U.S. communities with at-risk minority populations. Recommendations of the Advisory Council for the Elimination of Tuberculosis. *MMWR Recomm reports Morb Mortal Wkly report Recomm reports*. 1992;41(RR-5):1—11. <http://europepmc.org/abstract/MED/1314322>
29. Hill AN, Becerra JE, Castro KG. Modelling tuberculosis trends in the USA. *Epidemiol Infect*. 2012;140(10):1862-1872. doi:10.1017/S095026881100286X
30. Menzies NA, Cohen T, Hill AN, et al. Prospects for tuberculosis elimination in the United States:

- Results of a transmission dynamic model. *Am J Epidemiol*. Published online 2018. doi:10.1093/aje/kwy094
31. Shrestha S, Hill AN, Marks SM, Dowdy DW. Comparing Drivers and Dynamics of Tuberculosis in California, Florida, New York, and Texas. *Am J Respir Crit Care Med*. 2017;196(8):1050-1059. doi:10.1164/rccm.201702-0377OC
 32. LoBue PA, Mermin JH. Latent tuberculosis infection: the final frontier of tuberculosis elimination in the USA. *Lancet Infect Dis*. Published online 2017. doi:10.1016/S1473-3099(17)30248-7
 33. Centers for Disease Control and Prevention (CDC). NHANES 2011-2012 Overview.
 34. Grieco E, Acosta Y, de la Cruz G, et al. *The Foreign-Born Population in the United States: 2010.; 2012*. <https://www2.census.gov/library/publications/2012/acs/acs-19.pdf>
 35. Haddad MB. *Estimating the Changing Prevalence of Tuberculosis Infection in the United States, 1971-2015*. Emory University; 2019. <https://etd.library.emory.edu/concern/etds/0k225b954?locale=en>
 36. Lewinsohn DM, Leonard MK, Lobue PA, 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*. Published online 2017. doi:10.1093/cid/ciw694
 37. CDC. Latent TB Infection Testing and Treatment : Summary of U.S. Recommendations. Published 2018. Accessed August 9, 2019. <https://www.cdc.gov/tb/publications/ltbi/pdf/CDC-USPSTF-LTBI-Testing-Treatment-Recommendations-508.pdf>
 38. Council of State and Territorial Epidemiologists (CSTE). *Public Health Reporting and National Notification for Tuberculosis: CSTE Position Statement 09-ID-65*. <https://cdn.ymaws.com/www.cste.org/resource/resmgr/PS/09-ID-65.pdf>
 39. Mazurek GH, Jereb J, Vernon A, et al. Updated guidelines for using Interferon Gamma Release Assays to detect Mycobacterium tuberculosis infection - United States, 2010. *MMWR Recomm Reports*. Published online 2010.
 40. Farhat M, Greenway C, Pai M, Menzies D. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria? *Int Union Against Tuberc Lung Dis*. 2006;10(11):1192-1204.
 41. CDC. BCG Vaccine Fact Sheet. Published 2016. <https://www.cdc.gov/tb/publications/factsheets/prevention/bcg.htm>
 42. CDC. The role of BCG vaccine in the prevention and control of tuberculosis in the United States: a joint statement by ACET and the Advisory Committee on Immunization Practices. *MMWR*. 1996;59(No. RR-5).
 43. CDC. Development of new vaccines for tuberculosis: recommendations of the Advisory Council for the Elimination of Tuberculosis (ACET). *MMWR*. 1998;47(No. RR-13).
 44. Vozoris NT, Batt J. Change in the Prevalence of Testing for Latent Tuberculosis Infection in the United States: 1999-2012. *Can Respir J*. Published online 2016. doi:10.1155/2016/1850879
 45. Kimberlin DB, Jackson M. *Red Book: 2018 Report of the Committee on Infectious Diseases*.

- American Academy of Pediatrics; 2018.
46. Menzies D. Interpretation of Repeated Tuberculin Tests. *Am J Respir Crit Care Med*. 1999;159(1):15-21. doi:10.1164/ajrccm.159.1.9801120
 47. Chee CBE, Khinmar KW, Gan SH, et al. Tuberculosis treatment effect on T-cell interferon- γ responses to Mycobacterium tuberculosis specific antigens. *Eur Respir J*. 2010;36(2):355 LP - 361.
 48. Pollock NR, Kashino SS, Napolitano DR, et al. Evaluation of the effect of treatment of latent tuberculosis infection on QuantiFERON-TB gold assay results. *Infect Control Hosp Epidemiol*. 2009;30(4):392-395. doi:10.1086/596606
 49. Haddad MB, Raz KM, Lash TL, et al. Simple estimates for local prevalence of latent tuberculosis infection, United States, 2011–2015. *Emerg Infect Dis*. Published online 2018. doi:10.3201/eid2410.180716
 50. Mirzazadeh A, Kahn JG, Haddad MB, et al. State-level prevalence estimates of latent tuberculosis infection in the United States by medical risk factors, demographic characteristics and nativity. *PLoS One*. 2021;16(4):e0249012. doi:10.1371/journal.pone.0249012
 51. Tsang CA, Langer AJ, Steve Kammerer J, Navin TR. US tuberculosis rates among persons born outside the United States compared with rates in their countries of birth, 2012-2016. *Emerg Infect Dis*. Published online 2020. doi:10.3201/eid2603.190974
 52. Wilson JW, Kissner DG, Escalante P. Cascade of Care in the Management of Latent Tuberculosis Infection in the United States: A Lot to Improve and to Scale Up. *Ann Am Thorac Soc*. 2021;18(10):1620-1621. doi:10.1513/AnnalsATS.202106-722ED
 53. 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*. Published online 2016. doi:10.1016/S1473-3099(16)30216-X
 54. Holzman SB, Perry A, Saleeb P, et al. Evaluation of the Latent Tuberculosis Care Cascade Among Public Health Clinics in the United States. *Clin Infect Dis*. Published online April 1, 2022:ciac248. doi:10.1093/cid/ciac248
 55. US Preventive Services Task Force. Screening for Latent Tuberculosis Infection in Adults: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2016;316(9):962-969. doi:10.1001/jama.2016.11046
 56. Centers for Disease Control and Prevention (CDC). *Latent Tuberculosis Infection: A Guide for Primary Health Care Providers.*; 2020. <https://www.cdc.gov/tb/publications/ltbi/pdf/LTBIbooklet508.pdf>
 57. Kahwati LC, Feltner C, Halpern M, et al. Primary Care Screening and Treatment for Latent Tuberculosis Infection in Adults: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. 2016;316(9):970-983. doi:10.1001/jama.2016.10357
 58. Centers for Disease Control and Prevention (CDC). Tuberculosis associated with blocking agents against tumor necrosis factor- α --California, 2002-2003. *MMWR Morb Mortal Wkly Rep*. 2004;53(30):683-686.
 59. Wallis RS. Mathematical modeling of the cause of tuberculosis during tumor necrosis factor

- blockade. *Arthritis Rheum.* 2008;58(4):947-952. doi:10.1002/art.23285
60. Parriott A, Kahn JG, Ashki H, et al. Modeling the Impact of Recommendations for Primary Care-Based Screening for Latent Tuberculosis Infection in California. *Public Health Rep.* 2020;135(1_suppl):172S-181S. doi:10.1177/0033354920927845
 61. Tasillo A, Salomon JA, Trikalinos TA, Horsburgh CRJ, Marks SM, Linas BP. Cost-effectiveness of Testing and Treatment for Latent Tuberculosis Infection in Residents Born Outside the United States With and Without Medical Comorbidities in a Simulation Model. *JAMA Intern Med.* 2017;177(12):1755-1764. doi:10.1001/jamainternmed.2017.3941
 62. Mancuso JD, Diffenderfer JM, Ghassemieh BJ, Horne DJ, Kao TC. The prevalence of latent tuberculosis infection in the United States. *Am J Respir Crit Care Med.* Published online 2016. doi:10.1164/rccm.201508-1683OC
 63. Sullivan K, Pease C, Zwerling A, et al. Seven-year retrospective study understanding the latent TB infection treatment cascade of care among adults in a low incidence country. *BMC Public Health.* 2021;21(1):964. doi:10.1186/s12889-021-10733-9
 64. Kruse CS, Stein A, Thomas H, Kaur H. The use of Electronic Health Records to Support Population Health: A Systematic Review of the Literature. *J Med Syst.* 2018;42(11):214. doi:10.1007/s10916-018-1075-6
 65. U.S. Food and Drug Administration. *Use of Electronic Health Record Data in Clinical Investigations Guidance for Industry.*; 2018.
 66. National Institutes for Health. EHR Data. <https://www.nlm.gov/guides/data-glossary/ehr-data>
 67. Birkhead GS, Klompas M, Shah NR. Uses of Electronic Health Records for Public Health Surveillance to Advance Public Health. *Annu Rev Public Health.* Published online 2015. doi:10.1146/annurev-publhealth-031914-122747
 68. Jenks J, Garfein R, Zhu W, Hogarth M. Latent Tuberculosis Screening Using Electronic Health Record Data. *Emerg Infect Dis J.* 2020;26(9):2285. doi:10.3201/eid2609.191391
 69. Gershon AS, McGeer A, Bayoumi AM, Raboud J, Yang J. Health care workers and the initiation of treatment for latent tuberculosis infection. *Clin Infect Dis an Off Publ Infect Dis Soc Am.* 2004;39(5):667-672. doi:10.1086/422995
 70. Colson PW, Hirsch-Moverman Y, Bethel J, et al. Acceptance of treatment for latent tuberculosis infection: prospective cohort study in the United States and Canada. *Int J Tuberc Lung Dis Off J Int Union against Tuberc Lung Dis.* 2013;17(4):473-479. doi:10.5588/ijtld.12.0697
 71. Walker WL, Schmit KM, Welch EC, et al. Using the Food and Drug Administration Sentinel System for surveillance of TB infection. *Int J Tuberc Lung Dis.* 2022;26(12):1170-1176. doi:10.5588/ijtld.22.0259
 72. Prescribers' Digital Reference (PDR): Rifampin - Drug Summary. Published 2022. <https://www.pdr.net/drug-summary/Rifadin-rifampin-1036>
 73. Calderwood MS, Platt R, Hou X, et al. Real-time surveillance for tuberculosis using electronic health record data from an ambulatory practice in eastern Massachusetts. *Public Health Rep.* 2010;125(6):843-850. doi:10.1177/003335491012500611

74. Ronald LA, Ling DI, FitzGerald JM, et al. Validated methods for identifying tuberculosis patients in health administrative databases: systematic review. *Int J Tuberc Lung Dis Off J Int Union against Tuberc Lung Dis*. 2017;21(5):517-522. doi:10.5588/ijtld.16.0588
75. CDC. ICD-10-CM Codes for Tuberculosis (TB). Published 2019. <https://www.cdc.gov/tb/education/ICD-10.html>
76. Todd J, Puro J, Jones M, Oakley J, Vonnahme L, Ayers T. *Using Electronic Health Records to Describe TB in Community Health Settings: A Cohort Analysis in a Large Safety-Net Population.*; 2020.
77. Oregon Health Authority. Removing Barriers to Health Care Services. <https://www.oregon.gov/oha/HPA/HP-PCO/Pages/Safety-Net-Clinics.aspx>
78. National Association of Community Health Centers. *Community Health Center Chartbook.*; 2020. <https://www.nachc.org/wp-content/uploads/2020/01/Chartbook-2020-Final.pdf>
79. National Association of Community Health Centers. *America's Health Centers: 2022 Snapshot.*; 2022. <https://www.nachc.org/research-and-data/americas-health-centers-2022-snapshot/>
80. Carrasquillo O, Carrasquillo AI, Shea S. Health insurance coverage of immigrants living in the United States: differences by citizenship status and country of origin. *Am J Public Health*. 2000;90(6):917-923. doi:10.2105/ajph.90.6.917
81. Ku L, Matani S. Left out: immigrants' access to health care and insurance. *Health Aff (Millwood)*. 2001;20(1):247-256. doi:10.1377/hlthaff.20.1.247
82. Morales LS, Lara M, Kington RS, Valdez RO, Escarce JJ. Socioeconomic, cultural, and behavioral factors affecting Hispanic health outcomes. *J Health Care Poor Underserved*. 2002;13(4):477-503. doi:10.1177/104920802237532
83. Thamer M, Richard C, Casebeer AW, Ray NF. Health insurance coverage among foreign-born US residents: the impact of race, ethnicity, and length of residence. *Am J Public Health*. 1997;87(1):96-102. doi:10.2105/ajph.87.1.96
84. Bamrah S, Yelk Woodruff RS, Powell K, Ghosh S, Kammerer JS, Haddad MB. Tuberculosis among the homeless, United States, 1994-2010. *Int J Tuberc Lung Dis Off J Int Union against Tuberc Lung Dis*. 2013;17(11):1414-1419. doi:10.5588/ijtld.13.0270
85. Self JL, McDaniel CJ, Bamrah Morris S, Silk BJ. Estimating and Evaluating Tuberculosis Incidence Rates Among People Experiencing Homelessness, United States, 2007-2016. *Med Care*. 2021;59(Suppl 2):S175-S181. doi:10.1097/MLR.0000000000001466
86. Moss AR, Hahn JA, Tulskey JP, Daley CL, Small PM, Hopewell PC. Tuberculosis in the homeless. A prospective study. *Am J Respir Crit Care Med*. 2000;162(2 Pt 1):460-464. doi:10.1164/ajrccm.162.2.9910055
87. Readhead A, Barry P. Is language at medical visit or at home a good proxy for country of birth? In: *2019 National TB Conference.* ; 2019.
88. Readhead A, Flood J, Barry P. Health insurance, healthcare utilization and language use among populations who experience risk for tuberculosis, California 2014–2017. *PLoS One*. 2022;17(5):e0268739. <https://doi.org/10.1371/journal.pone.0268739>

89. OCHIN. *2021 Annual Report.*; 2022. <https://ochin.org/2021-annual-report>
90. OCHIN, Inc. <https://ochin.org/>
91. OCHIN. OCHIN: Overview. Accessed October 5, 2022. <https://static1.squarespace.com/static/5ade0eb85cfd79247926399a/t/61bb9db8963a180a03cd4968/1639685561178/OCHINOverview211216.pdf>
92. DeVoe JE, Gold R, Spofford M, et al. Developing a network of community health centers with a common electronic health record: Description of the Safety Net West Practice-based Research Network (SNW-PBRN). *J Am Board Fam Med*. Published online 2011. doi:10.3122/jabfm.2011.05.110052
93. OCHIN. *A Short Guide to the ADVANCE Clinical Research Network (CRN).*; 2022.
94. Miramontes R, Hill AN, Yelk Woodruff RS, et al. Tuberculosis Infection in the United States: Prevalence Estimates from the National Health and Nutrition Examination Survey, 2011-2012. *PLoS One*. 2015;10(11):e0140881. doi:10.1371/journal.pone.0140881
95. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372. doi:10.1136/bmj.n71
96. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw*. 2010;36(3 SE-Articles):1-48. doi:10.18637/jss.v036.i03
97. Yeats J. *Controlling Tuberculosis Among High Risk Populations in Los Angeles: Three Essays*. RAND Corporation PP - Santa Monica, CA; 2015. https://www.rand.org/pubs/rgs_dissertations/RGSD356.html
98. France AM, Grant J, Kammerer JS, Navin TR. A Field-Validated Approach Using Surveillance and Genotyping Data to Estimate Tuberculosis Attributable to Recent Transmission in the United States. *Am J Epidemiol*. Published online 2015. doi:10.1093/aje/kwv121
99. National Center for Health Statistics. *National Health and Nutrition Examination Survey: Analytic Guidelines, 2011-2012.*; 2013. https://www.nchs.gov/nchs/data/nhanes/2011-2012/analytic_guidelines_11_12.pdf.
100. Woodruff R, Miramontes R. Tuberculosis Infection among Non-US-Born Persons and Persons ≥ 60 Years of Age, United States, 2019–2020. *Emerg Infect Dis J*. 2023;29(7):1470. doi:10.3201/eid2907.230324
101. Haddad MB, Lash TL, Hill AN, et al. Robustness of NHANES Estimates of the US Prevalence of a Positive Tuberculin Skin Test. *Epidemiology*. 2020;31(2):248-258. doi:10.1097/EDE.0000000000001141
102. Centers for Disease Control and Prevention (CDC). *Reported Tuberculosis in the United States, 2021.*; 2022. <https://www.cdc.gov/tb/statistics/reports/2021/default.htm>
103. Institute of Medicine (US). *Americas's Health Care Safety Net: Intact but Endangered*. (Ein Lewin M, Altman S, eds.); 2000. doi:10.17226/9612
104. Arsenault C, Roder-DeWan S, Kruk ME. Measuring and improving the quality of tuberculosis care: A framework and implications from the Lancet Global Health Commission. *J Clin Tuberc Other Mycobact Dis*. 2019;16:100112. doi:<https://doi.org/10.1016/j.jctube.2019.100112>

105. Kruk ME, Gage AD, Arsenault C, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. *Lancet Glob Heal*. 2018;6(11):e1196-e1252. doi:[https://doi.org/10.1016/S2214-109X\(18\)30386-3](https://doi.org/10.1016/S2214-109X(18)30386-3)
106. California Department of Health. *California Adult Tuberculosis Risk Assessment California Adult TB Risk Assessment User Guide.*; 2018. [https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH/Document Library/TBCB-CA-TB-Risk-Assessment-and-Fact-Sheet.pdf](https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH/Document%20Library/TBCB-CA-TB-Risk-Assessment-and-Fact-Sheet.pdf)
107. Prescribers' Digital Reference (PDR). <https://www.pdr.net/>
108. Stockbridge EL, Loethen AD, Annan E, Miller TL. Interferon gamma release assay tests are associated with persistence and completion of latent tuberculosis infection treatment in the United States: Evidence from commercial insurance data. *PLoS One*. 2020;15(12):e0243102. doi:10.1371/journal.pone.0243102
109. Stockbridge EL, Miller TL, Carlson EK, Ho C. Tuberculosis Prevention in the Private Sector: Using Claims-Based Methods to Identify and Evaluate Latent Tuberculosis Infection Treatment With Isoniazid Among the Commercially Insured. *J Public Health Manag Pract*. 2018;24(4):E25-E33. doi:10.1097/PHH.0000000000000628
110. Stockbridge EL, Miller TL, Carlson EK, Ho C. Predictors of latent tuberculosis infection treatment completion in the US private sector: an analysis of administrative claims data. *BMC Public Health*. 2018;18(1):662. doi:10.1186/s12889-018-5578-3
111. Centers for Disease Control and Prevention (CDC). Interferon-Gamma Release Assays (IGRAs) – Blood Tests for TB Infection. Published 2016. <https://www.cdc.gov/tb/publications/factsheets/testing/igra.htm>
112. Migration Policy Institute. Language Diversity and English Proficiency in the United States. *Migr Inf Source*. Published online 2016. <https://www.migrationpolicy.org/article/language-diversity-and-english-proficiency-united-states-2015>
113. Centers for Disease Control and Prevention (CDC). CDC guidelines for treatment regimens for LTBI. <https://www.cdc.gov/tb/topic/treatment/ltbi.htm>
114. Nahid P, Dorman SE, Alipanah N, et al. Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. *Clin Infect Dis an Off Publ Infect Dis Soc Am*. 2016;63(7):e147-e195. doi:10.1093/cid/ciw376
115. Nahid P, Mase SR, Migliori GB, et al. Treatment of Drug-Resistant Tuberculosis. An Official ATS/CDC/ERS/IDSA Clinical Practice Guideline. *Am J Respir Crit Care Med*. 2019;200(10):e93-e142. doi:10.1164/rccm.201909-1874ST
116. Carr W, Kurbatova E, Starks A, Goswami N, Allen L, CA W. Interim Guidance: 4-Month Rifapentine-Moxifloxacin Regimen for the Treatment of Drug-Susceptible Pulmonary Tuberculosis — United States, 2022. *MMWR Morb Mortal Wkly Rep*. 2022;71:285-289. doi:[http://dx.doi.org/10.15585/mmwr.mm7108a1external icon](http://dx.doi.org/10.15585/mmwr.mm7108a1external%20icon)
117. Centers for Disease Control and Prevention (CDC). Treatment for TB Disease. Published 2022. <https://www.cdc.gov/tb/topic/treatment/tbdisease.htm>
118. Centers for Disease Control and Prevention (CDC). Provisional CDC Guidance for the Use of

- Pretomanid as part of a Regimen [Bedaquiline, Pretomanid, and Linezolid (BPAL)] to Treat Drug-Resistant Tuberculosis Disease. Published 2022.
<https://www.cdc.gov/tb/topic/drtb/bpal/default.htm>
119. Curry International Tuberculosis Center and California Department of Health. *Drug-Resistant Tuberculosis: A Survival Guide for Clinicians, 3rd Edition/2022 Updates.*; 2022.
<https://www.currytbcenter.ucsf.edu/products/view/drug-resistant-tuberculosis-survival-guide-clinicians-3rd-edition>
 120. Lexicomp. <https://online.lexi.com/>
 121. Conradie F, Bagdasaryan TR, Borisov S, et al. Bedaquiline–Pretomanid–Linezolid Regimens for Drug-Resistant Tuberculosis. *N Engl J Med.* 2022;387(9):810-823. doi:10.1056/NEJMoa2119430
 122. CDC. *Provisional CDC Guidance for the Use of Pretomanid as Part of a Regimen [Bedaquiline, Pretomanid, and Linezolid (BPAL)] to Treat Drug-Resistant Tuberculosis Disease.*; 2023.
<https://www.cdc.gov/tb/topic/drtb/bpal/default.htm>
 123. Hohman KH, Martinez AK, Klompas M, et al. Leveraging Electronic Health Record Data for Timely Chronic Disease Surveillance: The Multi-State EHR-Based Network for Disease Surveillance. *J Public Health Manag Pract.* 2023;29(2):162-173. doi:10.1097/PHH.0000000000001693
 124. Chak E, Taefi A, Li CS, et al. Electronic alerts increase screening for chronic hepatitis B: A randomized, double-blind, controlled trial. *Hepatology.* 2017;27(11):1352-1357.
doi:10.1158/1055-9965.EPI-18-0448
 125. Hsu L, Bowlus CL, Stewart SL, et al. Electronic messages increase hepatitis B screening in At-risk asian american patients: A randomized, controlled trial. *Dig Dis Sci.* Published online 2013.
doi:10.1007/s10620-012-2396-9