

Closing the U.S. Tuberculosis Care Gap: Artificial Intelligence, Video-Observed Therapy, and Specialist Workforce Alignment

Shreya Parimoo*

Submitted: 21 August 2026 Accepted: 10 September 2026 Publication date: 28 September 2026

DOI: 10.70671/3dr6ax32

Abstract: Tuberculosis (TB) remains a persistent, low-incidence, high-resource public health problem in the United States, an epidemiological profile that differs fundamentally from the high-burden settings where most TB technology was originally developed. This paper synthesizes three linked lines of evidence bearing on how the U.S. TB care cascade can be strengthened: (1) a systematic review, structured around PICO (Population, Intervention, Comparison, Outcome) criteria and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)-style search and screening principles, of artificial intelligence (AI) and machine-learning (ML) applications in TB diagnosis and risk prediction; (2) a systematic review of real-world U.S. implementation evidence for video-observed therapy (VDOT) in treatment monitoring; and (3) a county-level ecological analysis linking California TB burden to infectious disease (ID) and pulmonology specialist workforce density. Across the AI literature, diagnostic computer-aided detection (CAD) software for chest radiography showed consistently high pooled sensitivity (approximately 85%–94%) but more variable specificity (approximately 37%–69%) across three independently published meta-analyses, while a published meta-analysis of VDOT found a significant pooled improvement in medication adherence relative to in-person directly observed therapy (DOT) (relative risk 2.79, 95% CI 2.26–3.45). Predictive AI applications, spanning progression-risk and outbreak-cluster models, were too heterogeneous across primary studies for quantitative pooling and are synthesized narratively. A deeper, complementary review of ten real-world U.S. VDOT implementation studies found that all seven observational studies reporting an adherence or completion outcome favored VDOT or found it comparable to DOT; a purpose-built meta-analysis combining the only two eligible U.S. active-disease studies with comparable completed-doses data (one observational, one randomized crossover) found a small but statistically significant risk difference favoring VDOT (+2.1 percentage points, 95% CI 0.1 to 4.1) and a larger but highly heterogeneous, nonsignificant difference for latent TB infection (LTBI) preventive therapy (+14.5 points, 95% CI –3.9 to 33.0, $I^2 = 74\%$), indicating that large gains in observation-rate metrics should not be assumed to translate proportionally into treatment completion. Only one of the ten studies reported a formal sociodemographic subgroup analysis, which found lower VDOT adherence among U.S.- or Mexico-born patients. Separately, an ecological analysis of 39 California counties found that population-adjusted TB case rate and ID/pulmonology physician density were weakly correlated overall (Pearson $r = 0.26$) but moderately to strongly correlated once a single extreme outlier, Imperial County, was excluded ($r = 0.67$); Imperial County combines the state's highest TB rate (25.0 per 100,000) with among its lowest specialist densities (0.56 per 100,000), a gap equivalent to roughly sixteen-fold relative to a comparable neighboring county. Because the specialist workforce gap identified here cannot be closed on any clinically meaningful timeline through routine training pipelines alone, these findings support prioritizing AI and VDOT adoption specifically in specialist-scarce jurisdictions: both technologies absorb high-volume, nonspecialist components of the TB care cascade (screening triage, contact-cluster prioritization, routine dose observation), which can free scarce ID/pulmonology capacity for the complex, drug-resistant, and coinfecting cases that require it, though this mechanism is inferred from the three analyses' individual findings rather than directly tested. Read together, these findings indicate that AI and VDOT can meaningfully extend diagnostic reach and treatment-monitoring capacity in a low-incidence U.S. setting, but that technological gains complement rather than substitute for the specialist workforce required to act on them, and that both technological and workforce interventions remain under-evaluated for the demographic and geographic subgroups who carry a disproportionate share of the nation's remaining TB burden.

Author keywords: Tuberculosis; artificial intelligence; video-observed therapy; specialist workforce; low-incidence setting

*Corresponding Author: Shreya Parimoo.
Email: shreyaparimoo@gmail.com
Stockdale High School, Bakersfield, California, USA

Discussion period open till six months from the publication date. Please submit separate discussion for each individual paper. This paper is a part of the Vol. 1 of the International Journal of Bridge Engineering, Management and Research (© BER), ISSN 3065-0569.

Introduction

In 1989, the United States launched a formal strategic plan aiming toward tuberculosis (TB) elimination, defined as fewer than one case per million persons annually.¹ More than three decades later, that target remains distant: provisional CDC data indicate 10,260 TB cases in 2025, a rate of 3.0 cases per 100,000 population and only a modest one percent decline from 2024.² Roughly three-quarters of U.S. cases occur among non-U.S.-born persons, whose incidence rate (15.4 per 100,000) is nearly twenty times that of U.S.-born residents (0.8 per 100,000),² and cases cluster geographically: California, Texas, New York, and Florida together account for about half of all reported cases.³ California alone reports more TB cases annually than any other state, with a 2025 incidence of 5.4 per 100,000, nearly double the national rate, and an estimated 83% of its cases arising from progression of existing latent TB infection (LTBI) rather than new transmission.⁴

This epidemiological profile places the United States squarely within the category of a low-incidence, high-resource TB setting, one that behaves quite differently from the high-burden countries where most TB-related technology, including most AI diagnostic tools, was originally developed and validated. In high-burden settings, chest radiography interpretation is often the rate-limiting step in case-finding because trained radiologists are scarce relative to the volume of at-risk patients.⁵ In the United States, radiologist availability is comparatively high, but clinical suspicion for TB is low precisely because the disease is rare, and a chest radiograph may be reviewed by a radiologist who has seen few, if any, active TB cases in years. Directly observed therapy (DOT), the longstanding standard of care for treatment adherence, similarly strains a public health workforce that must serve patients spread across large geographic areas rather than concentrated case-finding sites.⁶

These structural differences motivate a distinct set of questions from the ones usually asked about TB and technology: not simply whether an algorithm can detect TB accurately, or whether a video platform can replace an in-person visit, but where in a low-incidence health system these tools add the most value, how strong the underlying quantitative evidence actually is, and whether technological capacity is matched by the human specialist workforce needed to act on it. This paper addresses these questions across three linked analyses. The first is a systematic review and, where data allow, meta-analytic synthesis of AI and ML applications across three domains of the TB care cascade: diagnostic applications (AI-assisted radiographic interpretation, genomic drug-resistance prediction, and automated case detection from clinical text), predictive applications (machine-learning models estimating individual progression risk or outbreak-cluster growth), and treatment-monitoring applications (AI-enabled video directly observed therapy, or VDOT). The second is a focused, deeper systematic review of real-world U.S. VDOT implementation evidence, addressing a gap left by prior randomized-trial-only reviews and by the necessarily brief treatment-monitoring domain of the first review: whether real-world, largely observational

U.S. programs show adherence and completion patterns consistent with randomized evidence, and whether sociodemographic subgroup differences in VDOT effectiveness have been examined. The third is a county-level ecological analysis asking a structurally different but related question: even where diagnostic and treatment-monitoring technology exists, is it deployed in places where the infectious disease (ID) and pulmonology specialist workforce needed to interpret results and manage complex cases is actually present?

Read together, these three analyses trace the U.S. TB care cascade from detection and treatment monitoring to the specialist capacity that ultimately determines whether a diagnosis or an adherence signal translates into effective care. The central argument that emerges is that technology's role in a low-incidence, high-resource setting is chiefly one of augmentation rather than substitution, extending the reach of an already substantial but geographically and demographically stretched public health and specialist workforce, rather than replacing scarce expertise as it more often does in high-burden countries. Where that human workforce is itself geographically absent, however, as this paper's ecological findings show most starkly in California's Imperial County, technology cannot by itself close the resulting gap in care.

Description, Methodology, or Proposed Research Approach

Overview and review framework

This paper synthesizes three methodologically distinct but thematically linked analyses, each retaining the approach appropriate to its own evidentiary base rather than being forced into a single uniform design. The first two are systematic reviews structured around the PICO (Population, Intervention, Comparison, Outcome) framework and informed by PRISMA 2020 reporting principles⁷; the third is a cross-sectional ecological analysis of publicly available surveillance and workforce data. [Table 1](#) summarizes the PICO framework guiding both review components. Both systematic reviews were conducted by a single reviewer using a general-purpose web-based literature discovery tool rather than a prospectively registered, dual-reviewer, exhaustive multi-database export; this departure from full PRISMA compliance is stated explicitly here and revisited under each component's own limitations, consistent with this paper's position that no methodological claim should be presented as stronger than the process that produced it.

AI applications in TB diagnosis and risk prediction: search strategy and eligibility

Following PRISMA 2020 reporting logic, searches were conducted using a web-based literature discovery tool spanning content indexed in PubMed/MEDLINE, Embase and Cochrane-participating journals, and IEEE Xplore, supplemented by direct queries of CDC surveillance and guidance publications; approximately 21 distinct search queries combined tuberculosis-related vocabulary with AI/ML vocabulary and, where relevant, low-incidence/U.S.

Table 1. Merged PICO framework guiding the AI/ML and VDOT systematic review components

Element	Definition for this review	Examples in the literature
Population	Individuals in low-incidence, high-resource U.S. health settings undergoing TB screening, diagnosis, risk stratification, or treatment for TB disease or LTBI	U.S.-bound migrants undergoing pre-migration screening; U.S. TB program patients; contacts of TB cases
Intervention	AI/ML tools applied to any stage of the TB care cascade; video-observed therapy (VDOT/VOT), live or asynchronous	CAD software for chest radiographs; genomic drug-resistance prediction; NLP case detection; ML risk models; live/recorded VDOT
Comparison	Standard human-based clinical practice or no AI tool; in-person, clinic- or field-based directly observed therapy (DOT)	Radiologist interpretation; phenotypic drug-susceptibility testing; in-person DOT
Outcomes	Diagnostic accuracy, predictive performance, treatment adherence/completion, cost, outbreak-detection performance, subgroup effects	Sensitivity/specificity; AUC; fraction of doses observed; completion %; cost; subgroup effect by birth country

setting terms. Coverage of Embase and Cochrane CENTRAL was achieved indirectly, through publisher sites and mirrored preprint content, rather than through direct institutional database access; PubMed/MEDLINE and IEEE Xplore were queried directly. Eligible studies described the development, validation, or implementation of an AI/ML method applied to any stage of the TB care cascade, in a U.S. or U.S.-bound cohort, reporting an extractable diagnostic-accuracy, predictive-performance, adherence/completion, cost-effectiveness, or outbreak-detection outcome, published in English in a peer-reviewed source; existing systematic reviews and meta-analyses of AI diagnostic tools were included and treated as primary evidence under an umbrella-review approach. Studies reporting only internal benchmark accuracy on public image datasets (e.g., Shenzhen, Montgomery, TBX11K) without clinical validation were excluded, as were studies outside the low-incidence/U.S. scope, non-English sources, and conference abstracts without peer-reviewed detail. Figure 1 summarizes the resulting study identification and selection process; because this review was not conducted through a reference-management platform with automated deduplication, the counts shown are an approximate single-reviewer reconstruction rather than an exact, independently reproducible export.

Risk of bias for the core diagnostic-accuracy literature was assessed using QUADAS-2 across four domains (patient selection, index test, reference standard, flow and timing); for the two published meta-analyses that reported their own QUADAS-2 assessments,^{5,8} this review reports their stated conclusions rather than re-deriving independent judgments. Where two or more sources reported comparable, homogeneous quantitative estimates on the same outcome and index test, this review presents those estimates side by side as an umbrella-level synthesis rather than performing new statistical pooling from primary-study raw data; this approach was used for AI chest-radiograph diagnostic accuracy (three published meta-analyses) and for VDOT treatment outcomes (one published meta-analysis, further extended in the

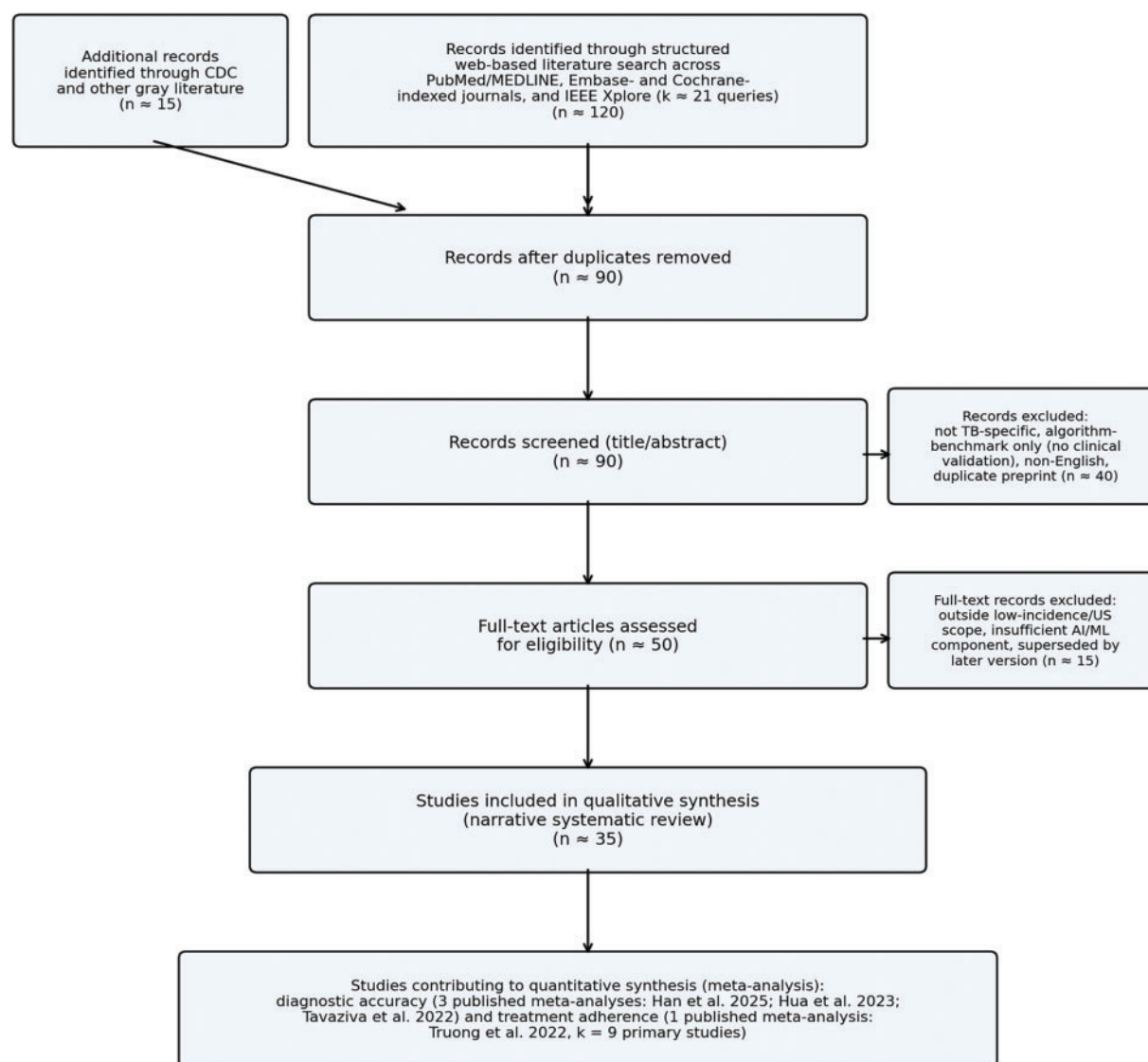
dedicated VDOT review below). No published meta-analysis with comparable pooled effect sizes was identified for the predictive-modeling domain, which is therefore synthesized narratively.

AI applications in TB diagnosis and risk prediction: findings

Diagnostic applications

The most extensively studied AI application in TB care is computer-aided detection (CAD) software that scores chest radiographs for features consistent with pulmonary TB. A 2019 systematic review catalogued CAD systems evaluated against microbiological reference standards and found generally strong discriminative performance, though with considerable heterogeneity in study design and threshold selection.⁵ More recently, a PROSPERO-registered systematic review and meta-analysis following PRISMA-DTA reporting standards pooled data from 21 studies through December 2024 and reported sensitivity/specificity ranges across five leading products: JF CXR-1 (86.0%/80.0%), qXR (90.0%/64.0%), Lunit INSIGHT CXR (90.0%/63.0%), CAD4TB (91.0%/60.0%), and InferRead DR Chest (89.0%/59.0%).⁹ Direct evidence on performance in a U.S.-relevant population comes from a multicenter evaluation, conducted by the International Organization for Migration with the CDC, of three CAD systems (CAD4TB, Lunit INSIGHT, qXR) against chest radiographs from the pre-migration health assessment database for migrants bound for the United States, a screening pathway that already relies on chest radiography for TB-infected visa applicants.^{10,11} Building on this evidence base, Qure.ai’s qXR platform has continued expanding its U.S. regulatory footprint, receiving additional FDA 510(k) clearances in 2026, reflecting a broader shift of CAD tools from screening use in high-burden countries toward point-of-care decision support in U.S. emergency and outpatient radiology.¹²

Figure 1. PRISMA-Style Flow Diagram of Study Identification and Selection



Note. Figures are an approximate, single-reviewer reconstruction of a web-search-based literature discovery process rather than an exact, reproducible export from institutional Embase, Cochrane CENTRAL, or IEEE Xplore interfaces. See Methods (Search Strategy) for details and limitations.

Figure 1. PRISMA-style flow diagram of study identification and selection for the AI/ML diagnostic and predictive review components

A second diagnostic frontier applies deep learning to whole-genome sequencing data to predict drug resistance directly from a bacterial isolate's genome, bypassing weeks-long culture-based phenotypic testing; one such model, TB-DROP, achieved a mean sensitivity of 0.90 and specificity of 0.87 for predicting resistance across multiple drugs.¹³ For a low-incidence country where drug-resistant TB cases are individually rare but carry outsized clinical and public-health stakes, genomic AI tools offer a way to obtain actionable resistance data faster than sequential phenotypic testing allows. A third, older but still relevant application uses natural language processing (NLP) to flag probable TB cases from unstructured clinical documentation;

early work paired an electronic medical record with a clinical event monitor and NLP system, finding a TB-culture-based rule that achieved 89% sensitivity and 96% positive predictive value,¹⁴ and more recent work has revisited this idea with transformer-based text classifiers such as Bio_ClinicalBERT.¹⁵ In a low-incidence setting where a busy clinician may not consider TB in a differential diagnosis, NLP-based case detection functions as a safety net, surfacing cases that might otherwise be closed out under a more common diagnosis.

Predictive applications

Because roughly 5%–10% of persons with LTBI will progress to active disease over their lifetime, and because LTBI is

estimated to affect around 13 million people in the United States, identifying which infected individuals are at highest risk of progression is a central problem for a low-incidence, elimination-oriented TB program.¹⁶ Several studies have applied machine-learning feature-selection methods, including LASSO regression, support vector machines with recursive feature elimination, and random forests, to gene-expression datasets to identify biomarker panels predictive of LTBI relapse or progression; one such study identified a three-gene signature (ALG2, FARS2, PGP) associated with relapse risk,¹⁷ while a separate model developed for people living with HIV used routinely collected clinical data to predict incident active TB and reportedly outperformed existing testing approaches.¹⁸ A second, more mature predictive application uses machine learning on national surveillance and molecular genotyping data to forecast which clusters of genetically related TB cases are likely to grow into larger outbreaks; using CDC National Tuberculosis Surveillance System data from 2009–2018, one analysis built a random-forest model finding that cluster-level characteristics were more predictive of unexpected growth than the social and demographic characteristics of individual cases.¹⁹ A related molecular surveillance analysis of 24 large U.S. outbreaks between 2014–2018 (518 cases) found these disproportionately affected U.S.-born racial and ethnic minorities and were associated with substance use, homelessness, and incarceration, underscoring that outbreak-prediction tools in a low-incidence setting are as much about directing scarce contact-investigation resources toward vulnerable populations as about raw detection accuracy.²⁰ These studies could not be quantitatively pooled because of heterogeneity in populations, gene panels, algorithms, and outcome definitions across the small number of available studies (Table 2).

Meta-analysis of diagnostic accuracy

Three independently conducted meta-analyses of AI chest-radiograph software were identified: Han et al.⁹ (2025, five products across 21 studies), Hua et al.⁸ (2023, a rapid review across eight studies meeting quality criteria), and, Tavaziva et al.²¹ (2022, an individual patient data meta-analysis of two products across 3,727 individuals self-referring for TB symptoms). Figure 2 displays pooled sensitivity and specificity by product across these three sources. A consistent pattern holds across all three: pooled sensitivity clusters in the 84%–94% range regardless of product, while pooled specificity is

both lower and considerably more variable, ranging from roughly 37% (CAD4TB v3.07) to roughly 69% (qXR v2). The overall pooled estimate reported by Hua et al.⁸ across all evaluated products was sensitivity 90.3% (95% CI 85.9–93.4) and specificity 52.6% (95% CI 40.9–64.1). This convergence across independently conducted meta-analyses, despite differing inclusion criteria and product versions, strengthens confidence that AI CXR software is best suited to a rule-out, triage role rather than as a stand-alone diagnostic; the wide spread in specificity, in particular the roughly 32-percentage-point range for CAD4TB alone, indicates that specificity is highly threshold- and version-dependent, a detail relevant to any U.S. deployment decision about which product and threshold to adopt for a given prevalence setting.

From meta-analysis to focused review: treatment-monitoring applications

Directly observed therapy (DOT), in which a health care worker watches a patient ingest each dose of TB medication, has long been the U.S. standard of care, but in-person DOT is resource-intensive.²² Video directly observed therapy (VDOT) replaces in-person observation with recorded or live-streamed video review, and on the strength of an accumulating U.S. evidence base, the CDC formally updated its clinical guidance in 2023 to recognize VDOT as an equivalent alternative to in-person DOT.²³ A published meta-analysis pooling nine largely international comparative studies found VDOT associated with significantly improved medication adherence (relative risk 2.79, 95% CI 2.26–3.45; $I^2 = 25\%$) and bacteriological resolution (RR 1.06, 95% CI 1.01–1.11), though very high heterogeneity for the doses-observed and treatment-completion outcomes ($I^2=96\%$ and 98% , respectively) limits confidence in those specific pooled estimates²⁴ (Fig. 3). A distinct and more recent development layers automated computer-vision ingestion verification onto the video-review model: an evaluation of the AiCure platform by the Los Angeles County Department of Public Health, using a Markov model built from a 43-patient pilot cohort and 71 historical DOT controls, found it cost-effective relative to in-person DOT in a jurisdiction whose TB incidence substantially exceeds the U.S. national average.²⁵ Because the Truong et al. pooled evidence base is largely international and reports only high-level outcomes, the remainder of this section presents a dedicated, deeper systematic review of real-world U.S. VDOT

Table 2. Individual (non-pooled) performance metrics for predictive-modeling studies

Study focus	Population/data source	Algorithm(s)	Reported performance
LTBI relapse risk (3-gene signature)	GSE54992 gene-expression dataset	SVM-RFE, LASSO, random forest	AUC reported; exact value not specified in the retrieved abstract
Active TB prediction in people with HIV	Routine clinical data, PWH cohort	Machine learning (unspecified ensemble)	Reported to outperform existing testing approaches
TB cluster/outbreak growth prediction	CDC NTSS surveillance data, 2009–2018	Random forest	Cluster-level characteristics more predictive than case-level demographics

Figure 2. Pooled Diagnostic Accuracy of AI Chest-Radiograph Software for Pulmonary Tuberculosis, Across Three Independently Published Meta-Analyses

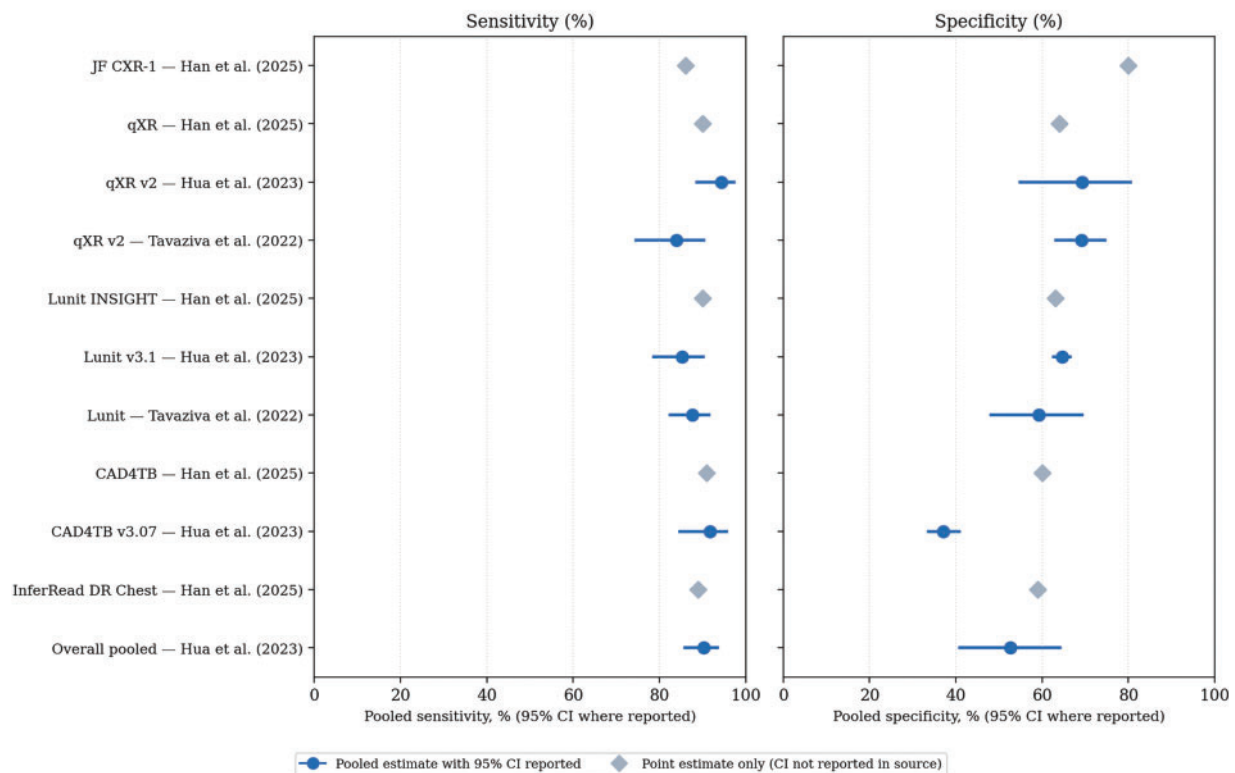


Figure 2. Pooled diagnostic accuracy of AI chest-radiograph software for pulmonary tuberculosis, compared across three independently published meta-analyses

Figure 3. Pooled Relative Risk of TB Treatment Outcomes, Video-Observed vs. In-Person Directly Observed Therapy (Truong et al., 2022; k = 9 studies)

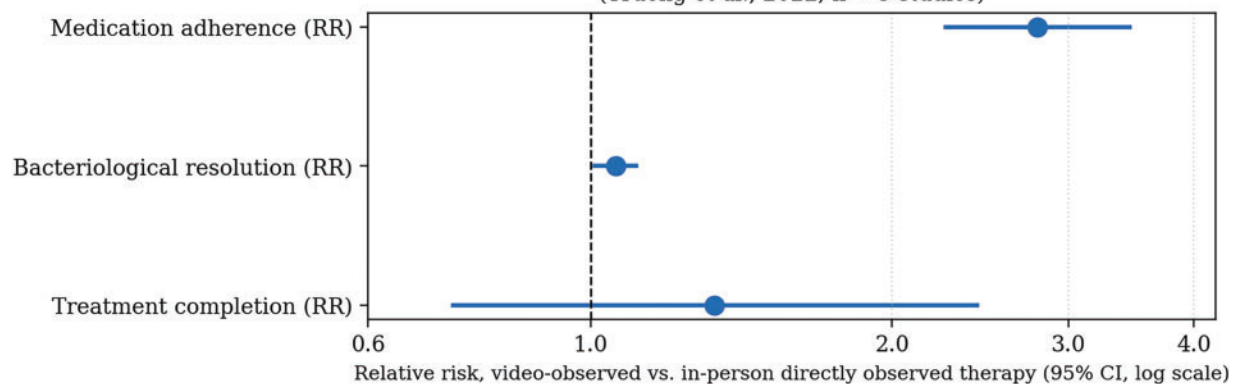


Figure 3. Pooled relative risk of TB treatment outcomes, video-observed versus in-person directly observed therapy²⁴

implementation evidence, addressing gaps in U.S. generalizability and sociodemographic subgroup reporting left by prior randomized-trial-focused reviews.

Real-world video-observed therapy implementation: search strategy and eligibility

A 2024 systematic review by Areas Lisboa Netto et al.,²⁶ restricted to randomized controlled trials with a control group, established VDOT's non-inferiority to in-person DOT for active TB treatment using four trials, only one of

which was conducted in the United States, while also identifying gaps in U.S. generalizability and sociodemographic subgroup reporting. This review addresses those gaps by systematically synthesizing real-world U.S. VDOT implementation studies across both active TB disease and LTBI preventive-therapy monitoring. The review question was structured using the PICO framework (Table 1); reporting was structured with reference to PRISMA 2020 principles, though because this project used a single-reviewer targeted search supplemented by citation tracking rather than a

prospectively registered, exhaustive multi-database review with duplicate screening, it does not claim full compliance with reporting elements that require those processes.

Studies were eligible if conducted in a U.S. health department, outpatient TB clinic, or affiliated public health program; directly compared VDOT (synchronous or asynchronous) to in-person DOT within the same program or study; enrolled patients receiving treatment for active TB disease or LTBI preventive therapy; reported at least one quantitative outcome (adherence/fraction of doses observed, treatment completion, cost, or a sociodemographic subgroup finding); and were published in a peer-reviewed journal or as CDC/MMWR guidance. Studies were excluded if they were conducted outside the United States, lacked an in-person DOT comparison group, compared VDOT to self-administered therapy rather than in-person DOT, were conducted in an inpatient or institutional setting,

used non-video remote monitoring only, or were conference abstracts or protocols without outcome data.

Literature was identified through targeted searches combining VDOT/VOT/eDOT terminology with U.S. setting and outcome terms, supplemented by review of PubMed/MEDLINE-indexed literature, CDC/MMWR publications, the reference list of the 2024 PLOS Neglected Tropical Diseases review, and CDC's 2023 VDOT guidance citation list, followed by citation tracking from these seed documents. This process identified ten eligible U.S. studies (eight observational, two randomized); Fig. 4 summarizes the selection process. Because the identified studies use non-identical adherence metrics (scheduled-session adherence, fraction of expected doses observed [FEDO], verified adherence, observable fraction), pooling all adherence measures into a single statistic was not appropriate. Treatment completion was the primary outcome selected for quantitative synthesis. For LTBI, two U.S. studies^{27,28} reported

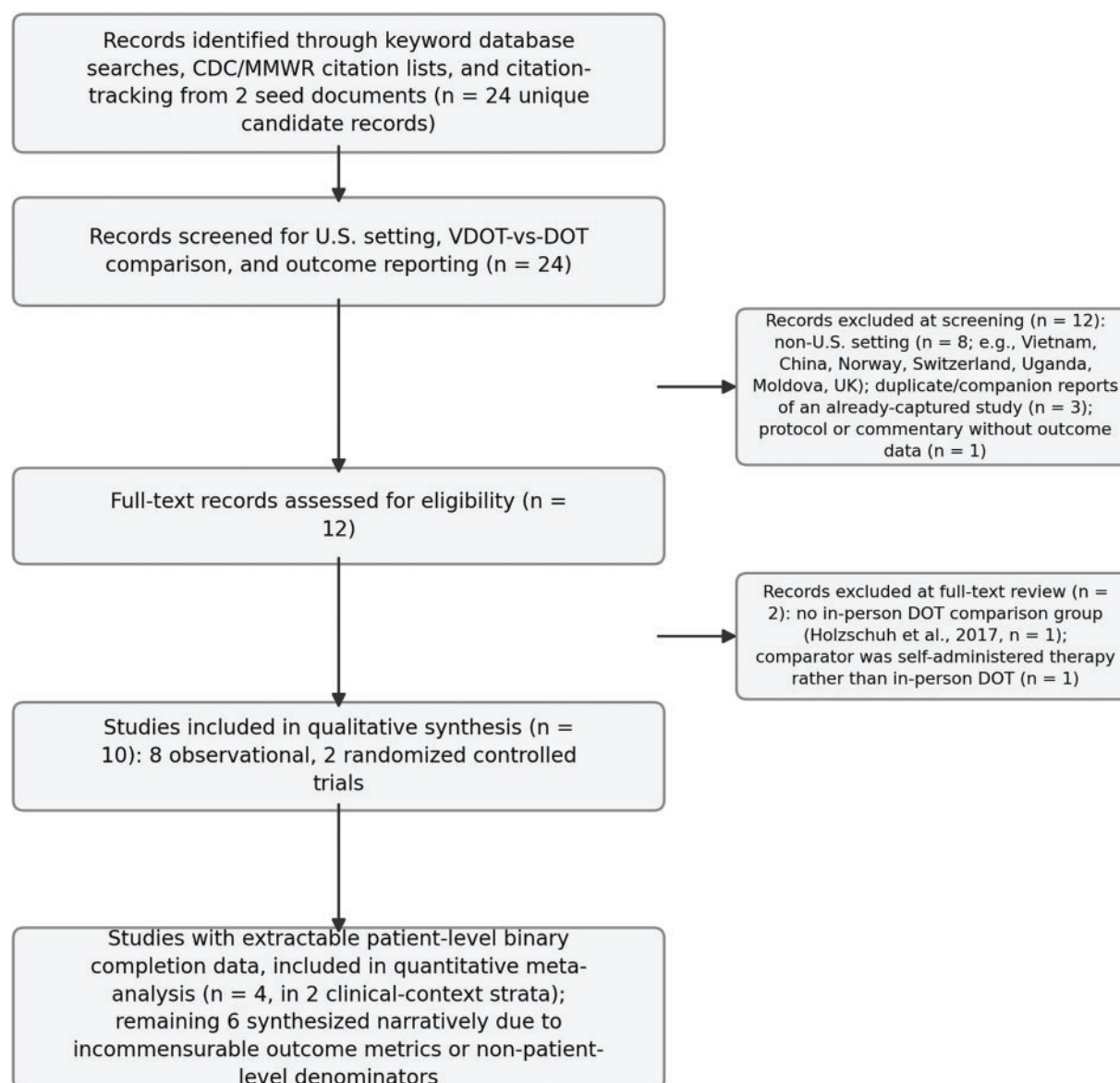


Figure 4. Study identification and selection, real-world U.S. VDOT implementation review

directly comparable patient-level completion counts by arm, allowing a standard random-effects DerSimonian–Laird meta-analysis of risk difference. For active disease, only one U.S. observational study²⁹ reported patient-level completion counts; Perry et al.³⁰ reported verified-dose observation only and, per CDC’s own subsequent evidence review, did not evaluate treatment completion, so it does not contribute a completion estimate here (see Findings). The only other eligible U.S. active-disease study reporting a directly quantifiable completed-doses outcome was the randomized crossover trial by Burzynski et al.,³¹ whose published effect estimate and 95% CI already account for the paired, within-patient structure of its crossover design via bootstrap resampling. Rather than treat individual doses as independent patient-level observations, which would understate variance, this review incorporates Burzynski et al.’s trial-level effect estimate directly into a generic inverse-variance random-effects meta-analysis alongside Chuck et al.’s risk difference, a standard approach for combining a parallel-group and a crossover design when only study-level effect estimates and confidence intervals are available. Studies reporting only adherence/FEDO, cost, or other non-comparable outcomes were synthesized narratively.

Real-world video-observed therapy implementation: findings

Adherence and completion: consistent direction, variable magnitude

Seven of the eight identified U.S. observational studies reported an adherence or completion comparison (Table 3); all seven found VDOT associated with adherence or completion at least as high as in-person DOT, though effect sizes varied by outcome definition, from a 4-point gap in scheduled-session adherence²⁹ to a 26.6-point gap in FEDO across five California districts.⁶ Because these measures are not interchangeable across studies, they are presented narratively in Table 3 rather than pooled; quantitative pooling was reserved for the comparable patient-level completion outcomes reported below.

Cost findings were consistently favorable to VDOT. Beeler Asay et al.,³² surveying 225 patients across New York City, San Francisco, and Rhode Island, found median patient costs of \$1.01 per live VDOT session versus \$34.53 for clinic-based DOT; Holzman et al.³³ estimated savings of approximately \$1,391 per patient for a standard six-month course; and Garfein et al.⁶ found VDOT cost 32% less than DOT overall. No identified U.S. study reported higher overall VDOT costs than in-person DOT, though cost estimates are setting-specific.

Meta-analysis of treatment completion

This review pools Chuck et al.’s²⁹ observational completion data with Burzynski et al.’s³¹ randomized crossover completed-doses effect using the generic inverse-variance method described in Methods, yielding +2.1 percentage points favoring VDOT (95% CI 0.1 to 4.1; $p = .04$; $I^2 = 8\%$; Table 4). One caveat: Chuck et al.’s outcome is binary completion of a full treatment course, while Burzynski et

al.’s is the percentage of a fixed dose block completed under a crossover design; the two are related but not identical, combined here because they are the only eligible U.S. active-disease studies with a quantifiable, comparable completed-doses effect. Perry et al.³⁰ does not contribute a completion estimate: it measured verified-dose observation only, and CDC’s 2023 evidence review states explicitly that this study “did not evaluate treatment completion or microbiologic resolution.” The corrected evidence indicates a small but statistically significant improvement in active-disease completion under VDOT, an order of magnitude smaller than the 20–30 percentage-point gaps seen on dose-observation-rate metrics; those larger gaps reflect how thoroughly doses were observed rather than a comparably large improvement in whether treatment was ultimately completed.

Pooling Lam et al.²⁸ and Garfein et al.,²⁷ the two U.S. LTBI studies with comparable completion counts, yielded a pooled risk difference of +14.5 percentage points favoring VDOT (95% CI −3.9 to 33.0; $p = .12$; $I^2 = 74\%$; Table 5). This estimate crosses the null and is based on only two studies with markedly different designs: Lam et al. used a historical DOT cohort and reported a 23.1-point difference, whereas Garfein et al. randomized participants and found a smaller, non-significant 4.2-point difference; the pooled LTBI estimate should therefore be regarded as hypothesis-generating rather than confirmatory.

The active-disease and LTBI pooled estimates read differently: the active-disease effect is small but reaches significance with low heterogeneity between its two contributing designs; the LTBI effect is larger but crosses the null and is driven almost entirely by one historical-control study, with Garfein et al.’s randomized estimate alone (+4.2 points; $p = .70$) the more conservative single-study benchmark. Observational cohorts can be affected by selection into VDOT and historical-control effects, and this review treats the larger observational adherence gaps described earlier as evidence of improved observation and implementation performance rather than proof of comparably large causal improvements in completion.

Sociodemographic subgroup findings

Consistent with the gap identified by the 2024 PLOS Neglected Tropical Diseases review, only one of the ten eligible studies, Garfein et al.,⁶ reported a formal subgroup analysis of factors associated with adherence under VDOT (Table 6); Perry et al.³⁰ reported a significant baseline age difference between groups (46 vs. 61 years; $p < .001$) but did not conduct a formal subgroup analysis, and this imbalance is treated as a potential confounder rather than a subgroup finding.

The finding that U.S.- or Mexico-born patients had significantly lower VDOT adherence than patients born elsewhere is the only formal sociodemographic subgroup signal identified across the literature reviewed here. It has not been replicated, and possible explanations such as digital literacy, work schedule, housing stability, or technical barriers remain hypotheses rather than demonstrated causes.

Table 3. Characteristics and key outcomes of eligible U.S. VDOT studies

Study (site)	Design/period	N (VDOT/DOT)	Key outcome	Notes
Chuck et al. ²⁹ (New York City)	Prospective observational, 2013–2014	61/329	Scheduled adherence 95% vs. 91% ($p < .01$); completion 96% vs. 97%	Active TB; used in the meta-analysis
Holzman et al. ³³ (Maryland)	Mixed-methods pilot, 2016–2017	28 total	Median adherence 94% vs. 98% ($p = .17$); cost saving ~\$1,391/patient	Active TB and LTBI; 3 clinics
Garfein et al. ⁶ (5 CA districts)	Observational, historical control, 2014–2015	274/159	Median FEDO 93.0% vs. 66.4%; 32% lower cost	Subgroup predictors reported
Lam et al. ²⁸ (NYC, LTBI/3HP)	Observational vs. historical DOT, 2015	50/302	Completion 88.0% vs. 64.9% ($p < .001$)	Used in LTBI meta-analysis
Perry et al. ³⁰ (Alameda County, CA)	Programmatic observational, 2018–2020	94/69	Verified doses 68% vs. 54% ($p < .001$); completion not reported	Age imbalance (46 vs. 61 yrs) noted as a confounder; per CDC, ²³ this study did not evaluate treatment completion
Lippincott et al. ⁴⁴ (Baltimore, MD)	Retrospective, 2019–2021	52 patients	Median verified adherence 86% vs. 59% ($p < .01$)	COVID-period cohort
Bachina et al. ⁴⁵ (Minneapolis, MN)	Retrospective, 2019–2021	Pre/post-COVID cohorts	Verified adherence 81% vs. 54.5% ($p = .001$)	Programmatic implementation
Beeler Asay et al. ³² (NYC/SF/RI)	Time-motion + cost survey, 2017–2018	225 patients	Patient cost \$1.01 vs. \$34.53/session	Cost only; no adherence endpoint
Burzynski et al. ³¹ (New York City)	Randomized crossover non-inferiority, 2017–2019	216 randomized	89.8% vs. 87.2% doses observed; non-inferior	Crossover RCT; effect estimate used in active-disease completion meta-analysis
Garfein et al. ²⁷ (San Diego, CA, LTBI)	Randomized controlled trial, 2016–2019	68/62	Completion 73.5% vs. 69.4% ($p = .70$)	Used in LTBI meta-analysis

Table 4. Generic inverse-variance meta-analysis of treatment completion, active-disease VDOT vs. DOT, combining an observational study (binary patient-level completion) and a randomized crossover trial (percentage of scheduled doses completed; see text for outcome-definition caveat). Perry et al.³⁰ is excluded because, per CDC,²³ it did not evaluate treatment completion

Study (design)	Outcome/effect measure	VDOT vs. DOT	Effect (95% CI)
Chuck et al. ²⁹ (observational)	Patient-level completion (binary)	59/61 (96.7%) vs. 319/329 (97.0%)	RD −0.2 pts (−5.1, 4.6)
Burzynski et al. ³¹ (RCT, crossover)	% scheduled doses completed (MITT)	89.8% vs. 87.2%	RD +2.6 pts (0.3, 4.8)
Pooled (generic inverse-variance, random-effects)	—	—	+2.1 pts (0.1, 4.1), $p = .04$, $I^2 = 8\%$

TB burden and specialist workforce alignment: data sources and analytic approach

Whether the geographic distribution of ID and pulmonology physicians, the specialists most directly responsible for TB diagnosis and management,³⁴ is aligned with county-level

TB burden has not been formally quantified for California, despite the state's markedly heterogeneous TB epidemiology and unevenly distributed physician workforce, with rural and border counties historically reporting substantially lower per-capita specialist density than major metropolitan

Table 5. Random-effects meta-analysis of treatment completion, LTBI VDOT vs. DOT

Study	VDOT completed/N (%)	DOT completed/N (%)	Risk difference (95% CI)
Lam et al. ²⁸	44/50 (88.0%)	196/302 (64.9%)	+23.1 pts (12.6, 33.6)
Garfein et al. ²⁷ (RCT)	50/68 (73.5%)	43/62 (69.4%)	+4.2 pts (−11.4, 19.7)
Random-effects pooled	—	—	+14.5 pts (−3.9, 33.0), p = .12, I ² = 74%

Table 6. Factors independently associated with lower fraction of expected doses observed (FEDO) under VDOT⁶

Factor associated with lower FEDO	Direction	Statistical significance
U.S. or Mexico birth (vs. other birth country)	Lower adherence	p < 0.05
Shorter VDOT duration on study	Lower adherence	p < 0.05
Patient-reported difficulty using VDOT	Lower adherence	p < 0.05
Frequently taking medication away from home	Lower adherence	p < 0.05
Video-recording technical problems	Lower adherence	p < 0.05

counties.³⁵ This cross-sectional ecological analysis linked 2025 provisional county-level TB case counts and rates⁴ with county-level estimates of actively working ID and pulmonology physicians³⁶ across 39 of California's 58 counties; the remaining 19, disproportionately small and rural counties (median population 26,599 vs. 443,956 for included counties) were excluded because HCAI's weighted survey estimate rounded to zero for both specialties, meaning no analyzable physician-density denominator could be constructed. City jurisdictions reporting TB cases independently (Berkeley, Long Beach, Pasadena) were merged into their parent counties (Alameda and Los Angeles, respectively) to align with county-level physician data, consistent with CDPH's own description of these jurisdictions as subsets of their parent counties.³⁷ ID and pulmonology counts were summed into a single combined specialist count per county, reflecting their overlapping clinical role in TB care; a sensitivity analysis examined each specialty separately.

Physician supply was expressed as a rate per 100,000 population. Pearson and Spearman correlations were computed between raw case/physician counts (population-confounded) and between population-adjusted TB rate and physician density, both including and excluding a pre-specified outlier county (Imperial), whose outlier status was formally confirmed using leverage and Cook's distance diagnostics from the full-sample regression model. A simple linear regression of physician density on TB rate (excluding Imperial County) was fit to quantify the population-adjusted relationship; because residuals showed right skew driven by a small number of high-density counties, a log–log correlation and a Huber robust regression were performed as sensitivity checks. Analyses were conducted in Python 3 using pandas, SciPy, and scikit-learn.

TB burden and specialist workforce alignment: findings

Statewide, TB burden and physician supply varied by more than an order of magnitude across the 39 included counties: TB case rates ranged from 0.74 per 100,000 (Humboldt) to 25.01 per 100,000 (Imperial), and physician density ranged from 0.35 per 100,000 (Merced) to 29.42 per 100,000 (San Francisco) (full county-level data provided in Supplemental Table S1). As expected, raw TB case counts and raw physician counts were very strongly correlated (Pearson $r = 0.973$; $p < .0001$), reflecting their shared dependence on county population rather than a substantive relationship. After adjusting both variables for population, the picture changed substantially: across all 39 counties, the population-adjusted correlation was weak and did not reach significance by Pearson's test ($r = 0.264$; $p = .104$), though the Spearman rank correlation remained moderate and significant ($\rho = 0.528$; $p < .001$), suggesting a real but outlier-distorted association (Table 7). When Imperial County was excluded, the Pearson correlation strengthened markedly ($r = 0.675$; $p < .0001$), as did the Spearman correlation ($\rho = 0.645$; $p < .0001$).

Imperial County's exclusion from the primary regression fit was not an arbitrary judgment call: in the full 39-county regression, Imperial had a leverage of 0.69 (nearly 14 times the average expected under an even distribution) and a Cook's distance of 17.6, more than 170 times the conventional concern threshold; no other county approached this level of influence (the next-highest, San Francisco, had a Cook's distance of 0.77). The regression of physician density on TB rate, excluding Imperial County, produced a slope of 1.47 physicians per 100,000 for every one-point increase in TB rate (95% CI 0.93–2.01; $R^2 = 0.46$); a Huber robust regression, downweighting the influence of San Francisco and Marin's unusually high physician density, produced a

Table 7. Correlation between TB case rate and ID/pulmonology physician density, 39 California counties (excluding Imperial rows use n = 38)

Comparison	Pearson r	p-value	Spearman ρ (p)
Raw case counts vs. raw physician counts (population-confounded)	0.973	<.0001	0.852 (<.0001)
TB rate/100k vs. physicians/100k—all 39 counties	0.264	.104	0.528 (<.001)
TB rate/100k vs. physicians/100k—excluding Imperial County	0.675	<.0001	0.645 (<.0001)

Table 8. Regression of physician density (per 100,000) on TB case rate (per 100,000), excluding Imperial County (n = 38)

Model	Slope	Slope 95% CI	Intercept	R ²
OLS, excluding Imperial County (primary)	1.47	(0.93, 2.01)	−0.78	0.46
Huber robust regression (sensitivity)	1.06	—	0.31	—

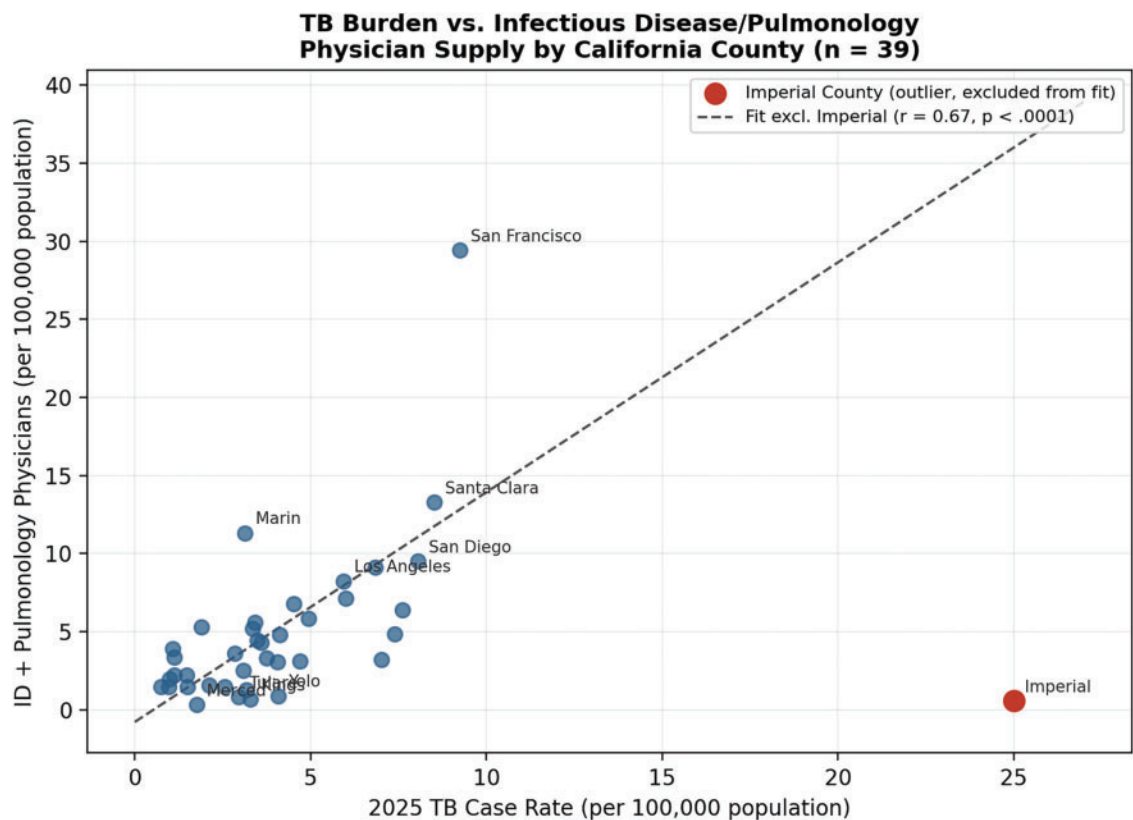


Figure 5. TB case rate (2025) plotted against combined ID/pulmonology physician density by California county. Dashed line is the linear fit excluding Imperial County

shallower but directionally consistent slope (1.06), indicating the qualitative finding is not an artifact of ordinary-least-squares fitting (Table 8; Fig. 5). A sensitivity analysis found the population-adjusted relationship was similar whether restricted to ID physicians alone ($r = 0.648$) or pulmonology physicians alone ($r = 0.673$), suggesting the combined-specialty relationship is not driven disproportionately by either specialty.

Fig. 5 illustrates the population-adjusted relationship: Imperial County (TB rate 25.01 per 100,000; physician density 0.56 per 100,000, equivalent to one physician for the

entire county) falls far outside the distribution defined by the remaining counties. Because Imperial's TB rate falls far beyond the range observed in the other 38 counties (maximum 9.25 per 100,000), the fitted line should not be extrapolated to predict a specific physician count for Imperial; a more defensible quantification uses a peer-county benchmark instead. San Diego County, a neighboring county with a broadly comparable population base and a substantial but much lower TB rate (8.05 per 100,000, less than a third of Imperial's), has a physician density of 9.53 per 100,000. Matching San Diego's density

would require approximately 17 ID/pulmonology physicians in Imperial County, an empirically grounded benchmark implying roughly a sixteen-fold gap relative to Imperial's current capacity. Several additional counties departed notably from the fitted relationship: San Francisco and Marin had substantially greater physician density than their TB burden alone would predict, consistent with their status as academic medical referral centers, while San Joaquin, Alameda, Yolo, Sacramento, Kings, and Solano combined meaningful TB burden with physician density below the statewide trend, identifying them as a secondary cluster of counties of concern, several of them predominantly rural or agricultural Central Valley counties with large migrant and seasonal farmworker populations recognized as being at elevated LTBI risk.³⁸

Discussions and Guidelines for Applications

Read together, these three analyses point to a consistent organizing theme: the value AI and digital-monitoring technology add to TB care depends on the resource profile of the setting, and the United States' status as a low-incidence, high-resource country changes that role in identifiable ways, while a parallel workforce-geography problem determines whether that value actually reaches the patients who need it.

In high-burden countries, CAD software for chest radiography was developed largely to substitute for a shortage of radiologists in active case-finding campaigns.⁵ In the United States, the more consistent pattern across the literature reviewed here is one of augmentation rather than substitution: CAD tools support panel physicians and civil surgeons conducting immigration medical screening for a population with a higher pretest probability of TB than the general U.S. population^{10,11}; genomic AI tools compress the time to actionable drug-susceptibility data for the individually rare but consequential drug-resistant case; predictive cluster models help a well-staffed but geographically stretched CDC surveillance apparatus decide where to direct contact investigations¹⁹; and AI-enabled VDOT extends the reach of local health department staff across large jurisdictions rather than replacing clinical judgment.²⁵ The convergence of pooled sensitivity estimates across three independently conducted CAD meta-analyses, despite differing inclusion criteria and product versions, is a reassuring signal about the reliability of the high-sensitivity, rule-out use case; the far greater spread in pooled specificity is an equally clear signal that false-positive burden is highly sensitive to threshold and product-version choice, a detail that matters for how any given U.S. jurisdiction calibrates a deployment decision.

The dedicated VDOT review adds a second, more operational layer to this picture. Health departments planning a VDOT rollout should distinguish between two related but distinct claims found in the literature: that VDOT substantially increases how thoroughly doses are observed (a claim well-supported by large, consistent gaps of 20–30 percentage points across nearly every observational study identified), and that VDOT substantially increases how often treatment is ultimately completed (a claim supported only modestly

for active disease, with a small but statistically significant pooled effect of +2.1 percentage points drawn from the sole eligible U.S. observational study and the sole eligible U.S. randomized trial). Programs that plan for VDOT to close a completion gap of the same magnitude as its observation-rate gap risk overestimating its causal effect on the outcome that ultimately matters most. The LTBI evidence is less settled still, with a pooled estimate that is both larger and highly heterogeneous, driven by the contrast between one historical-control study and one randomized trial; this should be treated as hypothesis-generating rather than a basis for programmatic assumptions.

A third theme, cutting across both the AI and VDOT reviews, is that the U.S. TB burden, and therefore the populations who stand to benefit most from these tools, is concentrated among specific groups: non-U.S.-born persons, whose incidence rate remains many times higher than that of U.S.-born residents, and U.S.-born racial and ethnic minorities affected by outbreaks linked to homelessness, substance use, and incarceration.^{2,20} Neither the diagnostic-accuracy meta-analyses nor the original VDOT meta-analysis reported subgroup estimates by these characteristics, and the dedicated VDOT review found only one study, out of ten, that formally examined sociodemographic subgroups at all, a finding (lower adherence among U.S.- and Mexico-born patients) that is provocative but isolated and unreplicated. This is a notable gap given the CDC's own stated equity goals for TB elimination,³⁹ and it means that neither this paper's AI findings nor its VDOT findings can speak with confidence to whether pooled accuracy or pooled treatment effects hold uniformly across the populations a U.S. elimination program most needs to reach.

The ecological workforce analysis supplies the piece that no technology-focused review can address on its own, and it sharpens, rather than merely qualifies, this paper's central argument. Across most of California, ID/pulmonology specialist density does scale with local TB burden, a broadly reassuring finding. However, Imperial County's combination of the state's highest TB rate and a specialist density equivalent to one physician for the entire county is a sharp, statistically extreme exception, and one with direct clinical implications: patients with suspected or confirmed TB, drug-resistant TB, or TB–HIV coinfection in a jurisdiction with a well-documented shortage of primary care and specialist providers relative to the population⁴⁰ likely face a substantial travel burden or extended wait times to access specialist evaluation. This requires either a referral to neighboring San Diego or Riverside Counties, or reliance on primary care and public health TB staff operating without specialist backup. The weight of evidence in TB care access literature links limited local specialist availability to diagnosis and treatment delay, though the relationship between geographic distance and outcomes is not uniform across all settings.⁴¹ Because the county-level ecological analysis excluded 19 smaller counties for lack of an analyzable physician-density estimate, and because those excluded counties are systematically smaller and more rural than the counties studied, the true extent of TB burden/specialist-supply misalignment

statewide is likely understated, not overstated, by these findings.

A gap of this size cannot close on a clinically meaningful timeline through routine workforce growth alone: producing one additional ID or pulmonology specialist requires a decade or more of training, and matching Imperial County's neighbor benchmark would require roughly seventeen-fold growth in local supply. This is precisely the condition under which this paper's diagnostic and treatment-monitoring findings carry the most practical weight, not because AI or VDOT can substitute for a specialist, but because both technologies demonstrably absorb high-volume, nonspecialist components of the TB care cascade: CAD chest-radiograph triage reduces how many studies require any physician review before a case can be ruled out^{8,9}; predictive cluster models direct scarce contact-investigation staff toward the clusters most likely to grow¹⁹; and VDOT shifts routine dose observation, which does not require ID or pulmonology training, away from field staff and toward a video-review workflow.^{6,24} None of these effects have been measured here as a change in ID/pulmonology caseload or wait times specifically, so this should be read as a plausible mechanism supported by the individual findings rather than a directly tested interaction. Understood this way, however, the case for AI and VDOT adoption is strongest in exactly the counties this paper identifies as most workforce-constrained: by absorbing tasks that do not require specialist judgment, these tools can concentrate Imperial County's single ID/pulmonology physician, and the comparably stretched capacity in the secondary cluster of underserved counties, on the complex, drug-resistant, and coinfecting cases only a specialist can manage.

For jurisdictions weighing where to prioritize investment, the guidance implied by these three analyses is therefore layered but convergent: AI diagnostic tools are best deployed as high-sensitivity, rule-out triage aids calibrated to local prevalence rather than as stand-alone diagnostics; VDOT rollouts should be planned around realistic, completion-specific effect sizes rather than the larger observation-rate gaps reported in most studies; and in specialist-scarce counties specifically, both technologies should be prioritized precisely because they can absorb cascade tasks that do not require an ID or pulmonology physician, freeing what specialist time exists for the cases that do. This is a case for adoption, not for substitution: no diagnostic or monitoring tool can supply the specialist judgment that complex TB cases ultimately require, and technology adoption should be pursued alongside, not instead of, direct workforce interventions such as telemedicine partnerships. Telemedicine-based specialist access is one concrete, near-term complement for counties like Imperial: Medi-Cal telehealth coverage dates to the 1996 California Telemedicine Advancement Act and was substantially expanded by the 2011 Telehealth Advancement Act (AB 415), and infectious disease telemedicine ("Tele-ID") programs elsewhere in the country have been associated with reduced transfer rates and shorter hospital stays at rural facilities without local ID coverage, with current evidence suggesting outcomes comparable to in-person specialist care,^{42,43} though reimbursement instability following the expiration of COVID-era federal telehealth

flexibilities remains an ongoing concern for the sustainability of such programs.

Recommendations for Future Work

Diagnostic and predictive AI. Prioritize prospective, U.S.-specific validation studies, particularly for predictive-risk models; report subgroup performance by birth country, race/ethnicity, and other equity-relevant characteristics; and standardize threshold reporting for CAD products so specificity trade-offs are transparent across deployment settings.

Video-observed therapy. Adopt standardized, prespecified completion definitions to allow future pooling; report baseline characteristics by monitoring strategy to identify confounding (as in the Perry et al. age imbalance); and prespecify sociodemographic subgroup analyses for age, sex, race/ethnicity, country of birth, and technology access, given that only one of ten studies reported any such analysis to date.

Workforce alignment. Prioritize Imperial County, and the secondary cluster of underserved Central Valley and border counties identified here, for feasibility assessment of Tele-ID or tele-pulmonology partnerships, loan-repayment or recruitment incentives for ID/pulmonology physicians, and prioritized AI/VDOT deployment specifically to offset nonspecialist workload; extend this ecological framework to the 19 excluded counties as improved workforce-survey methodology becomes available; and pursue program-level data directly linking specialist access and technology adoption to TB diagnostic delay and treatment outcomes, since the associations reported here are ecological and hypothesis-generating rather than causal.

Cross-cutting. Future research should directly test whether AI-enabled diagnostic and monitoring tools measurably reduce ID/pulmonology caseload in specialist-scarce jurisdictions, an interaction this paper's three analyses could not test jointly; and future reviews here should move toward prospectively registered, dual-reviewer protocols with full database access.

Conclusions

As the United States continues to pursue TB elimination against a backdrop of persistently low but non-negligible incidence, this paper's three linked analyses show that artificial intelligence, video-based treatment monitoring, and specialist workforce geography each play a distinct and only partially substitutable role in closing the remaining gap. AI tools are emerging across the diagnostic, predictive, and treatment-monitoring stages of TB care with meaningful, if domain-specific, quantitative support: consistently high pooled sensitivity for CAD chest-radiograph screening, a significant pooled improvement in medication adherence and bacteriological resolution under VDOT, and a growing but still heterogeneous predictive-modeling literature. A dedicated review of real-world U.S. VDOT implementation reinforces and refines this picture, finding directionally consistent, favorable evidence across ten studies

while cautioning that large gains in dose-observation metrics should not be assumed to translate proportionally into gains in treatment completion, particularly for LTBI preventive therapy.

Neither of these technology-focused analyses, however, can address a structurally separate question that this paper's ecological analysis makes concrete: whether the human specialist workforce needed to act on a diagnosis or an adherence signal is actually present where TB burden is highest. Across most of California, it broadly is; in Imperial County, a jurisdiction combining the state's highest TB rate with among its lowest specialist densities, it plainly is not, a gap on the order of sixteen-fold relative to a comparably burdened neighboring county, and one that routine workforce growth cannot close on any clinically meaningful timeline given the decade-plus training pipeline for ID and pulmonology specialists. This gap is precisely why the case for AI and VDOT adoption is strongest, not weakest, in the specialist-scarce counties this paper identifies: both technologies absorb the high-volume, nonspecialist components of the TB care cascade, chest-radiograph triage, contact-cluster prioritization, and routine dose observation, freeing what scarce specialist time exists for the complex, drug-resistant, and coinfecting cases that only a specialist can manage. This is a case for adoption rather than substitution, and for concentrating a scarce resource rather than replacing it: instead of serving the substitutive function it more often serves in high-burden countries, chiefly compensating for a scarcity of trained radiologists or field staff, AI and digital-monitoring technology in the U.S. context works by augmenting an already substantial but geographically and demographically stretched public health and specialist workforce, and by clearing the high-volume tasks that stand between that workforce and the cases that most need it. Closing the remaining gap in the U.S. TB elimination will likely require both continued technological investment of this kind, calibrated to realistic effect sizes and validated across the demographic subgroups who carry a disproportionate share of the nation's remaining burden, and targeted workforce interventions, including telemedicine partnerships and specialist recruitment incentives, in the specific counties where this paper's ecological analysis shows capacity has not kept pace with need.

Acknowledgements

The authors thank the California Department of Public Health, Tuberculosis Control Branch, and the California Department of Health Care Access and Information for maintaining the publicly available surveillance and workforce datasets underlying the ecological analysis in this paper.

Mentor

Rohina Swaroop, MD, served as research mentor for this study. Any errors or omissions in this manuscript are the authors' own.

Author Contributions

S.P. conceived the review questions and ecological analysis, conducted the literature searches, citation tracking, and data extraction, performed the meta-analyses and statistical analyses, and drafted the manuscript. R.S. provided mentorship, methodological guidance, and critical revision of the manuscript.

Funding

This work received no external funding, grant support, or other financial assistance.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability

All data used in the systematic review components are drawn directly from the cited published sources listed in the References section. County-level TB and physician workforce data are publicly available from CDPH and HCAI, respectively (see References). Meta-analysis and regression calculations were performed by the authors from extracted published counts and public datasets and are available upon request. No new primary patient-level data were generated.

Supplemental Materials

Available from the corresponding author upon request: (S1) full county-level TB case and physician-density data for all 39 included California counties, with each county's regression residual; (S2) the list of 19 California counties excluded from the ecological analysis for non-analyzable physician-density estimates; (S3) the approximately 21 search queries used in the AI/ML review; and (S4) extended, non-pooled predictive-modeling metrics beyond Table 2.

References

- [1] Centers for Disease Control and Prevention. Executive commentary: TB in the U.S., 2024. U.S. Department of Health and Human Services; 2025a. <https://www.cdc.gov/tb-surveillance-report-2024/executive-commentary/index.html>.
- [2] Center for Infectious Disease Research and Policy. *CDC Data Suggest Small Decline in US TB Cases*. University of Minnesota; March 26, 2026. <https://www.cidrap.umn.edu/tuberculosis/cdc-data-suggest-small-decline-us-tb-cases>.
- [3] Centers for Disease Control and Prevention. TB by reporting areas: 2023 and 2024. Reported Tuberculosis in the United States, 2004; 2025b. <https://www.cdc.gov/tb-surveillance-report-2024/data/reporting-areas.html>.
- [4] California Department of Public Health, Tuberculosis Control Branch. Tuberculosis cases, rates per 100,000 population,

- and rank according to rate: reporting jurisdictions in California, 2024–2025 (Table 4). Provisional data as of February 1, 2026; 2025a.
- [5] Harris M, Qi A, Jeagal L, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. *Plos One*. 2019;14(9):e0221339. doi:10.1371/journal.pone.0221339.
 - [6] Garfein RS, Liu L, Cuevas-Mota J, et al. Tuberculosis treatment monitoring by video directly observed therapy in 5 health districts. *Em Infect Dis*. 2018;24(10):1806–1815. doi:10.3201/eid2410.180459.
 - [7] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA, 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71.
 - [8] Hua D, Nguyen K, Petrina N, et al. Benchmarking the diagnostic test accuracy of certified AI products for screening pulmonary tuberculosis in digital chest radiographs: preliminary evidence from a rapid review and meta-analysis. *Int J Med Inf*. 2023;177(9):105159. doi:10.1016/j.ijmedinf.2023.105159.
 - [9] Han ZL, Zhang YY, Li J, et al. A systematic review and meta-analysis of artificial intelligence software for tuberculosis diagnosis using chest X-ray imaging. *J Thorac Dis*. AME Publishing Company. 2025;17(5):3223–3237. <https://jtd.amegroups.org>.
 - [10] Gelaw SM, Kik SV, Ruhwald M, et al. Diagnostic accuracy of three computer-aided detection systems for detecting pulmonary tuberculosis on chest radiography when used for screening: analysis of an international, multicenter migrants screening study. *PLOS Glob Public Health*. San Francisco, California, USA: PLOS. 2023;3(7):e0000402. doi:10.1371/journal.pgph.0000402.
 - [11] Centers for Disease Control and Prevention. Pre-arrival medical screening and interventions for newly arrived refugees, immigrants, and migrants. In: *CDC Yellow Book*. 2026. <https://www.cdc.gov/yellow-book/hcp/refugees-immigrants-migrants/pre-arrival-medical-screening-and-interventions.html>.
 - [12] Businesswire. Qure.ai nets six new indications cleared by the FDA, taking the chest X-ray crown in the radiology AI race. San Francisco, California, USA. February 26, 2026. <https://www.businesswire.com>.
 - [13] Wang Y, Jiang Z, Liang P, Liu Z, Cai H, Sun Q. TB-DROP: deep learning-based drug resistance prediction of Mycobacterium tuberculosis utilizing whole genome mutations. *BMC Genom*. London, United Kingdom: BioMed Central/Springer Nature. 2024;25:167. doi:10.1186/s12864-024-10066-y.
 - [14] Hripcsak G, Knirsch CA, Jain NL, Pablos-Mendez A. Automated tuberculosis detection. *J Am Med Inf Asso*. 1997;4(5):376–381. doi:10.1136/jamia.1997.0040376.
 - [15] Ahamed S, Karuppasamy S, Chinnaiyan P. Clinical text classification for tuberculosis diagnosis using natural language processing and deep learning model with statistical feature selection technique. *Informatics*. Basel, Switzerland: MDPI. 2025;12(3):64. doi:10.3390/informatics12030064.
 - [16] Landsman D, Abdelbasit A, Wang C, et al. Cohort profile: St. Michael's Hospital Tuberculosis Database (SMH-TB), a retrospective cohort of electronic health record data and variables extracted using natural language processing. *medRxiv*. 2020. doi:10.1101/2020.09.11.20192419.
 - [17] Wang C, Li G, Ruan H et al. A predictive model for evaluating the risk of latent tuberculosis relapse via machine learning. *BMC Infect Dis*. 2025;25:1470. doi:10.1186/s12879-025-11881-8.
 - [18] Bartl L, Zeeb M, Kälén M, et al. Machine learning-based prediction of active tuberculosis in people with human immunodeficiency virus using clinical data. *Clin Infect Dis*. 2025. doi:10.1093/cid/ciaf14910.1093/cid/ciaf149. Advance online publication.
 - [19] Althomsons SP, Winglee K, Heilig CM, et al. Using machine learning techniques and national tuberculosis surveillance data to predict excess growth in genotyped tuberculosis clusters. *Am J Epidemiol*. 2022;191(11):1936–1943. doi:10.1093/aje/kwac117.
 - [20] Raz KM, Talarico S, Althomsons SP, et al. Molecular surveillance for large outbreaks of tuberculosis in the United States, 2014–2018. *Tuberculosis*. Amsterdam, Netherlands: Elsevier (Churchill Livingstone imprint). 2022;136(3):102232. doi:10.1016/j.tube.2022.102232.
 - [21] Tavaziva G, Harris M, Abidi SK, et al. Chest X-ray analysis with deep learning-based software as a triage test for pulmonary tuberculosis: an individual patient data meta-analysis of diagnostic accuracy. *Clin Infect Dis*. 2022;74(8):1390–1400. doi:10.1093/cid/ciab639.
 - [22] Garfein RS, Collins K, Muñoz F, et al. Feasibility of tuberculosis treatment monitoring by video directly observed therapy: a binational pilot study. *Int J Tuberc Lung Dis*. 2015;19(9):1057–1064. doi:10.5588/ijtld.14.0923.
 - [23] Centers for Disease Control and Prevention. Recommendations for use of video directly observed therapy during tuberculosis treatment—United States, 2023. *MMWR recommendations and reports*; 2023.
 - [24] Truong CB, Tanni KA, Qian J. Video-observed therapy versus directly observed therapy in patients with tuberculosis. *Am J Prev Med*. 2022;62(3):450–458. doi:10.1016/j.amepre.2021.10.013.
 - [25] Salcedo J, Rosales M, Kim JS, Nuno D, Suen S, Chang AH. Cost-effectiveness of artificial intelligence monitoring for active tuberculosis treatment: a modeling study. *Plos One*. 2021;16(7):e0254950. doi:10.1371/journal.pone.0254950.
 - [26] Areas Lisboa Netto T, Dellatorre Diniz B, Odutola P, et al. Video-observed therapy (VOT) vs directly observed therapy (DOT) for tuberculosis treatment: a systematic review on adherence, cost of treatment observation, time spent observing treatment and patient satisfaction. *PLOS Negl Trop Dis*. 2024;18(10):e0012565. doi:10.1371/journal.pntd.0012565.
 - [27] Garfein RS, Liu L, Cepeda J, et al. Asynchronous video directly observed therapy to monitor short-course latent tuberculosis infection treatment: results of a randomized controlled trial. *Open Forum Infect Dis*. 2024;11(4):ofae180. doi:10.1093/ofid/ofae180.
 - [28] Lam CK, McGinnis Pilote K, Haque A, Burzynski J, Chuck C, Macaraig M. Using video technology to increase treatment completion for patients with latent tuberculosis infection on 3-month isoniazid and rifapentine: an implementation study. *J Med Int Res*. 2018;20(11):e287. doi:10.2196/jmir.9825.
 - [29] Chuck C, Robinson E, Macaraig M, Alexander M, Burzynski J. Enhancing management of tuberculosis treatment with video directly observed therapy in New York City. *Int J Tuberc Lung Dis*. 2016;20(5):588–593. doi:10.5588/ijtld.15.0738.
 - [30] Perry A, Chitnis A, Chin A, et al. Real-world implementation of video-observed therapy in an urban TB program in the United States. *Int J Tuberc Lung Dis*. 2021;25(8):655–661. doi:10.5588/ijtld.21.0170.
 - [31] Burzynski J, Mangan JM, Lam CK, et al. In-person vs electronic directly observed therapy for tuberculosis

- treatment adherence: a randomized noninferiority trial. *JAMA Netw Open*. 2022;5(1):e2144210. doi:10.1001/jamanetworkopen.2021.44210.
- [32] Beeler Asay GR, Lam CK, Stewart B, et al. Cost of tuberculosis therapy directly observed on video for health departments and patients in. *Am J Public Health*. 2020;110(11):1696–1703. doi:10.2105/AJPH.2020.305877.
 - [33] Holzman SB, Zenilman A, Shah M. Advancing patient-centered care in tuberculosis management: a mixed-methods appraisal of video directly observed therapy. *Open Forum Infect Dis*. 2018;5(4):ofy046. doi:10.1093/ofid/ofy046.
 - [34] Centers for Disease Control and Prevention. *Core Curriculum on Tuberculosis: What the Clinician Should Know*. 7th ed. Atlanta, Georgia, USA: Centers for Disease Control and Prevention / U.S. Department of Health and Human Services; 2021.
 - [35] Philip R, Lee Institute for Health Policy Studies, University of California, San Francisco. *California Physicians Almanac*. California Health Care Foundation; 2025.
 - [36] California Department of Health Care Access and Information. Physicians actively working by specialty and patient care hours: Health Workforce License Renewal Survey, reference date April 3, 2025. HCAI; 2025.
 - [37] California Department of Public Health, Tuberculosis Control Branch. *About the data–Tuberculosis*. CDPH; 2025b.
 - [38] Oren E, Fiero MH, Barrett E, Anderson B, Nuñez M, Gonzalez-Salazar F. Detection of latent tuberculosis infection among migrant farmworkers along the US-Mexico border. *BMC Infect Dis*. London, United Kingdom: BioMed Central. 2016;16(1):630. doi:10.1186/s12879-016-1959-3.
 - [39] Centers for Disease Control and Prevention. Tuberculosis elimination priorities. National Center for HIV, Viral Hepatitis, STD, and TB Prevention; 2025c. <https://www.cdc.gov/nchhstp/priorities/tuberculosis-elimination.html>.
 - [40] California Health Care Foundation. *Shortchanged: Health Workforce Gaps in California–San Diego and Imperial Counties*. CHCF; 2020.
 - [41] Robsky KO, Hughes S, Kityamuwesi A, et al. Is distance associated with tuberculosis treatment outcomes? A retrospective cohort study in Kampala. *Uganda BMC Infect Dis*. 2020;20(1):406. doi:10.1186/s12879-020-05099-z.
 - [42] Nabha L. *Telemedicine in Infectious Diseases: Opportunities, Challenges and the Path Forward*. Science Speaks Blog, Infectious Diseases Society of America; March 30, 2026.
 - [43] California Department of Health Care Services. *Medi-Cal & Telehealth*. DHCS; 2026.
 - [44] Lippincott CK, Perry A, Munk E, Maltas G, Shah M. Tuberculosis treatment adherence in the era of COVID-19. *BMC Infect Dis*. 2022;22(1):800. doi:10.1186/s12879-022-07787-4.
 - [45] Bachina P, Lippincott CK, Perry A, et al. Programmatic adoption and implementation of video-observed therapy in Minnesota: prospective observational cohort study. *JMIR Form Res*. 2022;6(8):e38247. doi:10.2196/38247.

About the Author



Shreya Parimoo is a 10th-grade student at Stockdale High School in Bakersfield, California, with research interests in healthcare, psychology, and the application of technology to advance patient care. She is the founder of ShreyaVocabLabs.com, a free vocabulary-learning platform for high school students grounded in principles of cognitive science, work profiled in Seven Oaks Greet Magazine (“Meet Shreya Parimoo: She Turned a Study Frustration into Something Brilliant,” May 2026). Her writing on technology, identity, and mental health has appeared in Teen Ink. She hopes to improve health care access and outcomes in medically underserved regions such as California’s San Joaquin Valley.