SCIENTIFIC REP♥RTS

Serial testing for latent tuberculosis using QuantiFERON-TB Gold In-Tube: A Markov model

Mark W. Moses[1], Alice Zwerling[1], Adithya Cattamanchi[2], Claudia M. Denkinger[3], Niaz Banaei[4], Sandra V. Kik[5], John Metcalfe[2], Madhukar Pai[6] & David Dowdy[1]

Healthcare workers (HCWs) in low-incidence settings are often serially tested for latent TB infection (LTBI) with the QuantiFERON-TB Gold In-Tube (QFT) assay, which exhibits frequent conversions and reversions. The clinical impact of such variability on serial testing remains unknown. We used a microsimulation Markov model that accounts for major sources of variability to project diagnostic outcomes in a simulated North American HCW cohort. Serial testing using a single QFT with the recommended conversion cutoff (IFN-g > 0.35 IU/mL) resulted in 24.6% (95% uncertainty range, UR: 23.8–25.5) of the entire population testing false-positive over ten years. Raising the cutoff to >1.0 IU/mL or confirming initial positive results with a (presumed independent) second test reduced this false-positive percentage to 2.3% (95%UR: 2.0–2.6%) or 4.1% (95%UR: 3.7–4.5%), but also reduced the proportion of true incident infections detected within the first year of infection from 76.5% (95%UR: 66.3–84.6%) to 54.8% (95%UR: 44.6–64.5%) or 61.5% (95%UR: 51.6–70.9%), respectively. Serial QFT testing of HCWs in North America may result in tremendous over-diagnosis and over-treatment of LTBI, with nearly thirty false-positives for every true infection diagnosed. Using higher cutoffs for conversion or confirmatory tests (for initial positives) can mitigate these effects, but will also diagnose fewer true infections.

Health care workers (HCWs) are at risk for exposure to tuberculosis (TB) and are routinely tested for latent TB infection (LTBI) as part of most infection control programs in North American hospitals[1–3]. Traditionally, HCWs have been screened with the Tuberculin Skin Test (TST), and preventive therapy provided for those testing positive[1,4], but over the last decade, interferon-gamma release assays (IGRAs) have been introduced as an alternative to the TST[5–7]. IGRAs–including the QuantiFERON-TB Gold In-Tube (QFT, Qiagen, Valencia, USA) and T-SPOT. TB (Oxford Immunotec, Oxfordshire, UK)–offer several advantages over the TST, including improved specificity (particularly in BCG vaccinated populations)[7], *in vitro* testing, and the requirement that patients make only a single visit. As a result, many hospitals in the United States have begun to use IGRAs for serial testing of HCWs; this practice has revealed unexpectedly frequent conversions and reversions among IGRA results[8–13], the clinical implications of which remain unclear.

IGRA results are known to have multiple sources of variability (unrelated to TB infection), which may contribute to the observed variation in test results[14]. The clinical implications of such variability in the setting of serial testing remain unknown. As a result, it is uncertain how best to implement IGRAs for serial testing of HCWs–for example, to reduce the occurrence of false positives among individuals subjected to numerous repeat tests, serial testing might warrant a different cutoff for a conversion or confirmatory testing of initial positive results. A recent systematic review[15] summarized the evidence on potential sources of IGRA variability; evidence was strongest for boosting of IGRA (and specifically, QFT) responses by previously placed TSTs, delays in sample processing prior to incubation, and variations in the blood volume drawn. The evidence base for quantifying variability in serial testing by T-SPOT.TB (which has fundamental methodological differences in performance, compared to QFT) was somewhat weaker, and sources of T-SPOT.TB variability are likely to be different from those related to

[1]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA. [2]Department of Medicine, UCSF, San Francisco, USA. [3]Tuberculosis and Hepatitis Programme, FIND, Geneva, Switzerland. [4]Department of Medicine, Stanford University, Palo Alto, USA. [5]Evidence Team, KNCV, The Hague, Netherlands. [6]Department of Epidemiology & Biostatistics and McGill International TB Centre, McGill University, Montreal, Canada. Correspondence and requests for materials should be addressed to M.P. (email: madhukar.pai@mcgill.ca) or D.D. (email: ddowdy1@jhmi.edu)
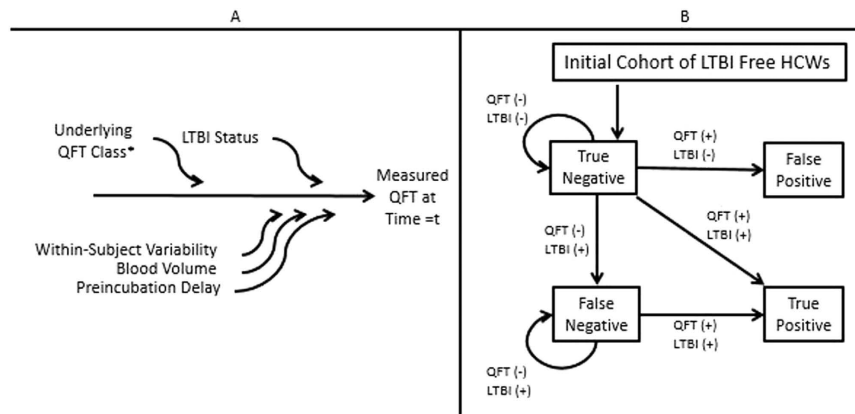
Figure 1. Markov model of serial QuantiFERON-TB (QFT) testing. Panel A illustrates the method by which each individual's measured QFT result was calculated at each year *t*. This result reflects the individual's initial underlying QFT "class", within-subject variability from year to year, and two additional sources of variability, as described in the methods. *Within-subject variability is incorporated through annual random draws from distributions based on each individual's initial underlying QFT class. Panel B shows the four-state Markov model. We assume that healthcare workers (HCWs) enter as true negatives and cycle between other states on an annual basis for 10 years based on yearly QFT result (according to different conversion cutoffs and testing strategies, as described in the Methods) and true latent tuberculosis infection (LTBI) status. QFT: Quantiferon-TB Gold IN-Tube test, LTBI: latent tuberculosis infection, TST: tuberculin skin test, HCW: health care workers.

QFT. We therefore constructed a simulation of hypothetical North American HCWs undergoing serial testing with QFT, taking into account best estimates of both within-subject variability (probabilities of conversion and reversion in populations at very low risk of TB infection) and two important additional sources of variability (incubation delays and variation in blood volume). We use this simulated population to project the outcomes of serial QFT testing over a period of ten years.

Methods

Overview. We used a microsimulation Markov model to project diagnostic outcomes in a simulated HCW cohort under different serial screening strategies. The simulated cohort comprised 10,000 hypothetical HCWs at a representative North American institution undergoing serial LTBI screening using QFT every year over a period of ten years. The model incorporates the major sources of variability in serial QFT results as well as different serial screening strategies, including different cutoffs for a QFT conversion and confirmatory testing with different degrees of correlation (modeled as Pearson correlation coefficients) between initial and confirmatory results (Fig. 1).

Modeling Approach. To accomplish the model objectives, the model first assigns each individual HCW to one of three "QFT classes" based on the IFN-gamma response (0.00–0.19 International Units (IU)/mL ["class 1"], 0.20–0.99 IU/mL ["class 2"], or ≥ 1.00 IU/mL ["class 3"]), reflecting the distribution of initial QFT results from a large, multi-center study of HCWs at low risk of TB infection who were tested serially under semi-controlled conditions[8]. This "class" then dictates the HCW's future "underlying" QFT results on a yearly basis under random within-subject variability alone, as follows. First, we take the initial QFT results from the aforementioned study, dividing these QFT results into the same three classes. Thus, for example, individuals in this study whose initial QFT result was 0.00–0.19 IU/mL would be assigned to "class 1". Then, we fit a gamma distribution to the repeat QFT results in each class, again using the serial testing data from the aforementioned study. (For example, if nine of ten patients whose initial QFT results were <0.20 IU/mL in the source study had a repeat value of 0.00 IU/mL, and one had a repeat value of 0.40 IU/mL, we would fit a gamma distribution to those values, to apply to the simulated HCWs in class 1.) Gamma distributions were selected because their lower bound is zero, but results are distributed in a right-skewed fashion with an infinite upper bound. The cutoffs for the three classes were selected based on natural cut-points in the data[8], and to be maximally conservative (i.e., reducing the probability of false conversions on serial testing). Finally, to account for within-subject variability within each class, we assign a different "underlying" QFT value to each simulated HCW on a yearly basis, each yearly value representing a random draw from the distribution corresponding to that HCW's class. We assume for simplicity that, at intake, all members of the cohort are free of actual TB infection.

In subsequent years, in addition to the within-subject variability as described above, an annual risk of true TB infection is applied. In the baseline scenario, we assume a high-risk HCW population in a North American healthcare institution in which 0.1% of the cohort acquires LTBI each year[16,17]. This annual risk of infection is varied in scenario analyses. The process of true infection is modeled by replacing the previous "underlying" QFT result with a new "underlying" value selected by randomly sampling from the distribution of positive QFT test results in the same study (i.e., assuming that the distribution of positive results in this study is similar to the distribution of true-positive results following true TB infection) [8].

After assigning an "underlying" QFT value to each simulated HCW in each year (to account for within-subject variability and the possibility of true infection)[8,10,18], the model incorporates additional variation in QFT results according to two additional sources of variability (Fig. 1A): (1) variation in blood volume: $1\,mL +/- 0.2\,mL$[19] (assuming variation in either direction on 5% of all blood draws, for a total of 10%); and (2) pre-analytical delay of six or 12 hours[20,21] (assuming each occurs in 5% of blood draws, for a total of 10%). We performed a sensitivity analysis that also included variability from TST boosting[22–24], assuming boosting in 2% of all blood draws. We considered TST boosting an unlikely source of variability in individuals enrolled in a QFT screening program and thus excluded it from all other analyses. Each of these sources of variability is modeled as a beta distribution (with defined upper and lower bounds); values and ranges for change in QFT results are presented in Table 1. For HCWs randomly selected to experience each source of variability, a value would be selected from the corresponding distribution and added (or subtracted) from the "underlying" QFT value to obtain the final measured QFT value in that year. Variability from blood volume, pre-analytical delay, and TST boosting is assumed to occur independently between each other and over time, such that the occurrence of one of these sources of variability in one year has no bearing on whether it or another source of variability occurs again in the current year or in a subsequent year.

The model can thus be represented mathematically as follows. Given an "underlying" QFT value (a_{st}) in each year, additional variability in QFT results is added according to Equation 1:

$$Measured\ QFT_t = a_{st} + \sum_{k=1}^{5} \lambda_{kt} x_{skt}$$

(1)

where a_{st} is the "underlying" QFT value (conditional on TB infection status) as described above, k represents the additional sources of variability for which sufficient data were available in the systematic review (pre-incubation delay, blood volume, plus true TB infection[15]), λ_{kt} is an indicator variable drawn from a binomial distribution corresponding to the probability that k occurred in year t (1 = yes, 0 = no), and x_{skt} is a value randomly drawn from a beta distribution fit to the measured variability in QFT results among people whose initial QFT value was in class s, and whose repeat tests were known to have been exposed to each item k, using data from the systematic review[15].

After incorporating these sources of variability and true LTBI infection, measured QFT results are truncated between 0 and 10 IU/mL (corresponding to manufacturer's guidelines), and a test result (positive or negative) is assigned at each year, according to different cutoffs for a positive result and different testing strategies (single test, or initial positive test requiring confirmation with a repeat test). Based on the test result and true LTBI status (incorporating this annual risk of TB infection) at each year, individuals in the population cycle between four Markov states: true negative, false negative, true positive, and false positive (Fig. 1B). All eligible (previously negative) individuals are assumed to undergo testing each year, and individuals who test positive (according to the selected cutoff and testing strategy) are assumed to be offered treatment for LTBI and therefore not subjected to further testing. QFT reversions are thus not explicitly modeled, as they have no bearing on outcome. Our primary outcome is the proportion of the population in each of the four states (with particular interest in the false-positive state) at the end of ten years. Key secondary outcomes include the number of QFT tests that would be performed over ten years and the proportion of people with true TB infection who would be properly identified as true positive in the first year after the infection occurred (i.e., those at greatest risk of progression to active disease and therefore greatest priority for preventive therapy, despite the relatively low accuracy of QFTs in predicting progression to active TB[15]).

Sensitivity Analyses.
We evaluated outcomes using the manufacturer's recommended conversion cutoff for a positive test result (≥ 0.35 IU/mL), as well as higher cutoffs at regular intervals (≥ 0.5, ≥ 0.75, and ≥ 1.0 IU/mL). In addition, we modeled the correlation between initial and confirmatory tests in separate analyses as a Pearson correlation coefficient of 0.0 (presumed independent tests), 0.5, and 0.9 (best reflection of available data)[25]. We evaluated a wide range of thresholds for defining the three QFT classes (not shown), choosing the thresholds above based on natural breaks in the data distribution and a desire to be conservative in our assumptions. To evaluate the impact of the different modeled sources of variability on the outcomes of serial testing, we performed additional model runs in which each source of variability was removed from the model. Finally, we ran 1,000 separate model simulations (of 10,000 individuals each) to construct 95% uncertainty ranges (95% URs) as the 2.5th and 97.5th percentile of results across all simulations.

Results
In the base case scenario (0.1% annual risk of TB infection), 0.9% (95%UR: 0.7–1.1%) of the population experienced a true incident TB infection within ten years. Under a testing strategy in which a QFT conversion was defined as an initial value of <0.35 IU/mL followed by a subsequent test ≥ 0.35 IU/mL–and with no confirmatory testing–24.6% (95%UR: 23.8–25.5%) of the population was projected to test false positive by the end of ten years, versus 0.85% (95% UR: 0.68–1.04%) true positives: a ratio of 29 false positives for every true positive TB infection diagnosed. Under this strategy, 23.5% (95%UR: 15.4 −33.7%) of patients with true incident infections were missed on the first test after infection. Including variability due to TST boosting nearly doubled the number of false positives, while removing variability due to incubation delays and blood volume (i.e., limiting variability to within-subject variability) did not materially change results (data not shown).

We then evaluated different strategies for reducing the false positive: true positive ratio, by instituting a higher cutoff for conversion and/or requiring confirmation of positive tests, under different assumptions about the correlation between the initial and confirmatory test (Fig. 2). Increasing the cutoff for a conversion to 1.0 IU/mL reduced the proportion of the population testing false-positive in ten years to 2.3% (95%UR: 2.0–2.6%)

Parameter	Value*	Distribution	Source
Distributions Used to Account for Within-Subject Variability:			
Initial QFT (to determine QFT class)	0.00 (0.05)	Gamma	8
Annual "Underlying" QFT Result			8,10,18
QFT class 0–0.19 IU/mL	0.00 (0.11)	Gamma	
QFT class 0.20 IU/mL–0.99 IU/mL	0.00 (0.61)	Gamma	
QFT class ≥1.00 IU/mL	10 (2.82)	Gamma	
Additional Sources of Variability:			
Probability of Event			Assumed
6-Hour Pre-incubation Delay	0.05	Binomial	
12-Hour Pre-incubation Delay	0.05	Binomial	
0.8 mL Blood Sample	0.05	Binomial	
1.2 mL Blood Sample	0.05	Binomial	
Annual Risk of TB Infection	0.01	Binomial	
Change in QFT Result According To:			
6 Hour Pre-analytical Delay			20,21
QFT class 0–0.19 IU/mL	0.02 [−0.14, 4.23]	Beta	
QFT class 0.20 IU/mL–0.99 IU/mL	−0.16 [−0.60, 0.19]	Beta	
QFT class ≥1.00 IU/mL	−2.03 [−7.97, 0.86]	Beta	
12 Hour Pre-analytical Delay			20,21
QFT class 0–0.19 IU/mL	0.01 [−0.13, 0.61]	Beta	
QFT class 0.20 IU/mL–0.99 IU/mL	−0.16 [−0.89, 0.06]	Beta	
QFT class ≥1.00 IU/mL	−2.85 [−7.16, 0.00]	Beta	
0.8 mL Blood Volume			19
QFT class 0–0.19 IU/mL	0.03 [−0.02, 0.33]	Beta	
QFT class 0.20 IU/mL–0.99 IU/mL	0.35 [−0.11, 1.36]	Beta	
QFT class ≥1.00 IU/mL	2.68 [−1.11, 7.32]	Beta	
1.2 mL Blood Volume			19
QFT class 0–0.19 IU/mL	0.01 [−0.11, 0.59]	Beta	
QFT class 0.20 IU/mL–0.99 IU/mL	−0.28 [−0.42, −0.05]	Beta	
QFT class ≥1.00 IU/mL	−0.97 [−2.13, 0.09]	Beta	
TST Boosting (excluded from base case)			8,22–24
QFT class 0–0.19 IU/mL	0.04 [−0.16, 7.46]	Beta	
QFT class 0.20 IU/mL–0.99 IU/mL	1.13 [−0.85, 9.22]	Beta	
QFT class ≥1.00 IU/mL	2.86 [−1.56, 8.59]	Beta	

Table 1. Parameter Distributions for a Markov Model of Serial QFT Testing. *Gamma distributions show mode (standard deviation), each rounded to the nearest 0.01. Binomial distributions show probabilities. Beta distributions show mode (minimum, maximum), assuming an alpha (shape) value of 4. QFT: Quantiferon, TST: Tuberculin skin test, TB: tuberculosis.

(Figs 2 and 3), but at the expense of reducing the percentage of true TB infections diagnosed within the first year after infection to 54.8% (95%UR: 44.6–64.5%).

The value of a confirmatory test depended strongly on the modeled correlation between initial and confirmatory test results. If the hypothetical correlation between initial and confirmatory test was high, outcomes were very similar whether or not confirmatory testing was performed. (See Fig. 2, "0.9" bars, and Fig. 3, middle and right-hand panels.) By contrast, if confirmatory tests were uncorrelated with the initial test result (Fig. 2, "0.0" bars, and Fig. 3, left panel), confirmatory testing even at a threshold of 0.35 IU/mL reduced the ten-year false positive percentage from 24.6% to 4.1% (95%UR: 3.7–4.5%). In this case, confirmatory testing also reduced the proportion of true infections detected within one year from 76.5% to 61.5% (95%UR: 51.6–70.9%).

At a higher annual risk of TB infection, the proportions of true positive tests increased, but the number of false positive tests remained the same, such that the false positive: true positive ratio declined (Table 2). In an extreme scenario of 1.0% annual risk of TB infection per year (1 of every 100 individuals tested actually has incident TB infection), assuming a correlation between initial and confirmatory testing of 0.5, a cutoff of 1.0 IU/mL, and only within-subject variability, the number of individuals testing true positive (7.95%) was much greater than the number testing false positive (0.55%). As a result, the post-test probability of a positive test representing true TB infection in this scenario was 0.94, versus 0.035 in the base case (0.1% annual risk of infection, no confirmatory testing, cutoff of 0.35 IU/mL).

Discussion

Using QFT for annual serial testing among HCWs in a low-incidence setting such as North America may lead to tremendous over-diagnosis of QFT conversions, and likely over-treatment of patients testing false-positive for
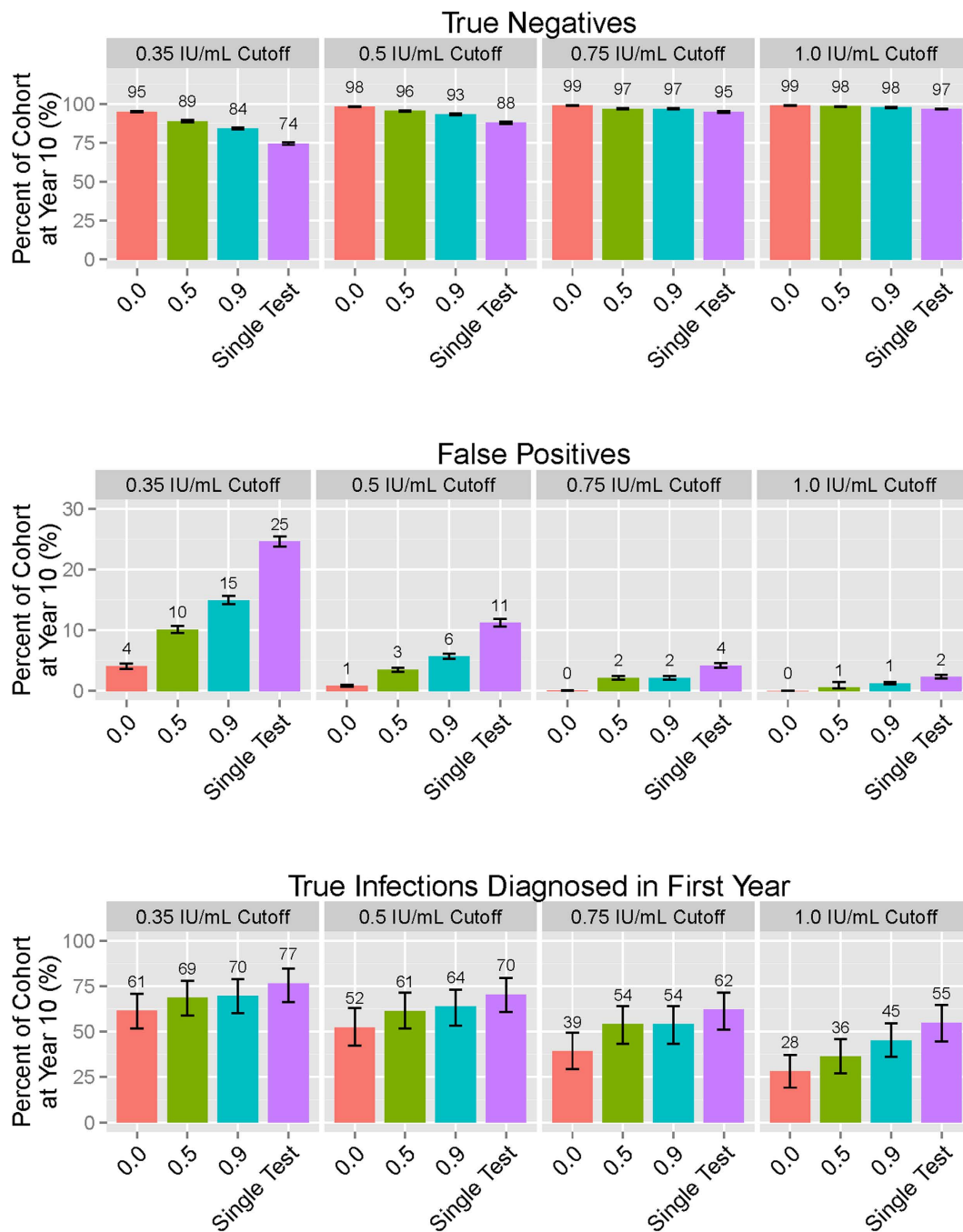
Figure 2. Cumulative Outcomes of Serial QuantiFERON-TB (QFT) Testing After Ten Years. The three panels respectively display proportions of the initial cohort classified as true negatives, false positives, and true infections diagnosed in the first year of infection at the end of ten years for each correlation-cutoff scenario. Correlations represent the Pearson correlation between initial test and confirmatory test (given if initial test was positive).

LTBI. Using the manufacturer's recommended diagnostic cutoff as the cutoff for conversion, and without a confirmatory test, resulted in 24.6% of HCWs being diagnosed with false positive conversions over 10 years of serial testing. If this number of individuals were unnecessarily treated, adverse events and costs associated with treatment, monitoring, and side effects would be substantial, given the millions of HCWs who undergo serial testing in North America. Based on the present analysis, serial testing in the absence of confirmatory tests or alternative conversion cutoffs should therefore be discouraged in low-incidence settings.

Using a higher cutoff for a QFT conversion, or confirming positive results with a second test (assuming low correlation between those tests) can substantially reduce the number of false-positive results, but at the consequence of diagnosing fewer people within the first year after true infection. Furthermore, in settings with a low annual risk of TB infection (1 per 1,000 per year or less), the positive predictive value of serial QFT testing, even
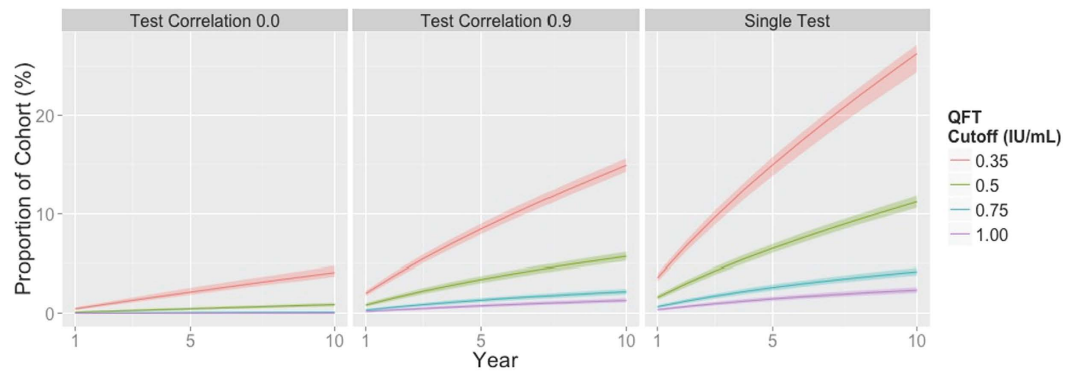
Figure 3. Proportion of a Hypothetical Population of Healthcare Workers Testing False-Positive for Tuberculosis Infection Using QuantiFERON-TB. Lines show the cumulative proportion of healthcare workers expected to test false-positive for latent tuberculosis infection (LTBI) if tested serially with QuantiFERON-TB on an annual basis, using different cutoffs for a positive test (shown in different colors) and different testing strategies (independent confirmatory test on the left, confirmatory test with Pearson correlation of 0.9 in the middle, and no confirmatory testing on the right).

Annual Risk of TB infection (%)	0.10%	0.50%	1.00%
Outcomes at 10 Years (%)			
True Negative	98.4 (98.2, 98.7)	94.6 (94.1, 95.0)	89.9 (89.3, 90.5)
False Negative	0.17 (0.09, 0.25)	0.82 (0.64, 0.98)	1.57 (1.35, 1.83)
True Positive	0.82 (0.66, 1.01)	4.04 (3.66, 4.45)	7.95 (7.45, 8.49)
False Positive	0.57 (0.44, 0.73)	0.56 (0.42, 0.72)	0.55 (0.4, 0.7)
Number of QFT Tests Performed*	99,800 (99,800, 99,900)	99,100 (98,900, 99,300)	98,100 (97,900, 98,300)
TB Infections Diagnosed in the First Year	36.8 (27.3, 45.6)	36.7 (32.3, 41.3)	36.7 (33.7, 39.8)

Table 2. Outcomes of Serial QFT Testing Assuming Only Within-Subject Variability. Assuming a cutoff for conversion of 1.0 IU/mL and a confirmatory test with a Pearson correlation of 0.5. *Tests required to evaluate 10,000 individuals, followed over 10 years.

with high cutoffs and independent confirmatory testing, can be 60% or lower. For example, if preventive therapy costs $430 per treatment course[26], and assuming that preventive therapy is taken as prescribed, a serial QFT strategy using a single annual test with a cutoff of 0.35 IU/mL (29 false-positives for every true-positive LTBI case) could cost $13,000 per true-positive LTBI case treated, in treatment costs alone–a figure comparable to the cost of treating a case of active TB[27]. In a population of 10,000 individuals, if all (2350) individuals testing false-positive under this strategy were to take preventive therapy, 87–115 people (3.7–4.9%) would be expected to experience an adverse event sufficient to result in permanent drug discontinuation[28]. Using a confirmatory test with correlation of 0.5 and a cutoff of 0.35 IU/mL could reduce the cost per true-positive LTBI case treated to $5,400, but at the expense of approximately two more cases of TB infection going undiagnosed (double the false negative rate of an annual single QFT serial screening strategy). Decisions of whether to adopt such higher cutoffs in the context of serial testing must therefore consider the underlying risk of TB infection, weighing the benefits of fewer false-positive tests against the potential harm of failing to promptly diagnose individuals with true TB infection.

Confirmatory testing has been discussed before[29] but, to date, no guidance exists regarding its appropriate implementation. Data to describe the true correlation between initial and confirmatory QFT test results and estimates of cost-effectiveness could be helpful in this regard. Despite this uncertainty, many hospitals have implemented confirmatory testing as a method to reduce false positive results and avoid unnecessary treatment[30,31]. Our model demonstrates the wide range of clinical outcomes possible, depending on the true correlation of results between initial and confirmatory testing. Given that confirmatory QFT testing is widely employed, more accurately characterizing this correlation should be a high priority for future research.

As with any modeling study, we are limited by the accuracy and appropriateness of the model and parameter assumptions employed. While we performed sensitivity analyses around the sources of variability, distributions employed, cutoffs for the underlying QFT categories and other parameters, we cannot exclude the possibility of having not modeled or not modeled accurately a particular unknown source of variability. In the assignment of initial QFT values and the distribution of within-subject variability, our analysis relied heavily on data from the largest study of IGRAs in North American HCWs [8]; however, this study collated QFT measurements from laboratories employing potentially diverse methodologies. Such pooling of data from different settings may increase uncertainty in the data underlying our results. In choosing categories and cutoffs we employed conservative estimates likely to be favorable to QFT; to the extent that we have underestimated the probability of QFT conversions under each source of variability, our results may underestimate the number of false-positive results that could be expected under each serial QFT testing strategy. Furthermore, as the focus of this modeling work was on serial

QFT testing, we did not consider serial TST or T-SPOT.TB algorithms; further work comparing serial QFT to serial TST or T-SPOT.TB testing–incorporating year-to-year variability in each test–could add further value.

This model demonstrates the potential pitfalls of blindly using cutoffs designed for single tests in the setting of serial testing, particularly in settings of low TB incidence (i.e., low positive predictive value of QFT conversions). Simulation can help us understand these consequences in the absence of large prospective cohort studies. In realistic settings of low risk of TB infection, even with high cutoffs and confirmation, the post-test probability of a positive QFT test on serial testing may still only approach 60%, while annual testing with a single QFT using the recommended diagnostic cutoff may cause over 25% of individuals to have a false-positive test over a span of ten years. The value of serial testing of HCWs should therefore be discouraged in the North American setting, where TB incidence is now at an all-time low.

References

1. Jensen, P. A., Lambert, L. A., Iademarco, M. F. & Ridzon, R. & Cdc. Guidelines for preventing the transmission of Mycobacterium tuberculosis in health-care settings, 2005. *MMWR. Recommendations and reports: Morbidity and mortality weekly report. Recommendations and reports/Centers for Disease Control* **54,** 1–141 (2005).
2. Linas, B. P., Wong, A. Y., Freedberg, K. A. & Horsburgh, C. R. Jr. Priorities for screening and treatment of latent tuberculosis infection in the United States. *American journal of respiratory and critical care medicine* **184,** 590–601, doi: 10.1164/rccm.201101-0181OC (2011).
3. Mazurek, G. H. *et al.* Updated guidelines for using Interferon Gamma Release Assays to detect Mycobacterium tuberculosis infection-United States, 2010. *MMWR. Recommendations and reports: Morbidity and mortality weekly report. Recommendations and reports/Centers for Disease Control* **59,** 1–25 (2010).
4. Targeted tuberculin testing and treatment of latent tuberculosis infection. American Thoracic Society. MMWR. Recommendations and reports: Morbidity and mortality weekly report. Recommendations and reports/Centers for Disease *Control* **49,** 1–51 (2000).
5. Pai, M. *et al.* Serial testing of health care workers for tuberculosis using interferon-gamma assay. *American journal of respiratory and critical care medicine* **174,** 349–355, doi: 10.1164/rccm.200604-472OC (2006).
6. Pai, M. & Menzies, D. The new IGRA and the old TST: making good use of disagreement. *American journal of respiratory and critical care medicine* **175,** 529–531, doi: 10.1164/rccm.200701-024ED (2007).
7. Pai, M., Zwerling, A. & Menzies, D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. *Ann Intern Med* **149,** 177–184 (2008).
8. Dorman, S. E. *et al.* Interferon-gamma release assays and tuberculin skin testing for diagnosis of latent tuberculosis infection in healthcare workers in the United States. *American journal of respiratory and critical care medicine* **189,** 77–87, doi: 10.1164/rccm.201302-0365OC (2014).
9. Fong, K. S. *et al.* Challenges of Interferon-gamma Release Assay Conversions in Serial Testing of Health Care Workers in a Tuberculosis Control Program. *Chest*, doi: 10.1378/chest.11-0992 (2012).
10. Joshi, M., Monson, T. P. & Woods, G. L. Use of interferon-gamma release assays in a health care worker screening program: experience from a tertiary care centre in the United States. *Canadian respiratory journal: journal of the Canadian Thoracic Society* **19,** 84–88 (2012).
11. LoBue, P. A. & Catanzaro, A. Effectiveness of a nosocomial tuberculosis control program at an urban teaching hospital. *Chest* **113,** 1184–1189 (1998).
12. Pai, M. & Elwood, K. Interferon-gamma release assays for screening of health care workers in low tuberculosis incidence settings: dynamic patterns and interpretational challenges. *Canadian respiratory journal: journal of the Canadian Thoracic Society* **19,** 81–83 (2012).
13. Zwerling, A. *et al.* Repeat IGRA Testing in Canadian Health Workers: Conversions or Unexplained Variability? *PloS one* **8,** e54748, doi: 10.1371/journal.pone.0054748 (2013).
14. Banaei, N., Gaur, R. L. & Pai, M. Interferon-gamma release assays for latent tuberculosis: what are the sources of variability? *Journal of clinical microbiology*, doi: 10.1128/JCM.02803-15 (2016).
15. Tagmouti, S. *et al.* Reproducibility of interferon gamma (IFN-gamma) release Assays. A systematic review. *Annals of the American Thoracic Society* **11,** 1267–1276, doi: 10.1513/AnnalsATS.201405-188OC (2014).
16. Blumberg, H. M. *et al.* Preventing the nosocomial transmission of tuberculosis. *Ann Intern Med* **122,** 658–663 (1995).
17. Lambert, L. A., Pratt, R. H., Armstrong, L. R. & Haddad, M. B. Tuberculosis among healthcare workers, United States, 1995–2007. *Infection control and hospital epidemiology* **33,** 1126–1132, doi: 10.1086/668016 (2012).
18. Gandra, S. *et al.* Questionable effectiveness of the QuantiFERON-TB Gold Test (Cellestis) as a screening tool in healthcare workers. *Infection control and hospital epidemiology* **31,** 1279–1285, doi: 10.1086/657336 (2010).
19. Gaur, R. L., Pai, M. & Banaei, N. Impact of blood volume, tube shaking, and incubation time on reproducibility of QuantiFERON-TB gold in-tube assay. *Journal of clinical microbiology* **51,** 3521–3526, doi: 10.1128/JCM.01627-13 (2013).
20. Doberne, D., Gaur, R. L. & Banaei, N. Preanalytical delay reduces sensitivity of QuantiFERON-TB gold in-tube assay for detection of latent tuberculosis infection. *Journal of clinical microbiology* **49,** 3061–3064, doi: 10.1128/JCM.01136-11 (2011).
21. Herrera, V., Yeh, E., Murphy, K., Parsonnet, J. & Banaei, N. Immediate incubation reduces indeterminate results for QuantiFERON-TB Gold in-tube assay. *Journal of clinical microbiology* **48,** 2672–2676, doi: 10.1128/JCM.00482-10 (2010).
22. Baker, C. A., Thomas, W., Stauffer, W. M., Peterson, P. K. & Tsukayama, D. T. Serial testing of refugees for latent tuberculosis using the QuantiFERON-gold in-tube: effects of an antecedent tuberculin skin test. *The American journal of tropical medicine and hygiene* **80,** 628–633 (2009).
23. Cummings, K. J. *et al.* Prospective comparison of tuberculin skin test and QuantiFERON-TB Gold In-Tube assay for the detection of latent tuberculosis infection among healthcare workers in a low-incidence setting. *Infection control and hospital epidemiology* **30,** 1123–1126, doi: 10.1086/644754 (2009).
24. Ritz, N. *et al.* Absence of interferon-gamma release assay conversion following tuberculin skin testing. *The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease* **15,** 767–769, doi: 10.5588/ijtld.10.0339 (2011).
25. Ringshausen, F. C. *et al.* Within-subject variability of Mycobacterium tuberculosis-specific gamma interferon responses in German health care workers. *Clinical and vaccine immunology: CVI* **18,** 1176–1182, doi: 10.1128/CVI.05058-11 (2011).
26. Oxlade, O., Schwartzman, K. & Menzies, D. Interferon-gamma release assays and TB screening in high-income countries: a cost-effectiveness analysis. *The International Journal of Tuberculosis and Lung Disease* **11,** 16–26 (2007).
27. Marks, S. M. *et al.* Treatment practices, outcomes, and costs of multidrug-resistant and extensively drug-resistant tuberculosis, United States, 2005–2007. *Emerg Infect Dis* **20,** 812–821 (2014).
28. Sterling, T. R. *et al.* Three months of rifapentine and isoniazid for latent tuberculosis infection. *New England Journal of Medicine* **365,** 2155–2166 (2011).
29. Daley, C. L. *et al.* A summary of meeting proceedings on addressing variability around the cut point in serial interferon-gamma release assay testing. *Infection control and hospital epidemiology* **34,** 625–630, doi: 10.1086/670635 (2013).

30. Slater, M. L., Welland, G., Pai, M., Parsonnet, J. & Banaei, N. Challenges with QuantiFERON-TB Gold assay for large-scale, routine screening of U.S. healthcare workers. *American journal of respiratory and critical care medicine* **188,** 1005–1010, doi: 10.1164/rccm.201305-0831OC (2013).
31. Thanassi, W. *et al.* Delineating a Retesting Zone Using Receiver Operating Characteristic Analysis on Serial QuantiFERON Tuberculosis Test Results in US Healthcare Workers. *Pulmonary medicine* **2012,** 291294, doi: 10.1155/2012/291294 (2012).

Acknowledgements

Author Contributions

D.D. and M.P. conceived the experiments. D.D. designed the experiments. M.W.M. conducted the experiments. M.W.M., A.Z. and D.D. contributed to data analysis. A.Z. led drafting of the manuscript. M.W.M., A.Z., A.C., C.M.D., N.B., S.V.K., J.M., M.P. and D.D. reviewed, provided feedback, and approved the submitted manuscript.

Additional Information