HHS Public Access

Author manuscript

Int J Tuberc Lung Dis. Author manuscript; available in PMC 2022 May 26.

Published in final edited form as:

Int J Tuberc Lung Dis. 2018 May 01; 22(5): 518–523. doi:10.5588/ijtld.17.0650.

Paradox of serial interferon-gamma release assays: variability width more important than specificity size

J. E. Stout[*], **R. Belknap**[†], **Y-J. Wu**[‡], **C. S. Ho**[§], **Tuberculosis Epidemiologic Studies Consortium**

[*]Division of Infectious Diseases and International Health, Department of Medicine, Duke University Medical Center, Durham, North Carolina,

[†]Denver Health, Denver, Colorado, Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

[‡]Northrop Grumman, Atlanta, Georgia, Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

[§]Field Services Branch, Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

SUMMARY

SETTING: Serial screening for latent tuberculous infection (LTBI) is commonly performed in certain populations, such as health care workers. The high apparent conversion rate in some studies of interferon-gamma release assays is puzzling given the claimed high specificity of these tests.

OBJECTIVE: To understand how test-retest variability, specificity, and underlying LTBI prevalence affect observed outcomes of repeated testing for LTBI.

DESIGN: Mathematical model assuming constant test sensitivity and specificity over time and no new infections.

RESULTS: Test-retest variability had a large effect on the observed proportion of conversions (initial negative test, followed by a positive test) and reversions (initial positive test, followed by a negative test). For example, a test with 70% specificity and 5% test-retest variability would be associated with a conversion rate of 3.7% and a reversion rate of 7.7%, while a test with 95% specificity but 10% test-retest variability would be associated with a conversion rate of 5.5% and a reversion rate of 57%, assuming that both tests are 80% sensitive and underlying LTBI prevalence was 5%.

Correspondence to: Jason E Stout, Division of Infectious Diseases and International Health, Department of Medicine, Duke University Medical Center, Box 102359-DUMC, Durham, NC 27710 USA. jason.stout@dm.duke.edu.

References in this manuscript to any specific commercial products, process, service, manufacturer, or company does not constitute their endorsement or recommendation by the US Government or the CDC. The findings and conclusions are those of the authors and do not necessarily represent the views of the CDC.

Conflicts of interest: none declared.

CONCLUSION: Test-retest variability is a key parameter that should be reported for tests used for serial screening for LTBI. Reducing test-retest variability can reduce false-positive and false-negative results.

RÉSUMÉ

Le dépistage en série de l'infection tuberculeuse latente (LTBI) est habituellement réalisé dans certaines populations, comme le personnel de santé. Le taux de conversion, apparemment élevé, dans certaines études des tests de libération de l'interféron gamma est surprenant, sachant que ces tests revendiquent une spécificité éleveé.

Comprendre comment la variabilité, la spécificité du test et de sa répétition, et la prévalence de la LTBI sous-jacente affectent les résultats observés des tests répétés pour la LTBI.

Un modéle mathématique supposant une sensibilité et une spécificité constantes du test dans le temps et l'absence de nouvelles infections.

La variabilité du test et de sa répétition a eu un effet important sur la proportion observée de conversions (test initial négatif suivi par untest positif) et de réversions (test initial positif suivi par un test négatif). Un test ayant une spécificité de 70% et une variabilité de 5% lors des répétitions serait associé avec un taux de conversion de 3,7% et un taux de réversion de 7,7%, tandis qu'un test ayant une spécificité de 95% mais 10% de variabilité lors des répétitions serait associé avec un taux de conversion de 5,5% et un taux de réversion de 57%, en supposant que les deux tests ont une sensibilité de 80% et que la prévalence de la LTBI sous-jacente a été de 5%.

La variabilité du test est un paramétre majeur qui devrait être pris en compte pour les tests utilisés dans le dépistage en série de la LTBI. Réduire la variabilité du test peut réduire les résultats faux positifs et faux négatifs.

RESUMEN

En determinados grupos poblacionales como los profesionales de salud se practica con frecuencia la detección seriada de la infección tuberculosa latente (LTBI). La alta tasa de conversión aparente en algunos estudios con pruebas de liberación de interferón γ es desconcertante dada la alta especificidad aducida de la prueba.

Comprender en qué medida la variabilidad al repetir la prueba, su especificidad y la prevalencia subyacente de LTBI influyen sobre los resultados de la detección seriada de la LTBI.

Se llevó a cabo una modelización matemática suponiendo una sensibilidad y una especificidad constantes de la prueba con el transcurso del tiempo y la ausencia de nuevas infecciones.

La variabilidad al repetir la prueba tuvo un efecto importante sobre la proporción observada de conversiones (prueba inicial negativa seguida de una prueba positiva) y de reversiones (prueba inicial positiva seguida de una prueba negativa). Por ejemplo, una prueba con una especificidad de 70% y una variabilidad al repetir la prueba de 5% se asociaría con una tasa de conversión de 3,7% y una tasa de reversión de 7,7%, pero una prueba con una especificidad de 95% y una variabilidad al repetir la prueba de 10% se asociaría con una tasa de conversión de 5,5% y una tasa de reversión de 57%, al suponer que ambas pruebas tienen una sensibilidad de 80% y que la prevalencia subyacente de LTBI es 5%.

La variabilidad al repetir la prueba constituye un parámetro fundamental y sería necesario notificarlo con las pruebas analı´ticas que se utilizan en la detección sistemática seriada de la LTBI. Cuando disminuye la variabilidad al repetir la prueba se pueden limitar los resultados positivos falsos y negativos falsos.

Keywords

tuberculosis; diagnostic tests; test-retest reliability; IGRAs

INTERFERON-GAMMA RELEASE ASSAYS (IGRAs) are increasingly being used as screening tests for latent tuberculous infection (LTBI). Two IGRAs are commercially available in the United States: T-SPOT®.*TB* (T-SPOT, Oxford Immunotec, Abingdon, UK) and QuantiFERON® Gold In-Tube (QFT, Qiagen, Hilden, Germany). One of the purported advantages of these tests is improved specificity over the tuberculin skin test (TST),[1] which can reduce unnecessary treatment of persons not truly infected with *Mycobacterium tuberculosis*. While the specificities of both the IGRAs and the TSTare reportedly high (over 95%) among predominantly non-bacille Calmette-Guérin (BCG) vaccinated persons in low-prevalence settings,[2] the specificity of the TST has been reported to be much lower (59% in one meta-analysis[3]) among BCG-vaccinated persons, while the IGRAs had a specificity of 93–99% in this population. However, several studies examining the use of IGRAs in serial screening of health care workers and others have noted a higher conversion rate, defined as the proportion of persons with a newly positive result after a previously negative result, with the IGRAs than with the TST.[4–8] For example, one large multicenter study reported conversion rates of 0.9% for the TST, 6.1% for QFT, and 8.3% for T-SPOT.[4] Conversions generally occurred among persons with no known exposure to a person with infectious tuberculosis (TB) and, when a third test was performed, the result was frequently negative.[4,6,7,9,10] This finding has led some authors to speculate that the actual specificity of the IGRAs in practice is not as high as has been reported in the literature.[4] To address this concern, we developed a mathematical model of serial testing that reconciled the apparent paradox of more specific tests being associated with more apparent false-positive results.

METHODS

We developed a mathematical model to examine the impact of test characteristics (sensitivity, specificity and test-retest variability) and LTBI prevalence on the results of serial testing. We used broad ranges of test characteristics that would include literature-based estimates of test sensitivity (reported at between 71% and 90% for the TST and IGRAs using the surrogate of active TB and 40–100% for these tests using the surrogate of subsequent progression to active TB after initial testing) and specificity (reported at 94–99% for IGRAs and TST in non-BCG-vaccinated populations, but as low as 59% in BCG-vaccinated populations; we went as low as 70% specificity for the purposes of illustration).[2,3,11] We specifically examined the concordance of two serially performed tests in the same population.

The key assumption of the model was that the sensitivity and specificity of the tests for the study population were identical for the first and second test. In making this assumption, we

understand that it incorporates the effects of multiple factors on test characteristics, such as changes in test characteristics due to the time of day the blood is drawn, transient changes in immune function, and differences in processing between the two tests, well described by Pai et al.[12] and Tagmouti et al.,[13] with the assumption that the average of all of these effects is zero. To assess the effect of relaxing this assumption on our model, we performed additional simulations in which test sensitivity and specificity were randomly varied around the base values with a binomial distribution so the test sensitivity and specificity of the second test could differ from the first test. As including these random effects did not change the average model predictions (see Appendix),* they were omitted from the final models.

A second assumption of the model was that no new infections occurred between the time of the first and second tests. Test-retest variability was characterized by a 'change proportion' variable, c, defined as the total proportion of tests that switched results (positive to negative plus negative to positive divided by the total number of tests performed) from the first to the second test. Given the key assumption of unchanging sensitivity and specificity over time, the number of tests that switched from positive to negative must therefore equal the number of tests that switched from negative to positive (see the Appendix for a mathematical justification of this statement). As an example, in an imaginary cohort of 100 persons, 20 have an initial positive test and 80 have an initial negative test. When the entire cohort is retested, 10 of those who had an initial positive test have a second test that is negative, and 10 of those who initially tested negative have a second test that is positive. For this cohort, the parameter c would equal (10 + 10)/100 = 0.2. The observed conversion rate would be 10/80 = 0.125, and the observed reversion rate (rate of persons with a positive result on the first test but a negative result on the second test) would be 10/20 = 0.5. The calculated combinations of true LTBI and test results are summarized in the Table. Note that if both q and r are 1, all persons will have the same test results for the first and second test, so $c = 0$ (perfect test-retest reliability).

Given the information in the Table, the observed frequency of conversion among the cohort would be the sum of rows 3 and 7 divided by the sum of rows 3, 4, 7 and 8. In other words, the conversion proportion equals the number who initially tested negative but then had a second test that was positive, divided by the total number who initially tested negative. Mathematically, this translates to the following calculation (P = prevalence, Sn = sensitivity, Sp = specificity):

$$\text{conversion proportion} = (c/2)/(P*(1-\text{Sn}) + (1-P)*(\text{Sp})) \quad (1a)$$

Similarly, the observed proportion of reversion (positive to negative) among the cohort would be the sum of rows 2 and 6, divided by the sum of rows 1, 2, 5 and 6. In other words, the reversion proportion equals the number who initially tested positive but then had a second test that was negative, divided by the total number who initially tested positive. Mathematically, this results in the following calculation:

*The appendix is available in the online version of this article, at http://www.ingentaconnect.com/content/iuatld/ijtld/2018/00000022/00000005/art00010

$$\text{reversion proportion} = (c/2)/(P * \text{Sn} + (1 - P) * (1 - \text{Sp})) \quad (1b)$$

We then examined the effect of changing one parameter at a time on the observed frequency of conversions and reversions. The mathematical equations described above may be manipulated further to express c as a function of conversions and reversions observed in a particular cohort. The equation can be stated as follows:

$$c = 1/[1/(2 * \text{conversion proportion}) + 1/(2 * \text{reversion proportion})] \quad (2)$$

We used this result to illustrate the estimation of c from studies in which only conversion and reversion were reported.

As the study did not involve human subjects, no ethics approval was required.

RESULTS

Figure 1 illustrates the effect of varying specificity from 70% to 99% on the observed frequency of conversions and reversions. Given an LTBI prevalence of 5%, a test sensitivity of 80%, and change proportion (c) of 0.05 (5%), the observed frequency of conversion ranged from 2.6% at a specificity of 99%, to 3.7% at a specificity of 70%. The reversion rate ranged from 50.5% at a specificity of 99%, to 7.7% at a specificity of 70%. The change proportion is mathematically limited by the specificity given in Equations 1a and 1b; values of the change proportion must be less than 2 × (1 − specificity) and also less than 2 × (1 − sensitivity) to keep the conversion/reversion proportions ⩾ 0 (negative values are not possible for these proportions).

Figure 2 illustrates the effect of varying the change proportion (c) between 0 and 0.15 on the observed frequency of conversions and reversions. Given an LTBI prevalence of 5%, a test sensitivity of 80%, and specificity of 70%, the conversion rate ranged from 0% at a c of 0 to 11.1% at a c of 0.15, with corresponding reversion rates of 0% and 23.1%. Given the same prevalence and sensitivity, but a higher specificity of 95%, the conversion rate ranged from 0% at a c of 0 to 5.5% at a c of 0.1, with corresponding reversion rates of 0% and 57%. Increasing the specificity to 99% and examining the range of c between 0 and 0.02, the conversion rate ranged from 0% to 1.1% and the reversion rate between 0 and 20%. The trade-off between a less specific but less variable test vs. a more specific but more variable test is illustrated respectively by the points marked by the open circle (specificity 70%, c=0.05) and the open triangle (specificity 95%, c = 0.10). The less specific but less variable test would be associated with a lower observed rate of conversions and reversions. Note that in the low-prevalence scenario outlined above, most conversions will represent false-positive tests, unless the specificity is very high. For example, with an LTBI prevalence of 5%, a test sensitivity of 80%, specificity of 70% and c of 0.05, 370 persons in a cohort of 10000 would be expected to convert from negative to positive on the second test. Even if all 100 persons in this cohort who had false-negative initial tests converted to positive on the second test, the rest of the persons converting (n = 270) would be individuals without LTBI with false-positive tests. If specificity is increased to 95%, 270 persons in a cohort of 10000

would be expected to convert, and a minimum of 170 of those would be individuals without LTBI.

Figure 3 illustrates the effect of changes in LTBI prevalence on observed conversions and reversions at defined specificity levels (70%, 95%, and 99%), with an assumed sensitivity of 80%. As illustrated in the Figure, the prevalence of LTBI has relatively little effect on the observed proportion of individuals who convert, but it has a larger effect on the observed rate of reversion.

To understand how test-retest variability can be derived from published reports, we used the example of a study of serial testing in health care workers in low-prevalence areas in the United States.[4] In that study, the reported percentage conversion for the T-SPOT test using a cut-off point of ⩾8 spots was 8.3%, and the percentage reversion was 63.9%. Substituting these values into Equation 2 above, one calculates a 0.147 test-retest change. In other words, in our study population, assuming that no true infection occurred (a reasonable assumption given that the study was conducted in a low-prevalence area over a relatively short time period), 14.7% of participants were estimated to have a different result on retesting (i.e., initially positive with a negative result on retesting or vice versa). The corresponding test-retest changes for the TST and QFT in our study were respectively 0.018 and 0.110. These results, in turn, have mathematical implications for the boundaries of the test characteristics of T-SPOT in the present study. Examining a range of prevalence from 1% to 10% and a range of sensitivities from 60% to 90%, possible values for the specificity of T-SPOT ranged from a low of 89.1% (at sensitivity 60%, prevalence 1%) to a high of 97.2% (sensitivity 90%, prevalence 10%). In contrast, a more recent study of T-SPOT among US health care workers reported a conversion rate of 0.8% and a reversion rate of 17.6%.[14] From Equation 2, the test-retest change for our study was calculated to be 0.015. Examining the same ranges of prevalence/sensitivity as for the first study above, the minimum specificity was 94.3% (assuming a sensitivity of 60%, prevalence 1%) and specificity approached 100% when sensitivity ⩾ 69% and prevalence ⩾ 9%.

DISCUSSION

Our model demonstrated that test-retest variability was at least as important as specificity in assessing serial testing of low-risk patients. Given the published test-retest variability of IGRAs in the literature,[4,15,16] our model explains the relatively high frequency of reversions that has been observed. It also provides a reasonable explanation for the relatively low rate of TST conversions in the same setting: although the TST is likely less specific than either IGRA, its lower test-retest variability would be associated with the lower conversion rates observed.

This model also explains why more specific tests would be associated with a higher reversion rate. Higher test specificity is associated with a higher reversion rate. In a low-prevalence setting, test specificity drives the reversion rate and, to a lesser extent, the conversion rate. This can be intuitively explained as the second test ‘correcting’ many of the false-positive results produced by the first test, and supports recent guidelines recommending repeating unexpected positive tests in persons at a low risk for LTBI.[17]

Of course, using the model presented here, one would expect a relatively lower rate of reversions and a higher rate of conversions in a high-prevalence setting. Some authors have used the strategy of increasing the cut-off for a positive test for serial testing, effectively increasing the specificity of the second test while reducing the sensitivity.[18–20] While we did not explicitly examine this strategy in our model, it would essentially reduce the number of observed conversions (rows 3 and 7 in the Table) and, as in a low-prevalence setting the number of false-positive conversions (row 7) greatly outweighs the number of true-positive conversions (row 3), the net effect would likely be to reduce unnecessary treatment with some reduction in detecting true infection. Moses et al. created a Markov model to examine the potential impact of serial testing as well as changes in IGRA cut-off. They found that while both of these measures were potentially helpful, they had limited impact in mitigating overtreatment of the LTBI associated with serial testing in a low-prevalence area.[21] Explicit understanding of the impact of changes in cut-offs on test characteristics, which is hampered by the absence of a gold standard for LTBI, would be required to quantitatively examine the effects of adjusting cut-offs for serial testing.

The main limitations of our model stem from the difficulty in proving the validity of the major assumption, given that there is no gold standard for LTBI. The first assumption of constant test sensitivity and specificity over time precludes the acquisition of new immunocompromising conditions between the first and second tests, which would be likely to reduce test sensitivity. It further precludes exposure to cross-reactive environmental antigens (e.g., *Mycobacterium kansasii*) in the environment between the first and second tests, which could reduce test specificity. In theory, test sensitivity could systematically decline with the passage of time alone due to reduced immune recall, but if that were the case one would expect a consistent overall decline in the rate of positive tests over time, which has not been observed in studies of serial IGRAs either in low-prevalence populations being screened for LTBI[6,7,9,10,22] or in persons being treated for active TB.[23] In addition, the point estimates of sensitivity and specificity may differ over time due to random effects, such as differences in specimen processing between the first and second tests, and the time of day blood was drawn, but, as the average of such random effects would be zero, we ignored these for model simplification purposes. Appendix Section 1 demonstrates the effect of permitting sensitivity and specificity to vary between the first and second tests; this basically adds variability but does not change the fundamental relationships described here. We also ignored invalid/indeterminate results, which certainly occur in practice, but are less relevant to understanding rates of conversions/reversions in serial testing. This illustrates the oft-quoted principle that ‘all models are wrong, some models are useful’.[24]

CONCLUSIONS

These findings have practical implications for the choice of test when serial testing is indicated. In clinical practice, as persons with a positive test are generally not tested again, reversions are not measured. Conversions are used both to identify individuals who may benefit from LTBI treatment and to identify possibly unidentified exposures to TB in health care and other congregate settings. False-positive conversions may result in unnecessary treatment on the individual level, as well as unnecessary investigation at the facility level. As shown, a test with lower test-retest variability but relatively low specificity may be

preferable to a test with higher test-retest variability but higher specificity, as fewer false conversions will be observed. As a corollary to this principle, any efforts that can be made to reduce test variability through specimen collection and laboratory standardization will also reduce false conversions and improve the ability to detect occult exposures to TB in health care facilities.[25] Potential sources of variability and potential ways to minimize variability stemming from these sources have been well-enumerated by Banaei et al.[26] At least one study demonstrated relatively low rates of conversions and reversions in serial testing of health care workers;[14] one could speculate that lower test variability may have played a role in these results. Finally, reporting of test-retest concordance along with sensitivity and specificity will be crucial to the evaluation of any new tests that may be used for serial tuberculosis screening.

Acknowledgements

This work was supported by a contract from the Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA. The authors thank A Hill for his thorough reading and thoughtful comments on the manuscript.

APPENDIX

1. Mathematical derivation of formulas

Table A.1

Consider the following 2×2 table:

	Test#1 result	
LTBI	**Positive**	**Negative**
Yes	A	B
No	C	D

A, B, C, and D represent the number of persons who fall into each category (i.e., A = number of persons with true latent tuberculous infection (LTBI) and a positive test, B = number with true LTBI and a negative test, etc.).

Table A.2

A similar table may be constructed for the second test:

	Test #2 result	
LTBI	**Positive**	**Negative**
Yes	A + a1	B + b1
No	C + c1	D + d1

A, B, C, and D are the same numbers and in the first table, and a1, b1, c1, and d1 represent the differences in the cell counts between the first and second tests.

We have three assumptions:

i. No new infections occur between the two tests, so the prevalence of LTBI is the same between the first and second tests. Mathematically, this means:

$$(A+B)/(A+B+C+D) = (A+a1+B+b1)/(A+a1+B+b1+C+c1+D+d1)$$

As the number of persons studied does not change between tests, this can be simplified:

$$A+B+C+D = A+a1+B+b1+C+c1+D+d1,$$

so

$$(A+B)/(A+B+C+D) = (A+a1+B+b1)/(A+B+C+D),$$

and

$$A+B = A+a1+B+b1.$$

Subtracting A+B from both sides of the equation,

$$a1+b1=0 \quad \text{or} \quad a1=-b1$$

Similarly, c1=−d1 by the same calculations on the second row.

ii. The sensitivities of the first and second tests are equal. Mathematically this means:

$$A/(A+B) = (A+a1)/(A+a1+B+b1).$$

Since a1 = − b1 from (a) above, this simplifies to

$$A=A+a1,$$

which means that a1 = 0, so b1 also is equal to zero.

iii. The specificities of the first and second tests are equal. Mathematically this means:

$$D/(C+D) = (D+d1)/(C+c1+D+d1).$$

As c1 = − d1 from (a) above, this simplifies to

$$D=D+d1,$$

which means that d1=0, so c1 is also equal to zero.

QED: Given the assumptions stated above, the number of individuals in each cell is the same between the first and second test.

As no person changes true LTBI status between the first and second tests, the number of participants with LTBI who switch from positive on test #1 to negative on test #2 must therefore equal the number of participants with LTBI

who switch from negative on test #1 to positive on test #2. A similar relationship holds for participants without LTBI.

2. Derivation of equations in Table 1.

Notation and assumptions:

- Let D refer to true LTBI status, so D=0 refers to an uninfected state, and D = 1 refers to an infected state
- Let T_1 and T_2 refer to the results of the first and second tests, respectively, so $T_1 = 0$ is a negative first test, $T_1 = 1$ is a positive first test, $T_2 = 0$ is a negative second test, and $T_2 = 1$ is a positive second test
- Let p equal the true prevalence of LTBI
- The notation $P[T_1 = 1|D = 1]$ refers to 'the probability that $T_1 = 1$ given that D = 1.'
- The sensitivity (Sn) is assumed to be equal for both tests:
  - $Sn = P[T_1 = 1|D = 1] = P[T_2 = 1|D = 1]$
- The specificity (Sp) is assumed to be equal for both tests:
  - $Sp = P[T_1 = 0|D = 0] = P[T_2 = 0|D = 0]$
- The probability of obtaining consecutive positive tests when infected is defined as follows:
  - $P[T_1 = 1,T_2 = 1|D = 1] = q$Sn, where q is an unknown between 0 and 1
- The probability of obtaining consecutive negative tests when uninfected is defined as follows:
  - $P[T_1 = 0,T_2 = 0|D = 0] = r$Sp, where r is an unknown between 0 and 1
- The expressions in Table 1 can then be derived as follows:
  - Row 1 = $P[T_1 = 1,T_2 = 1, D = 1] = pq$Sn
  - Row 2 = $P[T_1 = 1,T_2 = 0, D = 1] = p(1 - q)$Sn
  - Row 3 = $P[T_1 = 0,T_2 = 1, D = 1] = p(1 - q)$Sn
  - Row 4=$P[T_1 =0,T_2 =0, D=1]=p(1–Sn)–p(1 - q)Sn = p[1 – Sn – (1 – q)Sn]$
  - Row 5=$P[T_1=1,T_2=1, D=0]=(1–p)(1–Sp) –(1–p)(1–r)Sp=(1–p)[1–Sp–(1–r)Sp]$
  - Row 6 = $P[T_1 = 1,T_2 = 0, D = 0] = (1 - p)(1 - r)$Sp
  - Row 7 = $P[T_1 = 0,T_2 = 1, D = 0] = (1 - p)(1 - r)$Sp
  - Row 8 = $P[T_1 = 0,T_2 = 0, D = 0] = (1 - p)Sp – (1 - p)(1 - r)Sp = (1 - p)r$Sp

- As all rows in the table must be between 0 and 1 (inclusive), the equations above imply bounds on q and r as follows:
  - $1 - Sn - (1 - q)Sn \geqslant 0$ and $(1 - q) \geqslant 0$
    - $1 - Sn - Sn + qSn \geqslant 0$ and $-q \geqslant -1$
    - $q \geqslant 2 - (1/Sn)$ and $q \leqslant 1$
  - $1 - Sp - (1 - r)Sp \geqslant 0$ and $(1 - r) \geqslant 0$
    - $1 - Sp - Sp + rSp \geqslant 0$ and $-r \geqslant -1$
    - $r \geqslant 2 - (1/Sp)$ and $r \leqslant 1$
- The change proportion, c, represents the total proportion of patients whose tests change between the first and second tests (positive to negative or vice versa), which is equal to the sum of rows 2, 3, 6, and 7:
  - $p(1 - q)Sn + p(1 - q)Sn + (1 - p)(1 - r)Sp + (1 - p)(1 - r)Sp$
  - $= 2[p(1 - q)Sn + (1 - p)(1 - r)Sp]$
- The proportion converting from a negative first test to a positive second test equals the sum of rows 3 and 7 divided by the sum of rows 3, 4, 7, and 8:
  - $=[p(1-q)Sn+(1-p)(1-r)Sp]/[p(1-q)Sn+p(1 - Sn - (1 - q)Sn) + (1 - p)(1 - r)Sp + (1- p)rSp]$
  - $= (c/2)/[p(1 - Sn) + (1 - p)Sp]$
- Similarly, the proportion reverting from a positive first test to a negative second test equals the sum of rows 2 and 6 divided by the sum of rows 1, 2, 5, and 6:
  - $= [p(1 - q)Sn + (1 - p)(1 - r)Sp]/[pqSn + p(1 - q)Sn + (1 - p)(1 - Sp–(1 -r)Sp) + (1 -p)(1 - r)Sp]$
  - $= (c/2)/[pSn + (1 - p)(1 - Sp)]$
- It is not necessary to know the values of q and r to determine the conversion/reversion rates from the change proportion

3. Relaxing the assumptions of constant test sensitivity and specificity

If the sensitivity and specificity measured at the time of the second test are not the same as those measured at the time of the first test (i.e., there is stochastic variability, which is probably the case), the equations in Table A.1 look slightly different. We can use the expressions Sn_1 and Sn_2 to refer to the sensitivities of the first and second tests, respectively, and Sp_1 and Sp_2 to refer to the specificities of the first and second tests. We can then define the parameter u as the probability that both tests are false-negative, and the parameter v as the probability that both tests are false-positive. The corresponding equations for this scenario are in Table A.3 below.

Table A.3

Mathematical expressions for observed combinations of LTBI results and the results of two serial theoretical interferon-gamma release assays, discarding the assumption that sensitivity and specificity are the same for the first and second tests. The formulae in the rightmost column represent the proportion of all patients who have the

Row number	LTBI (gold standard)	Test 1 result	Test 2 result	Proportion of total patients
1	+	+	+	$p(Sn_1 + Sn_2 - 1 + u)$
2	+	+	−	$p(1 - Sn_2 - u)$
3	+	−	+	$p(1 - Sn_1 - u)$
4	+	−	−	pu
5	−	+	+	$(1 - p)v$
6	−	+	−	$(1 - p)(1 - Sp_1 - v)$
7	−	−	+	$(1 - p)(1 - Sp_2 - v)$
8	−	−	−	$(1 - p)(Sp_1 + Sp_2 - 1 + v)$

LTBI = latent tuberculous infection.

In this case, Sn and Sp represent the sensitivity and specificity, respectively, measured at the time of the first test. The sensitivity of the second test is equal to (row 1 + row 3)/(row 1 + row 2 + row 3 + row 4), which simplifies to the expression $Sn(q_1 - q_2 + 1)$. Similarly, the specificity of the second test is equal to (row 6 + row 8)/(row 5 + row 6 + row 7 + row 8), which simplifies to the expression $Sp(r_1 - r_2 + 1)$. The mathematical constraints on u and v are as follows:

- $0 < u < 1 - \max(Sn_1, Sn_2)$
- $0 < v < 1 - \max(Sp_1, Sp_2)$

In this case, the change proportion (c) is equal to the sums of rows 2, 3, 6, and 7, which simplifies to the expression $c = p[2 - 2u - (Sn_1 + Sn_2)] + (1 - p)[2 - 2v - (Sp_1 + Sp_2)]$. The proportion of patients with conversions and reversions between the two tests are therefore as follows:

$$\text{Conversion} = [p(1 - Sn_1 - u) + (1 - p)(1 - Sp_2 - v)]/[p(1 - Sn_1) + (1 - p)Sp_1]$$

$$\text{Reversion} = [p(1 - Sn_2 - u) + (1 - p)(1 - Sp_1 - v)]/[pSn_1 + (1 - p)(1 - Sp_1)]$$

To understand how using this model differs from the simpler model that assumes equal sensitivity/specificity over time, we performed a simple simulation as follows:

- Set $Sn_1 = 0.8$
- Sampled Sn_2 from a uniform distribution between 0.75 and 0.85 (mean = Sn_1 = 0.8)
- Varied Sp_1 between 0.7 and 0.95 in steps of 0.05

- Sampled Sp_2 from a uniform distribution with mean = Sp_1 and range of $Sp_1 \pm 0.025$
- Sampled u from a uniform distribution on $(0,1 - \max(Sn_1,Sn_2))$
- Solved for v given Sn_1, Sn_2, c, and u (holding c constant at 0.05).

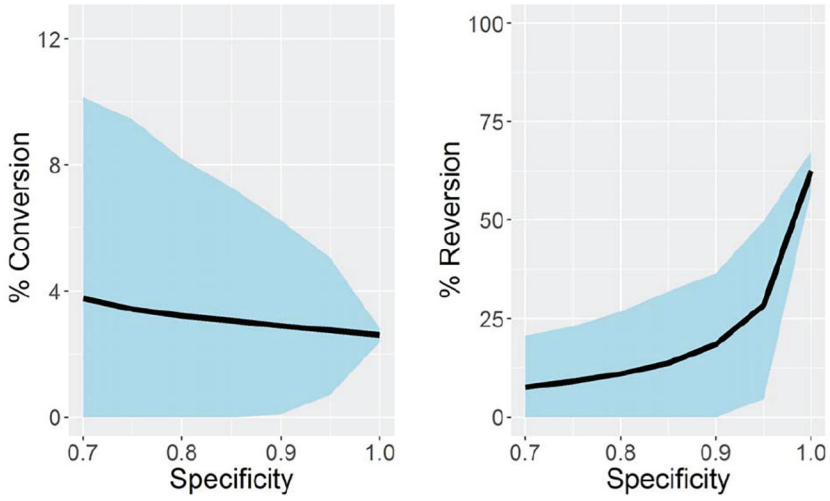

Figure A.1.
Conversion and reversion rates by specificity

Figures A.1 and A.2 correspond to Figures 1 and 2 in the manuscript. The sensitivity and specificity of the second test randomly vary around the sensitivity and specificity of the first test, so, on average, the sensitivities of the first and second test are equal, as are the specificities of the first and second test. The lines in the graphs represent the mean values, with the shaded ribbons encompassing 95% confidence intervals (CIs) around those mean values. If the sensitivity and specificity estimates had been permitted to vary more widely (i.e., higher variance), the CIs would be wider but again, by definition, the average values would be the same. Figure A.1 illustrates the CIs around conversion and reversion rates as specificity is varied between 70% and 95%. Figure A.2 illustrates the confidence intervals around conversion and reversion rates as the change proportion is varied between 0% and 15% (when specificity is fixed at 70%) and 0–10% (when specificity is fixed at 95%).

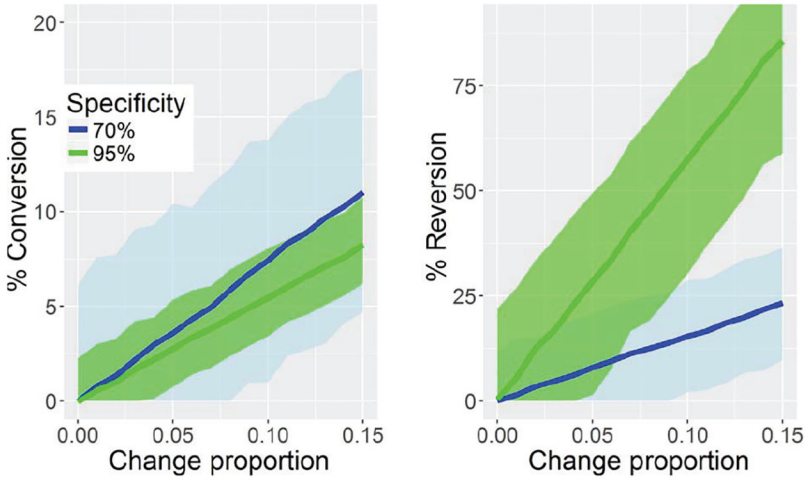

Figure A.2.
Conversion and reversion rates by change proportion.

References

1. Diel R, Goletti D, Ferrara G, et al. Interferon-gamma release assays for the diagnosis of latent *Mycobacterium tuberculosis* infection: a systematic review and meta-analysis. Eur Respir J 2011; 37: 88–99. [PubMed: 21030451]
2. Kahwati LC, Feltner C, Halpern M, et al. Screening for latent tuberculosis infection in adults: an evidence review for the US Preventive Services Task Force. US Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Rockville, MD, USA: US Preventive Services Task Force, 2016.
3. Pai M, Zwerling A, Menzies D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. Ann Intern Med 2008; 149: 177–184. [PubMed: 18593687]
4. Dorman SE, Belknap R, Graviss EA, et al. Interferon-gamma release assays and tuberculin skin testing for diagnosis of latent tuberculosis infection in healthcare workers in the United States. Am J Respir Crit Care Med 2014; 189: 77–87. [PubMed: 24299555]
5. Lucet JC, Abiteboul D, Estellat C, et al. Interferon-gamma release assay vs. tuberculin skin test for tuberculosis screening in exposed healthcare workers: a longitudinal multicenter comparative study. Infect Control Hosp Epidemiol 2015; 36: 569–574. [PubMed: 25682769]
6. Joshi M, Monson TP, Joshi A, Woods GL. IFN-gamma release assay conversions and reversions. Challenges with serial testing in US health care workers. Ann Am Thorac Soc 2014; 11: 296–302. [PubMed: 24446969]
7. Zwerling A, Benedetti A, Cojocariu M, et al. Repeat IGRA testing in Canadian health workers: conversions or unexplained variability? PLOS ONE 2013; 8: e54748. [PubMed: 23382955]
8. Zwerling A, van den Hof S, Scholten J, Cobelens F, Menzies D, Pai M. Interferon-gamma release assays for tuberculosis screening of healthcare workers: a systematic review. Thorax 2012; 67: 62–70. [PubMed: 21228420]
9. Slater ML, Welland G, Pai M, Parsonnet J, Banaei N. Challenges with QuantiFERON-TB Gold assay for large-scale, routine screening of US healthcare workers. Am J Respir Crit Care Med 2013; 188: 1005–1010. [PubMed: 23978270]
10. Gamsky TE, Lum T, Hung-Fan M, Green JA. Cumulative false-positive QuantiFERON-TB interferon-gamma release assay results. Ann Am Thorac Soc 2016; 13: 660–665. [PubMed: 26783649]
11. Herrera V, Perry S, Parsonnet J, Banaei N. Clinical application and limitations of interferon-gamma release assays for the diagnosis of latent tuberculosis infection. Clin Infect Dis 2011; 52: 1031–1037. [PubMed: 21460320]

12. Pai M, Denkinger CM, Kik SV, et al. Gamma interferonrelease assays for detection of *Mycobacterium tuberculosis* infection. Clin Microbiol Rev 2014; 27: 3–20. [PubMed: 24396134]
13. Tagmouti S, Slater M, Benedetti A, et al. Reproducibility of interferon gamma (IFN-gamma) release Assays. A systematic review. Ann Am Thorac Soc 2014; 11: 1267–1276. [PubMed: 25188809]
14. King TC, Upfal M, Gottlieb A, et al. T-SPOT.*TB* interferon-gamma release assay performance in healthcare worker screening at nineteen US Hospitals. Am J Respir Crit Care Med 2015; 192: 367–373. [PubMed: 26017193]
15. Metcalfe JZ,Cattamanchi A, McCullochC E, Lew JD, Ha NP, Graviss EA. Test variability of the QuantiFERON-TB Gold In-Tube assay in clinical practice. Am J Respir Crit Care Med 2013; 187: 206–211. [PubMed: 23103734]
16. van Zyl-Smit RN, Zwerling A, Dheda K, Pai M. Within-subject variability of interferon-g assay results for tuberculosis and boosting effect of tuberculin skin testing: a systematic review. PLOS ONE 2009; 4: e8517. [PubMed: 20041113]
17. Lewinsohn DM, Leonard MK, LoBue PA, et al. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention. Clinical practice guidelines: diagnosis of tuberculosis in adults and children. Clin Infect Dis 2017; 64: 111–115. [PubMed: 28052967]
18. Schablon A, Nienhaus A, Ringshausen FC, Preisser AM, Peters C. Occupational screening for tuberculosis and the use of a borderline zone for interpretation of the IGRA in German healthcare workers. PLOS ONE 2014; 9: e115322. [PubMed: 25541947]
19. Fong KS, Tomford JW, Teixeira L, et al. Challenges of interferon-gamma release assay conversions in serial testing of health-care workers in a TB control program. Chest 2012; 142: 55–62. [PubMed: 22796839]
20. Veerapathran A, Joshi R, Goswami K, et al. T-cell assays for tuberculosis infection: deriving cut-offs for conversions using reproducibility data. PLOS ONE 2008; 3: e1850. [PubMed: 18365006]
21. Moses MW, Zwerling A,Cattamanchi A,et al. Serialtestingfor latent tuberculosis using QuantiFERON-TB Gold In-Tube: a Markov model. Sci Rep 2016; 6: 30781. [PubMed: 27469388]
22. Aichelburg MC, Reiberger T, Breitenecker F, Mandorfer M, Makristathis A, Rieger A. Reversion and conversion of interferon-gamma release assay results in HIV-1-infected individuals. J Infect Dis 2014; 209: 729–733. [PubMed: 23911707]
23. Denkinger CM, Pai M, Patel M, Menzies D. Gamma interferon release assay for monitoring of treatment response for active tuberculosis: an explosion in the spaghetti factory. J Clin Microbiol 2013; 51: 607–610. [PubMed: 23175268]
24. Box GEP, Hunter JS, Hunter WG. Statistics for experimenters: design, innovation, and discovery. 2nd ed. Hoboken, NJ, USA: Wiley Interscience, 2005: xvii, 633.
25. Gaur RL, Pai M, Banaei N. Impact of blood volume, tube shaking, and incubation time on reproducibility of QuantiFERON-TB Gold In-Tube assay. J Clin Microbiol 2013; 51: 3521–3526. [PubMed: 23966505]
26. Banaei N, Gaur RL, Pai M. Interferon-gamma release assays for latent tuberculosis: what are the sources of variability? J Clin Microbiol 2016; 54: 845–850. [PubMed: 26763969]

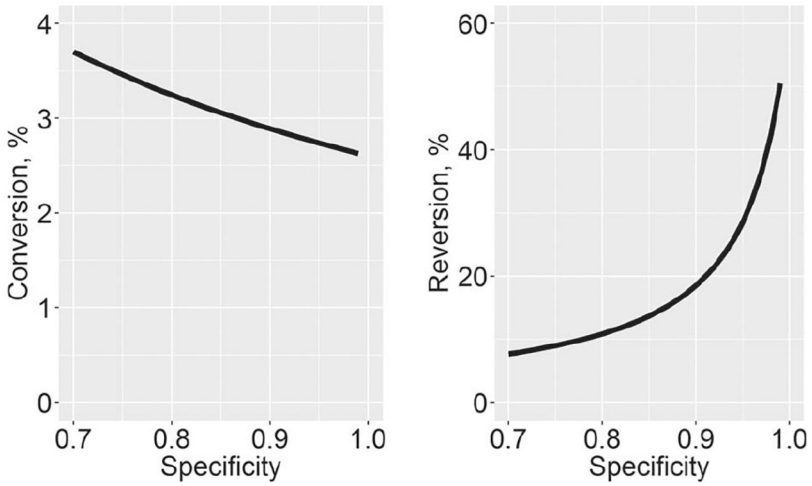

Figure 1.

Relationship between test specificity and proportion of observed conversions and reversions between a first and second test given a fixed sensitivity (80%), LTBI prevalence (5%), and proportion change (0.05). The change proportion cannot exceed twice the value of (1 – specificity) as higher values result in a negative value for row 5 in the Table, so the upper bound of specificity in this graph is 1 – (0.05/2) = 97.5%. LTBI = latent tuberculous infection.

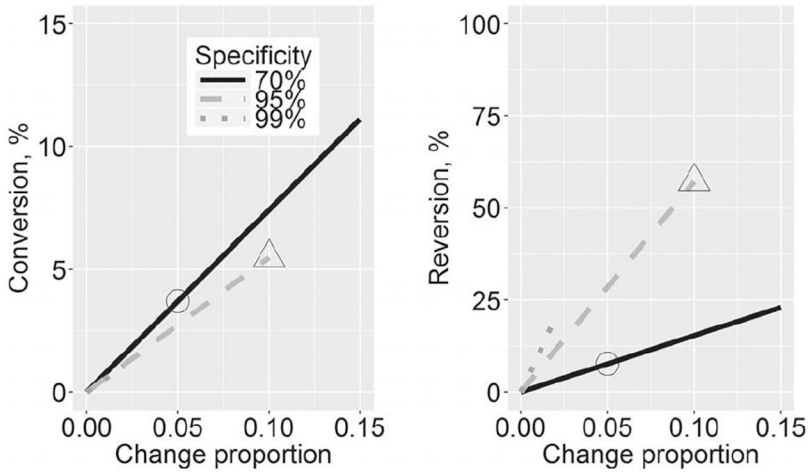

Figure 2.

Relationship between change proportion (*c*) between tests and proportion of observed conversions and reversions, given fixed sensitivity (80%), specificity at 70%, 95%, or 99%, and fixed LTBI prevalence (5%). The open circle represents a test with low specificity (70%) and low change proportion (0.05), while the open triangle represents a more specific test (95%) with a higher percentage change (10%). The change proportion is mathematically constrained by the equations in the Table to be ⩽2 × (1 – specificity) and, as the line demonstrating percentage conversion for a specificity of 99% is very similar to that for 95%, this is indistinguishable on the left-hand graph. LTBI = latent tuberculous infection.

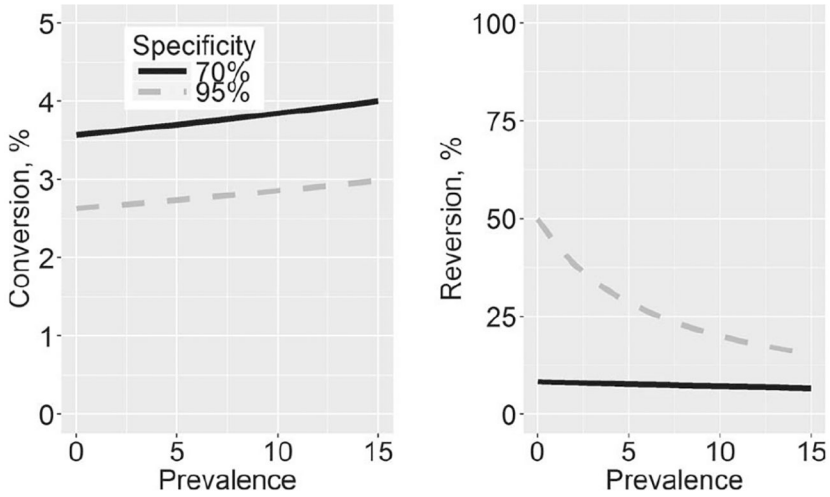

Figure 3.

Relationship between the underlying prevalence of LTBI and observed frequency of conversions and reversions, given a change proportion of 0.05, fixed sensitivity of 80%, specificity of either 70% or 95%, and LTBI prevalence ranging from 1% to 15%. LTBI = latent tuberculous infection.

Table

Mathematical expression for observed combinations of the LTBI result and the results of two serial theoretical interferon-gamma release assays (i.e., the same assay carried out twice) performed in the same patient population. The formula in the right column represents the proportion of all patients who have the combination of LTBI and the two test results in each row; by definition these sum to one

Row number	LTBI (gold standard)	Test 1 result	Test 2 result	Proportion of total patients
1	+	+	+	pqSn
2	+	+	−	$p(1-q)$Sn
3	+	−	+	$p(1-q)$Sn
4	+	−	−	$p(1-\text{Sn}-(1-q)\text{Sn})$
5	−	+	+	$(1-p)[1-\text{Sp}-(1-r)\text{Sp}]$
6	−	+	−	$(1-p)(1-r)$Sp
7	−	−	+	$(1-p)(1-r)$Sp
8	−	−	−	$(1-p)r$Sp

LTBI = latent tuberculous infection; p = prevalence, Sn = sensitivity; q = proportion of persons (among those with LTBI by the gold standard) who tested positive on the first test who also test positive on the second test; Sp = specificity, r = proportion of persons who tested negative on the first test (among those with LTBI by the gold standard) who also tested negative on the second test.