APPENDIX

Supplementary panel: Measures of diagnostic accuracy and test performance

Sensitivity is the probability that a patient with the target disease has a positive test result. Sensitivity is expressed as a percentage and calculated as follows:

*Sensitivity = (a / (a + c)) * 100*

	Diseased	**Non-diseased**
Test positive	*a (true-positive)*	*b (false-positive)*
Test negative	*c (false-negative)*	*d (true-negative)*

A test of high sensitivity will give high true-positive and low false-negative rates. Conversely, a test of low sensitivity will give low true-positive and high false-negative rates.

Specificity is the probability that a patient without the target disease has a negative test result. Specificity is expressed as a percentage and calculated as follows:

*Specificity = (d / (b + d)) * 100*

A test of high specificity will give high true-negative and low false-positive rates. Conversely, a test of low specificity will give low true-negative and high false-positive rates.

Positive likelihood ratio (PLR) is the likelihood that a positive test result wold be expected in a patient with the target disease compared to the likelihood that a positive test result would be expected in a patient without the target disease. It is calculated as follows:

PLR = (a / (a + c)) / (b / (b + d))

A PLR of >1 indicates the positive test result is associated with the presence of disease; a PLR of <1 indicates that the positive test result is associated with absence of disease. The further the PLR from 1, the stronger the association.

Negative likelihood ratio (NLR) is the likelihood that a negative test result wold be expected in a patient with the target disease compared to the likelihood that a negative test result would be expected in a patient without the target disease. It is calculated as follows:

NLR = (c / (a + c)) / (d / (b + d))

A NLR of >1 indicates the negative test result is associated with the presence of disease; a NLR of <1 indicates that the negative test result is associated with absence of disease. The further the PLR from 1, the stronger the association.

Positive predictive value (PPV) is the probability that a patient with a positive test result actually has the target disease. PPV is often expressed as a percentage and calculated as follows:

*PPV = (a / (a + b)) * 100*

Negative predictive value (NPV) is the probability that a patient with a negative test result truly does not have the target disease. NPV is often expressed as a percentage and calculated as follows:

*NPV = (d / (c + d)) * 100*

Supplementary checklist: STARD

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1, 5
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions	5
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	7-9
	4	Study objectives and hypotheses	8-9
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	9
Participants	**6**	Eligibility criteria	9
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	9
	8	Where and when potentially eligible participants were identified (setting, location and dates)	9
	9	Whether participants formed a consecutive, random or convenience series	9
Test methods	**10a**	Index test, in sufficient detail to allow replication	11 & Appendix pg 4-5
	10b	Reference standard, in sufficient detail to allow replication	10, 26 (table 1)
	11	Rationale for choosing the reference standard (if alternatives exist)	10, 26 (table 1 footnote a)
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	Appendix pg 4-5
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	10, 26 (table 1)
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	11
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	10
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	11-12
	15	How indeterminate index test or reference standard results were handled	11-12
	16	How missing data on the index test and reference standard were handled	11-12
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	11-12
	18	Intended sample size and how it was determined	11

RESULTS			
Participants	**19**	Flow of participants, using a diagram	12, Figure 1
	20	Baseline demographic and clinical characteristics of participants	12-13, 27-28 (table 2)
	21a	Distribution of severity of disease in those with the target condition	27-28 (table 2), 29 (table 3), Appendix pg 6 (S.table 1)
	21b	Distribution of alternative diagnoses in those without the target condition	29 (table 3)
	22	Time interval and any clinical interventions between index test and reference standard	10-11
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Appendix, pg 8-9 (S. tables 3 and 4)
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	13-15, 30-31 (table 4)
	25	Any adverse events from performing the index test or the reference standard	Not applicable
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	18-19
	27	Implications for practice, including the intended use and clinical role of the index test	15-19
OTHER INFORMATION			
	28	Registration number and name of registry	Not applicable
	29	Where the full study protocol can be accessed	9
	30	Sources of funding and other support; role of funders	12

Supplementary methods: Laboratory procedures

QFT-GIT was carried out according to the manufacturer's instructions (http://www.quantiferon.com/wp-content/uploads/2018/03/L1075115_QFT_EU_ROW_Rev005.pdf) and as described in Whitworth et al.[6] In brief, blood was incubated overnight (16-24 hours) at 37°C in collection tubes containing a pool of Mtb-specific antigens (ESAT-6, CFP-10 and Rv2654) and positive (mitogen) and negative controls. Plasma was separated and stored at 4˚C prior to measurement of IFN-γ released in response to antigen stimulation by enzyme-linked immunosorbent assay (ELISA). Optical density readings were determined using a microplate reader (Elx800 Absorbance reader, VIC, Australia), and interferon-gamma (IFN-γ) levels calculated against a series of standard concentrations. The test was considered positive if the IFN-γ level for TB-specific antigens was ≥0.35 IU/mL after subtracting the negative control reading. It was considered indeterminate (invalid) if the negative control reading was >8.0 IU/ml and/or the positive control was <0.5 IU/ml.

T-SPOT.TB was carried out by enzyme-linked immunospot (ELISPOT) assay on PBMCs isolated from heparinised whole-blood (using the Ficoll Paque density centrifugation method), as per the manufacturer's instructions (http://www.tspot.com/wp-content/uploads/2017/07/PI-TB-US-V6.pdf) and as described in Whitworth et al.[6] The second-generation and ESAT-6-free IGRAs used the same platform and methodology as T-SPOT.TB. In brief, freshly-isolated PBMCs were suspended in serum-free AIM-V at a concentration of 2.5 million cells per ml. Cells were incubated overnight (18 hours; 37°C) with T-SPOT.TB antigens (ESAT-6, CFP-10), novel antigens (Rv3615c, Rv3879c), and positive (phytohemagglutinin (PHA)) and nil (RPMI medium) controls in a 96-well plate, pre-coated with IFN-γ specific monoclonal capture antibodies. For the novel antigens, peptide pools comprised 15-mer peptides overlapping their adjacent peptides by 10 amino acids representing the full sequence of Rv3615c (n=19 peptides) and previously defined selected sequences from Rv3879c (n=17 peptides; covering amino acid residues 1-95),[1,13] as shown below.

Rv3879c/1	MSITRPTGSYARQML	Rv3615c/1	MTENLTVQPERLGVL
Rv3879c/2	PTGSYARQMLDPGGW	Rv3616c/2	TVQPERLGVLASHHD
Rv3879c/3	ARQMLDPGGWVEADE	Rv3615c/3	RLGVLASHHDNAAVD
Rv3879c/4	DPGGWVEADEDTFYD	Rv3615c/4	ASHHDNAAVDASSGV
Rv3879c/5	VEADEDTFYDRAQEY	Rv3615c/5	NAAVDASSGVEAAAG
Rv3879c/6	DTFYDRAQEYSQVLQ	Rv3615c/6	ASSGVEAAAGLGESV
Rv3879c/7	RAQEYSQVLQRVTDV	Rv3615c/7	EAAAGLGESVAITHG
Rv3879c/8	SQLVQRVTDVLDTCR	Rv3615c/8	LGESVAITHGPYCSQ
Rv3879c/9	RVTDVLDTCRQQKGH	Rv3615c/9	AITHGPYCSQFNDTL
Rv3879c/10	LDTCRQQKGHVFEGG	Rv3615c/10	PYCSQFNDTLNVYLT
Rv3879c/11	QQKGHVFEGGLWSGG	Rv3615c/11	FNDTLNVYLTAHNAL
Rv3879c/12	VFEGGLWSGGAANAA	Rv3615c/12	NVYLTAHNALGSSLH
Rv3879c/13	LWSGGAANAANGALG	Rv3615c/13	AHNALGSSLHTAGVD
Rv3879c/14	AANAANGALGANINQ	Rv3615c/14	GSSLHTAGVDLAKSL
Rv3879c/15	NGALGANINQLMTLQ	Rv3615c/15	TAGVDLAKSLRIAAK
Rv3879c/16	ANINQLMTLQDYLAT	Rv3615c/16	LAKSLRIAAKIYSEA

Rv3879c/17	LMTLQDYLATVITWH	Rv3615c/17	RIAAKIYSEADEAWR
		Rv3615c/18	IYSEADEAWRKAIDG
		Rv3615c/19	DEAWRKAIDGLFT

For each peptide, identity was confirmed by mass spectrometry, and purity exceeded 80%. Pooled peptides were diluted firstly in DMSO (25mg/ml) and secondly in RPMI. Each T-SPOT.TB antigen, novel antigen peptide pool or control was added to an individual well of the plate, with a final concentration for each T-SPOT.TB antigen or novel antigen peptide of 10μg/ml. The final concentration of DMSO per well ranged from 0.68% (for ESAT-6 and Rv3879) to 0.76% (for Rv3615c). After incubation, wells were washed with phosphate-buffered saline and an alkaline phosphatase-conjugated secondary IFN-γ specific monoclonal antibody was added. For a visible representation of the spots (spot-forming cells; SFCs) on the membrane, an alkaline-phosphatase chromogen substrate was added. SFCs were enumerated using an ELISPOT plate reader (AID ELISpot read system ELRIFL04, Advanced Imaging Devices GmbH, Straßberg, Germany). The test was considered positive if the number of SFCs for the TB antigen minus the negative control was ≥8. Where this difference was 5, 6 or 7 SFCs, the assay was deemed borderline. Results were classified as indeterminate (invalid) if the positive control produced <20 SFCs and/or the negative control produced >10 SFCs.

All samples were processed within eight hours of blood collection.

Supplementary table 1: Medication history. Column percentages for each medication are shown.

Medication, n (%)	Diagnosis as per Reference Standard[1]				Total
	Culture-confirmed TB N = 261	Highly-probable TB N = 102	Clinically indeterminate N = 43	Active TB excluded N = 439	N = 845
None	63 (24.1)	35 (34.3)	13 (30.2)	203 (46.2)	314 (37.2)
Chemotherapy	0	0	0	1 (0.2)	1 (0.1)
Corticosteroids ≥15 mg/day	20 (7.7)	5 (4.9)	5 (11.6)	20 (4.6)	50 (5.9)
Corticosteroids <15 mg/day	13 (5.0)	7 (6.9)	1 (2.3)	19 (4.3)	40 (4.7)
Corticosteroids unknown	1 (0.4)	1 (1.0)	0	0	2 (0.2)
Other immune suppressants	1 (0.4)	0	0	11 (2.4)	12 (1.4)
Other	191 (73.2)	64 (62.7)	30 (69.8)	233 (53.1)	518 (61.3)
Unknown	1 (0.4)	0	0	1 (0.2)	2 (0.2)

Supplementary table 2: Symptoms at presentation. Column percentages for each symptom are shown.

Symptom	Diagnosis as per Reference Standard[1]				Total
	Culture-confirmed TB N = 256	Highly-probable TB N = 99	Clinically indeterminate N = 43	Active TB excluded N = 429	N = 827[a]
Cough, n (%)	174 (68.0)	53 (53.5)	23 (53.5)	326 (76.0)	576 (69.6)
Fever, n (%)	126 (49.2)	49 (49.5)	14 (32.6)	195 (45.5)	384 (46.4)
Night sweats, n (%)	129 (50.4)	53 (53.5)	20 (46.5)	215 (50.1)	417 (50.4)
Weight loss, n (%)	154 (60.2)	54 (54.5)	21 (48.8)	211 (49.2)	440 (53.2)
Haemoptysis, n (%)	31 (12.1)	8 (8.0)	3 (7.0)	65 (15.2)	107 (12.9)
Lethargy, n (%)	133 (52.0)	56 (56.6)	23 (53.5)	222 (51.7)	434 (52.5)
Other, n (%)	163 (63.7)	59 (59.46)	25 (58.1)	202 (47.1)	449 (54.3)
Median no. of symptoms (range)	4 (1–10)	4 (1–8)	3 (1–7)	3 (1–10)	4 (1–10)

[a]Eighteen participants were recruited on the basis of abnormal clinical signs rather than symptoms.

Supplementary table 3: Cross-tabulation of T-SPOT.TB and QFT-GIT results for patients with active TB (active TB positive) and non-TB diagnoses (active TB-negative)

ACTIVE TB POSITVE		**T-SPOT.TB**					
		Positive	**Negative**	**Borderline**	**Indeterminate**	**Missing**	**Total**
	Positive	187	13	6	9	5	220
	Negative	49	41	8	7	2	107
QFT-GIT	**Indeterminate**	16	4	3	1	2	26
	Missing	1	0	0	0	9	10
	Total	253	58	17	17	18	363
ACTIVE TB NEGATIVE		**T-SPOT.TB**					
		Positive	**Negative**	**Borderline**	**Indeterminate**	**Missing**	**Total**
	Positive	37	30	3	3	1	74
	Negative	12	250	12	26	4	304
QFT-GIT	**Indeterminate**	2	36	1	8	0	47
	Missing	0	3	0	0	11	14
	Total	51	319	16	37	16	439

For patients with a definitive final diagnosis (categories 1, 2 and 4), there was 73% concordance in QFT-GIT and T-SPOT.TB positivity, and 77% concordance in negativity.

Supplementary table 4: Cross-tabulation of T-SPOT.TB and second-generation IGRA results for patients with active TB (active TB positive) and non-TB diagnoses (active TB-negative)

ACTIVE TB POSITVE		**Second-generation IGRA**					
		Positive[a]	**Negative**	**Borderline**	**Indeterminate**	**Missing**	**Total**
T-SPOT.TB	**Positive**	253	0	0	0	0	253
	Negative	16	33	9	0	0	58
	Borderline	4	0	13	0	0	17
	Indeterminate	0	0	0	17	0	17
	Missing	0	0	0	0	18	18
	Total	273	33	22	17	18	363
ACTIVE TB NEGATIVE		**Second-generation IGRA**					
		Positive	**Negative**	**Borderline**	**Indeterminate**	**Missing**	**Total**
T-SPOT.TB	**Positive**	51	0	0	0	0	51
	Negative	19	296	4	0	0	319
	Borderline	4	0	12	0	0	16
	Indeterminate	0	0	1	36	0	37
	Missing	0	0	0	0	16	16
	Total	74	296	17	36	16	439

[a]Of 20 additional TB cases detected by second-generation IGRA (compared to T-SPOT.TB), only three responses to ESAT-6 and one to CFP-10 were borderline.

Supplementary table 5: Response magnitudes to individual antigens included in T-SPOT.TB and second-generation IGRA

	Dosanjh category								Total
	1	2	3	4A	4B	4C	4D	4A-D	
Spot-forming cells, median (IQR)									
ESAT-6	14 (3–40)	13 (1–46)	0 (0–5)	0 (0–1)	2 (0–10)	0 (0–1)	0 (0–0)	0 (0–1)	1 (0–14)
CFP-10	18 (4–64)	13 (1–70)	1 (0–12)	0 (0–0)	1 (0–12)	0 (0–1)	0 (0–0)	0 (0–1)	1 (0–17)
Rv3615c	25 (6–59)	19 (2–60)	1 (0–23)	0 (0–1)	1 (0–9)	0 (0–2)	0 (0–1)	0 (0–2)	2 (0–27)
Rv3879c	2 (0–10)	1 (0–10)	0 (0–1)	0 (0–0)	0 (0–1)	0 (0–1)	0 (0–1)	0 (0–1)	0 (0–2)

IQR: Interquartile range

Supplementary table 6: Diagnostic accuracy of current and second-generation IGRAs for diagnosis of active TB among patients with HIV-infection. Sensitivity and specificity are presented as percentages.

Test performance	T-SPOT.TB[a]		QFT-GIT[a]		ESAT+ CFP10 + Rv3615c[a]		CFP10 + Rv3615c + Rv3879c[a]	
	n/N	Estimate (95% CI)	n/N	Estimate (95% CI)	n/N	Estimate (95% CI)	n/N	Estimate (95% CI)
Sensitivity for active TB								
All TB	12/19	63.2 (41.0–80.9)	13/23	56.5 (36.8–74.4)	12/17	70.6 (46.9–86.7)	11/16	68.8 (44.4–85.8)
Culture-confirmed TB[b]	7/11	63.6 (35.4–84.8)	8/13	61.5 (35.5–82.3)	7/9	77.8 (45.3–93.7)	7/9	77.8 (45.3–93.7)
Specificity for active TB								
Active TB excluded	71/76	93.4 (85.5–97.2)	80/87	92.0 (84.3–96.1)	70/76	92.1 (83.8–96.3)	67/77	87.0 (77.7–892.8)
Active TB excluded, TST-negative, no risk factors for LTBI	28/29	96.6 (82.8–99.4)	36/38	94.7 (82.7–98.5)	27/28	96.4 (82.3–99.4)	27/29	93.1 (78.0–98.1)

LTBI, latent tuberculosis infection; TST, tuberculin skin test.

[a]One QFT-GIT, and two T-SPOT.TB and second-generation IGRA results were missing due to blood draw difficulties, samples being unsuitable for testing, or samples being destroyed for laboratory reasons. Missing results were spread across all diagnostic categories.

[b]Thirty-three HIV-positive patients had indeterminate T-SPOT.TB results, and 22 had indeterminate QFT-GIT results. Indeterminate and borderline IGRA results were excluded from analyses.

Supplementary table 7: Diagnostic accuracy of current and second-generation IGRAs for diagnosis of active TB among patients with Diabetes. Sensitivity and specificity are presented as percentages.

Test performance	T-SPOT.TB[a]		QFT-GIT[a]		ESAT+ CFP10 + Rv3615c[a]		CFP10 + Rv3615c + Rv3879c[a]	
	n/N	Estimate (95% CI)	n/N	Estimate (95% CI)	n/N	Estimate (95% CI)	n/N	Estimate (95% CI)
Sensitivity for active TB								
All TB	16/24	66.7 (46.7–82.0)	15/27	55.6 (37.3–72.4)	18/21	85.7 (65.4–95.0)	18/21	85.7 (65.4–95.0)
Culture-confirmed TB[b]	13/20	65.0 (43.3–81.9)	12/22	54.5 (34.7–73.1)	14/17	82.4 (59.0–93.8)	14/17	82.4 (59.0–93.8)
Specificity for active TB								
Active TB excluded	38/47	80.9 (67.5–89.6)	37/47	78.7 (65.1–88.0)	34/48	70.8 (56.8–81.8)	34/47	72.3 (58.2–83.1)
Active TB excluded, TST-negative, no risk factors for LTBI	6/7	85.7 (48.7–97.4)	4/5	80.0 (37.6–96.4)	6/7	85.7 (48.7–97.4)	6/7	85.7 (48.7–97.4)

LTBI, latent tuberculosis infection; TST, tuberculin skin test.

[a]Two QFT-GIT,and four T-SPOT.TB and second-generation IGRA results were missing due to blood draw difficulties, samples being unsuitable for testing, or samples being destroyed for laboratory reasons. Missing results were spread across all diagnostic categories.

[b]Two diabetic patients had indeterminate T-SPOT.TB results, and four had indeterminate QFT-GIT results. Indeterminate and borderline IGRA results were excluded from analyses.

Supplementary table 8: Diagnostic accuracy of T-SPOT.TB and second-generation IGRAs for diagnosis of active TB using ≥5 vs ≥8 SFC cut-off criteria for scoring positive results. Sensitivity and specificity are presented as percentages (95% CI).

Test performance	T-SPOT.TB		ESAT+ CFP10 + Rv3615c		CFP10 + Rv3615c + Rv3879c	
	Borderline excluded[a] (cut-off ≥8)	**Borderline included[b] (cut-off ≥5)**	**Borderline excluded[a] (cut-off ≥8)**	**Borderline included[b] (cut-off ≥5)**	**Borderline excluded[a] (cut-off ≥8)**	**Borderline included[b] (cut-off ≥5)**
Sensitivity for active TB						
All TB	81.4 (76.6–85.3)	82.3 (77.8–86.1)	89.2 (85.2–92.2)	89.9 (86.2–92.7)	88.0 (83.8–91.2)	89.0 (85.1–91.9)
Culture-confirmed TB	84.9 (79.5–89.0)	85.9 (80.9–89.8)	94.0 (90.0–96.4)	94.4 (90.7–96.7)	93.4 (89.2–96.0)	94.0 (90.2–96.4)
Specificity for active TB excluded	86.2 (82.3–89.4)	82.6 (78.6–86.1)	80.0 (75.6–83.8)	76.5 (72.0–80.4)	79.6 (75.2–83.4)	76.3 (71.8–80.3)

[a]Analyses *exclude* borderline IGRA results, giving a cut-off for a positive test result of ≥*8* SFCs after subtraction of negative control.

[b]Analyses *include* borderline IGRA results, giving a cut-off for a positive test result of ≥*5* SFCs after subtraction of negative control.

NB: References referred to in these Supplementary Materials are listed in the References section of the primary manuscript.