CONFIDENTIAL

STATISTICAL ANALYSIS PLAN

Trial Title:

Two-site laboratory-based diagnostic accuracy and feasibility study of the Xpert MTB/XDR assay for detection of isoniazid, fluoroquinolone, ethionamide and second-line injectable anti-tuberculosis drug resistance

Short title:
TB-CAPT MTB/XDR Study

Version: 1.2

Date: 28 March 2021

Funder: EDCTP

Sponsor Name: FIND

Disease Programme: TB

Protocol number: 1.0

Study Statistician: Stefano Ongarello

Consultant on sample size calculation:
Thomas Keller (Acomed Statistik)

FIND
Campus Biotech
Chemin des Mines 9
1202 Geneva
Switzerland
T: +41 (0)22 710 05 90
F: +41 (0)22 710 05 99

Name and Function	**Date** *DD/MM/YYYY*	**Signature**
SAP Author		
Chad Centner Pathologist National Health Laboratory Service University of Cape Town	28-03-2021	
Supervisor of SAP author		
Helen Cox	28 March 2021	
Trial Manager		
Adam Penn-Nicholson	31 March 2021	
Head of Data Services and Biobanking (if not author)		
Stefano Ongarello	31MAR2021	

CONFIDENTIALITY STATEMENT:

Table of contents

1 Introduction

This document describes the statistical analysis plan for the following project: “Two-site laboratory-based diagnostic accuracy and feasibility study of the Xpert MTB/XDR assay for detection of isoniazid, fluoroquinolone, ethionamide and second-line injectable anti-tuberculosis drug resistance.”

1.1 Description of the study

The study aims to evaluate the Xpert MTB/XDR cartridge by estimating the diagnostic accuracy of the assay for the detection of resistance to isoniazid, fluoroquinolones, ethionamide and second-line injectable anti-tuberculosis drugs in clinical specimens positive for rifampicin-resistant *Mycobacterium tuberculosis* (*Mtb*) by Xpert MTB/RIF Ultra, using a composite reference standard. The composite reference standard comprises phenotypic susceptibility testing and whole genome sequencing (WGS) performed on culture isolates.

Secondary aims of the project are to estimate the diagnostic accuracy of Xpert MTB/XDR for detection of resistance to isoniazid, fluoroquinolones and second-line injectable anti-tuberculosis drugs compared to each method alone. Further secondary aims are to compare the diagnostic accuracy of the MTB/XDR assay to that of standard of care (SOC) testing, as well as compare the turnaround time (time-to-result) of Xpert MTB/XDR versus SOC. SOC at the two sites includes direct and indirect (culture-based) line probes assays (LPA)s.

Additionally, the study will investigate the feasibility of performing Xpert MTB/XDR on residual SR-sputum mix of rifampicin-resistant respiratory samples leftover from the Xpert MTB/RIF Ultra, under routine conditions.

1.2 Timing of the analysis

The final analysis will be conducted once all results are available (i.e. enrolment is complete and all phenotypic DST and genotypic testing results have been returned). All objectives will be analysed together at this time.

2 Statistical hypotheses and methods

2.1 Primary endpoints

1. The primary endpoint is the estimate of sensitivity and specificity of Xpert MTB/XDR assay vs. the composite reference standard for detection of resistance to:
   a. Isoniazid

b. Fluoroquinolones
c. Ethionamide
d. Second-line injectable anti-tuberculosis drugs

2.2 Secondary endpoints

1. Point estimated of sensitivity and specificity of Xpert MTB/XDR for detection of resistance to:
    a. Isoniazid
    b. Fluoroquinolones
    c. Second-line injectable anti-tuberculosis drugs

   compared to phenotypic anti-tuberculosis drug susceptibility testing and WGS separately

2. Point estimates of sensitivity and specificity of Xpert MTB/XDR vs. WGS for detection of mutations in the following gene targets:
    a. *inhA* promoter
    b. katG
    c. gyrA
    d. gyrB
    e. eis
    f. rrs

3. Point estimates of sensitivity and specificity of Xpert MTB/XDR compared to that of SOC testing against the composite reference standard for detection of resistance to:
    a. Isoniazid
    b. Fluoroquinolones
    c. Second-line injectable anti-tuberculosis drugs

4. Estimate of the proportion of participants for whom valid test results become available based on:
    a. Xpert MTB/XDR
    b. SOC testing:
        i. LPAs performed directly on specimen (“direct”)
        ii. LPAs performed on culture isolates (“indirect”)

5. Estimate of the median time-to-result (time in hours from sample receipt to finalised result):
    a. Separately for each drug class, and comparing:
        i. Xpert MTB/XDR
        ii. SOC testing:

1. Direct LPAs
2. Indirect LPAs

6. Feasibility endpoints:
    a. Estimate of the proportion of participants with sufficient residual SR-sputum mix (≥2 ml) available to perform Xpert MTB/XDR
    b. Median period between preparation of SR-sputum mix and start of Xpert MTB/XDR test
    c. Estimate of the proportion of specimens with non-determinate MTB/XDR results:
        i. Combined for all
        ii. Separately for each drug class
        iii. Separately for each target gene
        iv. Comparing specimens stored according to manufacturer recommendations (at 2-8ºC for ≤4 hours or 25-35ºC for ≤2.5 hours) and those exceeding those limits for operational reasons
        v. By subgroup:
            1. Sufficient specimen (yes/no)
            2. Smear status
            3. Age group

7. Exploratory endpoints:
    a. Point estimates of sensitivity and specificity of Xpert MTB/XDR assay for detection of resistance to each anti-tuberculosis drug class by subgroup:
        i. Sufficient specimen (yes/no)
        ii. Smear status
        iii. Age group
    b. Point estimates of sensitivity and specificity of Xpert MTB/XDR assay for detection of resistance to each anti-tuberculosis drug class by rifampicin susceptibility result concordance between Xpert MTB/RIF Ultra and LPA
    c. Estimates of the agreement (kappa statistic) between Xpert MTB/XDR and WGS for the detection of heteroresistance for each anti-tuberculosis drug class
    d. Description of discordance between phenotypic testing and WGS
    e. Description of discordance between testing methodologies by presence of heteroresistance

3 Trial population and analysis datasets

3.1 Criteria for eligibility, recruitment, withdrawal and follow-up

Inclusion:
- *Mtb*-positive, rifampicin-resistant respiratory specimen identified on Xpert MTB/RIF Ultra

Exclusion:
- Residual SR-sputum mix not retained or not found
- Patient previously included in the study

Exclusion for diagnostic accuracy and time-to-result endpoints:
- Insufficient residual SR-sputum mix remaining for Xpert MTB/XDR (<2 ml)
- Xpert MTB/XDR unsuccessful
- No second / follow-up specimen received
- Second / follow-up specimen culture-negative, contaminated or not available
- Reference standard uninterpretable (phenotyping or WGS)
    - Where phenotypic susceptibility testing results are uninterpretable, specimens will still be included in WGS comparison
    - Where WGS results are uninterpretable, specimens will still be included in phenotypic susceptibility testing comparison

3.2 Analysis datasets

Analysis populations:

- Intended to test (ITT): all samples intended to be tested
- Per protocol (PP): all samples with valid test results for the index test and reference standard

For the primary, diagnostic accuracy and time-to-result outcomes, the PP population will be analysed. For the feasibility endpoints, the ITT population will be analysed.

4 Description of statistical methods

4.1 Estimation of sensitivity and specificity

The following 2x2 table is defined:

Contingency table for analysis of diagnostic accuracy:

		RS		
		Positive	Negative	Total
IT	Positive	TP = a	FP = b	a+b

	Negative	FN = c	TN = d	c+d
	Total	a+c	b+d	a+b+c+d

RS=reference standard, IT=index test, TP=true positive, FP=false positive, FN=false negative, TN=true negative

The diagnostic accuracy measures and their confidence intervals are estimated according to the following formulae:

Formulas and rules to calculate diagnostic accuracy measures:

Measure	**Formula**	**Confidence interval**
Sensitivity	a/(a+c)	2-sided 95%-Wilson-Score-CI
Specificity	d/(b+d)	2-sided 95%-Wilson-Score-CI

4.2 Analysis endpoints

1. Estimates of sensitivity and specificity of Xpert MTB/XDR will be calculated compared to the reference standards defined below (for primary outcomes 1a-d):

DRUG RESISTANCE	**COMPOSITE REFERENCE STANDARD**
INH-resistant	NGS of *katG, fabG1, ahpC* and *inhA* detects ≥1 mutation associated with INH resistance or MGIT culture is INH drug-resistant at the critical concentration.
ETH-resistant	NGS of *inhA* detects ≥1 mutation associated with ETH resistance
FQ-resistant	NGS of *gyrA* and *gyrB* detects ≥1 mutation associated with MFX resistance or MGIT culture is levofloxacin drug-resistant at the critical concentration.
Injectable-resistant	NGS of rrs, eis detects ≥1 mutation associated with injectable resistance or MGIT DST is kanamycin-resistant
INH-susceptible	NGS of *katG, fabG1, ahpC* and *inhA* does NOT detect any mutation associated with INH resistance and MGIT culture is INH-susceptible at the critical concentration.
ETH-susceptible	NGS of *inhA* does NOT detect any mutation associated with ETH resistance
FQ-susceptible	NGS of *gyrA* and *gyrB* does NOT detect any mutation associated with MFX resistance and MGIT culture is MFX-susceptible at the critical concentration.
Injectable-susceptible	NGS of *rrs* and *eis* does NOT detect any mutation associated with injectable-resistance and kanamycin MGIT DST is susceptible

Secondary outcomes:

1. Estimates of sensitivity and specificity of Xpert MTB/XDR compared to WGS or phenotyping alone

2. Estimates of sensitivity and specificity of Xpert MTB/XDR to detect mutations in each gene target compared to WGS

3. Compare estimates of diagnostic accuracy of Xpert MTB/XDR to the estimates of diagnostic accuracy of SOC susceptibility testing against the composite reference standard using McNemar chi-square test

4. Compare proportion of participants for whom valid susceptibility testing results become available based on Xpert MTB/XDR to the proportion of participants for whom valid susceptibility testing results become available based on SOC testing

5. Time-to-result comparing Xpert MTB/RIF Ultra testing to first results from Xpert MTB/XDR using Wilcoxon rank-sum test or t-test.

6. Feasibility
    a. Proportion of participants with sufficient residual SR-sputum mix (≥2 ml) (total number of samples included as the denominator)
    b. Median period between processing for Xpert MTB/RIF (preparation of SR-sputum mix) and the start of the Xpert MTB/XDR test
    c. Determine the non-determinate rate of Xpert MTB/XDR (number of non-determinate result divided by total number of tests:
        i. For each drug separately and combined for all
        ii. Comparing specimens stored according to manufacturer recommendations (at 2-8ºC for ≤4 hours or 25-35ºC for ≤2.5 hours) and those exceeding those limits for operational reasons (Chi-squared test)
        iii. By Xpert MTB/RIF Ultra semi-quantitation result (chi-squared test)
        iv. By sub-group (sufficient specimen, smear status, age group) (chi-squared test)

7. Exploratory analyses

5 Baseline descriptive statistics

Descriptive statistics tables will be generated to summarize the characteristics of the participants whose samples have been included in the study. The number of samples that have been included and excluded will be reported. Among the included participants, the information will be broken down by sex, age and ward type.

Results will be reported either in absolute numbers or summarized by mean, median, standard deviation, minimum, maximum and quartiles.

6 Planned interim analyses

No interim analyses are planned.

7 Additional sub-group analyses

Subgroups analyses are described in section 2.2 (6c iv and v)

Definition of sub-groups:

- Sufficient specimen: Group with ≥2 ml SR-sputum mix available vs. group with <2 ml SR-sputum mix
- Smear status: Positive for acid-fast bacilli (any quantitation) vs. negative for acid-fast bacilli
- Age group: Age <12 vs. age ≥12 years

8 Multiple comparisons/multiplicity adjustments

No multiple testing of statistical hypotheses will be performed.

9 Exploratory analyses

Exploratory analyses will be performed post hoc.

10 Sample size

A sample size of 320 specimens will be used in the diagnostic accuracy study. The sample size is dictated by the resources allocated to the project, and it is expected that the data generated will be combined in meta-analyses with other studies currently underway.

The sample size calculations detail the uncertainty of estimation of sensitivity among reference standard (RS)-positives and specificity among RS-negatives, expressed by the width of the 2-sided 95% confidence intervals. For technical reasons, the Pearson-Clopper-CI have been used in sample size considerations; the differences compared to the Wilson-Score-CI planned to be used in statistical analysis are negligible. The widths of the 95%-Pearson-Clopper-intervals for sensitivities / specificities of 70%, 80% or 90% were dependent on the expected prevalence of resistance to a specific anti-tuberculosis drugs (between 11 and 50%) and sample sizes of 320, 500 or 750. The width ranges from >30% (70% sensitivity at a prevalence [for RF-positive] of 11% among 320 patients) down to 4.7% (90% specificity at a prevalence [or RF-negative] of 89% among 750 patients).

The table shows expected prevalence of resistance to anti-tuberculosis drug class and the related number of resistant and susceptible specimens among 320, 500 and 750 participants together with confidence intervals for proportions of 70%, 80% and 90%:

Type of resistance	Resistant	Prev.	Measure	D+ or D- among								
				320			500			750		
				N	lCL	uCL	N	lCL	uCL	N	lCL	uCL
Isoniazid resistance	yes	61%	n (Prev.)	196			305			458		
			70% Sens	137	63.0%	76.3%	214	64.5%	75.1%	321	65.6%	74.2%
			80% Sens	157	73.7%	85.4%	244	75.1%	84.3%	366	76.0%	83.6%
			90% Sens	176	84.9%	93.8%	275	86.1%	93.1%	412	86.9%	92.6%
	no	39%	n (1-Prev.).	125			195			293		
			70% Spec	88	61.2%	77.9%	137	63.0%	76.3%	205	64.4%	75.2%
			80% Spec	100	71.9%	86.6%	156	73.7%	85.4%	234	75.0%	84.4%
			90% Spec	113	83.4%	94.6%	176	84.9%	93.8%	264	86.0%	93.2%
Second-line injectable resistance	yes	15%	n (Prev.)	48			75			113		
			70% Sens	34	54.9%	82.5%	53	58.3%	80.0%	79	60.6%	78.3%
			80% Sens	38	65.9%	90.2%	60	69.2%	88.4%	90	71.4%	86.9%
			90% Sens	43	77.6%	96.9%	68	80.9%	95.7%	102	82.9%	94.9%
	no	85%	n (1-Prev.)	272			425			638		
			70% Spec	190	64.2%	75.4%	298	65.4%	74.3%	447	66.3%	73.5%
			80% Spec	218	74.7%	84.6%	340	75.9%	83.7%	510	76.7%	83.0%
			90% Spec	245	85.8%	93.3%	383	86.7%	92.7%	574	87.4%	92.2%
Fluoroquinolone resistance	yes	11%	n (Prev.)	36			55			83		
			70% Sens	25	52.3%	84.2%	39	56.1%	81.6%	58	58.9%	79.6%
			80% Sens	29	63.2%	91.5%	44	67.0%	89.6%	66	69.8%	88.0%
			90% Sens	32	75.2%	97.5%	50	78.9%	96.4%	75	81.3%	95.6%
	no	89%	n (1-Prev.).	285			445			668		
			70% Spec	200	64.3%	75.3%	312	65.5%	74.2%	468	66.4%	73.5%
			80% Spec	228	74.9%	84.5%	356	76.0%	83.6%	534	76.8%	83.0%
			90% Spec	257	85.9%	93.2%	401	86.8%	92.6%	601	87.5%	92.2%
Ethionamide resistance	yes	50%	n (Prev.)	160			250			375		
			70% Sens	112	62.3%	77.0%	175	63.9%	75.6%	263	65.1%	74.6%
			80% Sens	128	73.0%	85.9%	200	74.5%	84.8%	300	75.6%	83.9%
			90% Sens	144	84.3%	94.2%	225	85.6%	93.4%	338	86.5%	92.8%
	no	50%	n (1-Prev.)	160			250			375		
			70% Spec	112	62.3%	77.0%	175	63.9%	75.6%	263	65.1%	74.6%
			80% Spec	128	73.0%	85.9%	200	74.5%	84.8%	300	75.6%	83.9%
			90% Spec	144	84.3%	94.2%	225	85.6%	93.4%	338	86.5%	92.8%

D+ =presence of resistance, D- =absence of resistance, Prev=prevalence, y=yes, n=no, lCL=lower confidence level, uCL=upper confidence level
Red and green: widest and narrowest interval
Due to rounding errors, deviations of +/-1 case could occur for total N.

It is expected that a proportion of specimens will not be available for the diagnostic accuracy study because of an invalid Xpert MTB/XDR result, unavailability of a second or follow-up specimen and negative or contaminated culture. An additional 433 specimens will be required giving a final total of N=753:

Table: Number of specimens required to undergo Xpert MTB/XDR to achieve final sample size of N=320 for diagnostic accuracy study, and anticipated enrolment period

	UCT		WHC		Total
	%	N	%	N	N
Total specimens required		373		380	753
Available for feasibility study	100	373	100	380	753
Not available for diagnostic accuracy and time-to-result study	41	153	71	280	433
Insufficient specimen remaining for Xpert MTB/XDR	1	4	1	4	
Xpert MTB/XDR indeterminate	5	19	5	19	
No second / follow-up specimen	10	37	40	152	
Culture-negative or contaminated	20	75	20	76	
MGIT not available	5	19	5	19	
Available for diagnostic accuracy and time-to-result study	59	220	29	100	320
Enrolment period (@ 2 rifampicin-resistant specimens/day)		37 weeks		38 weeks	

11 Minimization of error and bias

Consecutive, unselected samples with rifampicin-resistant MTB detected on Xpert Ultra testing in the routine laboratory will be used. These will include samples from a wide variety of facility types, geographical areas and patient/clinical backgrounds. Patients will only be excluded if they have been included previously, and not for any patient related factors. The study population will therefore be directly representative of the future target population and context of the Xpert MTB/XDR test.

Reference standard phenotyping, WGS and downstream bioinformatics analysis will be performed blinded to the index test result, which will have been performed in a different laboratory several weeks or months earlier.

11.1 Enrolment and randomization procedures

Enrolment will be consecutive rifampicin-resistant samples identified on routine Xpert MTB/RIF Ultra testing. There will be no randomisation.

12 Case definitions

The composite reference standard comprises phenotypic susceptibility testing and WGS performed on culture isolates. For the composite reference standard, samples will be defined as “resistant” if they are resistant by either phenotypic susceptibility testing or WGS and “susceptible” if they tested susceptible by both methods.

13 Statistical software

The analysis will be performed using Stata (version 16.1 or higher) and Microsoft Excel (version 2012).

14 References

None

15 Document history

Version	Notes / Changes
1.0	Initial version
1.1	Addressed comments by SO
1.2	Further corrections and removed tracked changes