A critical review of the XTEND trial comparing Xpert MTB/RIF with smear microscopy for the diagnosis of tuberculosis under routine settings:

Why a highly sensitive test failed to improve patient mortality

Kerrigan Mary McCarthy

Original published work submitted to the
Faculty of Health Sciences, University of the Witwatersrand
in fulfilment of the requirements
for the degree of Doctor of Philosophy

Johannesburg 2019

Declaration

Student's contribution to articles and agreement of co-authors

I, Kerrigan McCarthy, student number 9000754Y, declare that this Thesis is my own work and that I contributed adequately towards research findings published in the articles stated below which are included in my Thesis.

Signature of Student Date 29 September 2019

Signature of Primary Supervisor .....Date 24th September 2019

Agreement by co-authors:

By signing this declaration, the co-authors listed below agree to the use of the article by the student as part of her Thesis.

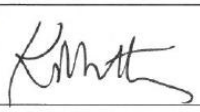
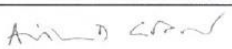
Article 1: Implementation and Operational Research: What Happens After a Negative Test for Tuberculosis? Evaluating Adherence to Tuberculosis Diagnostic Algorithms in South African Primary Health Clinics.
Reference: J. Acquir. Immune Defic. Syndr. 2016;71(5):e119-126.

Supervisor's comment: This paper was a secondary data analysis of data originating from the XTEND study. The candidate was the Research Scientist and Project Manager for the XTEND Study and managed all aspects of the XTEND study including data collection, cleaning and interpretation. Regarding this article, the candidate conceptualised the project together with PhD supervisors, identified and cleaned the data, conducted data analysis with the support of KL Fielding, wrote the first draft and compiled revisions in response to comments from co-authors.

Authors	Name	Signature	Date
1st author	KM McCarthy		26/09/2019
2nd author	AD Grant		22 September 2019
3rd author	V Chihota	V.Chihota	26 September 2019
4th author	S Ginindza		20-09-2019
5th author	L Mvusi		26.09.2019
6th author	GJ Churchyard		24/9/19
7th author	KL Fielding		20 September 2019

Article 2: Empiric tuberculosis treatment in South African primary health care facilities - for whom, where, when and why: Implications for the development of tuberculosis diagnostic tests.

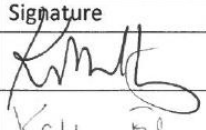
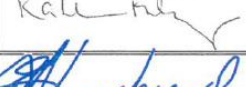
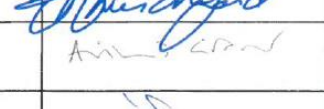
Reference: PLoS One. 2018;13(1):e0191608.

Authors	Name	Signature	Date
1st author	KM McCarthy		26/09/2019
2nd author	KL Fielding		20 September 2019
3rd author	GJ Churchyard		24/9/19
4th author	AD Grant		22 September 2019

Supervisor's comment: This paper was a secondary data analysis of data originating from the XTEND study. The candidate was the Research Scientist and Project Manager for the XTEND Study and managed all aspects of the XTEND study including data collection, cleaning and interpretation. Regarding this article, the candidate conceptualised the project together with PhD supervisors, identified and cleaned the data, conducted data analysis with the support of AD Grant, wrote the first draft and compiled revisions in response to comments from co-authors.

Article 3: A forgotten angle in impact evaluations of tuberculosis diagnostic tests: the assessment of integration and embedding of the Xpert MTB/RIF diagnostic technology into routine care using normalisation process theory PLOS ONE.

Reference: PLoSOne, submitted and accepted 9 September 2019

Authors	Name	Signature	Date
1st author	KM McCarthy		26/9/2019.
2nd author	KL Fielding		20 September 2019
3rd author	GJ Churchyard		24/9/19
4th author	AD Grant		22 September 2019
5th author	K Kielmann		23 September 2019

Supervisor's comment: This paper was conceptualised by the candidate and PhD supervisors, and involved a qualitative analysis of data collected by the candidate through semi-structured interviews. The candidate together with K Kielmann, developed a conceptual framework, coded and analysed data, wrote the first draft of the paper and compiled revisions in response to comments from co-authors.

To my parents,

Terence Sinclair McCarthy

and

Erna Louise McCarthy,

and

to my daughters,

Gemma Sutherland

and

Carys Sutherland,

with appreciation.

Presentations given by the candidate arising from this research project

1. 'Modelling public health responses to communicable disease: the story of TB diagnostics'. Keynote Address. South African Centre for Epidemiological Modelling and Analysis (SACEMA) Research Day. Stellenbosch, South Africa. 10 September 2019.
2. 'A qualitative evaluation of the implementation of Xpert MTB/RIF and the algorithm for tuberculosis diagnosis at primary health clinics in South Africa during the XTEND trial. ' Oral presentation at 47th Conference of the International Union against Tuberculosis and Lung Disease, Liverpool, United Kingdom. 13 October 2016.
3. 'Getting the most out of Xpert MTB/RIF: An evaluation of adherence to the algorithm for TB diagnosis in South African Primary Health Clinics during the XTEND trial'. Oral presentation at the South African HIV Clinician's Society Conference. Sandton, Johannesburg. South Africa. 14 April 2016.
4. 'A negative test for TB by Xpert MTB/RIF or smear microscopy: What does it mean for HCW and persons suspected of TB? A PhD in-progress through the Faculty of Health Sciences, University of the Witwatersrand'. Invited faculty Lecture, Queen Margaret University, Edinburgh, United Kingdom. 29 April 2015.
5. 'How do health care workers investigate persons living with HIV/AIDS for TB when initial sputum tests are negative? Factors associated with adherence to the RSA algorithm for the diagnosis of TB amongst PLWHA'. Oral presentation at 45th Conference of the International Union against Tuberculosis and Lung Disease, Barcelona, Spain. 31 October 2014.
6. 'What happens after a negative Xpert or smear test? Investigating adherence to the algorithm for TB diagnosis in the XTEND trial'. Oral presentation at South African Tuberculosis Conference. Durban, South Africa. 10 June 2014.

Publications by the candidate arising from this research project

1. McCarthy KM, Grant AD, Chihota V, et al. Implementation and operational research: What happens after a negative test for Tuberculosis? Evaluating adherence to TB Diagnostic algorithms in South African Primary Health Clinics. J. Acquir. Immune Defic. Syndr. 2016; 71(5):e119-12612.

2. McCarthy K, Fielding K, Churchyard GJ, Grant AD. Empiric tuberculosis treatment in South African primary health care facilities - for whom, where, when and why: Implications for the development of tuberculosis diagnostic tests. PLoS One. 2018; 13(1):e019160813.

3. McCarthy K, Grant AD, Fielding K, Churchyard GJ, Kielmann K. A forgotten angle in impact evaluations of tuberculosis diagnostic tests: the assessment of integration and embedding of the Xpert MTB/RIF diagnostic technology into routine care using normalisation process theory PLOS ONE. Submitted and accepted 9 September

Abstract

Beginning in 2011, a cartridge-based PCR assay with a two-hour turn around time known as Xpert MTB/RIF (hereafter 'Xpert') was rolled out as a replacement for smear microscopy for tuberculosis diagnosis in South Africa. The staged roll-out made it possible to conduct a pragmatic cluster-randomised trial (Xpert for TB – evaluating a new diagnostic, or 'XTEND') to determine the impact of Xpert on patient and programme outcomes. XTEND results indicated that Xpert identified one and a half times more tuberculosis cases. However, there was no difference in loss to follow up, the proportion of participants started on tuberculosis treatment or mortality rates across intervention (Xpert) and control (smear microscopy) arms. This PhD aimed to investigate reasons for the absence of a reduction in mortality by exploring contextual factors that may have impacted Xpert performance during the XTEND trial.

A literature review of impact evaluations of Xpert conducted between 2012-2019 and done as a part of this PhD suggests that the more systematically and rigorously tuberculosis diagnostic tests are conducted at patient enrolment and follow-up as part of study procedures, the less likely a study will find a mortality difference between intervention (Xpert) and control (smear) arms. Findings from the literature review suggested that a careful evaluation should be done of the extent of adherence to routine tuberculosis diagnostic algorithms in the XTEND study.

In a secondary data analysis of XTEND data, we showed poor levels of adherence to tuberculosis diagnostic algorithms. We identified greater levels of adherence to algorithms for tuberculosis diagnosis in the smear microscopy arm and amongst persons who were more ill, male or visited clinic repeatedly. In addition, we observed large between-clinic variation in adherence rates. In mapping the pathways to care for XTEND participants who were initiated empirically on tuberculosis treatment, we showed that empiric tuberculosis treatment was infrequently prescribed in the real-world setting of XTEND, where participating clinics were left to implement tuberculosis diagnostic algorithms without study intervention. Furthermore, we observed that a majority of empiric tuberculosis treatment initiations took place in the smear microscopy arm. In a qualitative evaluation at eight representative XTEND clinics, we identified large clinic-to-clinic variation with regard to the extent of Xpert implementation and embedding, and district health system strengths and weaknesses. We showed the impact of co-ordination of resources, creation and sustaining of referral networks and ownership of the tuberculosis programme by managers in supporting full implementation of and adherence to the algorithm for TB diagnosis using Xpert. Within clinics, we identified that motivated and well-trained facility managers, and a willing primary health clinic staff complement have the resources and motivation required to co-ordinate tuberculosis diagnosis, maintain referral networks, interpret Xpert results and translate tuberculosis testing into improved outcomes.

Drawing these findings together, we recognised that the XTEND trial did not measure certain factors that impacted on Xpert performance, including the degree of adherence to the algorithms for tuberculosis diagnosis, nor the rates of empiric tuberculosis treatment. The XTEND trial did not assess how tuberculosis diagnosis was implemented by health care workers. We showed that the provision of additional investigations to XTEND participants who were more ill decreased the mortality rates in both study arms. This may have led to a reduction in statistical power to detect a mortality difference between study arms. Further, we

demonstrated higher rates of empiric treatment amongst persons in the smear microscopy arm, which may have decreased the mortality difference between study arms. Our qualitative work revealed high clinic-to-clinic variation in the degree of integration of Xpert, which is unlikely to have introduced a systematic bias in mortality rates across study arm.

In conclusion, we found that the Xpert roll-out did not impact patient outcomes in the XTEND trial because increased sensitivity of Xpert and increased yield was offset by additional efforts to diagnose tuberculosis in participants who were at risk of death, and the more frequent provision of empiric TB treatment in the XTEND smear microscopy study arm. While Xpert may not have affected patient outcomes, a plethora of new diagnostic tests are under evaluation. The findings from this PhD highlight important methodological considerations in the design and conduct of impact evaluations of new tuberculosis diagnostic tests.

Acknowledgements

Research is never an individual pursuit. Even though the body of work, which I submit, contains my unique contributions, these are embedded in a collective output for which no single person can take credit. I owe an enormous debt of gratitude to all whom I had the privilege and pleasure of working alongside during and after the XTEND trial. The experience I have gained in every aspect of conducting research has and will continue to offer me resources for ongoing contributions to my professional and personal life.

I would like to thank Gavin Churchyard for the initial opportunity to work at the Aurum Institute and to Salome Charalambous, for making me feel so at home and welcome in my new role there. Violet Chihota was a continual source of calm considered pragmatic wisdom regarding strategic decision making as well as the day-to-day management of the XTEND trial. It was a pleasure to manage the XTEND study team and travel around the country with study co-ordinators Bandile Ndlazi, Siphokazi Mngcozelo and Nolundi Msheshwe. I have most enjoyable memories of conversations on the road with WHO colleague Christopher Dye, who introduced me to mathematical modelling and with Michael Kimmerling of the Bill and Melinda Gates Foundation, who opened my eyes to issues in global health. My Aurum PhD co-sojourners added so much fun and collective momentum to the journey. Thank you to Tendesyi Kufa-Chakedza, Yasmeen Hanifa, Sanj Karat and Noriah Matabane – I think I am the last to finish!

Each of my supervisors made the road easier, not the least through your steadfast encouragement over the recent three and most difficult years, but also through your friendship, which is a lasting gift. I have appreciated Katherine's dedication to explaining aspects of statistical analyses from simplest to complex, Gavin's down-to-earth and pragmatic focus, Alison's sense of humour and commitment to getting the message, facts and words exactly right and Karina's gift – bringing light and hope to my pursuit of science by imparting some skills in qualitative evaluation.

Lastly I'd like to thank all who made this work practically possible – to Kathy Kahn and Paul Bohloko and the team at the Wits School of Public Health for creating an enabling learning environment and to Lynn Morris, Lucille Blumberg, Irma Latsky and the team at the National Institute for Communicable Disease for generously allowing days off to complete this work, and for tirelessly smiling at repeated and seemingly endless mentions of 'I've got to do my PhD'. It is now finished.

Contents

List of Figures

List of Tables

List of Abbreviations

aOR	adjusted odds ratio
ART	anti-retroviral therapy
BMI	body mass index
CI	confidence interval
CRT	cluster-randomised trial
DR-TB	drug-resistant tuberculosis
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HCW	health care worker
HIV	human immunodeficiency virus
ICMJ	International Committee of Medical Journals
IPT	isoniazid preventive therapy
IQR	inter-quartile range
LED-FM	light-emitting-diode fluorescent smear microscopy
LTFU	loss to follow-up
NDoH	National Department of Health
NHLS	National Health Laboratory Service
NPT	normalisation process theory
OR	odds ratio
PCR	polymerase chain reaction
PHC	primary health care
QUADAS	Quality Assessment of Studies of Diagnostic Accuracy
RCT	randomised controlled trial
RSA	Republic of South Africa
STAG-TB	Strategic Advisory Group for Tuberculosis
STARD	Standards for the Reporting of Diagnostic accuracy studies
TB	tuberculosis
TCP	tuberculosis control programme
WHO	World Health Organization
XTEND	Xpert for TB – Evaluating a New Diagnostic

Part I: Accompanying narrative

Chapter 1: Introduction

1. 1 Introduction and background

The search for new improved tuberculosis diagnostic tests to accompany and make accessible the possibility of patient cure through antimycobacterial treatment, has been at the forefront of laboratory technology since the latter part of the 20th century[1]. Robert Koch and colleagues, in the late 19th century discovered that *Mycobacterium tuberculosis (M. tb)* was responsible for the disease known as 'consumption'. He and colleagues developed mycobacterial culture and microscopy techniques that, with a few modifications have remained relatively unchanged for over a hundred years[2].

In current programmatic use, both sputum smear microscopy and culture have limitations: smear microscopy results can be available within 12-24 hours, but the test has a threshold of detection of 10,000 bacteria/ml of sputum[3]. Culture, with a threshold of detection of 10-100 *M. tb* organisms, has a minimum turn-around time of 14 days using latest available technology[4]. It was only with the advent of molecular detection technologies using the polymerase chain reaction (PCR) that it became possible to detect small amounts of *M. tb* easily, early and economically. In 2010, an evaluation of a particular PCR-based technology – the cartridge-based Xpert MTB/RIF (hereafter 'Xpert') used on the GeneXpert platform (Cepheid Inc, Sunnyvale, California, United States of America) revealed promising results. The Xpert assay comprises a plastic cartridge with sample processing reagents and the GeneXpert instrument, which controls movements of reagents between cartridge compartments and which performs real-time PCR analysis. The PCR reaction in the Xpert cartridge amplifies a sequence of the *rpoB* gene specific to the *M. tuberculosis* complex, and probes for mutations within the 'rifampicin-resistance determining region' (RRDR) of the *rpoB* gene[5]. The test has a threshold of detection 100-1000 *M. tb* organisms and a turn-around-time of 2 hours[5]. Early laboratory-based multinational trials with this test in South Africa and elsewhere, demonstrated vastly improved sensitivity for the detection of *M. tb* over smear microscopy with the additional benefit of improved turn-around-times[6]. In 2011, the World Health Organization Strategic and Technical Advisory Group for Tuberculosis (WHO STAG-TB) published a strong recommendation supporting the use of Xpert as a first-line diagnostic test for persons with presumptive tuberculosis[7]. Many policy makers, programmatic experts and researchers held hopes that Xpert would provide tuberculosis control programmes with more accurate and rapid diagnosis, which would ultimately impact on patient outcomes, reduce tuberculosis transmission and control the epidemic[8,9].

In 2011, in response to these developments, the South African Minister of Health, Dr Aaron Motsoaledi made a decision to roll out Xpert as the first line diagnostic test for tuberculosis across the country in an attempt to control South Africa's rampant tuberculosis epidemic[10]. The implementation of the Xpert across the National Health Laboratory Service (NHLS) could not be done instantaneously. The phased roll-out afforded the opportunity to conduct a comparison of Xpert *vs* smear microscopy using a pragmatic (field) cluster-randomised study design to quantify the impact on patient and programme level outcomes. The study was undertaken from 2012-2014 through funding by the Bill and Melinda Gates foundation and was named 'XTEND' -Xpert for TB, Evaluating a New Diagnostic[11]. When the results of XTEND were analysed, Xpert identified one and a half times more tuberculosis cases. However, the improved sensitivity for tuberculosis

diagnosis did not translate into a reduction of all-cause mortality amongst persons investigated for tuberculosis, nor were more persons initiated on tuberculosis treatment[11]. Further, initial loss to follow-up rates were identical in intervention and control arms[11].

This PhD submission describes findings from two secondary data analyses and a qualitative investigation to better understand the reasons for the XTEND trial finding of 'no difference in mortality' between persons undergoing investigation for tuberculosis who were tested with Xpert compared with smear microscopy.

1. 2 An overview of this PhD

The PhD is divided into two parts. Part 1 comprises the chapters below

- Chapter 1 provides a brief background and overview of the PhD and a summary of the XTEND trial findings.
- Chapter 2 summarises the research question and lay out the aims and objectives of this PhD.
- Chapter 3 develops a contextual framework on which the objectives of this PhD are based. The contextual framework draws on theoretical approaches to the evaluation of new diagnostic tests, and specifically, new diagnostic tests for tuberculosis.
- Chapter 4 presents a narrative literature review of Xpert impact evaluations conducted between 2011-2019. focuses on contextual factors in study designs of the papers under review which may have impact on the reported mortality difference across Xpert and smear microscopy study arms. I use findings from this literature review to identify areas in the XTEND study that require close attention in terms of the broader aim of the PhD.
- Chapter 5 summarises findings from the paper entitled 'What Happens After a Negative Test for Tuberculosis? Evaluating Adherence to Tuberculosis Diagnostic Algorithms in South African Primary Health Clinics'. This paper describes findings of a secondary data analysis of XTEND trial participants and describes the extent of additional investigations for tuberculosis amongst persons with presumptive tuberculosis whose sputum tested negative for tuberculosis at enrolment.
- Chapter 6 summarises findings from the paper entitled 'Empiric tuberculosis treatment in South African primary health care facilities - for whom, where, when and why: Implications for the development of tuberculosis diagnostic tests'. This paper describes findings from a secondary data analysis of XTEND trial participants who were initiated on tuberculosis treatment without microbiological evidence during the six month follow-up period.
- Chapter 7 summarises findings from the paper 'A forgotten angle in impact evaluations of tuberculosis diagnostic tests: the assessment of integration and embedding of the Xpert diagnostic technology into routine care using normalisation process theory'. This paper describes a qualitative evaluation of how health care workers and tuberculosis control programme administrators implemented Xpert into tuberculosis diagnostic processes at eight representative primary care facilities which had participated in the XTEND trial.

- Chapter 8, draws the insights offered from the literature review and Papers 1, 2 and 3 together to describe contextual factors that may have impacted on Xpert performance during the XTEND trial, and I propose reasons for the absence of an impact on mortality.
- Chapter 9presents adiscussion of the limitations of the approach to this subject matter, and the limitations of the XTEND trial design that became evident through this study.
- Chapter 10concludes the study with a summary of the contribution of this PhD to knowledge pertaining to Xpert impact evaluations.

Part II contains the three papers submitted towards this PhD, and the XTEND study 'main paper' which is provided for context and the Appendices.

Papers submitted towards this PhD:

1. McCarthy KM, Grant AD, Chihota V, et al. Implementation and Operational Research: What Happens After a Negative Test for Tuberculosis? **Evaluating Adherence to TB Diagnostic Algorithms** in South African Primary Health Clinics. J. Acquir. Immune Defic. Syndr. 2016;71(5):e119-126[12].
2. McCarthy K, Fielding K, Churchyard GJ, Grant AD. **Empiric tuberculosis treatment** in South African primary health care facilities - for whom, where, when and why: Implications for the development of tuberculosis diagnostic tests. PLoS One. 2018;13(1):e0191608[13].
3. McCarthy K, Grant AD, Fielding K, Churchyard GJ, Kielmann K. A forgotten angle in impact evaluations of tuberculosis diagnostic tests: the assessment of **integration and embedding of the Xpert MTB/RIF diagnostic technology into routine care** using normalisation process theory PLOS ONE. *Submitted*

Papers arising from primary XTEND data and which provide important context for this PhD

4. Churchyard GJ, Stevens WS, Mametja LD, McCarthy KM, Chihota V, Nicol MP, Erasmus LK, Ndjeka NO, Mvusi L, Vassall A, Sinanovic E, Cox HS, Dye C, Grant AD, Fielding KL. Xpert versus sputum microscopy as the initial diagnostic test for tuberculosis: a cluster-randomised trial embedded in South African roll-out of Xpert. Lancet Glob Health. 2015;3(8):e450-457[11].

1. 3 Role of the PhD candidate in XTEND

I commenced work at the Aurum Institute in August 2011 in the role of XTEND Research Scientist and Project Manager. The grant application for the XTEND study had already been submitted and funding awarded by the Bill and Melinda Gates Foundation (OPP1034523). My role as XTEND Project Manager included recruitment, training and management of over 100 staff (primarily research nurses and study assistants, but also quality and data management staff) across 4 provinces and 40 study sites, co-ordination with the NHLS regarding the Xpert roll-out, communication and reporting to stakeholders, amendments to the protocol and submissions to national and international review boards, data cleaning, data review and interpretation, and protocol writing for sub-studies arising from XTEND. From 2011-2015, I worked with colleagues at Aurum on the XTEND trial, which was published in 2015 in Lancet Global Health[11].

In 2013 it became apparent that empiric diagnosis of tuberculosis in routine care could be explored using XTEND data. I developed a protocol to address the question of what happens to XTEND participants with negative initial sputum tests for tuberculosis. This was submitted and approved as a PhD proposal to the School of Public Health in the Faculty of Health Sciences, University of the Witwatersrand entitled '*A negative test for tuberculosis by Xpert MTB/RIF or smear microscopy – What does it mean for Health Care Workers and persons suspected of tuberculosis? A sub-study of the national evaluation of the use of Xpert vs smear microscopy for the diagnosis of tuberculosis by primary health clinics in South Africa*' (Ethics certification M140486, Part II, Appendix 1). In 2013, I commenced work on the PhD proposal, which evolved into the papers submitted as part of this PhD namely the secondary data analyses investigating adherence to the algorithms for tuberculosis diagnosis (paper 1) and the identification of XTEND participants who were initiated on empiric tuberculosis treatment (paper 2). With regard to the qualitative evaluation of Xpert implementation at a subset of XTEND clinics (paper 3), I identified the eight clinics for where the study was to take place. I developed and piloted the semi-structured interview format and worked on a conceptual framework. From 2014 until mid 2015, I travelled to the eight clinics where I conducted interviews with participants. I transcribed these and submitted field summaries of my findings to my supervisors by mid-2015.

In May 2015 when the XTEND study was officially completed, I left the Aurum Institute and took a position as Head of the Outbreak Response Unit at the National Institute for Communicable Diseases. Since then, I have continued to work on papers and this submission in my personal time. I conducted the primary and secondary data analyses for papers 1,2 and 3 from 2014-2016 with the support and assistance of my supervisors. I wrote and submitted papers 1, 2 and 3 from 2015 until 2019. I wrote the narrative of this PhD in December 2018-August 2019.

1. 4 The XTEND study - a summary of findings

In this section I provide a summary of the XTEND trial methodology and results, as these set the context for this PhD submission. The paper and supplementary appendices describing the main outcome of the XTEND trial are provided in Section II below.

1. 4. 1 XTEND research question and aims

The rationale of the XTEND study was to ascertain if Xpert, a test with better sensitivity and specificity than smear microscopy, would improve patient outcomes and tuberculosis control programme (TCP) indicators when used under routine conditions. The primary aim of XTEND was to measure the effectiveness of Xpert in reducing early mortality in persons with presumptive tuberculosis, defined as mortality six months after enrolment. The secondary objectives were to compare across study arms: 1) the proportion of participants with a positive result who did not start tuberculosis treatment within 28 days of enrolment, 2) the time from enrolment to start of treatment for drug susceptible and drug-resistant tuberculosis, and 3) the yield of confirmed tuberculosis cases across study arms.

1. 4. 2 XTEND methodology

XTEND was conducted as a pragmatic two arm parallel cluster randomised trial. A cluster was defined as a laboratory and two high burden primary health clinics. In collaboration with the National Department of Health (NDoH) and the NHLS, 20 laboratories in four provinces (Eastern Cape [8 laboratories], Free State [2], Gauteng [8] and Mpumalanga [2] were identified from medium-burden tuberculosis districts and randomised to the intervention or control arms. The two busiest clinics in each district were identified from which participants were to be recruited. As part of the roll-out, districts where Xpert was implemented were trained by the NDoH and the NHLS in the new Xpert algorithm for tuberculosis diagnosis[14] (Figure 1. 1A), while clinics in the control arm continued to use the smear microscopy algorithm (Figure 1. 1B)[15].

At participating clinics, sequential persons with presumptive tuberculosis (all or every second/third persons, depending on number requiring tuberculosis investigations per day) were invited to participate. Participants were investigated according to clinic practice using the NDoH algorithms (Figure 1. 1), and then offered enrolment by independent XTEND staff. Participants were then followed up by independent XTEND staff until six months after enrolment, when a follow up interview was conducted. At the post enrolment interview, NHLS laboratory results were obtained from their laboratory information system, case-records were obtained from patients' PHC file and reviewed, and if deceased, next-of-kin were asked about the circumstances leading to the participant's death.

Data analysis was conducted using a cluster-level analysis taking into account the stratified randomisation. A log-transformed cluster-level risk or rate was used to estimate geometric means for the two study groups. An approximate standard error for the log (risk or rate ratio) based on geometric means of cluster risks or rates was calculated by two-way ANOVA on stratum, group, and the interaction between stratum and group. Investigators also noticed a baseline imbalance between study groups in some factors associated with mortality, and adjusted for these individual level baseline factors. An individual-level risk factor analysis for mortality was also conducted using multivariable regression analysis.

The XTEND protocol was approved by the ethics committees of the University of the Witwatersrand; the University of Cape Town; the London School of Hygiene & Tropical Medicine; and the World Health Organization (WHO).

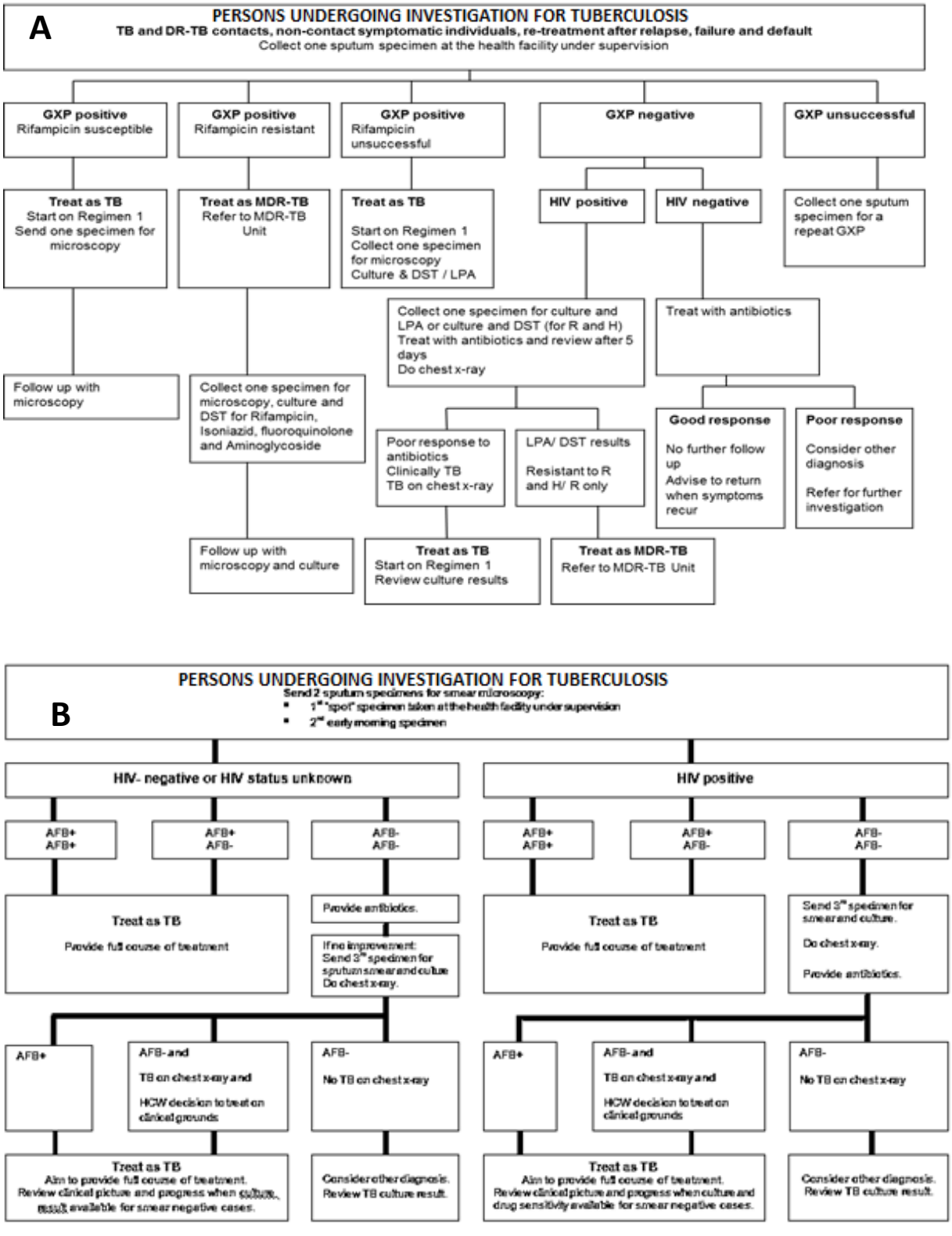


Figure 1. 1 Algorithms for the diagnosis of tuberculosis used in the XTEND intervention and control arms.

A: The algorithm for the diagnosis of tuberculosis used in the intervention arm of the XTEND trial, with Xpert as a first line sputum test for persons with suspected tuberculosis (South African National Guidelines for tuberculosis diagnosis and management, 2014). B: The algorithm for the diagnosis of tuberculosis used in the control arm of the XTEND trial, with smear microscopy as a first line sputum test for persons with suspected tuberculosis (South African National Guidelines for tuberculosis diagnosis and management, 2009)

1. 4. 3 XTEND results

Results of the primary and secondary outcomes are shown in Table 1 below and described in Paper 4. Analyses to identify individual-level risk factors for death identified the following risk factors: self-reported HIV-positive status amongst persons not on antiretroviral therapy (ART) (adjusted odds ratio (OR) 3·32, 95% confidence interval (CI) 2·03–5·41) or not knowing their HIV status (2·41, 1·47–3·98), *versus* being HIV negative; having a lower body mass index (BMI) (<18·5 kg/m² *vs* 18·5–24·9 kg/m²: aOR 1·40, 95% CI 0·94–2·07]), older age (<30 versus ≥50 years: aOR 2·53, 95% CI 1·59–4·04) and more tuberculosis symptoms (amongst current cough, night sweats, fever, or weight loss) (Table 2 in reference[11]). Tuberculosis treatment was initiated within a median of 7 days in the Xpert group and 10 days in the microscopy group. Initial loss to follow-up was documented for 16·0% (60/374) of participants with a positive index test result, which was similar by study group (Table 2 below).

Table 1. 1 Effect of Xpert on XTEND primary and secondary outcomes (Table 2 in Churchyard et al[11])

	Intervention		Control		Risk ratio (95% CI), p-value	
	Deaths/N	%*	Deaths/N	%*	Unadjusted	Adjusted[1]
Mortality risk over 6 months (n=4656)	91/2324	3. 9%	116/2332	5. 0%	0. 86 (0. 56-1. 28), p=0. 43	1. 10 (0. 75-1. 62), p=0. 61
Proportion with a positive result (n=4411)	200/2176	9. 2%	174/2235	7. 8%	1. 27 (0. 81-2. 00), p=0. 27	1. 49 (1. 00, 2. 23)[2], p=0. 05
Initial loss to follow-up (n=374)	34/200	17. 0%	26/174	14. 9%	0. 97 (0. 48-1. 96), p=0. 93	0. 96 (0. 48-1. 93)[3], p=0. 91
Proportion treated for TB (n=4656)	250/2324	10. 8%	291/2332	12. 5%	0. 88 (0. 60-1. 29), p=0. 48	1. 04 (0. 76-1. 43) [2], p=0. 79

CI=confidence interval; pyrs=person years; TB=tuberculosis *summary ignores cluster;
[1]adjusted for age group, sex, body mass index (BMI) group, number of symptoms & HIV; [2]adjusted for age group, sex, BMI group and number of symptoms; [3]adjusted for BMI group and number of symptoms; adding 0. 5 due to zero events in two clusters

1. 4. 4 XTEND conclusions

XTEND investigators noted the increased yield of microbiologically-confirmed tuberculosis, yet non-significant differences in mortality, initial loss to follow-up and treatment for tuberculosis. Regarding individual risk factors for mortality, they highlighted the association between untreated HIV, or unknown HIV status and death. They concluded that Xpert was unlikely to improve all-cause mortality in South Africa without strengthening HIV care and systems to retain in care persons undergoing investigation for TB and those initiated on TB treatment.

Chapter 2: The research question, aims and objectives

2. 1 Research question

Following the consistent demonstration of improved sensitivity of Xpert over smear microscopy in early laboratory-based evaluations[6,16], researchers began speculating on the longer term impact of improved diagnostics on the tuberculosis epidemic[17,18]. Improved case-finding including earlier detection of drug-resistant tuberculosis (DR-TB) was anticipated[19]. Reduction in time-to-treatment initiation with point-of-care application of Xpert was envisaged[20]. Mathematical modelling informed the research community that a reduction in morbidity and mortality due to tuberculosis was potentially possible following Xpert implementation when compared with smear microscopy[21].

However, as results from well-designed impact evaluations became available, including the XTEND trial findings[11,22-25], it was increasingly evident that the impact of Xpert testing on patient outcomes was less than hoped for. A number of reasons were offered, including the widespread practice of empiric tuberculosis treatment[26], failure to initiate patients on tuberculosis treatment following diagnosis and to retain them in care[27]. A careful narrative review of clinical trials of Xpert *vs* smear microscopy[28] identified weaknesses in trial design, trial conduct and the health system which the authors understood to contribute collectively to findings of no difference in patient mortality following Xpert implementation.

In the case of the XTEND trial, the sub-studies initiated in 2014 by the candidate and supervisors to investigate what happened to participants with negative sputum results presented an opportunity to better understand the reasons for no difference in mortality between cohorts of persons with presumptive tuberculosis who were tested using Xpert *vs* smear miscroscopy. We formulated a research question as follows:

> *What are the possible reasons for the XTEND trial finding of 'no difference in mortality between persons undergoing investigation for tuberculosis who were tested with Xpert compared with smear microscopy'?*

2. 2 Aim and objectives of this PhD submission

An aim statement

Diagnostic tests have potential to improve patient care through more accurate and timely identification of conditions requiring treatment[29]. However, diagnostic and treatment processes should support diagnostic testing so as to realise this potential[30]. Therefore we aimed

> *To better understand contextual factors that may have impacted on Xpert performance during the XTEND trial with regard to the management and investigation of persons with presumptive tuberculosis.*

Objectives

We identified three contextual areas of the XTEND trial for further investigation and set these out as objectives. We believed that these contextual areas were most likely to yield clues as to why a more sensitive test for tuberculosis did not translate into improved patient outcomes.

Our objectives were:

1. To establish the extent to which diagnostic algorithms for tuberculosis were adhered to in the XTEND trial;
2. To investigate how empiric tuberculosis treatment was prescribed in the XTEND trial;
3. To explore the extent to which Xpert and the algorithm for the diagnosis of tuberculosis was integrated and embedded into routine care during and after the XTEND trial.

Chapter 3: The conceptual framework for tuberculosis diagnosis and associated outcomes used in this evaluation

3. 1 Introduction and overview

In this chapter, I develop a contextual framework to support the objectives used in this evaluation. Firstly, I draw on literature describing theoretical approaches to the evaluation of the impact of diagnostic tests on patient outcomes. Secondly, I present the approach currently used by WHO in evaluating TB diagnostic tests. Thirdly, I utilize these theoretical approaches in a step-wise fashion to create a contextual framework. I identify key areas for evaluation and measurement in an impact evaluation of Xpert and specifically in the XTEND trial. Lastly, I explain how the contextual framework supports the objectives of this PhD.

3. 2 Evaluation of diagnostic tests

Diagnostic tests have potential to improve patient care through more accurate and timely identification of conditions requiring treatment[29]. However, diagnostic tests are rarely used in isolation, but rather as part of a test-treatment strategy within a health system[29]. Only in a system where diagnosis leads to changes in patient management, will improvements in patient outcomes be seen[31]. Therefore, a diagnostic test with better performance characteristics may not necessarily lead to better outcomes.

To ascertain the impact on patient outcomes, the 'downstream consequences'[31] of testing should be compared. Diagnostic testing with one group allocated to receive the newer/intervention test, and with long-term follow up of immediate and downstream consequences of testing is an evaluation strategy that is frequently used to compare diagnostic tests[29]. This study design is a randomised controlled trial (RCT) which has the potential to evaluate not only the diagnostic test, but the 'test-plus-treatment pathway'[29].

When setting up a RCT comparing impact of diagnostic tests on patient outcomes, the study designers should consider four design areas, namely; 1) they should be cognisant of the place of the new test in the diagnostic pathway – in other words, whether it will replace existing tests, be used as an 'add on' or 'triage test'[29]; 2) they should identify what the anticipated benefits of testing are and the pathways by which these occur; 3) they should incorporate quality criteria for conducting and analysing RCTs[32] and lastly 4) they should include strategies for qualitative assessment of the effectiveness of the intervention[33]. These four areas support the development of a conceptual framework to address reasons for the XTEND trial finding of 'no difference in mortality' between study arms.

3. 2. 1 The place of the test in the diagnostic pathway

The design of the RCT needs to accommodate the place of the new diagnostic test in the test-treatment pathway. Tests may replace older tests, be used as 'add-on' tests or as 'triage' tests to identify patients eligible for further testing. A test evaluation flow diagram can be used to display the existing test strategy

against the new test strategy[29]. This diagram can assist in identification of critical comparisons that may be measured and the pathways by which outcomes can be compared.

3. 2. 2 Elucidation of the pathways from diagnostic testing to improved patient outcomes

Tests may impact in patient health through a number of mechanisms[31], namely; 1) direct effects of the test (where a diagnostic procedure has therapeutic value, for example, where obtaining fluid for diagnostic purposes also alleviates swelling); 2) altering clinical decisions and actions; 3) changing timeframes of decisions and actions; and through 4) influencing patient and clinician perceptions. Each diagnostic test will have a unique pathway through which impact is achieved. A clear outline of these pathways will assist with identifying ways to measure impact along the test-treatment path.

3. 2. 3 Quality criteria for conducting and analysing trials of diagnostic accuracy.

Factors in the design and implementation of studies to evaluate diagnostic assays may impact or bias findings or their interpretation. Two tools were developed to support design and assessment of such studies. The QUADAS (Quality Assessment of Studies of Diagnostic Accuracy) tool assists designers and evaluators of trials to identify potential sources of bias in trial elements[34]. These include: composition of patient cohorts (spectrum composition); description of patient selection criteria; use of diagnostic reference standards; timing of follow-up in relation to disease progression and how diagnostic tests are applied to the intervention and control arms. These areas of a RCT, if not thoughtfully designed, could lead to partial or differential disease verification, test review bias, reference standard review bias and/or clinical review bias. Lastly, the tool requires designers to include a management strategy for patients with uninterpretable test results. The STARD (Standards for the Reporting of Diagnostic accuracy studies) tool was developed to improve conduct and the quality of reporting of diagnostic accuracy studies in general[32].

3. 2. 4 Qualitative assessment of intervention effectiveness

The health systems and programmes in which diagnostic assays are placed are socially constructed. Therefore many aspects of implementation and impact cannot be fully understood using quantitative methodologies[35]. The evaluation of diagnostic assays using the social sciences and other multidisciplinary knowledge paradigms yields richer, broader, deeper and more generalizable findings[35]. Qualitative evaluations examine how interventions are delivered or received, and guide researchers into a deeper understanding of who is implementing (or not) and under what conditions, and why it is (or is not) successfully implemented[33]. Useful components of qualitative evaluations include representative sampling and case selection, case studies, negative case analyses, triangulation of data from multiple sources, prolonged engagement with the subject of enquiry and the use of sociological or other theories in the development of conceptual frameworks[35].

3. 3 Evaluation of diagnostic tests for tuberculosis

3. 3. 1 The impact assessment framework for the evaluation of tuberculosis diagnostic tests

While tuberculosis researchers may apply the STARD tool and QUADAS criteria when designing and reporting on impact evaluations of tuberculosis diagnostic tests, guidance specific to tuberculosis diagnostic assays are more commonly used[18] . The evaluation of diagnostic tests for tuberculosis is guided by their stage in the test development continuum[1]. Early on in the continuum, immediately following test development and optimisation, laboratory-based assessments of test performance including test sensitivity and specificity are required[36]. The WHO STAG-TB committee evaluates results of these assessments using the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) methodology to inform their technical recommendations[18]. Later, evidence of impact on patient and programmatic outcomes including cost-effectiveness is required to support scaled implementation and national level adoption[36]. Presently the WHO advises the use of an impact assessment framework (Table 3.1) to support development of an evidence base to guide adoption of newer tuberculosis diagnostic tests[37]. The framework comprises five analytical layers, namely effectiveness, equity, health systems, scale-up and policy, each addressing the nature of evidence required by policy makers.

Table 3. 1 The impact assessment framework proposed by the World Health Organization for the evaluation of tuberculosis diagnostic tests[37]

Layer of assessment	Evidence required
Layer 1. Effectiveness analysis	• How well does the new tool work in terms of accuracy • How many additional cases will be identified who would otherwise not have been identified? • How many additional cases will actually start (and complete) treatment as a result of using the new tool
Layer 2. Equity analysis	• Who benefits from the new tool (ambulant *vs*. hospitalised, poor/less poor, men/women, adults/children)? • Why do these benefits accrue (level health system in which new diagnostic is deployed, change in time to issue of results, change in patient costs)
Layer 3. Health system analysis	• What are the human resource implications of introducing the new tool (training, number and care of staff)? • What are the infrastructure implications (equipment, laboratory layout, safety installations)? • What are the procurement implications (reagents, consumables, documentation)? • What are the implications for quality assurance (internal and external)?

Layer 4. Scale-up analysis	• What are the projected impacts of going to scale with the new tool?
	  ○ Cost savings to patients in relation to income
	  ○ Cost savings to health providers/the health system
	  ○ Effects on transmission of improved infection control as a result of the new tool

Layer 5. Policy analysis	• What other similar technologies are available or likely to become available • How do similar existing or emerging technologies compare in their projected performance within each of the layers above

3. 3. 2 Recent developments in impact assessment of tuberculosis diagnostic tests

Schumacher *et al* developed a helpful contextual framework for the evaluation of tuberculosis diagnostic tests[38] (Figure 3. 3). They identified four categories of appropriate outcome measures[38]. These are 1) technical performance (including diagnostic accuracy); 2) diagnostic impact (including time to diagnosis, number of confirmed diagnoses); 3) therapeutic impact (time to treatment initiation, number of patients placed on appropriate therapy); and 4) patient outcome impact (tuberculosis treatment outcome, mortality). When setting up an evaluation of a diagnostic test, tuberculosis researchers have advised that diagnostic and treatment pathways be mapped out, to better understand facilitators and barriers, and to identify meaningful end points. Researchers have cautioned against using surrogate end markers unless these have been validated[39].

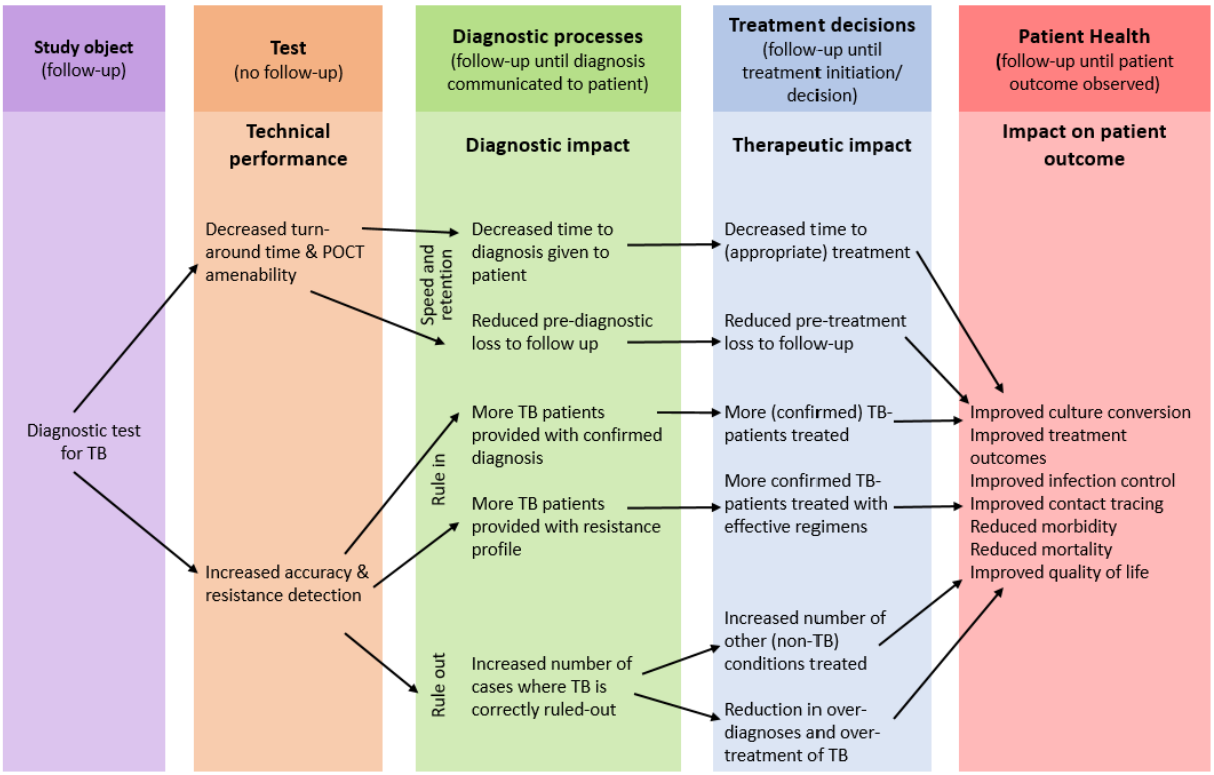


Figure 3. 3 A conceptual framework of outcome measures, outlining pathways through which improved tuberculosis diagnostics may lead to improved patient outcomes
POCT=point of care test. Figure adapted from Schumacher et al[38]

3. 4 A conceptual framework for evaluation of the impact of Xpert in the XTEND trial

We drew from literature presented above pertaining to the evaluation of diagnostic tests in general and tuberculosis diagnostic tests to draw up a conceptual framework to support the objectives of this PhD.

Firstly we mapped out the diagnostic pathways appreciating that Xpert was used as a replacement for smear microscopy[29].

Secondly we defined a 'good outcome' as 'survival after six months with tuberculosis treatment completion', and 'poor outcome' as 'death, or loss to follow-up, or non-completion of tuberculosis treatment'. We then

outlined patient pathways to good outcomes, as presented in Figure 3. 4. We drew from Schumacher *et al*'s conceptual framework[38] to appreciate that the primary assumption underlying pathways to good outcomes in the XTEND trial was that the increased accuracy of Xpert would lead to more tuberculosis patients being provided with confirmed results (statements B and C in Figure 3. 4), compared with the control arm, and that these patients would be initiated on and complete tuberculosis treatment. We recognised that the XTEND protocol assumed that failure to initiate tuberculosis treatment (empiric or following a laboratory-confirmed diagnosis) would result in poor outcomes (dotted lines E in Figure 3. 4), and that this applied across study arms. We recognized that XTEND measured the diagnostic yield of tuberculosis (statement B), the number of participants started on treatment (statement C) and the number lost to follow up (B-C) in the Xpert and control arms.

Thirdly, we recognised that good patient outcomes in XTEND were dependent on assumptions that were untested in the XTEND study[11], namely that additional investigations to rule-out false negative diagnoses were conducted (following statement A in Figure 3. 4), that persons with false negative tests were initiated on empiric tuberculosis treatment (statement D in Figure 3. 4), and that health care workers implemented Xpert and the new testing algorithm according to national guidelines (Box C in Figure 3. 4). We recognised that differential application of the testing algorithm in Xpert and control arms, or across clusters (Box C and specifically statement D) had potential to alter and therefore bias the survival difference in each study arm.

Fourthly, because XTEND and other trials of new TB diagnostic tests have focused primarily on the impact of improved sensitivity on outcomes, and because TB diagnostic tests generally have high specificity, we elected not to focus on the impact of false positive tests, nor on outcomes associated with these.

Lastly we recognised that qualitative rather than quantitative methods were better suited to evaluate how HCW implemented and integrated Xpert into routine diagnostic processes[35] (Box C).

3. 5 How this PhD's objectives are supported by the conceptual framework

The conceptual framework (Figure 3. 4) illustrates how tuberculosis investigative processes are key steps *en route* to good outcomes. The conceptual framework allows us to identify which of these key steps were measured in the XTEND study and which steps are unmeasured.

Tuberculosis diagnostic processes as required by South African national tuberculosis diagnostic algorithms (Figure 1. 1) include ascertainment of HIV status, chest radiography, sputum culture, provision of empiric antibiotics and revision of clinical response by a clinician. Referral to hospital may also be indicated if persons are seriously ill or if alternative diagnoses are suspected. The application of these processes may be collectively considered as 'contextual factors' supporting tuberculosis diagnosis using Xpert or smear microscopy.

In the XTEND study, allocation of clusters to the intervention (Xpert) *vs* control (smear microscopy) arm was determined by randomisation. Key XTEND processes in Box A and Box C, Figure 3. 4 were controlled and measured by the study staff. The XTEND trial[11] in each arm reported on

- the burden of tuberculosis diagnosed following enrolment sputum tests (I. e. statement B in Figure 3. 4),
- the rates of initial loss to follow up (a component of 'poor outcomes', landmark ix in Figure 3. 4),
- the proportions treated for tuberculosis (Statement C in Figure 3. 4) and
- mortality risk over six months by study arm (a component of 'poor outcomes', landmark ix in Figure 3. 4).

The conceptual framework also allows us to identify which key steps en route to good treatment outcomes were not measured in the XTEND trial. Whilst laboratories and clinics were trained in the use of Xpert and the new algorithm for tuberculosis diagnosis, the application of the intervention and control (Box A and Box C) was left entirely up to participating clinics. Sputum collection and submission to the laboratory was done by clinic staff using routine collection practices. Specimens were transported to NHLS laboratories using routine specimen transport. Results were returned to clinics and HCW using routine communication channels. XTEND neither monitored nor reported on tuberculosis investigative processes after submission of sputum for Xpert or smear testing nor compared these between clusters or study arms.

The conceptual framework (Figure 3. 4) identifies the following as key areas for evaluation and measurement in an impact evaluation of Xpert or any tuberculosis diagnostic test in addition to those measured in XTEND:

- Additional diagnostic testing in persons with false negative tests (pathway from landmarks iv-vii),
- Provision of empiric tuberculosis treatment (landmark vii),
- The degree to which additional diagnostic testing (Box C) is conducted and empiric tuberculosis treatment is prescribed in the intervention *vs* control arm, as a differential implementation may lead to differential case finding and bias study findings.
- Correct application of the tuberculosis diagnostic algorithm in each study arm (Box C as a whole)

Box B: outcomes following true negative or false positive tests are neither measured nor explored in the XTEND trial. False positive smear microscopy, culture and Xpert tests for tuberculosis are unlikely as the specificity of these tests is high. Further, clinician's underlying concern in high TB-HIV prevalent areas is to avert poor outcomes associated with undiagnosed untreated TB particularly following missed diagnoses from poorly sensitive tests.

This contextual framework supports my objectives by highlighting specific areas along the patient pathways to good outcomes for closer examination, namely: 1) the extent of adherence to diagnostic algorithms in XTEND (Objective 1) corresponds with the pathway from landmarks iv-vii in Box C, Figure 3. 4; 2) empiric tuberculosis treatment rates and pathways to empiric tuberculosis treatment initiation (Objective 2) corresponds with landmark vii, Figure 3. 4; and a qualitative evaluation of how the algorithm for tuberculosis diagnosis using Xpert was implemented during the XTEND trial (Objective 3) corresponds with box C, Figure 3. 4.

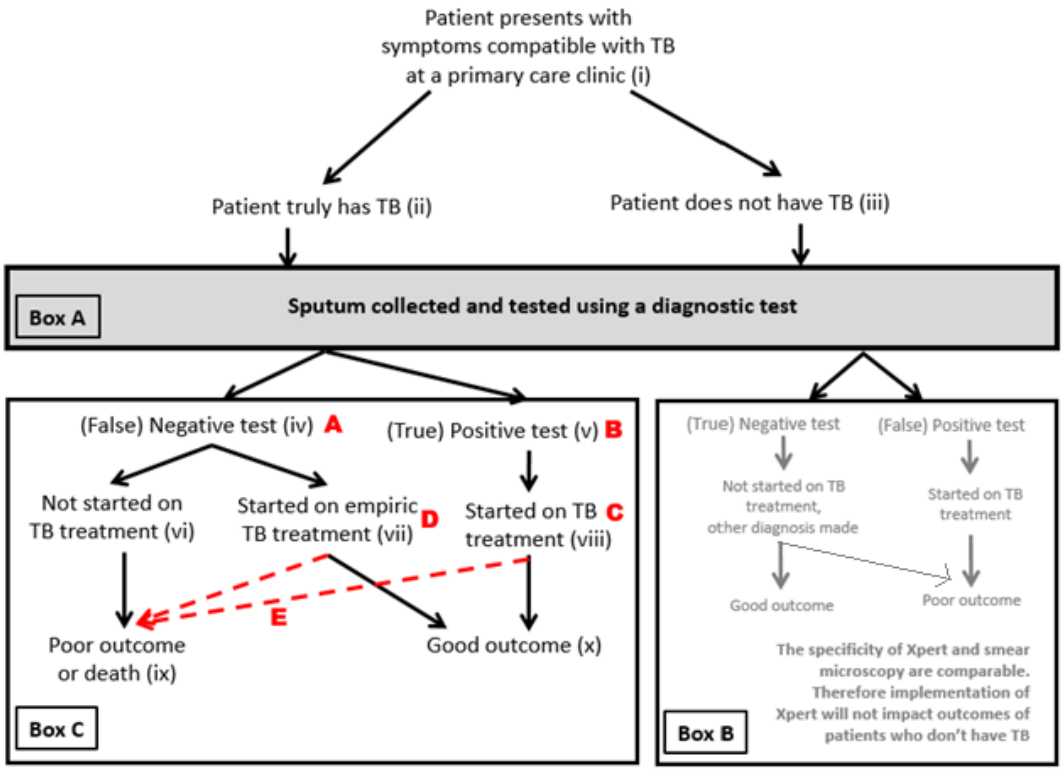


Figure 3. 4 Conceptual framework for evaluation of contextual factors influencing the ability of Xpert to impact positively on patient outcomes.
The figure represents landmarks (indicated by Roman numerals) on patient pathways to good outcomes (appropriate initiation of tuberculosis treatment AND treatment completion) or poor outcomes (failure to initiate tuberculosis treatment when it is required and/or death and/or loss to follow up).

3. 6 Conclusion

In this chapter, I developed a conceptual framework by drawing on theoretical approaches to the evaluation of diagnostic tests. I identified that the evaluation of diagnostic tests is aided by understanding the place of the diagnostic test in the test-treatment continuum, by mapping patient pathways, by utilising standardised checklists (QUADAS or STARD) and by including qualitative evaluations. I also drew on a conceptual framework of outcome measures that outlined pathways through which improved TB diagnostics may lead to improved patient outcomes. I identified that specific components of the test-treatment pathways in the XTEND trial were not measured in XTEND, and I showed how my objectives will describe and evaluate these components. Finally, I showed how findings from these objectives will support the aim of this submission - to better understand contextual factors that may have impacted on Xpert performance during the XTEND trial with regard to the management and investigation of persons with presumptive tuberculosis. Having satisfied this aim, I shall be able to answer the research question by proposing reasons why Xpert failed to reduce mortality amongst persons with presumptive tuberculosis.

Chapter 4: Literature review

4. 1 Introduction and rationale for literature review

In this chapter, I present a narrative review of impact evaluations of Xpert *vs* smear microscopy conducted from 2011-2019 for the diagnosis of tuberculosis. I use a recognised approach to conducting a narrative review [40], namely I objectively set out criteria for inclusion in the review and systematically identify appropriate studies. I summarise each study in narrative and tabular form, and focus specifically on factors relevant to this PhD, namely contextual factors (as defined above in Section 3. 5) that may impact on the outcomes of participants in Xpert and smear study arms. In the discussion, I draw these observations together to identify common themes across impact evaluations, and comment on how these align with the aim and objectives of this PhD.

The purpose of this narrative literature review is to systematically identify how contextual factors may have moderated the impact of Xpert on patient outcomes in other Xpert impact evaluations. We conducted the literature review with a specific focus on areas identified in our conceptual framework, namely the application of tuberculosis diagnostic processes after collection of sputum for Xpert or smear microscopy.

4. 2 Literature review methods

To identify impact evaluations of Xpert *vs* smear microscopy conducted from 01 January 2016 to 31 July 2019, I used search criteria reported in Auld *et al*[28]. I used the terms 'Xpert MTB/RIF', 'Xpert' or 'GeneXpert' and 'impact' or 'trial' or 'clinical trial'. I used Google Scholar and Pubmed search engines. I further refined the search using criteria reported by Auld *et al*[28], namely 1) that the evaluation was a controlled trial according to the International Committee of Medical Journals (ICMJ) with 'the prospective assignment of participants or groups to one or more health-related interventions to evaluate the effects of health outcomes'; 2) that the study utilised Xpert in an intervention arm; and 3) that a stated aim of the study was the assessment of how Xpert impacts on a patient-related outcome. In addition to the eight studies identified by Auld et al[11,22-25,41-43], we identified two further evaluations, namely the Chepetsa study[44], and a pragmatic evaluation of Xpert implementation in Botswana[45] which is available only as a conference abstract at the time of submission.

I extracted specific data on study methodology including the patient population, study design, randomisation level (cluster or individual), duration of participant follow-up, tuberculosis investigations at and after enrolment, the proportions of participants undergoing sputum culture and chest radiography, the group responsible for tuberculosis diagnosis (study or clinic staff), the follow-up methodology and a description of the study participants who had an opportunity to receive empiric tuberculosis treatment (Table 4. 1). I reported or calculated by study arm the yield of tuberculosis, the risk of tuberculosis over the duration of the study, the proportion receiving empiric tuberculosis treatment, and the outcome across intervention *vs* control arm (Table 4. 1). I classified these (excluding tuberculosis risk) arbitrarily as 'high', 'medium' or 'low' based on the distribution of values across the 10 studies in each category. I mapped each study's participant pathway through their investigative processes to their respective outcome classifications (high-medium-low) using a

different colour to represent each study (Figure 4. 1). I then drew inferences regarding the relationships between diagnostic processes and the impact of Xpert *vs* smear microscopy on patient outcomes reported by the study.

4. 3 Literature review results

4. 3. 1 Summary of methodologies used in studies assessing the impact of Xpert vs smear microscopy

Ten studies, including XTEND were identified meeting the inclusion criteria. Study methodologies (excluding XTEND) are summarized in Table 3 and below (excluding XTEND).

The TB-NEAT study[22] was a pragmatic, cluster randomised study conducted in five PHC facilities in southern or east Africa. Ambulatory presumptive tuberculosis patients were individually enrolled and randomised to receive point of care Xpert or smear microscopy. Two specimens were collected of which one was subject to Xpert or smear, and the other to culture. While awaiting results all participants underwent chest radiography. On the same day, participants were reviewed by a physician, and if smear or Xpert positive, they were commenced on tuberculosis treatment. If culture positive, tuberculosis treatment was commenced as soon as possible after the result was received. Outcome was assessed at day 56 following enrolment. Data were collected prospectively on an individual patient basis by study staff.

The Brazil study [23,41] was a stepped-wedge evaluation of the rollout of Xpert at laboratories servicing clinics in the cities of Rio de Janiero and Manaus, with a comparison of outcomes of persons in the cohorts before and after Xpert implementation. Data for the study was extracted from the laboratory information system and the tuberculosis control programme (TCP) electronic patient registers and linked using computer-generated algorithms. The Brazilian TCP recommends empirical treatment for tuberculosis if clinically indicated following negative sputum smear or microscopy and whilst awaiting culture results. The paper does not mention the role of chest radiography. These additional investigations for tuberculosis were conducted at the discretion of the attending clinicians, and details of these additional investigations were not collected nor reported on. Follow-up time was not specified.

The Zimbabwe RCT[25] was an individually randomised trial of Xpert *vs* smear microscopy conducted at a large ART referral centre in Harare. Symptomatic and asymptomatic persons with newly diagnosed HIV infection who were initiating ART were offered enrolment and followed up for three months. Following enrolment and submission of sputum, chest radiography was done at the discretion of the attending clinician. Participants had a standardized assessment conducted by a doctor at one and three months post enrolment, which included symptom screening, chest radiography and clinical examination. Sputum culture was not done. Outcomes were incident tuberculosis and mortality rates. Data were collected prospectively on an individual patient basis by study staff.

The RSA single clinic RCT[24] was a clustered cross-over design conducted in a single high TB/HIV burden area in Cape Town. Sputum smear microscopy or Xpert was offered on randomised weeks. Consecutive persons with presumptive tuberculosis were offered participation. Each gave two sputum specimens, the second of

which was submitted for culture if the participant was deemed at risk of DR-TB. Participants were followed up at 2-3 months to confirm tuberculosis treatment initiation (in those with positive Xpert, smear or culture results) and at 6 months. The role of chest radiography was not mentioned. Data was collected retrospectively through clinic records and the provincial tuberculosis database by study staff. Limited home follow-up was done by clinic staff for participants with missed appointments.

The Uganda pre-post trial[46] was conducted at a central hospital in Kampala comparing the outcomes of patients admitted with cough ≥ two weeks, but <six months duration before and after the implementation of Xpert. All persons were evaluated by physicians at enrolment with a history, physical examination and chest radiography. Two sputum specimens were collected for Xpert or smear microscopy and culture. Data were collected prospectively on an individual patient basis with telephonic follow up for 60 days post enrolment.

The RSA ICU[42] trial was a prospective, individually randomised comparison of Xpert *vs* smear microscopy conducted at three intensive care units across Cape Town, South Africa. Mechanically ventilated patients were eligible if they were HIV-positive, had pulmonary infiltrates on chest radiography or had constitutional symptoms compatible with tuberculosis prior to admission. Participants had tracheal aspirates submitted for Xpert or smear and culture. Patients were followed up until the end of hospitalisation or 90 days. Data collection methodology was not reported.

The Indonesia trial[43] compared treatment outcomes amongst cohorts of ambulant persons suspected of drug-resistant tuberculosis who presented to three primary health clinics in Java before and after the introduction of Xpert. All participants had sputum submitted for Xpert or smear microscopy and culture. Chest radiography was not mentioned. After three months, patients were followed up to record treatment start and culture results. Participant data was extracted from clinic and laboratory records and matched using algorithms. Mortality rates were not reported.

The Botswana trial[45] was a stepped wedge evaluation with three intervention arms – a retrospective cohort with data obtained from a period prior to interventions when smear microscopy was used, a prospective cohort with implementation of intensive case finding and retention in care interventions whilst using smear microscopy, and a second prospective cohort with Xpert replacing smear microscopy, whilst continuing with intensive case finding and retention in care interventions. The study was implemented at 22 ART clinics across the country where newly diagnosed HIV positive persons initiating ART were offered enrolment. Participants were followed up for one year. Study details are available only in oral abstract form, and details regarding additional tuberculosis diagnostic tests and data collection are not available at the time of submission. Outcomes in both prospective study arms (i. e. those with health system strengthening initiatives) were compared with the retrospective controls. Not reported in the table below is that mortality difference between the two prospective study arms was similar, whilst mortality difference between Xpert arm and the retrospective control in which smear microscopy was used was highly significant.

The Chepetsa study[44] was a cluster randomised trial of Xpert *vs* smear microscopy at 12 rural clinics in Malawi, prospectively conducted amongst newly diagnosed HIV positive persons. Following symptom assessment, persons were offered point of care Xpert or light-emitting-diode fluorescent smear microscopy (LED-FM), and staged for HIV infection. Persons with stage 3/4 HIV infection and positive sputa were initiated on ART and

TB treatment at prescribed time intervals. Persons with a negative symptom screen were initiated on isoniazid preventive therapy (IPT). Neither culture nor chest radiography were used and access to additional investigations was limited by the rural location of facilities. Participants were followed up for 12 months. Data were collected prospectively on an individual patient basis by study staff.

Additional tuberculosis diagnostic methodologies at and after enrolment

Four[22,42,43,46] of nine published studies conducted sputum culture on all participants as part of study procedures. Sputum culture was left to the discretion of the attending health care worker in the XTEND[11] trial (amongst participants who were HIV positive or had unknown HIV status, 18% had sputum culture[12]) and the Brazilian stepped wedge study[41] (rate not reported). Sputum culture was restricted to HIV-positive persons and persons suspected of DR-TB in the RSA ICU study[42] (47% of participants). Sputum culture played no part in tuberculosis diagnosis in the Chepetsa study[44] where it was restricted to persons with positive Xpert or smear results, and the Zimbabwe RCT[25] where it was not used at all.

Chest radiography was conducted on all participants in three studies[22,25,42], while in the Ugandan study[46], 81% of participants had chest radiography. In the XTEND study, the decision to conduct chest radiography was left to the attending nurse clinician, and by the study end, 3. 6% of participants living with HIV or HIV status unknown had undergone chest radiography. The role of chest radiography was not reported in three of nine studies[23,24,43].

The impact of Xpert on the tuberculosis diagnostic cascade and outcomes

All nine published studies reported an increased diagnostic yield of tuberculosis in the Xpert arm, and in only two studies[25,46] was the increase not statistically significant. The reported tuberculosis treatment initiation risk during follow-up was higher in the Xpert arm in seven of nine published studies, but only significantly greater in two studies[24,43]. In two studies[11,25] the proportion of participants starting tuberculosis treatment during follow up was higher in the microscopy arm. Empiric tuberculosis treatment was initiated more frequently in the microscopy arm in all eight published studies that reported on this outcome. This was statistically significant in the three studies[22,24,46] that provided p-values. The proportion deceased at the end of study follow-up was greater in the microscopy arm in six of eight studies that reported this outcome, but not statistically significant in any. However the Chepetsa study[44] reported a statistically significant and higher mortality difference amongst pre-defined sub-groups – namely males, persons >35 years of age and persons with WHO stage 3 or 4 HIV disease. In two studies mortality rates were identical across study arms[22,46].

4. 3. 2 The impact of the tuberculosis diagnostic cascade on yield, empiric treatment rates and patient outcomes

A visual analysis of the relationship between the extent of tuberculosis investigations and patient outcomes reveals a distinct trend (Figure 4. 1).

- Studies with more comprehensive investigations for tuberculosis at enrolment and follow up (TB-NEAT[22], the Uganda trial[46], the RSA-ICU trial[42]) tended to report higher yields of tuberculosis, high

empiric tuberculosis treatment rates, and no difference (or a lower trend to fewer deaths) in mortality rates between intervention and control arm.

- Conversely, studies with no or minimal additional investigations at or after enrolment (the Chepetsa study[44], and XTEND[11]) tended to report lower tuberculosis yield, lower rates of empiric tuberculosis treatment and significantly fewer deaths in sub-groups, or a trend to fewer deaths in the Xpert arm.

The Brazilian study[23] (which did not report on mortality) appears to be an exception, as TB diagnostic tests were left to the discretion of attending clinicians, but empiric TB treatment rates are high. However, data collection methodology in this study was through review of the electronic tuberculosis registers and the laboratory information system. Patient record reviews were not done. Therefore, it was not possible to ascertain the extent of additional investigations. Further, the authors acknowledge that persons with so-called 'empiric' tuberculosis treatment may equally have been patients for whom linking of sputum and culture results across the laboratory and TB databases failed.

Both studies of hospitalised patients[42,46] tended to have high tuberculosis diagnostic yields and higher rates of empiric tuberculosis treatment with non-significant differences in the Xpert *vs* smear microscopy arms. In both these studies, tuberculosis investigations were more rigorous compared with those in ambulant patients. Further, patient reviews were conducted by physicians in admitted patients, whereas in ambulant patients, reviews were most likely conducted by lower cadres of staff.

Amongst ambulant patients, the TB-NEAT[22] study had the most rigorous tuberculosis diagnostic investigations at enrolment, effectively ensuring that all cases of tuberculosis in both study arms were identified and treated early. By contrast, the Chepetsa trial[44] was conducted in a rural area of Malawi where access to tuberculosis investigations other than Xpert or smear microscopy was not possible. Empiric tuberculosis treatment was initiated in a minority of participants and only those who had sought care outside study clinics during the follow-up period. Drawing inferences from this, it appears that when Xpert or smear microscopy are not supported with additional tuberculosis investigations nor with empiric tuberculosis treatment, the use of Xpert as a first-line diagnostic test may improve patient outcomes amongst persons at higher risk of death.

The Chepetsa study[44] was conducted in an extremely poorly resourced environment, where both LED-FM and Xpert were newly implemented point-of-care assays, chest radiography was not available and culture was not used for diagnostic testing. The only additional intervention in this context was ART initiation amongst eligible participants. This was the only study where a significant impact on mortality was detected, and then only amongst males, persons over 35 years of age and persons with stage 3/4 HIV infection.

The Botswana[45] study reported of no difference in mortality between patients enrolled in the cohort that used smear microscopy alongside health system strengthening interventions *vs* patients enrolled in the cohort with the Xpert and the same health system strengthening interventions. The health system strengthening interventions included intensified case finding, and retention in care. This finding suggests that the addition of health system strengthening is sufficient to translate the benefits of TB testing by any methodology into improved outcomes amongst persons at high risk of tuberculosis-associated mortality.

4. 4 Literature review discussion

In this literature review of selected impact evaluations of Xpert from 2011-2019, I compare study methodologies, particularly contextual factors that may moderate the impact of Xpert on patient outcomes. I identify that health system support including additional diagnostic testing in patients at high risk of disease but with negative initial tests, patient review by physicians, patient follow-up, retention in care and initiation of empiric tuberculosis treatment, is essential to translate the benefits of improved diagnostics into improved patient outcomes[29].

Auld *et al*[28] identified factors related to trial design, trial conduct and health system weaknesses in the reviewed trials (except for Chepetsa[44] and the Botswana trial[45]) that may explain why these trials did not demonstrate morbidity and mortality reductions. With regard to trial design, higher empiric treatment rates in the microscopy arms of Xpert impact trials were likely a consequence of the lack of blinding and clinician awareness of lower negative predictive value of smear microscopy[26,28,47]. The selection of patient population in which to conduct the trial may also bias study outcomes: outpatients have lower mortality rates and hypothetically are less likely to benefit from improved tuberculosis testing, whilst inpatients are more likely to be treated empirically by attending clinicians[28]. This former factor may also apply to the Chepetsa study[44] which was conducted amongst outpatients. Further, an overestimate of mortality impact may lead to inadequate sample size and consequential under-powering[28]. A trial design limitation not mentioned by Auld *et al*[28] is the use of retrospective controls as in the Botswana trial[45], with poor data recording and reporting leading to inadequate linkage of cases across laboratory and treatment databases. With regard to trial conduct, inadequate ascertainment of the study endpoint of mortality is affected when loss to follow-up (LTFU) occurs[28]. Lastly, health system weaknesses may blunt the ability of impact studies to detect mortality differences, specifically sub-optimal HIV management and high LTFU before tuberculosis treatment initiation[28].

My observations go beyond Auld *et al*[28] to propose that the more systematically and rigorously tuberculosis diagnostic tests are conducted at patient enrolment and follow-up in each study arm, the less likely the study will find a mortality difference between intervention (Xpert) and control (smear microscopy) arms. Conversely, without additional diagnostic tests, more cases of tuberculosis may be missed in the smear microscopy arm relative to the Xpert arm, and this may lead to poorer outcomes in the control arm and a greater mortality difference. The findings from the Chepetsa[44] and Botswana[45] trials which were not available at the time of Auld's review, contribute meaningful observations towards this hypothesis. The Chepetsa study[44] was an impact assessment of Xpert where no additional diagnostic tests were conducted at enrolment nor at follow-up, nor was empiric treatment prescribed. This study design allowed an assessment of outcomes without the additional diagnostic or health system support in either arm. The Botswana study[45] showed a mortality difference between Xpert and the retrospective smear microscopy arm, but equivalent mortality rates between the arms with and Xpert or health system strengthening interventions and smear microscopy. This study design illustrates the importance of health system strengthening to translate diagnostic testing of any kind into measurable improvements in patient outcomes.

My findings identify a potential pitfall in study design of impact evaluations of Xpert *vs* smear microscopy, namely that in well supported environments with strong health systems, outcomes of persons tested with Xpert

vs smear microscopy may not be vastly different. The study design of number of Xpert impact evaluations (TB-NEAT[22], Uganda trial[46], the Botswana[45] trial) created a 'well supported environment', albeit that they were conducted under 'field' conditions. Conversely, in poorly supported environments, such as those where additional tests for tuberculosis are not available (such as the Chepetsa[44] study), or tuberculosis diagnostic algorithms are not widely adhered to (XTEND[11]), significantly improved diagnostic tests may offer health systems a way to improve tuberculosis outcomes.

4. 5 Conclusion to literature review

In this narrative review of selected impact evaluations of Xpert, findings suggest that the extent to which tuberculosis diagnostic testing is supported by contextual factors (additional investigations, health system strengthening interventions) will influence the magnitude of the difference in observed mortality between intervention and control arms. This suggests that to explain the absence of a significant reduction in mortality amongst XTEND participants across study arms, a careful evaluation of the extent to which tuberculosis diagnostic algorithms were adhered to, and the degree to which empiric tuberculosis treatment was prescribed in the XTEND trial should be conducted.

Findings from this literature review overlap with the areas for close evaluation that I highlighted in the conceptual framework, namely the provision of additional diagnostic testing in persons with false negative tests, provision of empiric tuberculosis treatment, the degree to which additional diagnostic testing is conducted and empiric tuberculosis treatment is prescribed in the intervention *vs* control arm, and a qualitative evaluation of the application of the tuberculosis diagnostic algorithm in each study arm.

I will now focus on each objective of the PhD in turn.

Table 4. 1. A comparison of methodology and outcome measures of clinical trials comparing Xpert with smear microscopy

Study name	Setting	Design	Sample size, inclusion criteria and follow-up time (days)	Standard of care	Xpert intervention	Procedure for TB diagnosis at enrolment	Procedure for TB diagnosis during follow-up	% of participants with sputum culture	% of participants with chest radiography	Group responsible for TB diagnosis	Participant follow up and data extraction	Who had opportunity for empiric TB treatment?	Comparisons			
													Measure	Micro %	Xpert %	p
TB-NEAT[22]	5 primary care clinics in RSA, Zimbabwe, Zambia and Tanzania	Pragmatic randomised, two-arm parallel group multicenter trial with individual randomisation	1502 ambulant adult patients with presumptive TB (56)	Same-day, onsite sputum smear microscopy by lab technician; one spot sputum/patient.	Nurse-performed POC Xpert at clinic on one spot sputum per patient	All participants had CXR and sputum. Sputum negative patients were reviewed by clinic staff	No study-related follow-up, but participants may have sought care on their own initiative	100% in each arm at enrolment	100% in each arm at enrolment	Non-study nurses and doctors; not blinded	Not reported	All participants with negative Xpert/ smear	Diagnostic yield	15	24	<0. 0001
													TB risk	42	43	NS
													Empiric Rx	26	17	0. 0001
													3m mortality	8	8	NS
XTEND[11]	20 labs and 40 primary care clinics in RSA	Pragmatic two-arm, parallel CRT, a cluster=a TB lab with 2 associated clinics	20 labs, 2 clinics per lab and 4656 patients with presumptive TB (182)	Sputum smear microscopy by lab technician; two sputa /patient	Technician-performed Xpert at laboratory; one spot sputum /patient	National TB diagnosis guidelines - participants with negative smear/Xpert had HIV testing, sputum culture CXR, +/- empiric TB treatment where appropriate.	No study-related follow-up, but participants may have sought care on their own initiative	Amongst HIV positive participants or HIV unknown status (n=2695), 18% had sputum culture performed (358, 13% in Xpert arm) by end of follow-up	Amongst HIV positive participants or HIV unknown status (n=2695), 3. 6% had CXR performed (62, 2. 3% in Xpert arm) by end of follow-up	Non-study clinic staff, not blinded	Study staff extracted data from patient files and directly contacted patients. Patients were traced to home	Participants at clinics where CXR was done and doctor available for patient review; else-during follow up, if participant presented at any district hospital	Diagnostic yield	7. 8	9. 2	0. 05
													TB risk	12. 5	10. 8	NS
													Empiric TB Rx	2. 6	1. 8	NA
													6m mortality	5. 0	3. 9	NS
Brazil stepped wedge[41]	14 primary care laboratories in Rio and Manaus, Brazil	Stepped wedge CRT, a cluster=a primary care lab	14 labs, all 24 227 presumptive TB patients who provided sputum during the study period (NA)	Sputum smear microscopy by lab technician; one or two sputa/patient	Technician-performed Xpert at lab; one spot sputum /patient	Clinic staff implemented national TB guidelines for TB diagnosis - patients with negative smear/Xpert had culture +/- initiation of empiric TB treatment	No study-related follow-up, but participants may have sought care on their own initiative	Not reported	Not reported	Non-study clinic staff, not blinded	No individual patient follow-up. Data was extracted from routine laboratory and TB notification databases and linked	Not reported; likely if participant presented at any district hospital for clinician review	Diagnostic yield	9. 7	14. 2	<0. 001
													TB risk	17. 5	20. 8	NA
													Empiric Rx	10. 4	9. 8	NA
													3m mortality	NA	NA	NA

*TB=tuberculosis; lab=laboratory; POC=point of care; Rx=treatment; m=month; Xpert=Xpert MTB/RIF (Cepheid, Sunnyvale, USA); CXR=chest X-ray; RSA=Republic of South Africa; CRT=cluster randomised trial; HIV=human immunodeficiency virus; ART=antiretroviral therapy; PHC=primary health care; DR-TB=drug-resistant tuberculosis; ICUs=Intensive care units; DST=drug susceptibility testing; ICF=intensive case finding; IPT=isoniazid preventive therapy; LED-FM=light-emitting diode fluorescent microscopy;

Table 4. 1 (cont). A comparison of methodology and outcome measures of clinical trials comparing Xpert with smear microscopy

Study name	Setting	Design	Sample size, inclusion criteria and follow-up time (days)	Standard of care	Xpert intervention	Procedure for TB diagnosis at enrolment	Procedure for TB diagnosis during follow-up	% of participants with sputum culture	% of participants with chest radiography	Group responsible for TB diagnosis	Participant follow up and data extraction	Who had opportunity for empiric TB treatment?	Measure	Comparisons		
														Micro %	Xpert %	p
Zimbabwe RCT[25]	A single ART referral clinic in Harare, Zimbabwe	Pragmatic randomised, two-arm parallel-group trial with individual randomisation	424 HIV-positive, ART-naïve symptomatic and asymptomatic patients presenting for ART initiation (182)	Sputum smear microscopy by lab technician; two spot sputa/patient	Technician-performed Xpert at lab; two spot sputa /patient	Following sputum collection, CXR was performed at discretion of attending doctor. Sputum results were only available 24 hours later.	Standardised assessment by study doctor at month 1 and 3 included CXR, clinical assessment and symptom evaluation	Sputum culture not used	All participants at month 1 and 3	Non-study clinic staff at enrolment who were not blinded. Study staff at follow up who were not blinded	Participants returned to clinic for study visits at 1 and 3 months; home tracing by study team. Study staff extracted data from clinic records reviewed	All participants at enrolment and at review at 1 and 3 months	Diagnostic yield	7	9	NS
													TB risk	21	20	NS
													Empiric TB Rx	15	11	NA
													6m mortality	10	6	NS
RSA single clinic CRT[48]	A single PHC in Khayelitsha, RSA	Single clinic, pragmatic, two-phase, crossover, CRT, a cluster=one PHC clinic with alternate randomisation to each arm	1985 adults presenting for TB investigations who were not receiving TB treatment (182)	Same-day, onsite sputum smear microscopy by lab technician; two spot sputa/patient.	Study staff-performed POC Xpert at clinic on one spot sputum per patient	HIV positive patients or patients with previous TB or suspected DR-TB or patients with negative smear/Xpert had culture and review by clinician	No study-related follow-up, but participants may have sought care on their own initiative	933/1985 (47%) had culture done at enrolment; 474 (24%) in Xpert arm	Role of CXR not mentioned	Clinic staff; not blinded	Study staff extracted data from clinic records, provincial databases with home visits for participants not initiated on TB treatment. RSA death registries were consulted.	Participants with negative sputum; Participants who self presented to clinic or hospital during follow-up.	Diagnostic yield	17	26	<0. 001
													TB risk	23	28	0. 013
													Empiric TB rx	9. 8	5. 2	0. 0025
													6m mortality	3. 8	3. 4	NS
Uganda pre-post trial[46]	A single referral hospital in Kampala Uganda	Single clinic, prospective pre-post study, non-randomised	477 consecutive hospitalized presumptive TB patients; (Excluded: death within 3 days, unable to produce sputum) (60)	On-site lab fluorescent smear microscopy; two spot and one morning sputa/patient, with remainder of sputum sent for culture.	Study-staff-performed POC Xpert at hospital; one spot sputum/patient.	Physician-led investigation of all admitted patients. All had sputum culture at enrolment	No study-related follow-up, but participants may have sought care on their own initiative	100% at enrolment	388/477 (81%) had CXR at enrolment.	Hospital staff; not blinded	Study staff contacted participants by telephone or did an in-person visit at 60 days after enrolment	All participants following physician review	Diagnostic yield	37	43	NS
													TB risk	81	85	NS
													Empiric TB Rx	15	7	0. 047
													2m mortality	17	17	NS
South Africa ICU RCT[42]	4 intensive care units in Cape Town hospitals, RSA	Prospective cohort at 4 ICUs with nested individual-level RCT substudy	341 patients admitted and mechanically ventilated patients with presumptive TB and from whom tracheal aspirate was obtained (90)	Same-day fluorescent smear microscopy of 1. 5-7. 5ml tracheal secretions performed at participating laboratory	Technician-performed Xpert of 1. 5-7. 5ml tracheal secretions at laboratory	Physician-led investigation of all admitted patients. All had tracheal aspirate culture at enrolment	Not stated	100% at enrolment	Role of CXR not mentioned	Hospital staff; not blinded	Patients were followed up to end of hospitalisation or death by study staff	All participants following physician review	Diagnostic yield	6	18	0. 012
													TB risk	14	22	NS
													Empiric TB treatment	7. 8	3. 6	NA
													3m mortality	42	32	NS

*TB=tuberculosis; lab=laboratory; POC=point of care; Rx=treatment; m=month; Xpert=Xpert MTB/RIF (Cepheid, Sunnyvale, USA); CXR=chest X-ray; RSA=Republic of South Africa; CRT=cluster randomised trial; HIV=human immunodeficiency virus; ART=antiretroviral therapy; PHC=primary health care; DR-TB=drug-resistant tuberculosis; ICUs=Intensive care units; DST=drug susceptibility testing; ICF=intensive case finding; IPT=isoniazid preventive therapy; LED-FM=light-emitting diode fluorescent microscopy;

Table 4. 1 (cont). A comparison of methodology and outcome measures of clinical trials comparing Xpert with smear microscopy

Studname	Setting	Design	Sample size, inclusion criteria and follow-up time (days)	Standard of care	Xpert intervention	Procedure for TB diagnosis at enrolment	Procedure for TB diagnosis during follow-up	% of participants with sputum culture	% of participants with chest radiography	Group responsible for TB diagnosis	Participant follow up and data extraction	Who had opportunity for empiric TB treatment?	Comparisons			
													Measure	Micro %	Xpert %	p
Indonesia pre-post trial[43]	3 PHC clinics in Java Indonesia	Pre-post trial at 3 provincial hospitals without randomisation	2417 persons at risk of drug-resistant TB according to Indonesian guidelines (NA)	Sputum smear microscopy by lab technician; two spot sputa/patient - second specimen for culture	Technician-performed Xpert at laboratory; one spot sputum /patient	Not reported. All patients had sputum culture and DST conducted	No study-related follow-up, but participants may have sought care on their own initiative	100% at enrolment	Role of CXR not mentioned	Clinic staff, not blinded.	Study staff extracted data from clinical, lab and TB treatment registers with linking of entries	Not stated	Diagnostic yield	65. 2	80. 2	<0. 001
													TB risk	39. 3	58. 5	<0. 001
													Empiric TB Rx	NA	NA	NA
													3m mortality	NA	NA	NA
Auld Botswana[45]	22 ART primary care clinics in Botswana	Pragmatic stepped wedge CRT with retrospective control component, cluster =an ART clinic	14,963 consecutive persons >=12 years initiating ART (365)	Sputum smear microscopy by lab technician; retrospectively extracted from registers	Additional staff to support ICF, tracing of missing appointments and Xpert replacing smear	Not reported in oral abstract	Not reported in oral abstract	Not reported in oral abstract	Not reported in oral abstract	Clinic staff, not blinded.	Not reported in oral abstract	Not reported in oral abstract	6m mortality (compared with retrospective cohort)	5. 1%	2. 8%	<0. 0001
Chepetsa Malawi[44]	12 rural primary care clinics in Malawi	Pragmatic CRT, a cluster=a PHC facility	1842 adults newly diagnosed HIV positive patients not on TB treatment, IPT or ART (350)	Same-day, onsite sputum fluorescent smear microscopy by lab technician; two spot sputa /patient.	Technician-performed Xpert on site; one spot sputum /patient	All participants underwent TB symptom screening and sputum for Xpert/ smear, and HIV staging and ART initiation when appropriate	Assessment by study staff every 3 months for TB symptoms, Xpert or LED-FM testing of sputum.	Only patients with positive smear or Xpert had culture	CXR not included in TB diagnostic algorithm; access to CXR was limited in rural clinics	Study staff; not blinded	Participants were asked to return every 3 months to study clinics for clinical evaluation by study staff; persons lost to follow-up were traced to home, and through clinic records	Participants who were started on TB treatment outside of study clinics	Diagnostic yield	1. 2	2. 4	0. 06
													TB risk	2. 4	3. 3	NS
													Empiric TB Rx	0. 8	0. 1	0. 04
													12m mortality	8. 6	6. 7	NS

*TB=tuberculosis; lab=laboratory; POC=point of care; Rx=treatment; m=month; Xpert=Xpert MTB/RIF (Cepheid, Sunnyvale, USA); CXR=chest X-ray; RSA=Republic of South Africa; CRT=cluster randomised trial; HIV=human immunodeficiency virus; ART=antiretroviral therapy; PHC=primary health care; DR-TB=drug-resistant tuberculosis; ICUs=Intensive care units; DST=drug susceptibility testing; ICF=intensive case finding; IPT=isoniazid preventive therapy; LED-FM=light-emitting diode fluorescent microscopy;

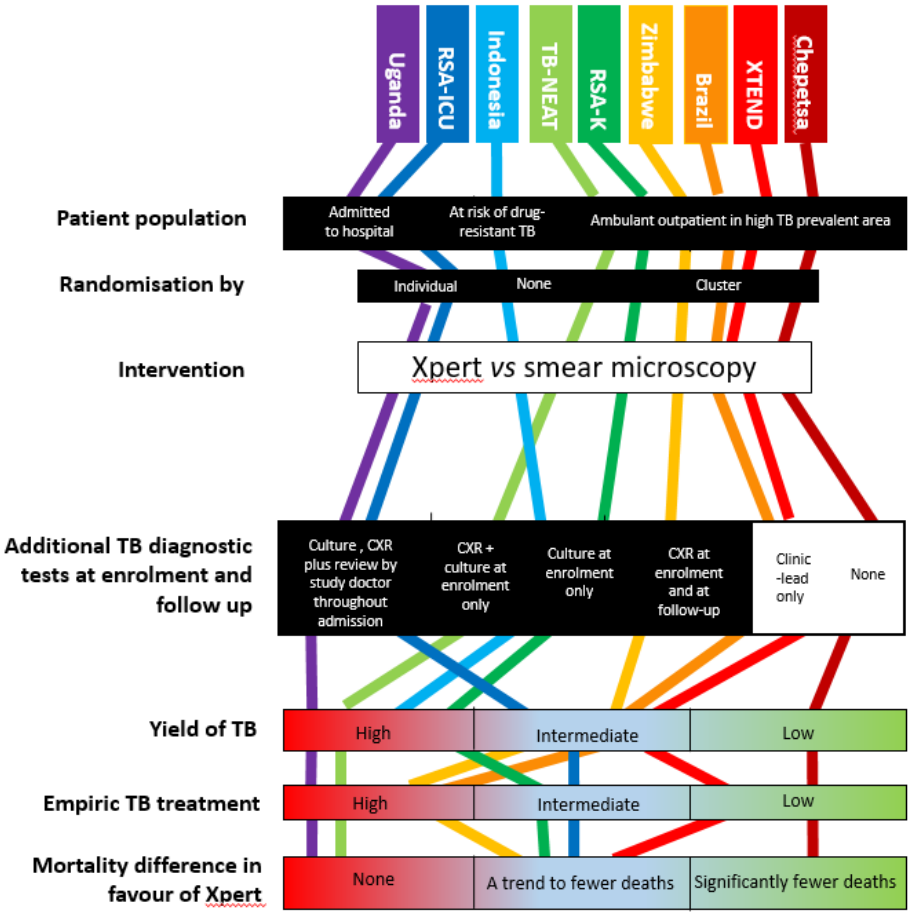


Figure 4. 1 A comparative flow chart illustrating the relationship between interventions, tuberculosis diagnostic methodology and outcomes (proportion tuberculosis yield, proportions on empiric tuberculosis treatment and mortality difference in favour of Xpert) in selected impact evaluations of Xpert.
With reference to 'Yield of TB', 'Empiric TB treatment' and 'Mortality diference in favour of Xpert', classifications as 'high-medium-low' or 'none, a trend to fewer deaths, significantly fewer deaths' were categorised heuristically after listing results from Table 4. 1 in sequence and arbitrarily designating categories based on the distribution of results. Colored lines are placed on categories in descending sequence from left to right. The Botswana trial[45] is not included as no detail pertaining to additional TB diagnostic investigations were available at the time of submission. (References as follows: Uganda[46], RSA-ICU[42], Indonesia[43], TB-NEAT[22], RSA-K[24], Zimbabwe[25], Brazil[23], XTEND[11], Chepetsa[44])

Chapter 5: Were diagnostic algorithms adhered to in the XTEND trial?

5. 1 Introduction

In this chapter, I address the first objective of this PhD, which is to establish the extent to which diagnostic algorithms for tuberculosis were adhered to in the XTEND trial. This objective is supported by the contextual framework and literature review, where I have shown how an investigation of the way tuberculosis diagnostic processes were applied will provide evidence to understand the reasons for the absence of an impact on mortality in the XTEND trial. I summarise findings of the paper 'Implementation and Operational Research: What Happens after a Negative Test for Tuberculosis? Evaluating Adherence to tuberculosis Diagnostic Algorithms in South African Primary Health Clinics' which was published in the Journal of the Acquired Immune Deficiency Syndrome 2016;71(5):e119-126[12] (see Part II of this PhD for the complete paper). I then explain how the findings of this paper support the aim of the PhD.

5. 2 An overview of Paper 1 'Evaluating adherence to tuberculosis diagnostic algorithms'

Paper 1 (Evaluating adherence to tuberculosis diagnostic algorithms) was a secondary data analysis of XTEND data[12]. Briefly, XTEND study documentation including case investigation forms, six month record reviews and laboratory records were assessed to ascertain levels of adherence to the tuberculosis diagnostic algorithm in the intervention (Xpert) and control (smear microscopy) arms. We defined adherence using key features of the RSA tuberculosis control programme diagnostic algorithms[14,15]. We recorded the presence of: i) collection of sputum for culture or ii) chest radiography or iii) referral to a higher level of care in any of the above data sources. We confined the assessment of adherence to participants with evidence of HIV infection (reported or documented) or evidence that the participant's HIV status was unknown or undocumented, because the diagnostic algorithms advise additional investigations in these patient populations. We compared adherence rates using methods appropriate for cluster randomised controlled trials with few clusters[49].

In this paper, we showed that 507/2155 (24%) HIV positive XTEND participants had sputum sent for culture, or chest radiography or referral to a higher level of care, indicating some level of adherence to the recommended tuberculosis guidelines. Participants in the Xpert arm were one third times less likely to have had additional tuberculosis investigations or referral (risk ratio 0. 32 (95% CI 0. 17-0. 62). Adherence rates between clinics (as indicated by the proportion of participants with any of chest radiography, tuberculosis culture, or referral) varied extensively from 0% to 86%. Whilst all clinics were able to request sputum culture from NHLS laboratories, seven clinics performed sputum culture in over 50% of persons with negative Xpert or smear microscopy results, and in contrast, 21 clinics requested

sputum culture on less than 5% of the same category of participants. Fourteen XTEND clinics had no record of chest radiography in patient files of HIV-positive XTEND participants with negative index sputum tests enrolled at their clinic.

We further identified that male patients, patients who were ill and patients who visited repeatedly were more likely to have additional investigations done. A higher odds of adherence to the algorithm was observed amongst persons with known HIV status, male gender, two or more previous clinic visits at the time of enrolment *vs* none, and more severe illness at enrolment indicated by more tuberculosis symptoms, a lower body mass index, lower Karnofsky score and CD4 count below 350 cells/mm^3. In a multivariable model, having three or more tuberculosis symptoms (adjusted odds ratio [aOR] 1. 59, 95% CI 1. 14-2. 23) and two or more prior visits to clinic (aOR 1. 42, 95% CI 0. 97-2. 10) remained associated with increased odds of having had further investigations. Being female (aOR 0. 66, 95% CI 0. 51-0. 88) was associated with decreased odds of having had further investigations. Participants with unknown HIV status were less likely to be investigated further ((59/540 (11%) *vs* 507/2155 (24%)) compared with participants with known HIV positive status.

5. 3 Implications of poor adherence to tuberculosis diagnostic algorithms in the assessment of mortality difference between smear microscopy and Xpert arms in the XTEND trial

In my conceptual framework, I identified that additional investigations following negative tests for tuberculosis are an essential part of the diagnostic-treatment pathway leading to good outcomes (Figure 3. 4). I also recognised that differential adherence to algorithms for tuberculosis diagnosis across study arms had potential to alter or bias the proportions of participants with good outcomes across study arms.

In my literature review, I demonstrated the relationship between contextual factors supporting tuberculosis diagnostic tests and good outcomes. I showed how health system support (and in the tuberculosis programme – additional diagnostic tests or retention in care) translates diagnostic tests into good outcomes in the test-treatment continuum. I hypothesised that in well supported environments, with adherence to algorithms and good follow-on diagnostic tests, mortality difference between Xpert and control arms may be minimal. I suggested that a careful evaluation of the extent to which tuberculosis diagnostic algorithms were adhered to in the XTEND trial should be done.

Paper 1 demonstrates that additional tuberculosis diagnostic tests were conducted in less than a quarter of XTEND participants. The XTEND study replicated field conditions exactly, as the implementation of tuberculosis diagnostic algorithms was left entirely up to routine PHC services. Further, this aspect of the XTEND trial is generalizable across sub-Saharan Africa where adherence to tuberculosis diagnostic algorithms is challenging. Amongst patients diagnosed in hospital, Loveday (South Africa)[50] and Tafuma (Botswana)[51] reported poor adherence to tuberculosis diagnostic algorithms, with a dependence on chest radiography and little use of sputum microscopy or culture. In an urban and rural setting in Uganda, Alamo et al[52] reported poor adherence under operational conditions to the WHO 2007 algorithm for smear-negative tuberculosis.

Paper 1 further demonstrated that additional diagnostic tests were differentially applied across study arms. Additional tuberculosis investigations including culture and chest radiography were just under three times more likely to have been conducted in the smear microscopy arm after adjustment for age, gender, number of tuberculosis symptoms and BMI[12]. The national tuberculosis control programme algorithm for use with smear microscopy, and used in the XTEND trial control arm, closely adheres to the WHO 2007 guidelines for smear negative tuberculosis amongst persons living with HIV[53]. A number of controlled trials and observational studies have demonstrated a significant improvement in survival amongst hospitalised and ambulant patients who undergo investigations according to these guidelines[52,54-57]. Therefore, the control arm (smear microscopy testing) was better supported than the intervention arm, with better potential for test results to be translated into good patient outcomes.

Paper 1 also showed that XTEND participants in both study arms who were more ill, and therefore more likely to have poorer outcomes, were more likely to undergo additional investigations. An individual-level analysis of XTEND participants identified that risk factors for death included male gender, age >/= 50 years, increasing number of tuberculosis symptoms, lower BMI, unknown HIV status, and HIV positive status but not on ART[11]. A number of these risk factors were identical to characteristics of persons more likely to have additional investigations, namely male gender, having more tuberculosis symptoms or a lower body mass index. Therefore, in both XTEND study arms, persons more likely to die were also better supported along the pathway to improved outcomes. This may have decreased the mortality across both study arms relative to other contexts where additional investigations are not done, or not available and may explain why mortality rates in XTEND were lower than anticipated[28].

5. 4 Conclusions: Contributory observations from Paper 1

A careful evaluation of the extent of adherence to additional investigations amongst XTEND participants showed poor levels of adherence to tuberculosis diagnostic algorithms, and identified greater levels of adherence to algorithms for tuberculosis diagnosis in the smear microscopy arm and amongst persons who were more ill, male or visited clinic repeatedly. In addition, we observed large between-clinic variation in adherence rates.

Additional tuberculosis diagnostic tests are key supportive steps on the diagnosis-treatment continuum. Therefore, XTEND participants who were more ill, male or visited clinic repeatedly were more likely to have additional investigations and potentially had better outcomes. The differential performance of additional tuberculosis diagnostic tests across study arms may have increased the potential for better outcomes amongst participants in the smear microscopy arm, and lead to smaller difference in mortality across XTEND study arms.

Chapter 6: How was empiric tuberculosis treatment prescribed in the XTEND trial?

6. 1 Introduction

In this chapter, I address the second objective of this PhD, which is to investigate how empiric tuberculosis treatment was prescribed in the XTEND trial. This objective is supported by the contextual framework and literature review where I have shown that the differential prescription of empiric tuberculosis treatment across XTEND study arms has potential to alter the mortality difference across study arms. In this chapter, I summarise the paper 'Empiric tuberculosis treatment in South African primary health care facilities - for whom, where, when and why: implications for the development of tuberculosis diagnostic tests' which was published in PLoS One 2018;13(1):e0191608[13] (see Part II of this PhD for the complete paper). I then explain how the findings of this paper support the aim of the PhD.

6. 2 An overview of Paper 2 'Empiric tuberculosis treatment in XTEND'

Paper 2 was a secondary data analysis of the XTEND cohort, focusing specifically on participants who initiated tuberculosis treatment during six months of follow up and who tested negative by Xpert or smear microscopy within four days of enrolment[13]. XTEND data sources, including case investigation forms, six month record reviews and laboratory records were used to reconstruct pathways to tuberculosis treatment initiation. Reasons for tuberculosis treatment start were inferred according to criteria reported in Table 1, McCarthy *et al*[13]. Briefly, reasons for tuberculosis treatment start were classified as 'positive laboratory test'; 'compatible radiological signs'; 'extra-pulmonary tuberculosis, 'referral to a higher level of care' and 'clinical assessment' (clinical signs and symptoms without radiological or microbiological evidence).

We determined which additional investigations had been conducted, the timing of tuberculosis treatment initiation relative to enrolment and sputum collection, the place and reason for tuberculosis treatment start, HIV and ART status and outcome at six months. We described patient demographic, clinical and treatment characteristics, tuberculosis investigations, health seeking behaviour, reason for tuberculosis treatment start and outcome amongst three overlapping cohorts, namely all XTEND participants with negative sputum tests at enrolment who were later started on tuberculosis treatment (group A), group A XTEND participants who had microbiological confirmation of tuberculosis at some point during follow up, and group A XTEND participants who met the definition of empiric tuberculosis treatment as defined above.

Group A comprised 167 XTEND participants who had negative sputum tests at enrolment were initiated on tuberculosis treatment during six months of follow up. This constituted 30% of all 541 XTEND participants who were commenced on tuberculosis treatment and 3. 5% of all 4665 XTEND participants. We assessed that 82/167 (49%) were initiated 'empirically' without microbiological confirmation at the

time of tuberculosis treatment initiation. This equated to 15% (82/541) of all persons initiated on tuberculosis treatment in the XTEND trial. The majority of participants (61/82, 74%) started on empiric tuberculosis treatment were in the smear microscopy arm. Relative to all tuberculosis treatment initiations, empiric treatment was started on 21/250 (8. 4%) and 61/291 (21%) participants in the Xpert *vs* smear microscopy arms respectively. Microbiological confirmation of tuberculosis after treatment initiation was obtained within six months in 60/167 (36%) participants. Findings compatible with tuberculosis on chest radiography was the most common reason for empiric tuberculosis treatment initiation with 63/82 (77%).

We described how the pathway to tuberculosis treatment initiation was prolonged and highly variable if initial sputum tests were negative. Each person initiated on tuberculosis treatment after the enrolment period had a median of 3 visits to their health care provider (Inter-quartile range (IQR) 2-4), and was initiated a median of 5. 7 weeks (IQR 2. 1-12. 1) after enrolment. Only 87/152 (52%) had chest radiography performed, and sputum was submitted for culture amongst 64/167 (38%). Tuberculosis treatment was initiated amongst 70/167 (42%) at hospital, of whom 38/167 (23%) were admitted. However, only 21/70 (30%) had a documented referral in their clinic file, indicating self-referral. We also observed that 12/167 (7%) persons died within 6 months of enrolment, of whom 11 were in the smear microscopy arm.

6. 3 Implications of infrequent empiric tuberculosis treatment in the assessment of mortality difference between smear microscopy and Xpert arms in the XTEND trial

In my conceptual framework, I identified that the initiation of empiric tuberculosis treatment is essential part of the diagnostic-treatment pathway leading to good outcomes (Figure 3. 4). I also recognised that differential initiation of empiric tuberculosis treatment across study arms had potential to alter or bias the proportions of participants with good outcomes across study arms.

In my literature review, I demonstrated the relationship between contextual factors supporting tuberculosis diagnostic tests and good outcomes. I showed how health system support (including additional diagnostic tests and empiric tuberculosis treatment) translates diagnostic test results into good outcomes in the test-treatment continuum. I hypothesised that in well supported environments, with adherence to algorithms and good follow-on diagnostic tests, mortality difference between Xpert and control arms may be minimal. I suggested that a careful evaluation of how empiric tuberculosis treatment was prescribed in the XTEND trial will allow conclusions to be drawn regarding the failure of Xpert to reduce mortality amongst persons with presumptive tuberculosis.

In paper 2, we described how empiric TB treatment accounted for a small proportion of TB treatment initiations in the XTEND trial and that empiric TB treatment was more frequently prescribed in the smear microscopy arm.

Empiric tuberculosis treatment is advised by WHO guidelines in resource-limited settings when ambulant HIV-positive persons with two negative sputum smear microscopy tests and chest radiography findings compatible with tuberculosis do not respond to broad-spectrum antimicrobial

therapy[53]. Adherence to this algorithm in routine practice leads to frequent use of empiric tuberculosis treatment, with up to two thirds of tuberculosis cases initiated empirically[55]. Empiric tuberculosis treatment is frequently prescribed in hospital and outpatient settings. Loveday *et al* reported that only 108/642(17%) hospitalised patients initiated on treatment for pulmonary tuberculosis in a tertiary care facility in KwaZulu-Natal Province, South Africa had sputum submitted for testing[50]. In a routine care PHC setting, amongst 1195 symptomatic outpatients with negative smear microscopy tests for tuberculosis, Dimario *et al* reported that 218 (18. 2%) were commenced on tuberculosis treatment of which 127 (59%) were smear-and culture-negative[57]. These data show that in the XTEND trial, even in the XTEND smear-microscopy arm, empiric tuberculosis treatment was infrequently prescribed.

Our review of Xpert impact evaluations confirmed that empiric tuberculosis treatment was initiated more frequently in the microscopy arm in all eight published studies which reported on this outcome. In part, this may be due to adherence to tuberculosis diagnostic algorithms. However, in all of the reviewed studies, clinicians were not blinded and may have been more likely to prescribe empiric tuberculosis treatment in smear-negative *vs* Xpert negative presumptive tuberculosis patients[28,44].

Tuberculosis researchers have reflected that Xpert has failed to impact on programmatic and patient outcomes because symptomatic patients who test sputum smear-microscopy negative, are prescribed empiric tuberculosis treatment[26,28,58]. This hypothesis concurs with our conceptual framework, which suggests that the differential support for tuberculosis diagnostics across study arms, including additional tuberculosis investigations (as shown in Chapter 5) and empiric tuberculosis treatment (as described in this chapter), may bias outcomes.

An important conclusion of paper 2 was that in nurse-driven, PHC-based tuberculosis control programmes where empiric tuberculosis treatment is infrequently prescribed, highly sensitive tuberculosis diagnostic tests are an essential component of tuberculosis diagnosis and management. This concurs with the contextual framework and literature review findings, which suggest that highly sensitive tuberculosis diagnostic tests may improve clinical outcomes in unsupportive clinical environments – especially those with weak health systems, poor access to additional tuberculosis diagnostic tests and high loss-to-follow-up rates.

6. 4 Conclusions: Contributory observations from Paper 2

In paper 2, we showed that empiric tuberculosis treatment was infrequently prescribed in the real-world setting of XTEND, where participating clinics were left to implement tuberculosis diagnostic algorithms without study intervention. Therefore a key step in the diagnosis-treatment continuum was under-utilized, and diagnostic tests in both the intervention and control arms were relatively unsupported. Furthermore, a majority of empiric tuberculosis treatment initiations took place in the smear microscopy arm. This potentially biased endpoint (outcome) evaluation, as the provision of empiric tuberculosis treatment may have saved lives, and led to fewer deaths in the smear microscopy arm relative to the Xpert arm.

Chapter 7: How was Xpert and the algorithm for tuberculosis diagnosis implemented during the XTEND trial?

7. 1 Introduction

In this chapter, I address the third objective of this PhD, which is to explore the extent to which Xpert and the algorithm for the diagnosis of tuberculosis was integrated and embedded into routine care during and after the XTEND trial. This objective is supported by the contextual framework and the literature review. In the contextual framework I have shown that the differential adherence to the algorithm for the diagnosis of tuberculosis across XTEND study arms has potential to alter the mortality difference across study arms. In the literature review I have shown that health system context in which diagnostic testing takes place supports translation of test results into good outcomes. In this chapter, I summarise the paper 'A forgotten angle in impact evaluations of tuberculosis diagnostic tests: the assessment of integration and embedding of the Xpert MTB/RIF diagnostic technology into routine care using normalisation process theory PLOS ONE' which has been submitted to and accepted by PLoS One (PONE-D-19-20783), and is awaiting publication (see Part II of this PhD for the complete paper). I then explain how the findings of this paper support the aim of the PhD.

7. 2 An overview of Paper 3 'Integration and embedding of Xpert into routine care'

Paper 3 (Integration of Xpert into routine care) was a qualitative evaluation to identify barriers and facilitators of adherence to tuberculosis diagnostic algorithms at eight representative clinics[59] in order to better understand the health system structures that support or undermine tuberculosis diagnostic processes. In a previous study, we had observed that there was marked variation in adherence to tuberculosis diagnostic algorithms across participating XTEND clinics[13]. To identify eight facilities spanning the spectrum of adherent and non-adherent clinics, urban *vs* rural and intervention *vs* control arm, we used a heuristic evaluation of data from all 40 participating XTEND clinics

We drew up and piloted a semi-structured interview to guide interviews of tuberculosis control programme staff, HCW and administrators at PHC facilities. The interview elicited HCW reflections on the burden of tuberculosis in their community and how HCW used Xpert in routine care. We gauged this by inviting respondents to narrate events from actual patients whose results were documented in the tuberculosis case investigation register. We invited respondents to tell the story of Xpert introduction to the clinic and their role in establishing its use. We asked participants to propose an 'ideal diagnostic test for tuberculosis'.

We used and adapted a 'middle-range' sociological theory (normalisation process theory) to interpret our interview findings, and to build a coherent explanatory narrative regarding the degree of Xpert integration into routine care. We conducted 41 interviews at eight clinics of all cadres of health care workers including sub-district tuberculosis co-ordinators, facility managers, PHC nurses, tuberculosis

nurses, community workers and data capturers. By the time the interviews were conducted, all facilities had switched from smear microscopy to Xpert.

Almost all implementers reported favourable experiences using Xpert. Reasons cited were improved tuberculosis case finding related to the increased sensitivity of Xpert, earlier and improved diagnosis of drug-resistant tuberculosis (DR-TB) and easier procedural administration as only a single sputum specimen was required in the new algorithm. Our study identified that each clinic had consistent ways of conducting tuberculosis investigations that were variously adherent, partially adherent or non-adherent to guidelines for tuberculosis diagnosis. We referred to these patterns of tuberculosis investigation as 'clinic lore' to reflect the 'ritualised' and 'static' manner clinics diagnosed tuberculosis using Xpert. A summary of clinic 'lore' regarding tuberculosis diagnosis at eight representative clinics participating in the XTEND trial, 2012-4, as assessed from health care worker interviews and clinic observations is tabulated in Table 4 of the paper[59] (Part II).

Using qualitative methodology, we observed how district health services supported adherence to tuberculosis diagnostic algorithms through provision of human and material resources. In districts with stronger health systems, tuberculosis co-ordinators and facility managers felt supported by the system, and confidently and authoritatively co-ordinated activities in the clinic and district to implement Xpert. We showed how integration of Xpert into routine care was enhanced when HCW correctly understood the Xpert algorithm and were able to interpret Xpert results. We showed how joint ownership of and shared responsibility for tuberculosis diagnosis by facility managers, tuberculosis nurses and PHC staff facilitated co-operation of tuberculosis diagnostic processes and integration of Xpert.

We observed that when a clinic was situated in a district with a weak or poorly functioning district health system, staff were less able to fulfil the requirements of the tuberculosis diagnostic algorithm due to human resource or material constraints. For example, referral for chest radiography was not possible if transport was not available and retention in care was compromised when community health worker contracts were not honoured at district level. Staff were often demoralised and felt disempowered. At facilities that did not adhere to diagnostic algorithms, implementation of tuberculosis diagnosis was often poorly co-ordinated between general PHC adult curative services and tuberculosis control programme nurse. Responsibility for tuberculosis diagnosis was frequently devolved to lower cadres of staff such as community health workers or 'DOT supporters'. We observed that tuberculosis diagnostic processes were driven forward by patient demand for additional tests when the initial Xpert result was negative. Staff at these clinics interpreted negative Xpert results as excluding tuberculosis. These staff would on occasion turn patients away who had ongoing tuberculosis symptoms and negative tests. Collective problem solving to support facility-based tuberculosis diagnostic processes and procedures was absent as were referral networks for patients requiring additional investigations.

7. 3 Implications of weak Xpert integration into routine diagnostic services in the assessment of mortality difference between smear microscopy and Xpert arms in the XTEND trial

In my conceptual framework, I identified that the manner of implementation of Xpert and the tuberculosis diagnostic algorithm is a critical element of the diagnostic-treatment pathway leading to good outcomes (Figure 3. 4). My literature review, demonstrated the relationship between contextual factors supporting tuberculosis diagnostic tests and good outcomes. I showed how health system support (including additional diagnostic tests and empiric tuberculosis treatment) translates diagnostic test results into good outcomes in the test-treatment continuum. I suggested that a careful evaluation of how the Xpert test was implemented and integrated into routine tuberculosis diagnostic processes in the XTEND trial would allow conclusions to be drawn regarding the failure of Xpert to reduce mortality amongst persons with presumptive tuberculosis.

Through the use of qualitative methodology, and specifically through the theoretical conceptual framework normalisation process theory, Paper 3 identifies areas in the health system that are essential to support tuberculosis diagnosis using Xpert beyond the mere provision of hardware and human resources. These are the co-ordination of district and clinic resources, creation and sustaining of referral networks, co-ordination within PHC facilities between adult curative services and the tuberculosis programme, and motivated, engaged facility managers, PHC nurses and tuberculosis programme staff.

Three qualitative evaluations of the tuberculosis diagnostic and management continuum at PHC level complement our findings[60-62]. Loveday et al[62] observed that district level ownership of the tuberculosis programme allowed realignment of district services in support of TCP activities and integration with other competing district health management team priorities. She reported that failure of district and facility managers to take responsibility for DR-TB management lead to partial implementation and delays in key logistical services such as patient transport. Cattamanchi *et al*[61] in Uganda reported that poor self-efficacy regarding tuberculosis diagnosis amongst HCW due to time and resource constraints made adherence to guidelines difficult. The PROVE-IT study identified the role of patient agency in ongoing efforts to secure a diagnosis in the context of weaker health systems[60].

7. 4 Conclusions: Contributory observations from Paper 3

Our findings in Paper 3 offer explanatory reasons for the poor adherence to tuberculosis diagnostic algorithms across study arms, especially the wide variation in adherence from clinic to clinic. While the tuberculosis diagnostic algorithm is static, the context into which it is implemented differs from district to district with regard to health system strengths and weaknesses, and from clinic to clinic. At district level, co-ordination of resources, creation and sustaining of referral networks, ownership of the TCP by managers are required in order to support full implementation of and adherence to the algorithm. Within clinics, motivated and well trained facility managers, and a willing PHC staff complement have the

resources and motivation required to co-ordinate tuberculosis diagnosis, maintain referral networks, interpret Xpert results and translate tuberculosis testing into improved outcomes. Varying combinations of health system strengths at clinic and district level in the XTEND trial are likely to have lead to variations in the ability to support the diagnosis of tuberculosis along the test-treatment continuum. In turn, this may have impacted on patient outcomes at cluster level. However, the lack of systematic factors influencing support for TB diagnosis at the level of study arm means that XTEND study findings are unlikely to have been systematically biased.

Chapter 8: Explaining XTEND findings

8. 1 Drawing together of findings from Papers 1, 2 and 3

In Chapter 3, I used the contextual framework to identify key processes *en route* to good study outcomes that were not measured in the XTEND study. The objectives delineated how these processes should be investigated. In chapter 4, my literature review suggested that contextual support (in the form of additional diagnostic tests or health system strengthening initiatives such as improved retention in care) is necessary to translate the benefits of tuberculosis testing into improved outcomes. Together, the objectives and literature review findings suggested that a review of adherence to diagnostic algorithms for tuberculosis during the XTEND trial may provide explanatory evidence for the finding of no difference in mortality across intervention (Xpert) and control (smear microscopy) study arms.

In the literature review, I identified that XTEND study design was closer to the Chepetsa[44] study design, in that participant follow-up, including provision of additional diagnostic testing and retention in care, was left entirely to PHC facilities. The XTEND study design was further from the TB-NEAT[22] study design, in which the entire range of diagnostic tests including sputum culture and smear microscopy was implemented as part of study procedures. Following on from this observation, one would have expected to observe a significant mortality difference between the Xpert and smear microscopy arms in the XTEND trial, as the XTEND study design did not include additional diagnostic testing, nor interventions to support retention in care. In this PhD, I have identified several factors that may have worked against the translation of more sensitive Xpert testing into improved patient outcomes. These factors are listed in Table 8. 1.

In this chapter, I review each of the factors that worked against translation Xpert results into improved patient outcomes, and draw these together in a single explanatory framework.

Table 8. 1 Key processes *en route* to good outcomes in the tuberculosis diagnostic test-treatment continuum, observations pertaining to these in the XTEND study, and their impact on mortality

Key process *en route* to good patient outcomes	Observation in XTEND study		Impact on test-treatment cascade	Impact on mortality	Impact on mortality difference between XTEND study arms
	Smear microscopy	Xpert			
Additional investigations for tuberculosis	More likely amongst persons who were ill	More likely amongst persons who were ill	More persons likely to be treated empirically across both arms	Fewer deaths in both study arms	May have decreased statistical power to detect a difference between study arms
	More commonly done	Less commonly done	More persons likely to be treated empirically in smear microscopy arm	Fewer deaths in smear microscopy arm	May have decreased the difference, and rendered mortality across study arms equivalent
Empiric tuberculosis treatment	More commonly prescribed	Less commonly prescribed	More persons treated empirically in smear microscopy arm	Fewer deaths in smear arm	May have decreased the difference, and render mortality across study arms equivalent
Implementation of Xpert diagnostic algorithm	Implementation is highly context dependent, and more adherent in districts where district managers co-ordinate resources and referral networks, in facilities where managers co-ordinate TB diagnostic processes and staff are motivated, trained and engaged		Highly variable from clinic to clinic	Highly variable from clinic to clinic	Unlikely to have created a systematic bias across study arms

8. 2 The impact of infrequent additional investigations for tuberculosis on XTEND participant outcomes

Our findings from Paper 1 indicated that additional clinic-initiated investigations in line with tuberculosis diagnostic guidelines were conducted by clinic staff in relatively few XTEND participants. Less than a quarter of persons known to be HIV positive or have unknown HIV status had any additional investigations conducted[12] and less than 5% of participants underwent chest radiography[13]. However, we observed in Paper 1 that participants across study arms who were at greatest risk of death in XTEND were also more likely to have additional investigations for tuberculosis. This observation may have minimised the difference in mortality between study arms in two ways. Firstly, it may have been responsible for the reduced number of deaths across both study arms compared with other Xpert impact trials. This may have lead to a reduction in statistical power and a type 2 error (failure to detect a difference when there is one) as suggested by Auld et al[28], and DiTanna et al[63]. Secondly, the provision of additional investigations for those at greatest risk of death supported those who needed it most to progress along the test-treatment continuum. Therefore although the XTEND study design did not support diagnostic testing, clinician intervention at participating PHC facilities in a limited subset of patients may have improved outcomes in those at greatest risk of death.

8. 3 The impact of empiric tuberculosis treatment on XTEND participant outcomes

In paper 2, we reported that XTEND participants in the smear microscopy arm were prescribed empiric TB treatment more frequently than participants in the Xpert arm. It is highly likely that this contributed to a reduction in mortality amongst XTEND participants in the smear microscopy arm. The provision

of empiric tuberculosis treatment is directly on the path to good patient outcomes in the tuberculosis test-treatment continuum. The negative predictive value of smear microscopy[28] is low compared with Xpert, and may have lead clinicians to prescribe empiric treatment in the XTEND control (smear microscopy) arm.

8. 4 The impact of context-dependent adherence to tuberculosis diagnostic algorithms on XTEND participant outcomes

Our qualitative evaluation of the integration and embedding of Xpert in a sub-set of XTEND clinics presented in Paper 3, revealed that adherence to the algorithm for tuberculosis diagnosis is highly context dependent. Factors supporting adherence are more complex than the mere availability of human or material resources. They include district-level co-ordination of resources, co-ordination of services within clinics, and levels of HCW motivation, training and engagement. It is also evident from the wide range of adherence to algorithms from clinic to clinic reported in paper 1[12], and the differences in participant mortality by XTEND cluster[11] that the health system support varied by cluster and facility. While it is likely that the mortality in better supported clinics was lower, the XTEND study did not measure this, nor did our evaluation in paper 3 extend to all XTEND facilities. Therefore it is not possible to ascertain the degree to which our observations impacted on final outcome across study arm.

8. 5 Explaining XTEND findings

The XTEND trial was a pragmatic cluster-randomised evaluation of Xpert *vs* smear microscopy for the diagnosis of tuberculosis in ambulant persons with presumptive tuberculosis. In the XTEND trial, after enrolment and collection of sputum for smear microscopy or Xpert, implementation of algorithms for the diagnosis of tuberculosis across study arms were left in the hands of clinic and health system staff. After six months of follow-up, XTEND determined the burden of tuberculosis, the initial loss to follow up rates, the proportion of persons initiated on tuberculosis treatment and participant outcomes. In the final analysis, while one and a half times more tuberculosis cases were identified in the Xpert *vs* the smear microscopy arms, there was no difference by study arm in the numbers of participants lost to follow up, the proportion started on tuberculosis treatment and mortality rates. The XTEND trial did not ascertain the degree of adherence to the algorithms for tuberculosis diagnosis, nor the rates of empiric tuberculosis treatment nor assess how tuberculosis diagnosis was implemented by health care workers. We showed that the provision of additional investigations to XTEND participants who were more ill across study arms, lead to a reduction in statistical power to detect a difference in mortality. We also demonstrated higher rates of empiric treatment amongst persons in the smear microscopy arm. These factors collectively contributed to decreasing the mortality difference between XTEND study arms. We also observed that adherence to the diagnostic algorithm for Xpert, which most likely also applied to diagnostic algorithms for smear microscopy, was highly context dependent and varied from clinic to clinic amongst the eight clinics in which in-depth evaluations were conducted. These observations are likely generalizable across all 40 XTEND clinics, and are unlikely to have introduced a systematic bias across XTEND study arms.

Chapter 9: Reflecting critically on methodology – strengths and limitations of studies towards this PhD submission and XTEND,

9. 1 Limitations of evaluations contributing to PhD findings

My investigation into reasons that the XTEND trial did not find a difference in mortality between persons suspected of tuberculosis who were tested with Xpert *vs* smear microscopy was structured around a conceptual framework which elucidated stages on the test-treatment continuum, focusing specifically on the impact of false negative results. I centred the investigations around three aspects of this continuum, namely adherence to algorithms for tuberculosis diagnosis, prescription of empiric tuberculosis treatment, and a qualitative evaluation of how Xpert was implemented. However, there were several limitations in my approach.

Firstly, I did not test or measure all the pathways in my conceptual framework. One of the assumptions in XTEND was that untreated tuberculosis progresses to death in persons with presumptive tuberculosis, and conversely, treatment of tuberculosis leads to good patient outcomes. My evaluation would be well complemented by an investigation of deaths amongst XTEND participants. It would be helpful to establish the distribution of deaths by tuberculosis and HIV treatment status, and where possible, to ascertain the contributions of the health system and patient to poor outcomes. This investigation may still be conducted, as verbal autopsies of family members of deceased participants were conducted as part of another study[64]. The XTEND analysis may have been biased if more deaths occurred in tuberculosis-treated patients in the Xpert arm. Theoretically, this is unlikely, as a recent meta-analysis identified that tuberculosis treatment outcomes in patients diagnosed by Xpert *vs* smear microscopy were no different[65]. Furthermore, deaths in patients on tuberculosis treatment are usually related to late presentation and concomitant advanced HIV infection, and if randomisation was successful in XTEND, these patients should have been equally distributed across study arms.

Secondly my conceptual framework was highly simplified in order to highlight what I thought were essential components of the tuberculosis test-treatment continuum. I omitted to include HIV testing and treatment which are important parts of integrated tuberculosis-HIV care and important determinants of outcome on tuberculosis treatment[66]. In addition, the XTEND trial identified that untreated HIV infection was a risk factor for death[11]. My evaluation would be complemented by an assessment by study arm of the implementation of HIV testing and treatment, and the degree to which this affected participant outcomes. However, if randomisation was successful, these services should have been offered to equivalent participant numbers across study arms, with equal impact on participant outcomes.

Thirdly, in Paper 3, I evaluated Xpert implementation using qualitative methodology in eight representative clinics. However, by the time interviews were conducted, all clinics were implementing Xpert. It was therefore not possible to draw conclusions about how health care workers conducted tuberculosis investigations in the control (smear microscopy) arm. However, what emerged from Paper 3 was that Xpert implementation was highly context dependent. Given that the algorithms for

tuberculosis diagnosis using Xpert and smear microscopy are similarly complex, it is likely that the findings hold true for both arms.

In addition to the limitations listed above, Papers 1, 2 and 3 which are presented as part of this PhD, each reported on their own study limitations.

9. 2 Limitations of the XTEND study

9. 2. 1 Limitations acknowledged by XTEND investigators

The XTEND study design was influenced to a large degree on logistic considerations, in that the national roll-out of Xpert could not be accomplished instantaneously. It was essential that Xpert implementation be done rapidly, efficiently and with minimal disruption to clinical services. Furthermore, in order to assess the impact of Xpert under routine conditions, the study design was chosen so as to limit study-staff intervention in tuberculosis diagnosis and management. Acknowledged limitations in XTEND study design include: 1) the absence of a gold standard (culture) in all participants and the consequently inability to measure the true prevalence of tuberculosis; 2) the inability to determine the impact of Xpert amongst DR-TB cases because of the low prevalence of DR-tuberculosis amongst participants and 3) the potential that study staff follow-up activities influence patient outcomes (although this impact would have been operational across study arms). In the final analysis, the investigators reported a baseline imbalance in disease severity amongst participants, with those in the Xpert arm tending to be slightly healthier. Statistical adjustments were made for this observation. However, this may have impacted on study outcomes in unknown ways.

This PhD submission, through the conceptual framework, literature review and findings, has highlighted additional limitations in the XTEND study design.

9. 2. 2 Limitations highlighted by this PhD

Patient diagnostic and treatment pathways

The XTEND trial study design and protocol did not draw out the test-treatment pathways to good outcomes following Xpert or smear microscopy testing. Consequently the study design did not incorporate nor report on measurements of important processes *en route* to good outcomes including the relationship between HIV treatment and care, and the tuberculosis diagnostic processes. Unmeasured variables, which fortunately could be extracted from XTEND data and which are presented in Papers 1,2 and 3 of this submission include the proportions of participants undergoing additional investigations and proportions treated empirically for tuberculosis. Further, the study design did not include an assessment of how tuberculosis diagnostic procedures were implemented by HCW. This is best done using qualitative methodology. Paper 3 in this PhD provides such an assessment. Lastly, the study design of XTEND did not allow investigation of false positive results, nor the impact of these on patient outcomes.

Quality criteria for diagnostic testing (QUADAS and STARD)

The XTEND trial protocol and final paper did not refer to the QUADAS[34] and STARD[32] tools, and it is not clear how these tools influenced XTEND study design. The QUADAS tool is helpful to systematically identify potential sources of bias including partial or differential verification, test review bias, reference standard review bias and clinical review bias. These use of these tools may have prompted a more thoughtful review of study design regarding measurement of variables on the test-treatment continuum.

Chapter 10: Conclusion

10. 1 The contribution of this PhD to Xpert impact evaluations

Many researchers have pondered over the failure of Xpert to impact on patient outcomes. Proposed reasons include the use of empiric tuberculosis treatment, inadequate power to ascertain a difference between study arms, the evaluation of Xpert in non-target populations, inadequate end-point ascertainment through loss to follow-up and health system weaknesses. This PhD provides an in depth evaluation of a particular impact trial -the XTEND trial -to establish the specific reasons for the absence of impact on mortality. In addition, the papers submitted towards this PhD have each made a contribution to the broader knowledge base pertaining to tuberculosis diagnosis in routine contexts.

Paper 1 illustrated how tuberculosis diagnostic algorithms are not adhered to in routine PHC settings in South Africa, a finding which has not been widely reported on to date. This finding opens the way for targeted training and health system strengthening in the South African tuberculosis control programme, and globally, alerts tuberculosis policy makers and managers to be more attentive to these programmatic aspects of tuberculosis control. Paper 1 drew attention to the differential rates of additional investigations across study arm in XTEND, and highlighted that these were more likely to be done in persons who were more ill. Both observations may have stunted the ability of XTEND to detect an impact on mortality. Finally, Paper 1 illustrated how additional tuberculosis investigations are important components of the test-treatment continuum, which should be considered when designing and conducting impact evaluations

Paper 2 reported that the frequency of empiric tuberculosis prescription under routine conditions in South African PHC settings is low. The study concluded that there is extensive reliance on diagnostic testing in the South African tuberculosis control programme, which is largely nurse-driven. This observation underscores the importance of highly sensitive diagnostic testing for tuberculosis especially those tests with point-of-care application. The more frequent prescription of empiric tuberculosis treatment in the smear microscopy arm in XTEND contributed to a reduced mortality difference across study arms. Paper 2 confirmed that empiric tuberculosis treatment is an important confounder in impact evaluations of tuberculosis diagnostic tests, and this should be considered in future impact evaluations.

Finally, Paper 3 provided a theory-based sociological assessment of integration and embedding of a diagnostic test into routine care. The paper is the first description of the use of normalisation process theory (NPT) to investigate a new tuberculosis diagnostic test. The findings from Paper 3 illustrate the significant contribution of health systems to tuberculosis diagnostic processes, and identified specific areas for strengthening. The paper further illustrated how qualitative assessments can be integrated into impact assessments to complement and explain quantitative findings by providing supporting explanatory evidence.

This PhD draws all these findings together into a coherent explanatory narrative to ascertains reasons for lack of impact on mortality in the XTEND trial.

10. 2 The contribution of this PhD to my personal and professional growth

On a professional level, this PhD allowed me to acquire new skills in the area of data management, Stata® coding and the statistical analysis of cluster randomised trials. With regard to qualitative research, I learned how to develop conceptual frameworks, develop interview guides, conduct, record and transcribe interviews, use NVivo to support coding of interviews and analysis of qualitative data. In the process of writing papers and the narrative, I learned to summarise and track important papers by maintaining an excel spreadsheet. I gained experience in the rigour of scientific writing, honed by over 35 numbered versions of all three papers.

On a personal level, participation with the team of XTEND investigators and study staff gave me the opportunity to work with a highly committed and able group of people, whilst at the same time receiving mentorship from colleagues vastly more experienced in all aspects of trial development, implementation, analysis and reporting. This PhD is the fruit of their enthusiasm and passion for research, their scientific integrity and their support and encouragement, particularly following my move from Aurum to the NICD.

10. 3 Thoughts on the way forward

Multiple impact evaluations[11,22,24,42,44-46], systematic reviews[28] and meta-analyses[65,67] have failed to demonstrate an improvement in patient outcomes amongst persons with presumptive tuberculosis who were tested with Xpert as opposed to smear. The findings from this PhD pinpoint reasons for this failure in a single impact evaluation, and illustrate important methodological considerations in the design and conduct of impact evaluations of new tuberculosis diagnostic tests. While Xpert may not have impacted on patient outcomes, a plethora of new diagnostic tests are becoming available. Each may have a place in the test-treatment continuum. The many lessons learnt from XTEND and other impact evaluations will most certainly contribute to the study design and impact assessments of these.

Chapter 11. References

1. WHO. *Pathways to better diagnostics for tuberculosis: a blueprint for the development of TB idiagnostics by the new diagnostics working group of the Stop TB Partnership.* Geneva, Switzerland: WHO;2009.
2. Weyer K. Discovery, innovation, and new frontiers in tuberculosis diagnostics: reflections and expectations. *J Infect Dis.* 2015;211 Suppl 2:S78-80.
3. Rouillon A, Perdrizet S, Parrot R. Transmission of tubercle bacilli: The effects of chemotherapy. *Tubercle.* 1976;57(4):275-299.
4. van Zyl-Smit RN, Binder A, Meldau R, et al. Comparison of quantitative techniques including Xpert MTB/RIF to evaluate mycobacterial burden. *PLoS One.* 2011;6(12):e28815.
5. Helb D, Jones M, Story E, et al. Rapid detection of Mycobacterium tuberculosis and rifampin resistance by use of on-demand, near-patient technology. *J Clin Microbiol.* 2010;48(1):229-237.
6. Boehme CC, Nabeta P, Hillemann D, et al. Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med.* 2010;363(11):1005-1015.
7. WHO. Policy statement: "Automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF system. 2011.
8. Weyer K, Mirzayev F, Migliori GB, et al. Rapid molecular TB diagnosis: evidence, policy making and global implementation of Xpert MTB/RIF. *Eur Respir J.* 2013;42(1):252-271.
9. Keeler E, Perkins MD, Small P, et al. Reducing the global burden of tuberculosis: the contribution of improved diagnostics. *Nature.* 2006;444 Suppl 1:49-57.
10. Keeton C. New diagnostic test changes tuberculosis landscape. *Bulletin of the World Health Organisation.* 2013;91(3):163-164.
11. Churchyard GJ, Stevens WS, Mametja LD, et al. Xpert MTB/RIF versus sputum microscopy as the initial diagnostic test for tuberculosis: a cluster-randomised trial embedded in South African roll-out of Xpert MTB/RIF. *The Lancet Global health.* 2015;3(8):e450-457.
12. McCarthy KM, Grant AD, Chihota V, et al. Implementation and Operational Research: What Happens After a Negative Test for Tuberculosis? Evaluating Adherence to TB Diagnostic Algorithms in South African Primary Health Clinics. *J Acquir Immune Defic Syndr.* 2016;71(5):e119-126.
13. McCarthy K, Fielding K, Churchyard GJ, Grant AD. Empiric tuberculosis treatment in South African primary health care facilities - for whom, where, when and why: Implications for the development of tuberculosis diagnostic tests. *PLoS One.* 2018;13(1):e0191608.
14. National Department of Health. National Tuberculosis Management Guidelines. 2014.
15. National Department of Health. National Tuberculosis Management Guidelines. Tshwane 2009.
16. Denkinger CM, Schumacher SG, Boehme CC, Dendukuri N, Pai M, Steingart KR. Xpert MTB/RIF assay for the diagnosis of extrapulmonary tuberculosis: a systematic review and meta-analysis. *Eur Respir J.* 2014;44(2):435-446.
17. Lawn SD, Nicol MP. Xpert(R) MTB/RIF assay: development, evaluation and implementation of a new rapid molecular diagnostic for tuberculosis and rifampicin resistance. *Future Microbiol.* 2011;6(9):1067-1082.
18. Lessells RJ, Cooke GS, Newell ML, Godfrey-Faussett P. Evaluation of tuberculosis diagnostics: establishing an evidence base around the public health impact. *J Infect Dis.* 2011;204 Suppl 4:S1187-1195.
19. Talbot EA, Pape JW, Sundaram L, Boehme CC, Perkins MD. Transforming TB diagnosis: can patients and control programs afford to wait? *Am J Trop Med Hyg.* 2012;87(2):202-204.
20. Hanrahan CF, Clouse K, Bassett J, et al. The patient impact of point-of-care vs. laboratory placement of Xpert((R)) MTB/RIF. *Int J Tuberc Lung Dis.* 2015;19(7):811-816.
21. Menzies NA, Cohen T, Lin HH, Murray M, Salomon JA. Population health impact and cost-effectiveness of tuberculosis diagnosis with Xpert MTB/RIF: a dynamic simulation and economic evaluation. *PLoS Med.* 2012;9(11):e1001347.
22. Theron G, Zijenah L, Chanda D, et al. Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. *Lancet.* 2014;383(9915):424-435.

23. Trajman A, Durovni B, Saraceni V, et al. Impact on Patients' Treatment Outcomes of XpertMTB/RIF Implementation for the Diagnosis of Tuberculosis: Follow-Up of a Stepped-Wedge Randomized Clinical Trial. *PLoS One.* 2015;10(4):e0123252.
24. Cox HS, Mbhele S, Mohess N, et al. Impact of Xpert MTB/RIF for TB diagnosis in a primary care clinic with high TB and HIV prevalence in South Africa: a pragmatic randomised trial. *PLoS Med.* 2014;11(11):e1001760.
25. Mupfumi L, Makamure B, Chirehwa M, et al. Impact of Xpert MTB/RIF on Antiretroviral Therapy-Associated Tuberculosis and Mortality: A Pragmatic Randomized Controlled Trial. *Open forum infectious diseases.* 2014;1(1):ofu038.
26. Boyles TH. Why do clinical trials of Xpert(R) MTB/RIF fail to show an effect on patient relevant outcomes? *Int J Tuberc Lung Dis.* 2017;21(3):249-250.
27. Albert H, Nathavitharana RR, Isaacs C, Pai M, Denkinger CM, Boehme CC. Development, roll-out and impact of Xpert MTB/RIF for tuberculosis: what lessons have we learnt and how can we do better? *Eur Respir J.* 2016;48(2):516-525.
28. Auld AF, Fielding KL, Gupta-Wright A, Lawn SD. Xpert MTB/RIF - why the lack of morbidity and mortality impact in intervention trials? *Trans R Soc Trop Med Hyg.* 2016;110(8):432-444.
29. Lord SJ, Irwig L, Bossuyt PM. Using the principles of randomized controlled trial design to guide test evaluation. *Med Decis Making.* 2009;29(5):E1-e12.
30. Dowdy DW, Gounder CR, Corbett EL, Ngwira LG, Chaisson RE, Merritt MW. The ethics of testing a test: randomized trials of the health impact of diagnostic tests for infectious diseases. *Clin Infect Dis.* 2012;55(11):1522-1526.
31. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ.* 2012;344:e686.
32. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527.
33. Van Belle S, van de Pas R, Marchal B. Towards an agenda for implementation science in global health: there is nothing more practical than good (social science) theories. *BMJ global health.* 2017;2(2):e000181.
34. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003;3:25.
35. Gilson L, Hanson K, Sheikh K, Agyepong IA, Ssengooba F, Bennett S. Building the field of health policy and systems research: social science matters. *PLoS Med.* 2011;8(8):e1001079.
36. Cobelens F, van den Hof S, Pai M, Squire SB, Ramsay A, Kimerling ME. Which new diagnostics for tuberculosis, and when? *J Infect Dis.* 2012;205 Suppl 2:S191-198.
37. Mann G, Squire SB, Bissell K, et al. Beyond accuracy: creating a comprehensive evidence base for TB diagnostic tools. *Int J Tuberc Lung Dis.* 2010;14(12):1518-1524.
38. Schumacher SG, Sohn H, Qin ZZ, et al. Impact of Molecular Diagnostics for Tuberculosis on Patient-Important Outcomes: A Systematic Review of Study Methodologies. *PLoS One.* 2016;11(3):e0151073.
39. Pai M, Schumacher SG, Abimbola S. Surrogate endpoints in global health research: still searching for killer apps and silver bullets? *BMJ global health.* 2018;3(2):e000755.
40. Green BN, Johnson CD, Adams A. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *J Chiropr Med.* 2006;5(3):101-117.
41. Durovni B, Saraceni V, van den Hof S, et al. Impact of replacing smear microscopy with Xpert MTB/RIF for diagnosing tuberculosis in Brazil: a stepped-wedge cluster-randomized trial. *PLoS Med.* 2014;11(12):e1001766.
42. Calligaro GL, Theron G, Khalfey H, et al. Burden of tuberculosis in intensive care units in Cape Town, South Africa, and assessment of the accuracy and effect on patient outcomes of the Xpert MTB/RIF test on tracheal aspirate samples for diagnosis of pulmonary tuberculosis: a prospective burden of disease study with a nested randomised controlled trial. *The Lancet Respiratory medicine.* 2015;3(8):621-630.
43. van Kampen SC, Susanto NH, Simon S, et al. Effects of Introducing Xpert MTB/RIF on Diagnosis and Treatment of Drug-Resistant Tuberculosis Patients in Indonesia: A Pre-Post Intervention Study. *PLoS One.* 2015;10(6):e0123536.
44. Ngwira LG, Corbett EL, Khundi M, et al. Screening for Tuberculosis With Xpert MTB/RIF Assay Versus Fluorescent Microscopy Among Adults Newly Diagnosed With Human Immunodeficiency Virus in Rural Malawi: A Cluster Randomized Trial (Chepetsa). *Clin Infect Dis.* 2019;68(7):1176-1183.

45. Auld AF, Agizew T, Mathoma A, Boyd R. Effect of TB screening and retention interventions on early ART mortality in Botswana. Paper presented at: Conference on Retroviruses and Opportunistic Infections (CROI); March 8-11, 2018; Boston, Massachusetts, USA.
46. Yoon C, Cattamanchi A, Davis JL, et al. Impact of Xpert MTB/RIF testing on tuberculosis management and outcomes in hospitalized patients in Uganda. *PLoS One.* 2012;7(11):e48599.
47. Lawn SD, Nicol MP, Corbett EL. Effect of empirical treatment on outcomes of clinical trials of diagnostic assays for tuberculosis. *Lancet Infect Dis.* 2015;15(1):17-18.
48. Cox H, Dickson-Hall L, Ndjeka N, et al. Delays and loss to follow-up before treatment of drug-resistant tuberculosis following implementation of Xpert MTB/RIF in South Africa: A retrospective cohort study. *PLoS Med.* 2017;14(2):e1002238.
49. Hayes RJ, Moulton LH. *Cluster randomised trials.* Boca Raton, FL: Chapman & Hall; 2009.
50. Loveday M, Thomson L, Chopra M, Ndlela Z. A health systems assessment of the KwaZulu-Natal tuberculosis programme in the context of increasing drug resistance. *Int J Tuberc Lung Dis.* 2008;12(9):1042-1047.
51. Tafuma TA, Burnett RJ, Huis In 't Veld D. National guidelines not always followed when diagnosing smear-negative pulmonary tuberculosis in patients with HIV in Botswana. *PLoS One.* 2014;9(2):e88654.
52. Alamo ST, Kunutsor S, Walley J, et al. Performance of the new WHO diagnostic algorithm for smear-negative pulmonary tuberculosis in HIV prevalent settings: a multisite study in Uganda. *Trop Med Int Health.* 2012;17(7):884-895.
53. WHO. Improving the diagnosis and treatment of smear-negative pulmonary and extra-pulmonary tuberculosis among adults and adolescents: Recommendations for HIV-prevalent and resource-constrained settings. . Geneva, Switzerland2007.
54. Holtz TH, Kabera G, Mthiyane T, et al. Use of a WHO-recommended algorithm to reduce mortality in seriously ill patients with HIV infection and smear-negative pulmonary tuberculosis in South Africa: an observational cohort study. *Lancet Infect Dis.* 2011;11(7):533-540.
55. Abebe G, Deribew A, Apers L, et al. Evaluation of the 2007 WHO guideline to diagnose smear negative tuberculosis in an urban hospital in Ethiopia. *BMC Infect Dis.* 2013;13:427.
56. Walley J, Kunutsor S, Evans M, et al. Validation in Uganda of the new WHO diagnostic algorithm for smear-negative pulmonary tuberculosis in HIV prevalent settings. *J Acquir Immune Defic Syndr.* 2011;57(5):e93-100.
57. Dimairo M, MacPherson P, Bandason T, et al. The risk and timing of tuberculosis diagnosed in smear-negative TB suspects: a 12 month cohort study in Harare, Zimbabwe. *PLoS One.* 2010;5(7):e11849.
58. Theron G, Peter J, Dowdy D, Langley I, Squire SB, Dheda K. Do high rates of empirical treatment undermine the potential effect of new diagnostic tests for tuberculosis in high-burden settings? *Lancet Infect Dis.* 2014;14(6):527-532.
59. McCarthy KM, Grant A, Fielding K, Kielmann K. A forgotten angle in impact evaluations of tuberculosis diagnostic tests: the assessment of integration and embedding of the Xpert MTB/RIF diagnostic technology into routine care using normalisation process theory. Submitted. *PLoS One.* 2019.
60. Naidoo P, van Niekerk M, du Toit E, Beyers N, Leon N. Pathways to multidrug-resistant tuberculosis diagnosis and treatment initiation: a qualitative comparison of patients' experiences in the era of rapid molecular diagnostic tests. *BMC Health Serv Res.* 2015;15:488.
61. Cattamanchi A, Miller CR, Tapley A, et al. Health worker perspectives on barriers to delivery of routine tuberculosis diagnostic evaluation services in Uganda: a qualitative study to guide clinic-based interventions. *BMC Health Serv Res.* 2015;15:10.
62. Loveday M, Padayatchi N, Wallengren K, et al. Association between health systems performance and treatment outcomes in patients co-infected with MDR-TB and HIV in KwaZulu-Natal, South Africa: implications for TB programmes. *PLoS One.* 2014;9(4):e94016.
63. Di Tanna GL, Khaki AR, Theron G, et al. Effect of Xpert MTB/RIF on clinical outcomes in routine care settings: individual patient data meta-analysis. *The Lancet Global health.* 2019;7(2):e191-e199.
64. Maraba N, Karat AS, McCarthy K, et al. Verbal autopsy-assigned causes of death among adults being investigated for TB in South Africa. *Trans R Soc Trop Med Hyg.* 2016;110(9):510-516.
65. Agizew T, Boyd R, Auld AF, et al. Treatment outcomes, diagnostic and therapeutic impact: Xpert vs. smear. A systematic review and meta-analysis. *Int J Tuberc Lung Dis.* 2019;23(1):82-92.
66. Colebunders R, Bastian I. A review of the diagnosis and treatment of smear-negative pulmonary tuberculosis. *Int J Tuberc Lung Dis.* 2000;4(2):97-107.

Part II: Published or submitted papers arising from the XTEND trial

Papers submitted towards this PhD

Paper 1: What Happens After a Negative Test for Tuberculosis? Evaluating Adherence to tuberculosis Diagnostic Algorithms in South African Primary Health Clinics. (Published in JAIDS Operational Research)

Paper 2: Empiric tuberculosis treatment in South African primary health care facilities -for whom, where, when and why? (Published in PLoSONE)

Paper 3: A forgotten angle in impact evaluations of tuberculosis diagnostic tests: the assessment of integration and embedding of the Xpert MTB/RIF diagnostic technology into routine care using normalisation process theory (Submitted to PLoSOne July 2019)

Xpert versus sputum microscopy as the initial diagnostic test for tuberculosis: a cluster-randomised trial embedded in South African roll-out of Xpert. (Published in Lancet Global Health)

Appendices

Appendix 1. Human Research Ethics Committee approval for PhD submission

R14/49 Dr Kerrigan McCarthy et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M140486

NAME: (Principal Investigator)	Dr Kerrigan McCarthy et al
DEPARTMENT:	School of Public Health Aurum Health Institute Eastern Cape, Gauteng, Free State and Mpumalanga Provinces
PROJECT TITLE:	A Negative Test for TB by XPert MTB/RIF or Smear Microscopy - What does it mean for Health Care Workers and Persons suspected of TB?
DATE CONSIDERED:	25/04/2014
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Prof Gavin Churchyard
APPROVED BY:	Professor P Cleaton-Jones, Co-Chairperson, HREC (Medical)
DATE OF APPROVAL:	08/09/2014

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

Appendix 2. Turnitin report

a0035612:KerrigansPhDnarrative_final_20190913a_forTurnitin.d.

ORIGINALITY REPORT

8%	6%	7%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	journals.plos.org Internet Source	2%
2	Gavin J Churchyard, Wendy S Stevens, Lerole D Mametja, Kerrigan M McCarthy et al. "Xpert MTB/RIF versus sputum microscopy as the initial diagnostic test for tuberculosis: a cluster-randomised trial embedded in South African roll-out of Xpert MTB/RIF", The Lancet Global Health, 2015 Publication	1%
3	gatesopenresearch.org Internet Source	1%
4	tbevidence.org Internet Source	1%
5	academic.oup.com Internet Source	<1%
6	www.teachepi.org Internet Source	<1%
7	bmchealthservres.biomedcentral.com	

20	Violet N. Chihota, Sibuse Ginindza, Kerrigan McCarthy, Alison D. Grant, Gavin Churchyard, Katherine Fielding. "Missed Opportunities for TB Investigation in Primary Care Clinics in South Africa: Experience from the XTEND Trial", PLOS ONE, 2015 Publication	<1%
21	Anna Vassall, Mariana Siapka, Nicola Foster, Lucy Cunnama et al. "Cost-effectiveness of Xpert MTB/RIF for tuberculosis diagnosis in South Africa: a real-world cost analysis and economic evaluation", The Lancet Global Health, 2017 Publication	<1%
22	"Abtracts from the 30th Annual Meeting of the Society of General Internal Medicine", Journal of General Internal Medicine, 2007 Publication	<1%
23	www.who.int Internet Source	<1%
24	qmro.qmul.ac.uk Internet Source	<1%
25	www.news24.com Internet Source	<1%
26	www.edgehill.ac.uk Internet Source	<1%
27	Ziegler, A., I. König, and P. Schulz-Knappe. "Challenges in planning and conducting diagnostic studies with molecular biomarkers", DMW - Deutsche Medizinische Wochenschrift, 2013. Publication	<1%
28	www.medicineandreligion.com Internet Source	<1%
29	scholar.sun.ac.za Internet Source	<1%
30	www.hepatologie.org Internet Source	<1%
31	lse.ac.uk Internet Source	<1%
32	www.inahta.org Internet Source	<1%
33	repository.up.ac.za Internet Source	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		