Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Nicholas Ripper Date

Analysis of Relationship between Participant Characteristics and Phenotypic Resistance to Second-Line Anti-Tuberculosis Drugs in KwaZulu-Natal, South Africa

By

Nicholas Ripper

Master of Science in Public Health

Global Epidemiology

_____ [Chair's signature]

N. Sarita Shah, MD MPH

Committee Chair

_____ [Member's signature]

James C.M. Brust, MD

Committee Member

Analysis of Relationship between Participant Characteristics and Phenotypic Resistance to Second-Line Anti-Tuberculosis Drugs in KwaZulu-Natal, South Africa

By

Nicholas Ripper

B.S.

Allegheny College

2021

Thesis Committee Chair: N. Sarita Shah, MD MPH

Thesis Committee Member: James C.M. Brust, MD

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Global Epidemiology

2023

Abstract

Analysis of Relationship between Participant Characteristics and Phenotypic Resistance to Second-Line Anti-Tuberculosis Drugs in KwaZulu-Natal, South Africa
By Nicholas Ripper

Background: Resistance to second-line drugs used to drug-resistant tuberculosis (SLDR-TB) is a public health threat in South Africa, both because of its severity and the difficulty of successful treatment. Although studies have examined risk factors of second-line drug resistance, most this work has been on older second-line regimens and little is known about newer medications and regimens. We therefore examined risk factors for resistance to second-line drugs (capreomycin, moxifloxacin or levofloxacin, linezolid, bedaquiline, and clofazimine) in the province of KwaZulu-Natal, South Africa from 2018-2022, when more current second-line regimens were being used.
Methods: Phenotypic drug-susceptibility testing (pDST) for each second-line drug was performed on *M. tuberculosis* isolates from individuals with SLDR-TB in KwaZulu-Natal, and participant characteristics were recorded for individuals within a subregion of the province. We calculated unadjusted odds ratios for sex and age with resistance to each drug for all study participants and unadjusted and adjusted odds ratios for sex, age, HIV status, alcohol use, and income for individuals within the subregion using binary logistic regression. We also assessed interaction between these variables and previous DR-TB treatment.
Results: Among participants in the entire study population (n=580) and within the subregion of interest (n=189), resistance varied across different second-line drugs. None of the exposures of interest were associated with resistance to any second-line drugs, and there was no significant interaction between previous treatment for DR-TB and the exposures.
Discussion: While we could not identify an association between the risk factors of interest with phenotypic resistance to second-line TB drugs, this may have been the result of the parent study population or because most resistant TB in KwaZulu-Natal is transmitted rather than acquired. Future studies could build on this work by examining the association of risk factors with resistance to other second-line drugs and focusing on differing effects between individuals with acquired and transmitted SLDR-TB.

Analysis of Relationship between Participant Characteristics and Phenotypic Resistance to Second-Line Anti-Tuberculosis Drugs in KwaZulu-Natal, South Africa

By

Nicholas Ripper

B.S.

Allegheny College

2021

Thesis Committee Chair: N. Sarita Shah, MD MPH

Thesis Committee Member: James C.M. Brust, MD

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Global Epidemiology

2023

Acknowledgements

More than anything else, this manuscript is the result of a team effort. Throughout every stage of producing this thesis, I have received so much help from an incredible group of individuals and mentors, and I want to take this space to thank them individually for their feedback, guidance, and encouragement. First, I want to thank my committee members, Dr. James C.M. Brust and Dr. N. Sarita Shah. I have learned so much from them, not just about the subject matter of this thesis, but also how to develop a research question, apply analytic methods to answer that research question, and interpret and communicate the results of this research. I am grateful for the time and effort that they have devoted to this thesis and to my personal and professional development.

I also want to thank the CONTEXT study team for their support. Fay Willis has played an instrumental role in initiating the abstraction of phenotypic drug susceptibility test (pDST) results and in training and assisting me to be involved in these study efforts. Dr. Keeren Lutchminarain of the National Health Laboratory Service not only has helped me with the interpretation and abstraction of pDST results but has also taught me much about SLDR-TB and the diagnostic processes for identifying drug resistance, for which I am grateful. Additionally, I want to thank members of the Emory team, including Nichole Evans, Angie Campbell, Kayleigh Nerhood, and Carmen Wagener, and members of the Centre for the AIDS Programme of Research in South Africa (CAPRISA) team in Durban, KwaZulu-Natal, South Africa, including Mfesane Kunju, Dr. Kogie Naidoo, Resha Boodhram, Sicelo Mthimkhulu, Senzo Hlathi, Nomthy Mbatha, and Anricia Padayachee, for the guidance that they have provided me on study processes.

Finally, I want to thank my friends and family for the support that they have given me while writing this manuscript; words cannot describe how grateful I am to have all of your love and support.

Table of Contents

Chapter I: Background on Drug-Resistant Tuberculosis (DR-TB), Including Extensively Drug-Resistant Tuberculosis (XDR-TB), Worldwide and in South Africa

Part I: Tuberculosis, Its Drug-Resistant Forms, and Its Global Public Health Importance

Tuberculosis (TB) is an infection and disease caused by bacteria belonging to the *Mycobacterium tuberculosis* (Mtb) complex, which includes the mycobacterial species *Mycobacterium tuberculosis*, *M. bovis*, and *M. avis*, among others (1). Individuals who are infected with *M. tuberculosis* often do not experience symptoms or transmit the infection (i.e., latent TB infection, or LTBI), while those who are symptomatic most commonly experience respiratory disease. Extrapulmonary manifestations of disease, such as in the lymph nodes or nervous system, also occur, but are less common (2; 3). The burden of disease attributable to TB on a global scale is considerable: in 2021, there were 10.6 million new cases of TB worldwide and 1.6 million deaths (an increase of about 300,000 deaths from the previous year) (3). In 2019, TB was the 13th most common cause of death worldwide, and, before the emergence of COVID-19, was the leading cause of death from a single infectious pathogen (3). TB is especially problematic in areas of high prevalence of human immunodeficiency virus (HIV), given the elevated risk of active TB disease among individuals living with HIV (3-5).

Treatment for TB is available, but regimens must include multiple drugs (also known as combination therapy), as resistance to one specific drug could lead to the failure of a single-drug regimen, a phenomenon that was first observed after attempts to treat TB with only streptomycin in the 1940s (2; 6). Currently, the standard treatment for individuals with TB is a six-month course of the drugs isoniazid and rifampicin, with two additional drugs, pyrazinamide and ethambutol, taken during the first two months (2; 3; 7). As noted above, however, drug resistance can prevent

effective treatment of TB infection. Multidrug-resistant tuberculosis (MDR-TB), defined as resistance to isoniazid and rifampicin, requires treatment to shift away from first-line drugs (isoniazid, rifampicin, pyrazinamide and ethambutol) to second-line drugs, such as fluoroquinolones (e.g., levofloxacin), injectable drugs (e.g., capreomycin), bedaquiline, and linezolid, among others (2-4; 8; 9). Globally, there were an estimated 450,000 new cases of MDR-TB or rifampicin-resistant TB (RR-TB) in 2021 (3).

Individuals may also experience resistance to second-line drugs as well. Until 2021, resistance to isoniazid, rifampicin, and at least one fluoroquinolone or one of the second-line injectable drugs (i.e., amikacin, kanamycin, or capreomycin) was classified as pre-extensively drug-resistant TB (pre-XDR-TB), while extensively drug-resistant TB (XDR-TB) was defined as resistance to fluoroquinolones AND injectables as well as isoniazid and rifampicin (10; 11). Although this study uses these definitions of pre-XDR-TB and XDR-TB, the WHO redefined both terms in 2020 due to the increasing importance of regimens with other second-line drugs, like bedaquiline, in treatment of drug-resistant TB (DR-TB) (8; 12-19). Under the new definitions, pre-XDR-TB is defined as resistance to a fluoroquinolone (plus isoniazid and rifampicin), while XDR-TB is now defined as resistance to at least one fluoroquinolone plus resistance to either bedaquiline or linezolid (12; 13; 20). Over 25,000 cases of pre-XDR-TB or XDR-TB (using the new definition) were reported in 2021 (3), but this is likely a considerable underestimate, given that many TB-endemic countries lack the resources for routine second-line drug resistance testing.

Although the incidence of DR-TB comprises a small share of total TB cases, its importance as a global public health issue cannot be overstated. While estimated detection of DR-TB has increased in recent years, improvement is still necessary; worldwide, only 71% of individuals with bacteriologically confirmed TB were tested for rifampicin resistance in 2021 (3). DR-TB,

especially pre-XDR-TB and XDR-TB, is difficult to treat, particularly for vulnerable populations, such as those with less social support (3; 11; 21). As we describe below, it was especially difficult to treat DR-TB with older regimens, which needed to be administered over longer periods of time (18-24 months), could cause severe side effects (e.g., kidney toxicity), and were less effective (14). Even though newer, safer, and more effective regimens have been developed, there are still disparities in treatment success between individuals with DS-TB, MDR-TB, and SLDR-TB (8; 15; 17-19). In 2021, while 86% of individuals with drug-susceptible TB (DS-TB) achieved a successful treatment outcome, only 60% of individuals treated for MDR-TB, pre-XDR-TB, and XDR-TB experienced treatment success (3).

Moreover, previous studies have identified greater difficulties in treatment of pre-XDR-TB and XDR-TB compared to MDR-TB (not just DS-TB): one study of individuals with DR-TB (n=1407) in South Korea from 2000-2002 found that survival was worse for individuals with XDR-TB (mean survival = 61.7-72.2 months) than for individuals with MDR-TB (mean survival = 89.2 months), and treatment success for individuals with pre-XDR-TB and resistance to ofloxacin (a fluoroquinolone) (35.8%) was lower than for MDR-TB (47.6%) (22). Another study conducted from 2005-2007 in the province of KwaZulu-Natal, South Africa likewise found a lower median survival time for individuals with XDR-TB (n=382, 28.5 days) than for individuals with MDR-TB (n=272, 60 days) (23). More recently, treatment success in South Africa in 2021 was less common for patients with XDR-TB (57%) than for those with MDR-TB (66%) (24), and a similar distribution of successful treatment outcomes appeared for individuals being treated with pre-XDR or XDR-TB (54%) and MDR-TB (60%) globally (25). Finally, longer hospitalization was reported for individuals with XDR-TB (n=7, mean=202 days) than for individuals with MDR-TB (n=177, mean=123 days) in Germany (26). Given the clinical difficulties and general well-

being that are at stake for individuals with DR-TB, especially pre-XDR-TB and XDR-TB, characterizing resistance to specific second-line drugs and investigating their potential determinants and risk factors is critical.

Part II: Genetic Mechanisms of Second-Line Drug Resistance

As noted above, a variety of first- and second-line drugs are used to treat TB (2). These drugs target specific cellular processes to kill or weaken the bacteria, such as development of the cell wall (isoniazid), DNA supercoiling (fluoroquinolones), transcription (rifampicin), protein synthesis (second-line injectable drugs, linezolid), development of ATP (bedaquiline) and the electron transport chain (clofazimine) (2; 27-31). Resistance of *Mycobacterium tuberculosis* to specific drugs is caused by changes to genes in the bacteria that code for these cellular processes (2; 32). Mutations in the *katG* and *inhA* genes, for example, confer resistance to isoniazid, and mutations in the *rpoB* gene are associated with rifampicin resistance (2; 5; 33). Resistance to fluoroquinolones like moxifloxacin and levofloxacin is linked to mutations in the *gyrA* and *gyrB* genes, and mutations in the *rrs* gene have been tied to resistance to second-line injectable drugs, like capreomycin (2; 5; 34). Resistance to linezolid, which disrupts protein synthesis, is associated with mutations in the *rplC* gene (31; 35). Finally, resistance to bedaquiline is associated with the mutations of the *Rv0678, atpE,* and *pepQ* genes. Mutations in *Rv0678* and *pepQ* upregulate efflux pumps which can expedite the removal of bedaquiline from the cell; certain mutations in each of these are also associated with clofazmine resistance (2; 29; 33; 36-39).

Part III: Risk Factors for Second-Line Drug Resistance

Previous studies with a variety of study settings, population sizes, and designs have sought to determine the risk factors associated with phenotypic resistance to specific second-line drugs, but the risk factors examined and the direction and significance of their association with resistance have generally been inconsistent or conflicting. Here, we describe some of the significant risk factors that have been previously identified by the literature. One study of individuals seeking treatment for TB at a single hospital in Beijing (n=3546) found significant associations between several risk factors, including previous exposure to fluoroquinolones, a diagnosis of COPD, and having been previously treated for TB, with resistance to ofloxacin (the prevalence of which was 8.6%) (40). In the PETTS study, another cohort study of individuals with MDR-TB across eight countries (including in Eastern Europe, East Asia, and South Africa) from 2005-2008, nearly 44% of participants were resistant to at least one second-line drug (41). Previous treatment for MDR-TB and receiving inpatient care were both significantly associated with resistance to fluoroquinolones (ciprofloxacin and ofloxacin) and second-line injectable drugs (capreomycin and amikacin), while unemployment, use of alcohol, and use of tobacco were associated with resistance to second-line injectable drugs (41).

Similarly, in a study conducted in South Africa from 2015 to 2019, resistance to bedaquiline among individuals about to start treatment with the drug (n=2023) was found to be associated with resistance to clofazimine (largely because of mutations in the *Rv0678* gene) (38), fluoroquinolones, and second-line injectable drugs (36). Positive HIV status, treatment with second-line drugs, and age 25-44 years were associated with acquired second-line injectable drug resistance in one study conducted in the United States from 1993-2008 (n=2274) (42), and treatment with other second-line drugs (e.g., moxifloxacin) were associated with acquired

resistance to capreomycin or ofloxacin in one study in Russia conducted between 2005 and 2008 (n=117) (43).

Other studies also considered potential risk factors for XDR-TB more broadly. First, medical status and history, especially in relation to TB, was identified as important in studies conducted across multiple geographic sites and time periods. Previous infection with TB was significantly associated with XDR-TB compared to MDR-TB in studies conducted in Taiwan from 2000-2006 (n=2625) (44; 45) and DS-TB in the United States from 1993-2007 (n=183,536) (46). Similarly, previous treatment with second-line injectable drugs was significantly associated with XDR-TB (compared to MDR-TB) in Russia (n=608) (47), Peru from 1997-2007 (n=1989) (48), and the multicountry PETTS study (41), and a history of previous TB treatment was significantly associated with XDR-TB compared to DS-TB in Pakistan from 2014-2019 (n=580) (49) and with MDR- or XDR-TB compared to DS-TB in Lithuania, Latvia, and Estonia from 2009-2012 (n=1041) (50) and in Japan in 2002 (n=2837) (51). Another important indicator of health, HIV status, was positively associated with XDR-TB compared to MDR-TB in the U.S. (46), Japan (51), and Portugal (1999-2007, n=132) (52), although a study in China did not find such an association (53).

Key socioeconomic and demographic characteristics are also important. In the eThekwini district of KwaZulu-Natal, South Africa (n=132), lower income was associated with XDR-TB, and geographic areas with likely high levels of XDR-TB transmission tended to have higher levels of unemployment (54). Alcohol use was significantly associated with acquired XDR-TB or MDR-TB in the Baltic states, and with XDR-TB alone in Estonia from 2003-2005 (n=1163) (50; 55), although not in Delhi from 2007-2010 (n=611) (56). Other studies in the U.S., the Baltic States, Pakistan, and Japan found an association between younger age and XDR-TB (46; 49-51; 57).

Interestingly, the relationship between sex and XDR-TB is inconsistent throughout the literature. While one study in South Korea (n=250) and the multi-country PETTS study found that XDR-TB was significantly associated with female sex (41; 58), a study in Iran (n=146) (59) and in the Baltic States (50) actually found a significant association with *male* sex, and other studies in the U.S., Japan, and Russia (n=75) found no significant association with sex at all (46; 51; 60).

Part IV: Diagnosis of Second-Line Drug Resistance

In the laboratory, drug resistance can be identified in several ways. Genotypic testing can identify the presence of specific genetic mutations conferring resistance to first- and second-line drugs, like the mutations described above (32; 61). For instance, the Genotype MTBDR*sl* test can identify mutations in the *gyrA* and *gyrB* genes (conferring fluoroquinolone resistance) and *rrs* and *eis* genes (conferring resistance to the injectable drugs) (61). Similarly, the GeneXpert MTB/RIF assay can identify rifampicin resistance by detecting mutations associated with rifampicin resistance in the *rpoB* gene (62). Other methods assess *phenotypic* resistance to specific drugs by examining the effect of drugs on bacterial growth *in vitro* in different media, such as agar and in liquid culture (via mycobacterial growth incubator tubes, or MGIT), which are collectively known as *phenotypic drug susceptibility testing*, or pDST (32). Identifying resistance to any first- and second-line drugs allows clinicians to create an effective regimen for an individual with DR-TB (61).

Part V: Treatment of DR-TB

The specific regimens used to treat various forms of DR-TB have changed over time. Previously, treatment options for DR-TB, especially XDR-TB, were incredibly lengthy, typically

lasting over eighteen months (2; 14). Successful treatment outcomes were less common, as well. Because of the issues that these treatments presented, the safety and efficacy of newer regimens and drugs have been a major focus of research. For instance, second-line injectable drugs were once used for treating MDR-TB, but have since been removed from recommended regimens due to their associated risk of adverse effects and poorer treatment outcomes (2; 8; 63). Drugs that were either newly developed or repurposed for TB, such as bedaquiline, clofazimine, linezolid, delamanid, and pretomanid, are now part of the most up-to-date and recommended regimens for DR-TB (2; 8; 17; 18; 29; 33; 36; 64). These new regimens include a 9-11 month treatment consisting of bedaquiline (which replaced second-line injectables), clofazimine, levofloxacin, and linezolid for MDR-TB and RR-TB (first implemented in South Africa in 2018), the 6-9 month BPaL (bedaquiline, pretomanid, and linezolid) regimen for MDR- and pre-XDR-TB, and the BPaLM regimen (the previous regimen with the fluoroquinolone moxifloxacin) for MDR-TB (8; 9; 15; 17-19).

Part VI: Pre-XDR-TB and XDR-TB in the Context of South Africa

South Africa has among the highest burdens of TB worldwide; in 2021, there were 304,000 new cases of TB (nearly 54% of whom were HIV-positive), accounting for nearly three percent of global incidence (3; 24). South Africa also has one of the highest burdens of DR-TB, with 7,106 reported cases of RR-TB and MDR-TB in 2021, 725 of whom had pre-XDR-TB and XDR-TB (3; 24; 65). In the face of this public health threat, South Africa has had a vigorous response. In 1994, the National TB Programme (NTP) was founded, which contributed to an expansion of activities to address TB by the government and related partners, including formal declarations of a public

health crisis and the expansion and improvement of clinics, diagnostic services, and data management related to care of individuals experiencing disease from TB (4).

South Africa's strong response to TB applies to its drug resistant forms, as well. Clinics dedicated to MDR-TB began appearing as early as 2001 (4). During the first three years of its use of the diagnostic test (2011-2013), South Africa accounted for a majority of global use of the GeneXpert MTB/RIF assay (4). Moreover, the country aggressively implemented the treatment of DR-TB (including MDR-TB) with bedaquiline in the 2010s, and eligibility for treatment with the 9-11 month regimen that replaced injectable drugs with bedaquiline expanded to include individuals with RR-TB in June 2018 (8; 16). This regimen would later be recommended by the WHO in 2019 (8). Ultimately, a majority of the individuals treated with bedaquiline worldwide are in South Africa (8; 29; 36) and many recent clinical trials that tested the safety or efficacy of shorter second-line regimens against DR-TB have had study sites in South Africa (15; 17-19).

DR-TB, especially SLDR-TB, has played a key role in the history of the TB epidemic in South Africa. In fact, one of the earliest descriptions of XDR-TB was from a rural area of KwaZulu-Natal in 2006 (7), although mutations conferring resistance to second-line drugs (*gyrA* and *rrs*) for the most common strain causing XDR-TB likely appeared before this outbreak, in the early 1990s (5). Other studies subsequently examined the extent of second-line drug-resistance in South Africa, with one research team finding that about 6% and 14% of MDR-TB isolates sent to a national lab in Johannesburg, South Africa from 2005 to 2007 were XDR-TB and pre-XDR-TB, respectively, although this study did exclude KwaZulu-Natal and the province of Mpumalanga from the analysis (66).

Investigations into the initial outbreak of XDR-TB in KwaZulu-Natal showed that a considerable proportion of individuals did not previously have TB and were receiving care at the

same hospital. This suggested that drug-resistant bacteria could be transmitted, rather than acquired, for many of the participants (7). Subsequent research found that a considerable proportion of individuals with XDR-TB in the province had Mtb isolates which were highly genetically related to other individuals with XDR-TB, indicating that transmission, rather than acquisition, has been the predominant driver of XDR-TB in KwaZulu-Natal (67; 68). Individuals with XDR-TB often traveled long distances (often to different *districts*, the geographical subdivisions of KwaZulu-Natal) to receive care (69), and the majority (84%) of so-called "genomic links" between individuals suggesting transmission occurred between districts, especially with the district of eThekwini, where the major city of Durban is located (70). These findings suggest that, in KwaZulu-Natal, XDR-TB is primarily *transmitted*, not acquired, and that transmission likely occurs in the community outside of more intimate social networks ("casual networks") and in the context of *migration* from the more rural districts of KwaZulu-Natal to the more heavily urbanized district of eThekwini (67-71).

Part VII: Gaps in the Literature and Study Aims

Understanding previous research on the epidemic of SLDR-TB in KwaZulu-Natal can lay the groundwork and guide assumptions for future studies of individuals with SLDR-TB in the province. Our study examines recent trends in phenotypic resistance to second-line drugs and associated risk factors among individuals with SLDR-TB in KwaZulu-Natal. While this has been done in previous studies (40; 42; 43), some of which were conducted in South Africa (36; 41; 72), most of those studies did not examine resistance to newer second-line drugs that are now recommended in current guidelines for treating DR-TB. For instance, previous studies have focused on older fluoroquinolones, such as ciprofloxacin and ofloxacin, but not on the newer

fluoroquinolones being used in therapies recommended by the WHO, like the six-month BPaLM regimen containing moxifloxacin and the nine-month regimen for RR-TB containing levofloxacin (8; 9; 19; 40-43). Resistance to other drugs being used in these regimens, like linezolid, similarly is not considered in these studies (8; 9; 17; 18; 40-43). Even those studies that do examine risk factors for resistance to newer drugs like bedaquiline and linezolid have limitations. For instance, a study examining risk factors for bedaquiline in South Africa was conducted from 2015-2019, when bedaquiline was still in use in the country but mainly before its widespread rollout in 2018 to most individuals with RR-TB (not just MDR-TB and SLDR-TB) (8; 36). Another study on phenotypic resistance to linezolid in South Africa not only had a small sample size (n=39), but also used data that had been collected in an earlier period (2010-2017) when the context of linezolid resistance in the country may have been different (72). Thus, a study examining risk factors for resistance to drugs currently being used to treat DR-TB in South Africa in a population for whom *M. tuberculosis* isolates have been more recently collected can provide an improved understanding of second-line drug resistance that is potentially more relevant to the current context in South Africa.

Chapter II: Analysis of Relationship between Participant Characteristics and Phenotypic Resistance to Second-Line Anti-Tuberculosis Drugs in KwaZulu-Natal, South Africa

Introduction

Tuberculosis (TB), a primarily respiratory disease caused by the bacterium *Mycobacterium tuberculosis*, is a global public health threat. Among infectious diseases, it is one of the largest contributors to the global burden of disease, with an estimated 10.6 million incident cases and 1.6 million deaths having occurred worldwide in 2021, an increase from previous years (3). Efforts to address this threat are complicated by numerous barriers, including drug-resistance, which come in several forms. Multidrug-resistant TB (MDR-TB) involves resistance to the first-line drugs isoniazid and rifampicin, and there were an estimated 450,000 cases globally in 2021 (3). Second line drug-resistant TB (SLDR-TB) involves resistance to the second-line drugs that are used to treat DR-TB. It includes extensively drug-resistant TB (XDR-TB), which was defined as resistance to at least one drug from both the fluoroquinolone and second-line injectable classes before 2020 but has since been redefined as resistance to at least one fluoroquinolone and either bedaquiline or linezolid, and pre-XDR-TB, or resistance to *either* a fluoroquinolone *or* bedaquiline *or* linezolid (10; 12; 13). SLDR-TB is an especially concerning type of TB. In 2021, 25,038 cases of XDR-TB or pre-XDR-TB were reported globally (3), and despite the recent rollout of more effective treatments of DR-TB (8; 9; 14; 17; 18), successful treatment outcomes for individuals with SLDR-TB are less common compared to individuals with drug-sensitive TB (DS-TB) or MDR-TB (3; 22-25).

SLDR-TB may develop after starting anti-tuberculosis treatment (acquisition), or via transmission of resistant strains (43); accordingly, possible risk factors for SLDR-TB exist within

these mechanisms, especially acquisition. There is a large body of research that has specifically examined the associations of numerous physiological, behavioral, and socioeconomic characteristics with SLDR-TB. However, the literature addressing this issue differs in the outcome of interest (which may be a diagnosis of XDR-TB or pre-XDR-TB or resistance to specific second-line drugs), the risk factors that are examined, and the study period, geographic location, and population (36; 40-43; 46; 49-52; 54; 55; 57; 58; 72). Moreover, the direction of the association of these risk factors with these outcomes related to SLDR-TB, and their significance, was often discrepant across studies.

Some of these studies were fully or partially conducted in South Africa (36; 41; 72), a country with 304,000 incident cases of TB in 2021 (24), and one of the highest case counts of XDR-TB worldwide (65). Treatment for DR-TB in South Africa, however, has undergone many major changes just in the past 10 years, and many studies were either conducted entirely or partially before the introduction or more widespread use of newer regimens for DR-TB, such as the rollout of bedaquiline for all people with RR-TB starting in 2018 (8; 41). Not only did one study focus on risk factors to older second-line drugs not currently prioritized for use (41), but even studies that did focus on more current second-line drugs, such as bedaquiline or linezolid, may not have captured the effects of the major changes in DR-TB treatment in South Africa in the past several years (36; 72). For that reason, this study seeks to examine the association of certain risk factors with phenotypic resistance to second-line drugs among individuals in the province of KwaZulu-Natal, South Africa, over the more recent study period 2018-2022. Although the incidence of SLDR-TB is dominated by transmission in KwaZulu-Natal (67-71), and we hypothesize that these risk factors may be related to second-line drug resistance through acquisition, quantifying these

associations for all individuals with SLDR-TB is still important, as doing so provides a more complete picture of the determinants to second-line resistance in KwaZulu-Natal.

Methods

Data Source

Parent Study Population and Data Collection

The Role of Casual Contact and Migration in XDR Transmission in South Africa: A Geospatial, Genomic, and Social Network Study (CONTEXT) was conducted between 2019-2023 and aims to characterize the role of casual contact in the transmission of SLDR-TB in KwaZulu-Natal, South Africa. We screened all consecutive Mtb isolates undergoing either second-line line probe testing or phenotypic drug susceptibility testing (pDST) at Inkosi Albert Luthuli Central Hospital (IALCH), in Durban, KwaZulu-Natal. Patients whose isolates were resistant to either second-line injectable drugs, fluoroquinolones, or both were eligible for inclusion in the CONTEXT study. Isolates were collected from 2018-2022, and multiple isolates could be associated with one participant. Date of birth and sex were abstracted from the lab report for the first ("diagnostic") isolate screened for a participant. Routine pDST was performed at IALCH and/or the National Institute for Communicable Diseases in Johannesburg on cultured *M. tuberculosis* isolates using agar-based methods or the mycobacterial growth indicator tube (MGIT; Becton Dickinson) system for isoniazid, rifampicin, capreomycin, low- and high-level moxifloxacin, levofloxacin, linezolid, bedaquiline, and clofazimine, although results were not available for all drugs for each participant. Abstraction of pDST results and each isolate's date of collection was prioritized for isolates with whole genome sequencing complete (WGS), or the diagnostic isolate if WGS was not completed.

Participants who lived in one of four districts of KwaZulu-Natal (eThekwini, Ugu, uMgungundlovu, and iLembe) were contacted by study staff and consented to an interview in which they provided information on demographic and medical characteristics. This group of participants, the "eThekwini cohort," is a subset of the entire "provincial population" that was screened at IALCH. Any participant from this cohort who did not know their HIV status or who had tested negative more than three months before the interview was offered HIV testing. All data for the provincial population and eThekwini cohort were collected via case report forms and stored in a REDCap database via double data entry.

Variables of Interest

The outcome of interest in this analysis was phenotypic resistance or susceptibility to capreomycin, any fluoroquinolone (moxifloxacin or levofloxacin), bedaquiline, clofazimine, and linezolid, where participants with resistant results were classified as the index group. Sex, age, HIV status, frequency of alcohol consumption, and monthly household income per capita were the exposures of interest. We selected these exposures to examine if their associations with phenotypic resistance existed in the current context of SLDR-TB in KwaZulu-Natal, and because the literature identified potential pathways from these exposures to second-line drug resistance, which we hypothesized existed via acquisition of SLDR-TB. Discrepant associations were sometimes identified in the literature between the exposure, such as HIV and sex, and resistance, suggesting that these variables' relationship with second-line drug resistance was not definitive. We also included unemployment status, marital status, highest level of education received, and former incarceration in the previous twelve months as confounders between at least one exposure and the

outcome. We also included previous DR-TB treatment to be part of an interaction term with each exposure, as described below.

Because the parent study is cross-sectional, and the interview process for the eThekwini cohort occurred after collection of the sputum sample(s) for culture and pDST, we made assumptions to justify using certain variables as potential exposures or confounders for drug resistance. Participants were asked about their employment status within the past two years, which was assumed to be long enough to extend before the collection of the isolate with pDST. We assumed that education (73), marital status, HIV status, and alcohol consumption patterns were static over time, and that reporting incarceration within the past twelve months would not exclude individuals who had been incarcerated shortly before this cutoff or include individuals incarcerated after date of collection. Finally, income can fluctuate over time (73), so we categorized this variable, as participants may be less likely to experience fluctuations large enough to place one's income during the interview in a different category than when the isolate with pDST was collected.

Data Cleaning

All data cleaning and analysis was conducted in RStudio (version 2022.12.0, build 353). We extracted two datasets from the CONTEXT dataset that correspond to participants with any pDST results in the provincial population and eThekwini cohort, respectively. Individuals who did not consent, withdrew, were lost to follow-up from the cohort, or were found not to meet eligibility criteria were excluded. If data entry was incomplete for a participant, they were also excluded.

If a participant was interviewed, we used the date of birth they reported in their interview if this date was discordant with the date of birth listed in their lab report. Age was defined as the difference between the date of collection of the isolate with pDST results from the participant's

date of birth and was categorized into the following groups: 0-19, 20-34, 35-54, and 55 or more years.

We dichotomized monthly household income divided by the number of individuals that income supported using South Africa's per-capita monthly food poverty line in 2022 (R663/month)(74). We synthesized interview questions on whether a participant ever started TB treatment, if and when a participant started MDR-TB treatment, and if the participant had previously received treatment for XDR-TB or SLDR-TB to create a variable indicating previous treatment for DR-TB. Education was categorized as "Primary School or Less," "Secondary School (No Matric)," or "Matric or Higher." Alcohol consumption was categorized as "Never," "4 or Less Times per Month," and "More than Once per Week." Finally, for marital status, the "Single" and "Widowed" strata were combined, in addition to a "Married" category. Any strata including participants who did not know or were unable to answer were categorized as "Don't know."

Statistical Analysis

For both the provincial population and the eThekwini cohort, we calculated summary statistics of pDST results and demographic and medical characteristics for participants with at least one pDST result. We calculated unadjusted odds ratios for each exposure and for resistance to each second-line drug using binary logistic regression, although odds ratios could only be estimated for age and sex in the provincial population.

We assessed each of our models for evidence of collinearity and removed covariates as needed. We then assessed whether the exposures interacted with previous treatment for DR-TB if the interaction term was not removed during the collinearity assessment. We included this interaction term because we hypothesized that the pathway of each exposure to the outcome was

via acquisition of resistance. Because acquisition of resistance can occur during previous treatment for DR-TB (42; 43; 75), we predicted that the association of the main exposures and second-line drug resistance would differ for those who were previously treated for DR-TB and those who were not. In other words, because the pathway from the exposure to resistance through acquisition may not exist for individuals who were never treated for DR-TB, previous treatment could modify the association between each exposure and phenotypic second-line drug resistance.

For each model with an interaction term, we used the Likelihood Ratio Test and an alpha level of $\alpha = 0.05$ to assess whether inclusion of the interaction term was statistically significant. The interaction term was included in the final model if statistically significant. When the interaction assessment was completed, we calculated adjusted odds ratios for each model and identified p-values using the likelihood ratio test, with an alpha level of $\alpha = 0.05$ indicating statistical significance.

Ethical considerations

The CONTEXT study was approved by the institutional review boards (IRBs) of all institutions involved in the study, including Emory University and the University of KwaZulu-Natal. All participants who were interviewed in the eThekwini cohort provided full written informed consent to participate in the study. Consent was waived for members of the provincial population as data collection was restricted to accessing previously collected records and did not involve contact with the participants themselves.

Results

Characteristics of Provincial Population and eThekwini Cohort

A total of 654 participants had confirmed second-line drug resistance and were enrolled into the provincial population. Among these participants, 580 had at least one pDST result available and were included in the current analysis. The mean age of the study participants was 35.5 years (SD=12.7) and a majority (309/580, 53.3%) were male (Table 1a; Figure S1a). The distribution of pDST results for the provincial population are shown in Figure 1a. pDST results were most frequently available for bedaquiline (n=402) and were the least available for capreomycin (n=76). Among those with available phenotypic resistance test results, 60/76 were resistant to capreomycin (78.9%), 163/296 were resistant to any fluoroquinolone (55.1%), 54/402 were resistant to bedaquiline (13.4%), 50/346 were resistant to clofazimine (13.0%), and 5/389 were resistant to linezolid (1.3%). Most participants with pDST results for both bedaquiline and clofazimine had concordant results for both drugs (n=336, 98.8%); all four individuals with discordant results were resistant to bedaquiline and susceptible to clofazimine.

Among the 654 participants in the provincial population, 316 participants were in one of four subdistricts of KwaZulu-Natal covered by the eThekwini cohort. Among these participants, 257 participants were successfully contacted, met eligibility requirements, and consented to the study, of whom 234 participants had complete interview and HIV test records. Of these 234 participants in the eThekwini cohort, 189 (81%) had at least one pDST result; the distribution of pDST results is shown in Figure 1b. Approximately half of these participants were male (n=99, 52.4%) and the mean age was 36.7 years (SD=10.9) (Table 1b, Figure S1b). Ninety-eight participants (51.9%) did not initiate any treatment for MDR-TB or SLDR-TB before the collection of the isolate with pDST results. Most participants reported that they had a positive HIV test

(n=139, 73.5%). When asked about alcohol use, 122 participants (64.6%) reported that they never drank alcoholic beverages (Table 1b). The median monthly household income per capita was R400 (IQR: R250, R667), and its distribution was skewed to the right (Table 1b; Figures S2a-b).

Among participants in the eThekwini cohort, pDST results were most frequently available for bedaquiline (n=157), while capreomycin (n=12) was the least frequently tested. All participants whose isolates underwent pDST for capreomycin were resistant (n=12, 100%), 39/80 participants were resistant to any fluoroquinolone (48.8%), 25/157 participants were resistant to bedaquiline (15.9%), 20/141 participants were resistant to clofazimine (14.2%), and 3/138 participants were resistant to linezolid (2.2%) (Table 1d). Discordant results for bedaquiline and clofazimine among individuals with results for both drugs (n=139) were rare (n=2, 1.4%).

Unadjusted Analyses

Unadjusted odds ratios were calculated for each exposure (sex, age, alcohol consumption, income, and HIV status) and phenotypic resistance to capreomycin, fluoroquinolones, linezolid, bedaquiline, and clofazimine. Because age and sex were available for the provincial population and eThekwini cohort, unadjusted odds ratios were calculated for both groups, while odds ratios for alcohol consumption, income, and HIV status could only be calculated for the eThekwini cohort. No unadjusted or adjusted analysis was completed for capreomycin in the eThekwini cohort, as none of the individuals in this group had isolates which were susceptible.

Tables 2a and 2b display all odds ratios and 95% confidence intervals (provincial population [PP] and eThekwini cohort [EC]). We found no significant association between female sex and resistance to any second-line drugs, including capreomycin (OR_{PP}=0.64, 95% CI: 0.21, 1.99), fluoroquinolones (OR_{PP} = 1.33, 95% CI: 0.84, 2.11; OR_{EC} = 1.0, 95% CI: 0.41, 2.40),

linezolid (OR_{PP} = 1.74, 95% CI: 0.29, 10.51; OR_{EC} = 0.53, 95% CI: 0.46, 51.87), bedaquiline (OR_{PP} = 1.1, 95% CI: 0.62, 1.96; OR_{EC} = 0.95, 95% CI: 0.40, 2.24), and clofazimine (OR_{PP} = 1.22, 95% CI: 0.65, 2.28; OR_{EC} = 1.2, 95% CI: 0.47, 3.09). When odds ratios could be calculated and there were no strata of age that only included resistant or susceptible individuals, individuals aged 0-19 did not have significantly different odds of resistance to capreomycin, fluoroquinolones, linezolid, bedaquiline, or clofazimine compared to individuals aged 20-34 years, 35-54 years, and 55 or more years, for the provincial population or the eThekwini cohort.

Among individuals above South Africa's per-capita monthly food poverty line in 2022 (R663), the odds of resistance to fluoroquinolones (OR_{EC} =2.46, 95% CI: 0.84, 7.20), linezolid (OR_{EC} =5.7, 95% CI: 0.50, 64.90), bedaquiline (OR_{EC} =1.17, 95% CI: 0.44, 3.06), and clofazimine (OR_{EC} = 0.92, 95% CI: 0.31, 2.76) were not significantly different from the odds of resistance to these drugs for individuals below the poverty line. When comparing individuals with and without HIV, we found no significant difference in the odds of having resistance to fluoroquinolones (OR=1.10, 95% CI: 0.39, 3.06), linezolid (OR=0.59, 95% CI: 0.05, 6.69), bedaquiline (OR=2.36, 95% CI: 0.66, 8.43), and clofazimine (OR=3.10, 95% CI: 0.68, 14.2). Compared to participants who never drank alcohol, participants who drank four or less times a month and more than once a week did not have significantly different odds of resistance to any of the four drugs under consideration.

Adjusted Analyses

We then analyzed these associations adjusted for potential confounders. Before calculating adjusted odds ratios, we conducted a collinearity assessment for each adjusted model, which required collinear covariates, including the interaction term with previous DR-TB treatment in

some cases, to be removed. After removal of these covariates, we conducted collinearity assessments and removed variables with severe collinearity until there were no collinearity issues present. At the conclusion of this process, all models besides the model assessing the association between age and resistance to linezolid had no serious collinearity issues. For this exception, the only variable with collinearity issues was the main exposure, age; given that it would be inappropriate to drop the main exposure, no changes were made to the model.

Next, we performed an interaction assessment if the interaction term was not removed during the collinearity assessment. The inclusion of an interaction term between previous treatment for DR-TB and any of the exposures was not statistically significant for any second-line drugs. We therefore dropped the interaction term from all final adjusted models.

Results of adjusted analyses based on these final models for the eThekwini cohort are shown in Tables 3a-3d. The odds of resistance among female participants compared to male participants were not significant for fluoroquinolones (OR=1.04, 95% CI: 0.43, 2.53), linezolid (OR=0.53, 95% CI: 0.05, 6.13), bedaquiline (OR=0.39; 95% CI: 0.07, 2.14) and clofazimine (OR=1.25, 95% CI: 0.46, 3.38). We found no significant difference in the odds of resistance for participants above and below the monthly food poverty line for all four drugs (fluoroquinolones: OR=2.85, [95% CI: 0.74, 10.92]; linezolid: OR=5.45 [95% CI: 0.48, 62.50]; bedaquiline: OR=1.0 [95% CI: 0.34, 2.97]; clofazimine: OR=0.58 [95% CI: 0.15, 2.23]). The odds of resistance for participants with HIV were also not significantly different compared to those without HIV (fluoroquinolones: OR=0.93 [95% CI: 0.25, 3.48]; linezolid: OR=0.89 [95% CI: 0.07, 11.69]; bedaquiline: OR=3.32 [95% CI: 0.79, 13.95]; clofazimine: OR=4.24 [95% CI: 0.75, 23.93]) Finally, we did not observe a significant association between alcohol use and resistance to the four

drugs under consideration, nor did we find any significant association between age and second-line drug resistance.

Discussion

In this study, we attempted to quantify the association of sex, age, HIV status, alcohol use, and income with phenotypic resistance to capreomycin, fluoroquinolones, linezolid, bedaquiline, and clofazimine in KwaZulu-Natal, South Africa. This question is important to answer, as other studies have not examined the risk factors for second-line drug resistance since the recent implementation of major changes to the second-line regimens used to treat DR-TB in South Africa (8; 16). However, we did not find any significant association between our exposures of interest and second-line drug resistance. These findings were surprising. Because these exposures of interest have been identified by previous studies as possible risk factors for unsuccessful outcomes of treatment for DR-TB (76-84), which, itself, can lead to acquisition of second-line resistance (50), we had hypothesized that these risk factors would have an effect on the occurrence of resistance among study participants through acquisition.

We acknowledge that, despite our hypothesis these exposures may be related to second-line resistance through acquisition, the incidence of SLDR-TB in KwaZulu-Natal is mainly driven by transmission, not acquisition (67-71). We therefore recognize that we may not have observed significant associations in this study because of the important role that transmission plays for SLDR-TB in KwaZulu-Natal. Nevertheless, we justify examining the associations of interest for all participants with SLDR-TB instead of just those who may have acquired resistance for multiple reasons. First, although self-reported previous treatment for DR-TB was available, an individual having undergone treatment for DR-TB in the past does not necessarily indicate whether their

episode of SLDR-TB at the time of the study is a result of acquisition, rather than transmission. Second, while we hypothesized that acquisition was the pathway between the exposures of interest and second-line resistance based on what was available from the literature, we could not definitively assume that these variables are not associated with SLDR-TB via transmission. In fact, if we had excluded participants whose SLDR-TB is due to transmission, any question about the overall association between these risk factors and resistance to second-line drug resistance in KwaZulu-Natal would have been left partially unanswered.

The literature on risk factors for SLDR-TB or for phenotypic resistance to specific second-line drugs varies considerably. Risk factors (e.g., alcohol use) and outcomes of interest (e.g., XDR-TB, or resistance to a specific second-line drug) differed across studies. Moreover, the associations identified within the literature are inconsistent. Like our analysis, some studies were unable to find an association between several exposures (e.g., sex, age, HIV status, and alcohol use), with second-line drug resistance (36; 41; 56; 72). However, other studies did identify a significant association between these variables and second-line drug resistance (41; 42; 46; 49-52; 55; 57; 58). Finally, although some studies were conducted in South Africa (36; 41; 72), others were conducted elsewhere, including in the United States (42; 46; 57), Eastern Europe (50; 55), and Pakistan (49).

There are several possible explanations for the discordance between findings in the literature and our study. First, many of these studies were conducted in an earlier time period than our study or they focused on resistance to second-line drugs that are not routinely used in treatment of DR-TB. For instance, several studies examined resistance to the fluoroquinolones ciprofloxacin and ofloxacin rather than the currently recommended moxifloxacin and levofloxacin, which were examined in this study (40-42; 85). Within South Africa and globally, there have been major changes in treatment regimens for DR-TB (8; 14; 16). These newer treatments are generally

provided for a shorter period of time, are safer, and are more successful than older regimens (14; 15; 17; 18; 85). The effect of risk factors that may be related to acquisition of second-line resistance because of unfavorable DR-TB treatment outcomes may be weaker in settings like this study where newer regimens with better outcomes are in greater use. Second, most studies that observed significant associations had larger sample sizes, with exceptions (52; 54), and were possibly better powered. Third, our study was unique in that the entire study population had an SLDR-TB diagnosis, while other studies compared individuals with second-line resistance to individuals with MDR-TB or drug-sensitive TB (DS-TB) (41; 42; 46; 49-52; 55; 57; 58). If certain risk factors are related to second-line drug resistance, then individuals with any second-line resistance may share more similar characteristics with one another than with other groups, such as individuals who have DS-TB, which could result in clearer contrasts between these different comparison groups. Finally, our methodology differed from other studies, some of which only identified risk factors using unadjusted analyses (41; 46; 51; 54; 57). Other studies adjusted estimates on covariates that were included in the model on the basis of p-value thresholds from the unadjusted analyses (42; 58), even though this strategy alone may potentially introduce bias (86-88), and only one study included confounders on an *a priori* basis as we did (42). These differences in methodology may also account for our discordant results.

This study has several limitations. First, sample size was limited because pDST was not done for every drug and for every participant, as testing priorities may have changed over time (89), and many variables were only available for the smaller eThekwini cohort. Second, the nature of the available data may have also impacted our results. We hypothesized that heavy alcohol consumption would be associated with second-line drug resistance (50; 55), but only 7.4% of the study population consumed alcohol more than once a week (Table 1b). Although we predicted that

low income would be associated with resistance, and income was dichotomized by a meaningful socioeconomic threshold (the food poverty line), this contrast in income status may not have been sensitive enough to detect an association. Additionally, certain strata of our exposures and outcomes, like resistance to linezolid, were uncommon, which may have contributed to very high standard error in certain analyses or prevented us from quantifying certain associations. Third, bias, such as confounding not predicted by our DAG, differential selection of participants by exposure and availability of pDST results, issues of recall, and misclassification due to social desirability bias surrounding potentially sensitive information (90; 91) may have affected our adjusted associations.

Another important limitation concerns the study data's cross-sectional nature. We made assumptions to use certain variables as exposures or confounders when their temporality with respect to the outcome was uncertain. These assumptions were reasonable for relatively static variables or where interview questions established that the exposure occurred before the outcome, but were harder to make for variables without clear temporality or that fluctuate. We also used a ranking method to prioritize removal of covariates when we found collinearity, which is admittedly subjective but also avoids other discouraged techniques, including dropping terms that are not significant from the model (88). In addition, we did not consider interactions between variables besides the exposure and previous DR-TB treatment because this was not this study's primary interest and this ran the risk of collinearity issues (88). Finally, our findings may not be generalizable beyond South Africa. Because South Africa has a high burden of DR-TB and SLDR-TB compared to other countries (3; 24; 65; 92), and transmission plays an important role for SLDR-TB in KwaZulu-Natal (67-71), it may be inappropriate to apply this study's findings to a low-burden setting or one where acquired resistance is more dominant.

Identifying risk factors can guide interventions that seek to improve the well-being of individuals who are disproportionately likely to be impacted by second-line drug resistance. The exposures in this study did not have an association with resistance to second-line drugs, particularly given our emphasis on the role of acquisition of resistance and the importance of transmission for SLDR-TB in KwaZulu-Natal. Still, we have filled a gap in knowledge about the association of second-line drug resistance with various participant characteristics in KwaZulu-Natal following a change in guidance about recommended treatment of DR-TB. Future studies in KwaZulu-Natal and elsewhere examining possible risk factors for second-line drug resistance could explore resistance to other second-line drugs used in newer regimens (e.g., pretomanid), stratify associations with risk factors among well-defined groups of individuals who have SLDR-TB due to acquisition or transmission, and use analytic and study design strategies to reduce the impact of bias and random error.

Chapter III: Public Health Implications

In this study, we aimed to identify potential risk factors for resistance to second-line drugs used to treat DR-TB. Examining the determinants of second-line drug resistance is important, because such work can not only uncover associations with drug resistance that have not been observed but can also inform interventions that specifically address the needs of at-risk groups (46). This is especially important in the context of diseases with high mortality like DR-TB. After all, despite the introduction of more effective regimens for DR-TB (8; 14; 17; 18), successful treatment outcomes are still less common for individuals with XDR-TB than for individuals with DS-TB in South Africa (24). Moreover, the treatment of DR-TB constitutes a considerable proportion of South Africa's resources dedicated to addressing TB (93). Thus, if risk factors are identified, then they could serve as the basis for interventions that could address the substantial burden of disease caused by SLDR-TB.

Surprisingly, we did not find any variables that were associated with resistance to any second-line drugs. Nevertheless, this study still has a place in the literature and can inform public health efforts against TB, either in research or elsewhere. First, in this study, the estimates of each association were not significant and were usually not precise. However, while imprecise estimates may not appear to be "meaningful" for an individual study, some scholars have stressed that they are still important to report; compiling many imprecise estimates of an association from the literature over time, including via a meta-analysis, can produce a more precise value that may be more meaningful, assuming that bias was addressed (94). The process of identifying an effect is often not immediate but instead iterative (94). Thus, this study is still contributing to the larger body of research on risk factors for second-line drug resistance, and hopefully can contribute to

future efforts to identify who is most at risk for this outcome, and thus serve as a basis for intervention.

Second, this study is important in that it highlights what is left unknown about risk factors for second-line drug resistance in KwaZulu-Natal. For instance, fewer variables were collected for participants outside of the eThekwini cohort. While we could conduct unadjusted analyses for age and sex in the provincial population, analyses involving other participant characteristics, such as unadjusted and adjusted analyses for the association of alcohol use, income, and HIV status, could not be done for the entire provincial population. Next, this study compared individuals with and without resistance to specific second-line drugs, but all participants had SLDR-TB. Because all participants had SLDR-TB, this study cannot answer whether there are significant differences in the distribution of certain variables among individuals with resistance to second-line drugs and individuals who do not have SLDR-TB, such as those with MDR-TB or DS-TB. Furthermore, we did not have phenotypic resistance data for other second-line drugs that are becoming increasingly important for treating DR-TB, like pretomanid (9; 17) Finally, this study has several additional limitations, including possible uncontrolled bias, and that data on other variables were not collected prospectively with respect to pDST results in this study.

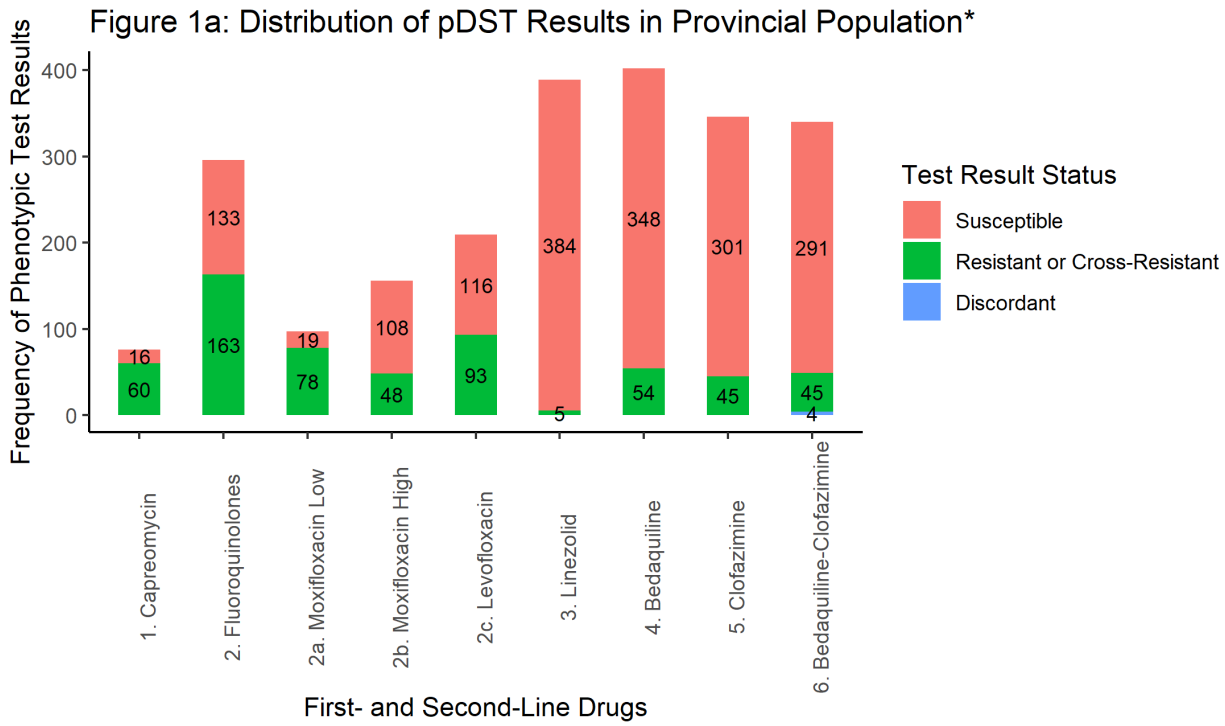
Although we did not identify any associations between any of the risk factors and drug resistance, our study findings are not definitive, and future studies can build on this work and further investigate risk factors for second-line drug resistance in KwaZulu-Natal. As noted above, identifying potential determinants of second-line drug resistance is important for addressing the burden of DR-TB and improving the well-being of those who are affected by it. Such a study could collect data on the main exposures of interest and their potential confounders for the entire study population and ensure consistent pDST for the drugs of interest for all participants. In addition to

controlling for confounding as done in this study, future studies could further address bias in their study design, such as introducing validation sub-studies that correct for potential misclassification of variables collected during interviews. For instance, if there were concerns about social desirability bias affecting self-reports of alcohol use, future studies could adjust this variable with sensitivity and specificity estimates from studies conducted in similar populations (90). Next, future studies could stratify analyses of risk factors among individuals whose resistance to second-line drugs is attributed to acquisition or transmission. For instance, previous studies have used pDST at multiple timepoints to determine whether second-line resistance was acquired (42; 43). Combined with genomic and epidemiologic data that can be used to determine whether transmission of SLDR-TB occurred, this could allow researchers to differentiate between acquired and transmitted cases and thus stratify their analyses on this status.

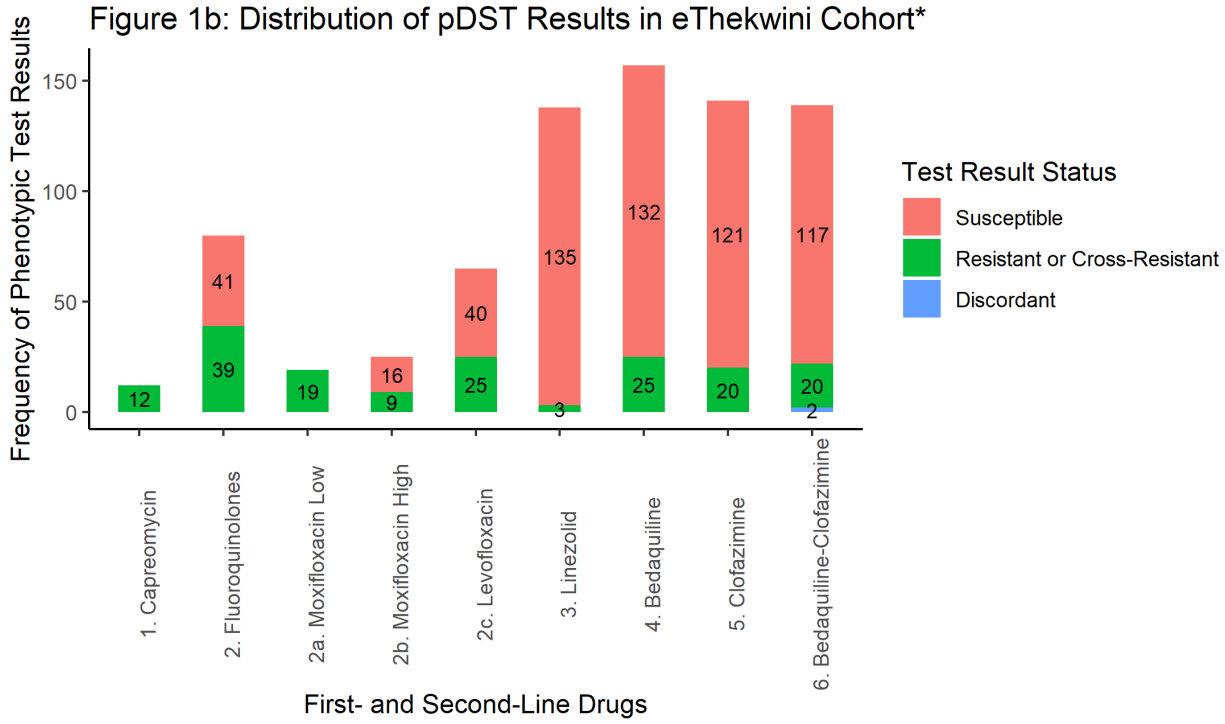
Future studies could also collect data prospectively, such as examining medical records for data before diagnosis with SLDR-TB, or use different comparison groups (e.g., individuals with MDR-TB and SLDR-TB) to identify estimates of variables that may influence resistance to second-line drugs. Finally, if available, future studies could examine risk factors for phenotypic resistance to newer drugs like pretomanid. All these strategies would help to identify the determinants of second-line resistance. Doing do could help contribute to the global struggle against TB, and to efforts to improve the well-being of the people and communities who are affected by this widespread and severe disease.

Figures and Tables

Descriptive Statistics



*Among those with available pDST results.



*Among those with available pDST results.

Table 1a: Distribution of Study Population Characteristics across Provincial Population (n=580)

	Overall (N=580)
Age (years)	
Mean (SD)	35.5 (12.7)
Median (Q1, Q3)	35.0 (28.0, 42.0)
Missing	3 (0.5%)
Sex	
Male	309 (53.3%)
Female	271 (46.7%)

Table 1b: Distribution of Study Population Characteristics across eThekwini Cohort (n=189)

	Overall (N=189)
Age (years)	
Mean (SD)	36.7 (10.9)
Median (Q1, Q3)	35.0 (30.0, 42.0)
Sex	
Male	99 (52.4%)
Female	90 (47.6%)
Previous Treatment for DR-TB	
No Previous DR-TB Treatment	98 (51.9%)
Previous DR-TB Treatment	70 (37.0%)
Unknown	21 (11.1%)
HIV Serostatus	
Seronegative	41 (21.7%)
Seropositive	139 (73.5%)
Indeterminate/Unknown	9 (4.8%)
Frequency of Alcohol Consumption	
Never	122 (64.6%)
Four or Less per Month	50 (26.5%)
More Than Once a Week	14 (7.4%)
Don't Know	3 (1.6%)
Marital Status	
Single	154 (81.5%)
Cohabitating/Married	35 (18.5%)
Don't Know	0 (0%)
Monthly Household Income per Capita (Rand)	
Mean (SD)	728 (1930)
Median (Q1, Q3)	400 (250, 667)
Missing	11 (5.8%)
Highest Level of Education	
Primary School or Less	32 (16.9%)
Secondary School (No Matric)	101 (53.4%)
Matric or Higher	55 (29.1%)
Don't Know	0 (0%)
Missing	1 (0.5%)
Employment in Previous Two Years	
Not Employed in Past Two Years	103 (54.5%)
Employed in Past Two Years	74 (39.2%)
Unknown	12 (6.3%)
Incarcerated in Previous Year	
Not Incarcerated in Past Year	182 (96.3%)
Incarcerated in Past Year	7 (3.7%)
Don't Know	0 (0%)
Refused to Answer	0 (0%)

Table 2a: Unadjusted Odds Ratios for Resistance to Second-Line Drugs of Interest in the Provincial Population

Outcome	*Sample Size*	*Exposure*	*Index Group*	*Referent Group*	*Odds Ratio*	*95% Confidence Interval*	*p-value*
Capreomycin	76	Sex	Female	Male	0.64	(0.21, 1.99)	0.44
		Age	20-34	0-19	-	-	0.27
			35-54	0-19	-	-	
			55+	0-19	-	-	
Fluoroquinolones	296	Sex	Female	Male	1.33	(0.84, 2.11)	0.22
		Age	20-34	0-19	1.36	(0.60, 3.12)	0.38
			35-54	0-19	0.91	(0.40, 2.07)	
			55+	0-19	0.78	(0.24, 2.51)	
Linezolid	389	Sex	Female	Male	1.74	(0.29, 10.51)	0.54
		Age	20-34	0-19	-	-	0.6
			35-54	0-19	-	-	
			55+	0-19	-	-	
Bedaquiline	402	Sex	Female	Male	1.1	(0.62, 1.96)	0.74
		Age	20-34	0-19	1.73	(0.38, 7.87)	0.71
			35-54	0-19	2.13	(0.48, 9.51)	
			55+	0-19	1.78	(0.30, 10.59)	
Clofazimine	389	Sex	Female	Male	1.22	(0.65, 2.28)	0.54
		Age	20-34	0-19	3.1	(0.39, 24.46)	0.39
			35-54	0-19	3.73	(0.48, 29.07)	
			55+	0-19	1.83	(0.15, 21.63)	

Note: a dash indicates that at least one stratum of the exposure (either the index or referent group) did not include any susceptible or resistant individuals, and an odds ratio thus could not be calculated.

Table 2b: Unadjusted Odds Ratios for Resistance to Second-Line Drugs of Interest in eThekwini Cohort

Outcome	*Exposure*	*Index Group*	*Referent Group*	*Odds Ratio*	*95% Confidence Interval*	*p-value*
Fluoroquinolones	Sex	Female	Male	1	(0.41, 2.40)	1
	Age	20-34	0-19	1.6	(0.23, 10.94)	0.43
		35-54	0-19	1.16	(0.17, 7.73)	
		55+	0-19	6	(0.35, 101.57)	
	Income	Above food poverty line	Below food poverty line	2.46	(0.84, 7.20)	0.09
	Alcohol	Drink 4 or less times/month	Never drink alcohol	1.18	(0.43, 3.25)	0.2
		More than once a week	Never drink alcohol	0.89	(0.18, 4.39)	
		Don't know	Never drink alcohol	-	-	
	HIV	Seropositive	Seronegative	1.1	(0.39, 3.06)	0.33
		Indeterminate/Unknown	Seronegative	4.89	(0.46, 51.87)	
Linezolid	Sex	Female	Male	0.53	(0.05, 5.90)	0.59
	Age	20-34	0-19	-	-	0.62
		35-54	0-19	-	-	
		55+	0-19	-	-	
	Income	Above food poverty line	Below food poverty line	5.7	(0.50, 64.90)	0.15
	Alcohol	Drink 4 or less times/month	Never drink alcohol	-	-	0.47
		More than once a week	Never drink alcohol	-	-	
		Don't know	Never drink alcohol	-	-	
	HIV	Seropositive	Seronegative	0.59	(0.05, 6.69)	0.78
		Indeterminate/Unknown	Seronegative	-	-	
Bedaquiline	Sex	Female	Male	0.95	(0.40, 2.24)	0.91
	Age	20-34	0-19	1.15	(0.12, 10.65)	0.82
		35-54	0-19	1.26	(0.14, 11.35)	
		55+	0-19	0.5	(0.03, 9.46)	
	Income	Above food poverty line	Below food poverty line	1.17	(0.45, 3.06)	0.76
	Alcohol	Drink 4 or less times/month	Never drink alcohol	1.48	(0.60, 3.68)	0.17
		More than once a week	Never drink alcohol	-	-	
		Don't know	Never drink alcohol	-	-	
	HIV	Seropositive	Seronegative	2.36	(0.66, 8.43)	0.36
		Indeterminate/Unknown	Seronegative	2.13	(0.18, 24.76)	
Clofazimine	Sex	Female	Male	1.2	(0.47, 3.09)	0.71
	Age	20-34	0-19	0.6	(0.06, 6.11)	0.35
		35-54	0-19	0.8	(0.08, 7.86	
		55+	0-19	-	-	
	Income	Above food poverty line	Below food poverty line	0.92	(0.31, 2.76)	0.88
	Alcohol	Drink 4 or less times/month	Never drink alcohol	1.88	(0.71, 4.98)	0.14
		More than once a week	Never drink alcohol	-	-	
		Don't know	Never drink alcohol	-	-	
	HIV	Seropositive	Seronegative	3.1	(0.68, 14.2)	0.26
		Indeterminate/Unknown	Seronegative	3.1	(0.23, 40.89)	

Note: a dash indicates that at least one stratum of the exposure (either the index or referent group) did not include any susceptible or resistant individuals, and an odds ratio thus could not be calculated.

Table 3a: Adjusted Odds Ratios for Resistance to Any Fluoroquinolone in eThekwini Cohort (n=80)

Exposure	*Confounders in Model*	*Index Group*	*Referent Group*	*OR (Point Estimate)*	*95% CI*	*P-value*
HIV Serostatus	Sex, incarceration in past year, employment in past two years, education, alcohol use, age, marital status, previous DR-TB treatment	Seropositive	Seronegative	0.93	(0.25, 3.48)	0.83
		Indeterminate/Unknown	Seronegative	2.23	(0.13, 38.69)	
Alcohol consumption	Age, sex, education, incarceration in past year, household income per capita, previous DR-TB treatment	Drink 4 or less times/month	Never drink alcohol	1.12	(0.35, 3.55)	0.39
		More than once a week	Never drink alcohol	0.81	(0.10, 6.49)	
		Don't know	Never drink alcohol	-	-	
Income per capita	Employment in past two years, age, sex, education, previous DR-TB treatment	Above food poverty line	Below food poverty line	2.85	(0.74, 10.92)	0.12
Age	Previous DR-TB treatment	20-34	0-19	1.68	(0.23, 12.19)	0.47
		35-54	0-19	1.23	(0.18, 8.54)	
		55+	0-19	5.88	(0.34, 101.05)	
Sex	Previous DR-TB treatment	Female	Male	1.04	(0.43, 2.53)	0.94

Note: a dash indicates that at least one stratum of the exposure (either the index or referent group) did not include any susceptible or resistant individuals, and an odds ratio thus could not be calculated.

Table 3b: Adjusted Odds Ratios for Resistance to Linezolid in eThekwini Cohort (n=138)

Exposure	*Confounders in Model*	*Index Group*	*Referent Group*	*OR (Point Estimate)*	*95% CI*	*P-value*
HIV Serostatus	Sex, alcohol use	Seropositive	Seronegative	0.89	(0.07, 11.69)	0.83
		Indeterminate/Unknown	Seronegative	-	-	
Alcohol consumption	Sex	Drink 4 or less times/month	Never drink alcohol	-	-	0.41
		More than once a week	Never drink alcohol	-	-	
		Don't know	Never drink alcohol	-	-	
Income per capita	Employment in past two years	Above food poverty line	Below food poverty line	5.45	(0.48, 62.50)	0.16
Age	Previous DR-TB treatment	20-34	0-19	-	-	0.54
		35-54	0-19	-	-	
		55+	0-19	-	-	
Sex	Previous DR-TB treatment	Female	Male	0.53	(0.05, 6.13)	0.6

Note: a dash indicates that at least one stratum of the exposure (either the index or referent group) did not include any susceptible or resistant individuals, and an odds ratio thus could not be calculated.

Table 3c: Adjusted Odds Ratios for Resistance to Bedaquiline in eThekwini Cohort (n=157)

Exposure	*Confounders in Model*	*Index Group*	*Referent Group*	*OR (Point Estimate)*	*95% CI*	*P-value*
HIV Serostatus	Sex, incarceration in past year, employment in past two years, education, alcohol use, age, marital status, previous DR-TB treatment	Seropositive	Seronegative	3.32	(0.79, 13.95)	0.18
		Indeterminate/Unknown	Seronegative	6.05	(0.32, 114.28)	
Alcohol consumption	Age, sex, education, incarceration in past year, household income per capita, previous DR-TB treatment	Drink 4 or less times/month	Never drink alcohol	1.26	(0.44, 3.64)	0.34
		More than once a week	Never drink alcohol	-	-	
		Don't know	Never drink alcohol	-	-	
Income per capita	Employment in past two years, age, sex, education, previous DR-TB treatment	Above food poverty line	Below food poverty line	1	(0.34, 2.97)	1
Age	Previous DR-TB treatment	20-34	0-19	0.9	(0.09, 9.07)	0.87
		35-54	0-19	0.81	(0.08, 8.03)	
		55+	0-19	0.38	(0.02, 7.87)	
Sex	Previous DR-TB treatment	Female	Male	0.95	(0.39, 2.32)	0.9

Note: a dash indicates that at least one stratum of the exposure (either the index or referent group) did not include any susceptible or resistant individuals, and an odds ratio thus could not be calculated.

Table 3d: Adjusted Odds Ratios for Resistance to Clofazimine in eThekwini Cohort (n=141)

Exposure	*Confounders in Model*	*Index Group*	*Referent Group*	*OR (Point Estimate)*	*95% CI*	*P-value*
HIV Serostatus	Sex, incarceration in past year, employment in past two years, education, alcohol use, age, marital status, previous DR-TB treatment	Seropositive	Seronegative	4.24	(0.75, 23.93)	0.11
		Indeterminate/Unknown	Seronegative	21.63	(0.52, 897.27)	
Alcohol consumption	Age, sex, education, incarceration in past year, household income per capita, previous DR-TB treatment	Drink 4 or less times/month	Never drink alcohol	1.71	(0.53, 5.54)	0.53
		More than once a week	Never drink alcohol	-	-	
		Don't know	Never drink alcohol	-	-	
Income per capita	Employment in past two years, age, sex, education, previous DR-TB treatment	Above food poverty line	Below food poverty line	0.58	(0.15, 2.23)	0.42
Age	Previous DR-TB treatment	20-34	0-19	0.47	(0.04, 5.52)	0.46
		35-54	0-19	0.52	(0.05, 5.94)	
		55+	0-19	-	-	
Sex	Previous DR-TB treatment	Female	Male	1.25	(0.46, 3.38)	0.66

Note: a dash indicates that at least one stratum of the exposure (either the index or referent group) did not include any susceptible or resistant individuals, and an odds ratio thus could not be calculated.

Appendix I: Supplementary Figures and Tables

Figure S1a: Distribution of Age in Provincial Population (n=580)

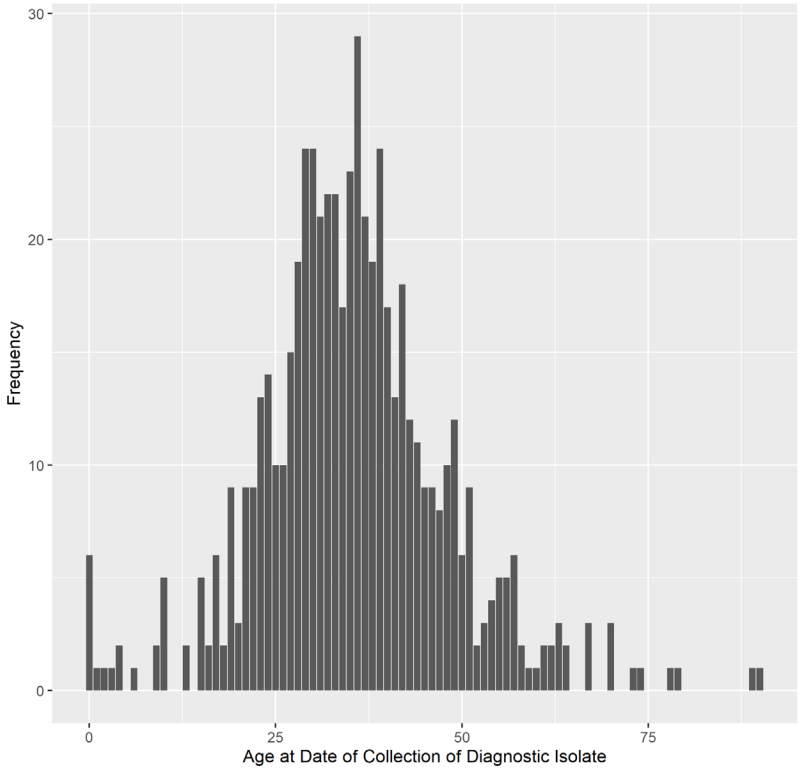


Figure S1b: Distribution of Age in eThewini Cohort (n=189)

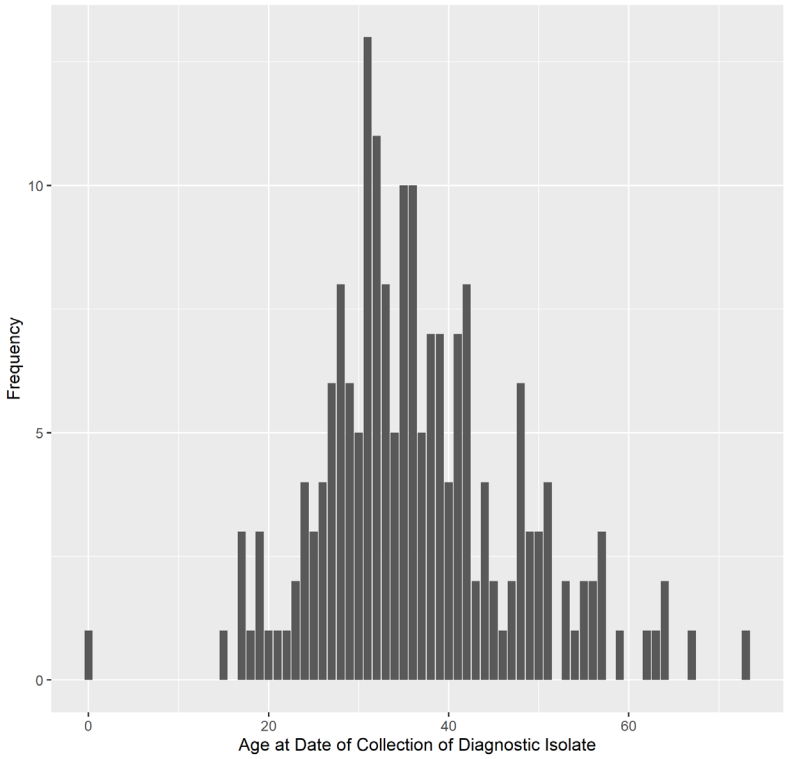


Figure S2a: Distribution of Monthly Household Income Per Capita

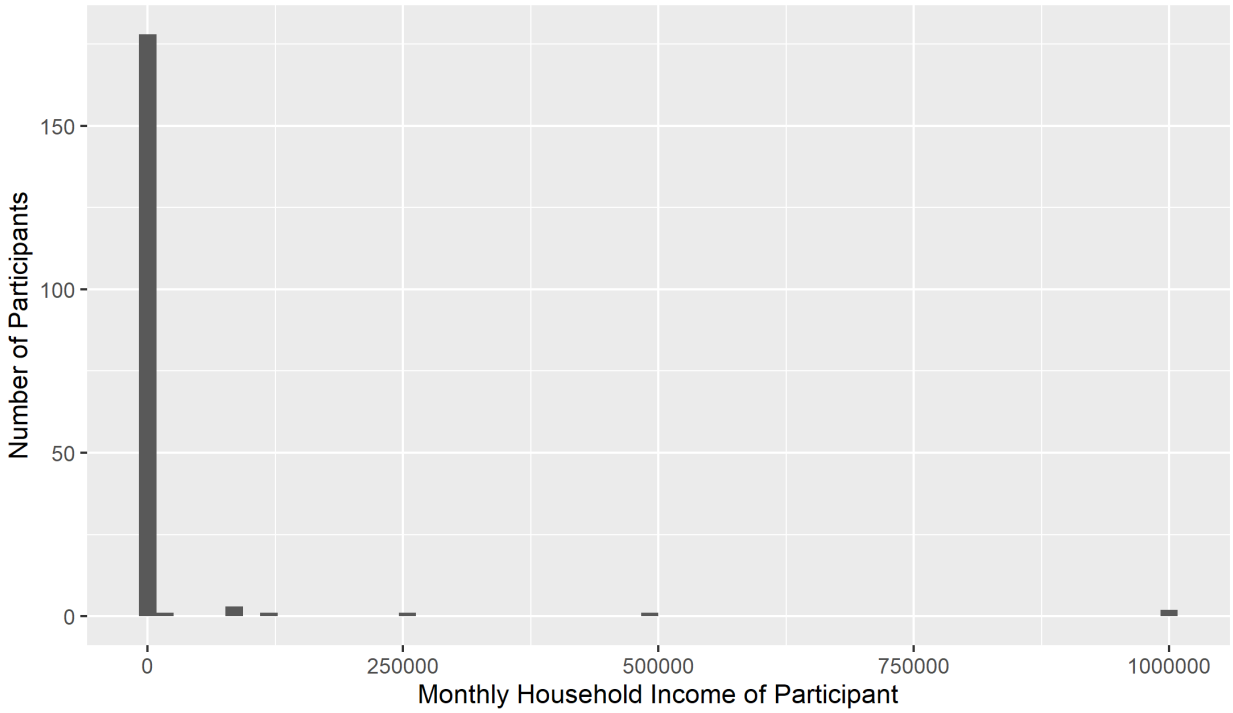
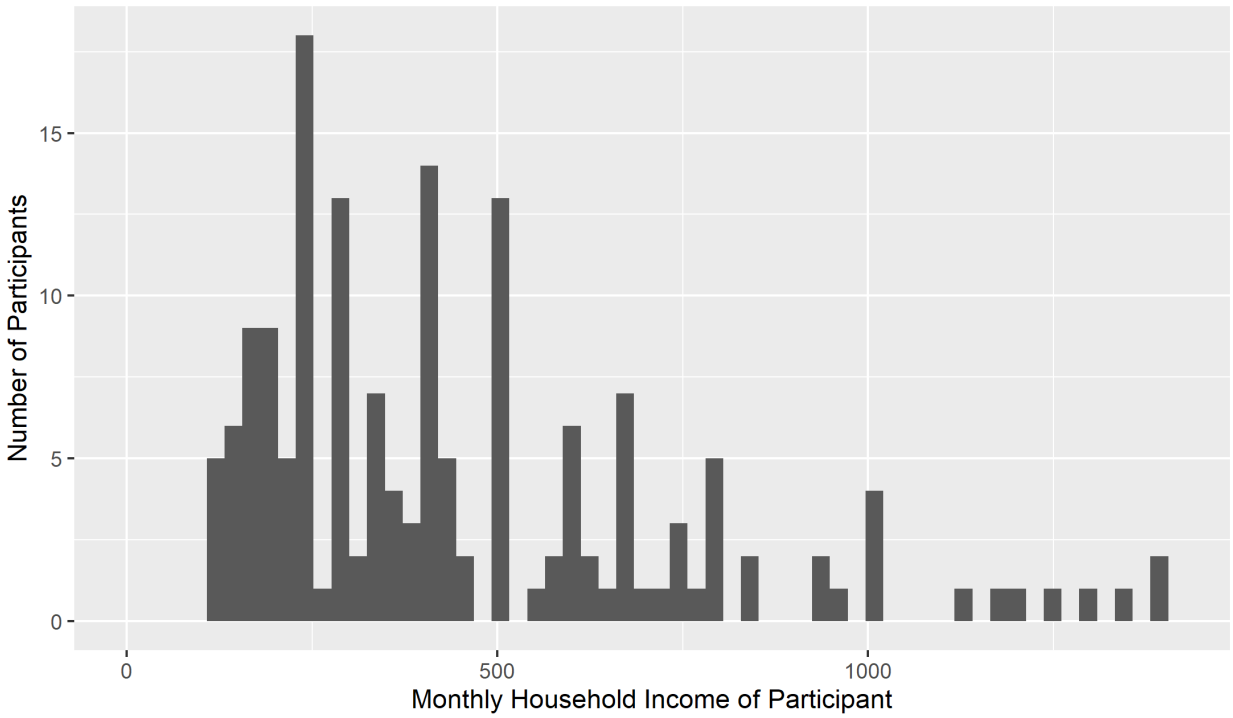
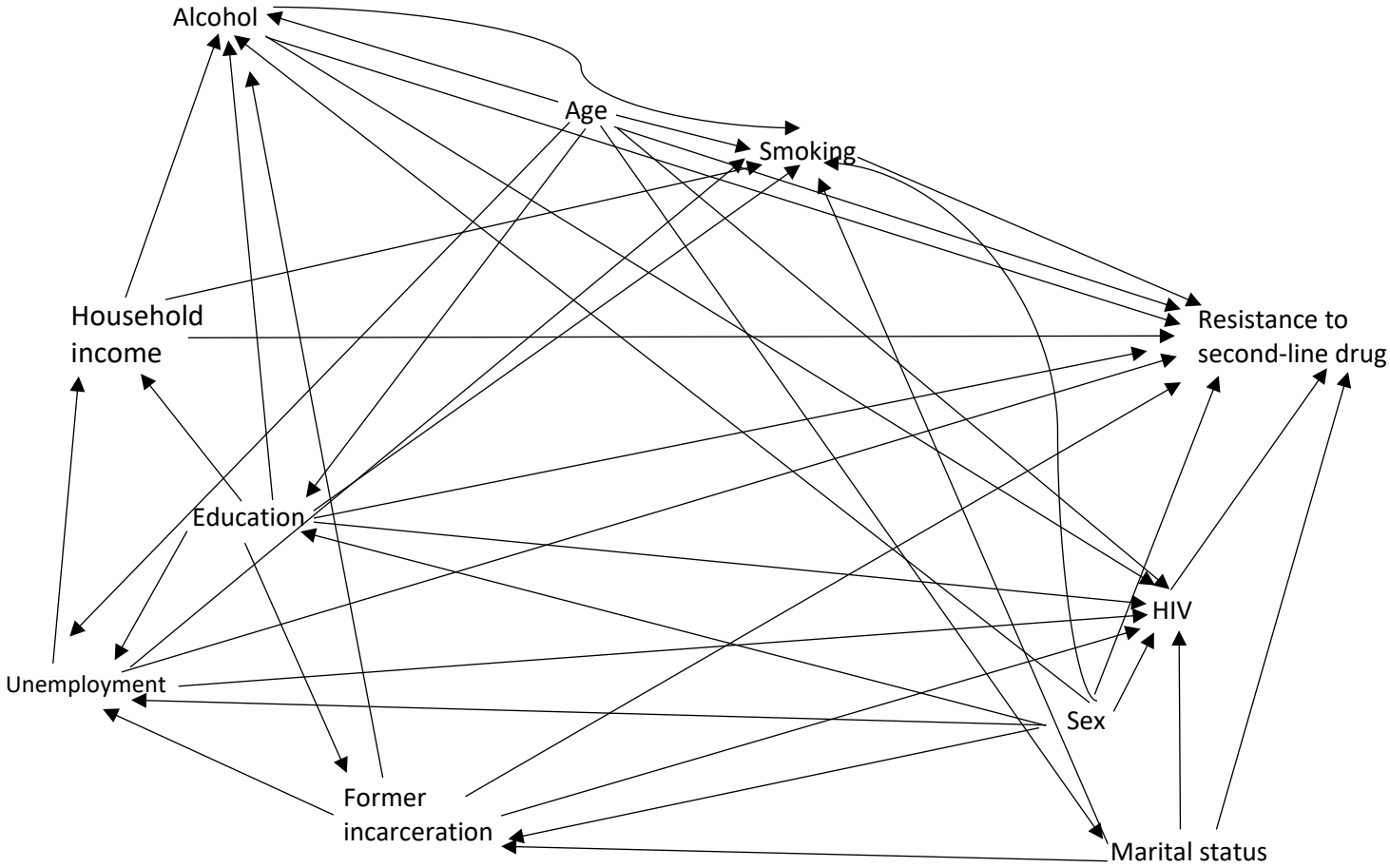


Figure S2b: Distribution of Monthly Household Income Per Capita Below Upper Poverty Line



Because the distribution of income was so skewed for this population, we included this second histogram (Figure 3b) that shows a more refined picture of the distribution of incomes under the upper-level poverty line, R1417/month per capita, where most incomes were located.

Figure S3: Directed Acyclic Graph (DAG) Depicting Hypothesized Relationships between Exposure Variables, Second-Line Drug Resistance, and Other Confounders



Note: "resistance to second-line drug" refers to resistance to any of the drugs in this study (moxifloxacin or levofloxacin, capreomycin, linezolid, bedaquiline, or clofazimine).

Table S1: Summary of Literature Supporting Pathways Represented in Directed Acyclic Graph

Relation	Literature Supporting Direction	Relation	Literature Supporting Direction
Age -> Alcohol	(95)	Household income -> SLD Resistance	(54; 56; 81)
Age -> Education	Assumption made by authors	Household income -> Smoking	(96; 97)
Age -> Marital status	Assumption made by authors	Marital status -> HIV	(98)
Age -> SLD Resistance	(46; 51; 57; 77; 99)	Marital status -> SLD Resistance	(40)
Age -> Smoking	(97; 100)	Marital status -> smoking	(97)
Age -> Unemployment	(101; 102)	Sex -> Alcohol	(95; 103)
Alcohol -> HIV	(104-106)	Sex -> Education	(107)
Alcohol -> SLD Resistance	(38; 41; 55; 77)	Sex -> HIV	(108; 109)
Alcohol -> CD4 Count	(103)	Sex -> SLD Resistance	(41; 58; 77; 83; 92)
Education -> Alcohol	(103; 110)	Sex -> Smoking	(100; 111)
Education -> HIV	(112)	Sex -> Unemployment	(113)
Education -> Household income	(114)	Alcohol -> Smoking	(103; 115)
Education -> SLD Resistance	(54; 116)	Smoking -> SLD Resistance	(117-119)
Education -> Smoking	(96)	Unemployment -> HIV	(98)
Education -> Unemployment	(102; 107)	Unemployment -> Household income	(114)
HIV -> SLD Resistance	(46; 51; 52; 55; 77; 82; 99; 120)	Unemployment -> SLD Resistance	(41; 54; 56; 77; 81; 116)
Household income -> Alcohol	(95; 103; 110)	Unemployment -> Smoking	(115)
Age -> HIV	(109)	Former incarceration -> SLD Resistance	(84)
Former incarceration -> Unemployment	(121)	Former incarceration -> HIV	(122)
Education -> Former incarceration	(123; 124)	Sex -> Former incarceration	(123)
Former incarceration -> Alcohol	(121)	Marital status -> Former incarceration	(124)

Appendix II: Supplementary Methods

A. Data Source

A.1: Parent Study Population

This analysis uses data collected from the ongoing Role of Casual Contact and Migration in XDR Transmission in South Africa: A Geospatial, Genomic, and Social Network Study (from here on referred to as the "CONTEXT study"). The central aims of this study are to characterize the role of casual (as opposed to prolonged) contact between individuals in the transmission of second-line drug-resistant tuberculosis, including individuals with XDR-TB and pre-XDR-TB (SLDR-TB). The study population includes individuals with a positive tuberculosis culture who exhibit resistance to second-line drugs (specifically, fluoroquinolones or second-line injectable drugs) who were diagnosed in the province of KwaZulu-Natal, South Africa. Data collection for this study began in January 2019 and has been ongoing and will finish through 2023.

Participants enter the study through a screening procedure that occurs at Inkosi Albert Luthuli Central Hospital (IALCH), in Durban, KwaZulu-Natal. Specifically, an individual was included in the provincial population if they were determined to have second-line drug resistance from results of second-line line probe assays (LPA). Copies of LPA results were regularly provided to study staff. Participants were also screened if resistance was detected via phenotypic drug susceptibility tests (pDST) before second-line LPA test results were available. The first isolate that was screened was referred to as the "diagnostic" isolate. However, multiple isolates could be tracked for a single participant, and the earliest isolates under consideration were collected in 2018 If the participant was located in one of four districts of KwaZulu-Natal close to or containing Durban—eThekwini, Ugu, iLembe, and uMgungundlovu—the participant was then contacted and offered to participate in an interview in which they would provide information on demographic

and medical characteristics, information on locations where they live or visit, and their routine contacts. This group of participants who consented to and completed an interview will be referred to as the "eThekwini cohort," and is a subset of the entire "provincial population" that was screened at IALCH.

A.2: Parent study data collection

In the CONTEXT study, pDST results were recorded for every participant in the provincial population. Phenotypic drug susceptibility tests were conducted at IALCH and, in some cases, at the National Institute of Communicable Diseases (NICD) in Johannesburg, South Africa. Phenotypic resistance testing was performed on cultured M. *tuberculosis* isolates (typically from sputum samples) using either agar-based methods or using the mycobacterial growth indicator tube (MGIT; Becton Dickinson) system. pDST results from IALCH and NICD were abstracted from laboratory reports onto a paper case report form, along with the date of collection of the isolate. Specifically, results for two first-line drugs, isoniazid and rifampicin, and several second-line drugs (capreomycin, low- and high-level moxifloxacin, levofloxacin, linezolid, bedaquiline, and clofazimine) were abstracted, although a variety of factors (test failure, national priorities for testing, etc.) (89) meant that participants usually did not have pDST results for all of these drugs.

We prioritized an isolate that had already undergone whole genome sequencing (WGS) by the time of pDST abstraction for results. If no isolates had undergone WGS, the diagnostic isolate was chosen for pDST abstraction, and if this isolate had no pDST results, we used the results of another isolate collected from the participant, which was typically collected within twelve months of the diagnostic. The case report form was uploaded to the study's shared drive, underwent quality control processes, and was entered into an online REDCap database used to store data collected

for the CONTEXT study. Information on the participant's sex and date of birth was also available from the laboratory report associated with the diagnostic isolate; this information was also abstracted onto case report forms, and then underwent quality control and double data entry into REDCap.

This analysis also includes information that was collected from the aforementioned interviews conducted for participants in the eThekwini cohort. Similarly, responses from participants were recorded on a paper case report form, which then underwent quality control, was uploaded to the study's shared drive, and underwent double data entry into the CONTEXT REDCap online database. A participant may have also undergone HIV testing (via the Abbott or Uni-Gold rapid test kits) if the participant did not have a prior HIV test, did not know the results of prior HIV tests, or tested negative more than three months before the interview, and if the participant consented to an HIV test. However, these scenarios were not common, and thus HIV testing was only done rarely as part of the study for the eThekwini cohort.

Ultimately, all variables used for analysis are stored in the CONTEXT REDCap database, and regular data compares between both data entry groups are conducted to minimize erroneous results entered. For this analysis, the entire CONTEXT dataset as entered by the first data access group was exported as an Excel file from REDCap in November 2022. A more in-depth description of the variables included in the analysis will be described in the "Data Cleaning" section.

B. Theoretical Framework for Research Question

B.1: Outcomes of Interest

The outcome of interest in this analysis is phenotypic resistance or susceptibility to the following second-line drugs: capreomycin, fluoroquinolones at large (moxifloxacin or

levofloxacin), bedaquiline, clofazimine, and linezolid. For each drug, individuals exhibiting resistance are considered the index group, and individuals who exhibited susceptibility were considered the reference group. This outcome has precedence in the literature (36; 40-43; 72). Previous studies, such as the PETTS study, considered associations of certain exposures to resistance to broad classes of drugs, namely fluoroquinolones and second-line injectable drugs (41). Here, phenotypic resistance is defined as the outcome of the authorized result of phenotypic drug susceptibility tests undertaken on Mycobacterial Growth Indicator Tube (MGIT)- or agar-based cultures of SLDR-TB isolates.

B.2: Selection and Justification of Regression Model

We constructed separate binary logistic regression models for phenotypic resistance to each of the second-line drugs of interest listed in Section I. There is precedence in using logistic regression in multivariable models where phenotypic resistance to second-line drugs was the outcome (36; 40; 42). Additionally, the outcome of interest is dichotomous and is thus well suited for binary logistic regression.

B.3: Selection of Exposures and Theoretical Framework for Causal Exposure-Outcome Relationship

To identify potential exposures that may have an effect on phenotypic resistance to the second-line drugs listed above, we reviewed the literature for risk factors for resistance to second-line drugs or XDR-TB, characteristics of the XDR-TB epidemic in South Africa (and specifically in KwaZulu-Natal), and relationships between risk factors for second-line drug resistance and their confounders. We then developed a causal framework for the relationship of a variable to resistance

to a specific second-line drug. This framework was centered on determining how these variables were related to resistance through the forces that drive second-line drug-resistant TB: acquisition and, importantly in the context of KwaZulu-Natal, transmission (67; 68; 70). Acquisition of second-line resistance can occur during treatment of drug-resistant tuberculosis with second-line drugs (42; 43; 75). Indeed, many studies have found that having been previously treated for DR-TB or with second-line drugs for TB was significantly associated with resistance to fluoroquinolones (40; 41), second-line injectable drugs (41), bedaquiline (36), and XDR-TB (47; 48). In particular, acquired resistance to second-line drugs is associated with certain outcomes from treatment for DR-TB; one study in the Baltic states found that treatment failure or interruption was associated with a greater risk of XDR-TB compared to MDR-TB (50). Such a finding aligns with research that indicated that treatment interruption was also a risk factor for resistance to first-line drugs (43). Although another study found that loss to follow-up was not a risk factor for XDR-TB, no participant in its study population had received second-line drugs in their previous treatment for TB (125). Meanwhile, transmission of the strains with resistance to a specific second-line drug may be disproportionately occurring among people in greater proximity to communities, individuals, or networks of people who likely are experiencing greater unfavorable treatment outcomes for reasons related to acquisition, a point that was previous articulated by authors of a 2015 paper (81).

We selected a range of biological variables (sex, age, and HIV status), a behavioral variable (frequency of alcohol consumption), and a socioeconomic variable (per capita monthly household income) as the main exposures of interest. We selected these variables because the literature identified an association between second-line drug resistance and these exposures, and we were interested in examining whether these associations existed in the more current context of SLDR-

TB in KwaZulu-Natal. In many cases, different associations were identified in the literature between the exposure, such as HIV and sex, and resistance, suggesting that these variables' relationship with second-line drug resistance was not definitive. We thus were interested in determining what the association would look like in KwaZulu-Natal.

Using the framework described above, we hypothesized how each exposure might be related to resistance to any of the second-line drugs under consideration, both through acquisition and transmission. We hypothesized that sex could have an effect on second-line drug resistance through acquisition of resistance; previous studies, including those conducted in South Africa, have identified that outcomes like loss to follow-up during treatment of DR-TB are associated with male sex (76-78; 83; 84), which could lead a greater risk of developing resistance (50). Age could also lead to the acquisition of drug resistance, as individuals who are younger have previously been found to also experience outcomes like loss-to-follow-up during DR-TB treatment that have been associated with development of resistance (43; 50; 77-80). Positive HIV status has previously been associated with loss to follow-up during treatment for DR-TB, which could lead to acquisition of second-line drug resistance (50; 76; 77). Greater consumption of alcohol is a risk factor for second-line drug resistance (41; 50) or for interruption or loss to follow-up of treating DR-TB (77; 80), and one study that found an association between alcohol use and XDR-TB hypothesized that higher levels of alcohol consumption could interrupt individuals' treatment for DR-TB, opening a pathway to acquisition of second-line drug resistance (55). Finally, we hypothesized income to have an effect on acquisition of resistance, based on interviews conducted with individuals in South Africa who experienced loss to follow-up or treatment failure during treatment for XDR-TB. This study's participants described the economic pressures of remaining in care (81), suggesting that individuals with lower incomes may be at greater risk of experiencing loss to

follow-up and potentially acquiring second-line drug resistance as a result. Moreover, in another study conducted in Indonesia, income significantly differed between individuals who had successful treatment for DR-TB and for individuals who were lost to follow-up, with lower-income individuals disproportionately in the lost-to-follow-up group (21).

To visualize these hypothesized relationships, a directed acyclic graph, or DAG, was constructed (86). This DAG was informed both by the literature that we had examined and, in certain cases, by our own assumptions, such as the assumption that age is a contributing factor to highest level of education. Using this DAG, confounders between the main exposures of interest and the outcome were identified, and any variables that could introduce a spurious association (also known as "colliders") were flagged (86). Using the DAG, we then identified minimally sufficient sets of covariates, or the fewest necessary variables that we hypothesized could control for confounding (86), to include in multivariable models to control for the relationship between each exposure and every set of outcomes. These sets of covariates exclude any potential mediators between the exposure and outcome and ensure that no spurious association or collider stratification bias as hypothesized by the DAG is introduced (86). These models are listed in the "Regression Analysis" section.

B.4: Additional Assumptions Necessary for Analysis

Given the structure and nature of the data collected for CONTEXT, we made certain assumptions before beginning analysis. First, we assumed that the risk factors of one participant would not affect whether an individual was resistant to a specific second-line drug. Second, because the parent study, CONTEXT, is cross-sectional, we needed to make assumptions to justify using a variable collected by CONTEXT as an exposure or confounder that may be linked to drug

resistance. For an individual who is in the eThekwini cohort, the interview process usually occurs after the isolate with pDST results is collected, and for that reason, it cannot be taken for granted that the exposures and confounders occur before the outcome.

For certain variables, such as age, sex, and whether someone was previously treated for DR-TB, this concern is less relevant. The two variables collected for employment determine whether someone was currently employed or, if unemployed, whether an individual was employed within the past two years of the interview, which is likely long enough to have occurred before collection of the isolate with pDST results. Education is a fairly static socioeconomic measure (73), and it would be fair to assume that education level has not changed between the date of collection of the isolate and the time of the interview. (This may not be the case for younger participants, although age is hypothesized to be a confounder of education in this study's DAG.) The interview's question on incarceration asks whether an individual has been in prison in the past twelve months. Of course, nothing is known about the length of imprisonment or if someone was imprisoned before this period, so temporality can be established to an extent—but not fully without making an assumption that it would be rare for a period of incarceration to finish shortly before the "12-month cutoff," and that an individual was incarcerated after date of collection of the isolate with results.

For HIV status, we assumed that the answer that the participant provided matched their status before the date of collection of the isolate with pDST results. We also assumed that, when the participant is asked about their current marital status, this status (like education) is static over time, at least for the period from date of collection to the interview date. Likewise, because the participant is simply asked about their habitual alcohol consumption patterns, we assumed that these behaviors have not varied greatly over time. Finally, per capita monthly household income

is a variable that can certainly fluctuate over time (73). For this reason, we categorized this variable, as described below. While there may be fluctuations in income over time, people may be less likely to experience a change in income so dramatic that one's income is in an entirely different category before collection of the isolate with pDST results than when the interview was conducted.

C. Data Cleaning, Creation of Analytic Datasets, and Descriptive Statistics

C.1: Data Cleaning and Creation of Analytic Datasets

All data cleaning, descriptive statistics, and regression analyses (as described below) were conducted with the R programming language in RStudio (version 2022.12.0, build 353). As noted above, the complete CONTEXT dataset was exported from REDCap on November 22, 2022. We extracted two smaller datasets from this larger file: one in which each observation corresponds to a participant with at least one pDST result in the provincial population, and another dataset in which each observation corresponds to a participant with at least one pDST result in the eThekwini cohort. Creating the latter dataset required linking pDST and interview variables by the participant's study ID. We calculated the frequency of individuals who could not be reached, did not consent, withdrew, were lost to follow-up from the cohort, were later found to not meet eligibility criteria, or had incomplete REDCap instruments for the variables being used. These individuals were then excluded from both datasets and from any further analysis. Variables that would not be used in the analysis were also removed. Finally, there were rare instances in which the date of birth and sex were missing in the provincial population dataset. We returned to REDCap to retrieve this information if it had been populated since the last download of the dataset in November, and, if the individual was in the eThekwini cohort, to see if the participant provided

this information in their interview. All missing information on sex and all but three instances of missing date of birth were recovered.

Additional variables were created for both the provincial population- and eThekwini cohort-specific datasets. One variable indicated if there was resistance to any fluoroquinolone listed on the CRF (low- or high-level moxifloxacin low and levofloxacin). Another variable provided an indication of the concordance between pDST results for bedaquiline and clofazimine, including whether the participant was sensitive to both drugs, resistant to both drugs, or had discordant results. Additionally, age of the participant was calculated by finding the difference between the date of collection of the isolate with pDST results and the participant's date of birth. In certain instances, the date of birth that was abstracted from the isolate laboratory report (for the entire provincial population) differed from the date of birth provided by the participant in the interview. In those instances, the date of birth reported by the participant themself was prioritized for calculations of age. Finally, a categorized age variable was created, grouping individuals together in the following categories: 0-19, 20-34, 35-54, and 55 or more years.

Basic summary statistics were then run to ensure that there were no problematic values in the dataset, including frequencies of categorical variables that were cross-referenced with the possible values listed in the CONTEXT Data Dictionary and identification of extreme values of continuous variables. No implausible values were identified, in part because of the internal validation of data entry fields in the CONTEXT REDCap database. Once this was completed, any continuous or date variable listed as a string of "9"s (i.e., 9999) was changed to "NA", as this was the convention used to report missing values on case report forms (according to the aforementioned data dictionary).

Four additional variables in the eThekwini cohort dataset were then derived from the original variables collected in the parent study. First, a new HIV variable was created as a composite of information obtained from the interview (in which participants reported their status) and the result of an HIV test conducted as part of the study for the scenarios described above. A new employment variable combined two variables on current employment and, if currently unemployed, employment in the past two years to indicate whether an individual was ever employed in the past two years (including at the time of interview). Income was standardized by dividing the monthly household income by the number of individuals supported partially or fully on the household income, and was then categorized based on the per-capita monthly food poverty line set by the government of South Africa in 2022 (R663/month) (74).

Finally, a "previous treatment for DR-TB" variable was created, indicating whether an individual had undergone treatment for MDR-TB, XDR-TB, or SLDR-TB before collection of their isolate with pDST results. This variable synthesized interview questions on whether the participant had started any TB treatment, if and when (i.e., month and year) a participant started MDR-TB treatment, and if the participant had received any prior treatment for XDR-TB or SLDR-TB. Unlike the wording of the questions posed about XDR or SLDR-TB treatment, it was not guaranteed that if the participant indicated that they had started MDR-TB treatment, they had started it before collection of their isolate with pDST results. For that reason, an individual was only considered to have had a previous MDR-TB treatment if month and year of the start of any of their MDR-TB regimens preceded the month and year of the date of collection of the isolate with pDST results.

Given the sparsity of data in original strata of the education, alcohol consumption, and marital status variables, these variables were collapsed into broader but still meaningful categories.

For education, “No Formal Schooling” and “Primary School” were collapsed into a “Primary School or Less” variable, the original “Secondary School (No Matric)” category was left alone, and the “Matric” and “University or Higher” categories were collapsed into a “Matric or Higher” category. For alcohol consumption, “Never” was left as its own category, “Monthly or Less” and “2-4 times per month” were combined into a “4 or Less Times per Month” category, and “2-3 Times a Week” and “4 or more times a week” were combined into a “More than Once per Week” category. Finally, the “Single” and “Widowed” strata were combined for marital status. Any strata including participants who did not know or were unable to answer questions about these variables were left as a “Don’t know” category (instead of setting these observations as missing), as this could potentially bias the results of the study (126).

Finally, five datasets from the provincial population and four datasets from the eThekwini cohort were created, each of which was restricted to include only individuals with results available for the second-line drug of interest (capreomycin, fluoroquinolones, bedaquiline, clofazimine, and linezolid); no dataset was created for capreomycin within the eThekwini cohort, as all participants with pDST results were resistant. In this process, variables that were used to create the composite variables were removed, and only the variables that would be incorporated into analysis were included. These new datasets were saved as CSV files.

C.2: Descriptive Statistics

For all participants with at least one pDST result in both the provincial population and the eThekwini cohort, the frequency and proportion of individuals with susceptible and resistant results for each first- and second-line drug, as well as the distribution of demographic variables of interest (e.g., age, sex, HIV status) were calculated. These datasets were transposed and

summarized to create stacked bar charts showing the distribution of pDST results, while a histogram was constructed for continuous age and per-capita monthly income variables (all of which used the "ggplot" package in R). The "table1" package in R was used to construct tables summarizing the characteristics of both study populations.

D. Regression Analysis

D.1: Presentation of Regression Models

Our objective was to analyze the association of five main exposures—sex, age, HIV status, alcohol use, and income, each with their own set of confounders as identified by the DAG described above—with resistance or susceptibility to five second-line drugs: capreomycin, fluoroquinolones (moxifloxacin and/or levofloxacin), linezolid, bedaquiline, and clofazimine. Below are the final regression models with the exposure, the minimally sufficient set of confounders, and an interaction term between the exposure and previous treatment for DR-TB that were first used for analysis, with "Resistance to Second-Line Drug" shorthand for the specific drug as the outcome of interest. Note that many of these variables are nominal and were represented as indicator variables in the actual models being analyzed.

$$\text{Logit(Resistance to Second-Line Drug)} = \beta_1\text{HIV} + \gamma_1\text{Sex} + \gamma_2\text{Incarceration} + \gamma_3\text{Employment} + \gamma_4\text{Education} + \gamma_5\text{Alcohol} + \gamma_6\text{Age} + \gamma_7\text{MaritalStatus} + \gamma_8\text{PreviousTreatment} + \delta_1\text{PreviousTreatment*HIV}$$

$$\text{Logit(Resistance to Second-Line Drug)} = \beta_1\text{Alcohol} + \gamma_1\text{Age} + \gamma_2\text{Sex} + \gamma_3\text{Incarceration} + \gamma_4\text{Education} + \gamma_5\text{Income} + \gamma_6\text{PreviousTreatment} + \delta_1\text{PreviousTreatment*Alcohol}$$

$$\text{Logit(Resistance to Second-Line Drug)} = \beta_1\text{Income} + \gamma_1\text{Employment} + \gamma_2\text{Sex} + \gamma_3\text{Age} + \gamma_4\text{Education} + \gamma_5\text{PreviousTreatment} + \delta_1\text{PreviousTreatment*Income}$$

$$\text{Logit(Resistance to Second-Line Drug)} = \beta_1\text{Age} + \gamma_1\text{PreviousTreatment} + \delta_1\text{PreviousTreatment*Age}$$

$$\text{Logit(Resistance to Second-Line Drug)} = \beta_1\text{Sex} + \gamma_1 + \delta_1\text{PreviousTreatment*Sex}$$

D.2: Unadjusted Estimates

First, we used unadjusted models without interaction terms or confounders to find the unadjusted odds ratios between each of the five exposures and resistance to each of the five drugs of interest. We used the datasets for the eThekwini cohort for the unadjusted odds ratios for alcohol, income, and HIV status, and both the eThekwini cohort and the entire provincial population datasets for age and sex.

D.3: Collinearity

After calculating the unadjusted odds ratios, we constructed a fully adjusted logistic regression model for each exposure and outcome of interest, using the interaction term between the exposure and previous treatment for DR-TB and the full set of confounders listed above. These models only incorporated data from the eThekwini cohort dataset, as these were the only individuals with the set of variables necessary for analysis. We assessed whether there was any collinearity in these full models using the collinR macro. If the diagnostic for the model could not be run, we followed one recommended approach to rank which covariates were absolutely essential to include in the model, and which terms were less important and could be eliminated if necessary to reduce harmful collinearity (88). Because we wanted to minimize confounding as much as possible, we removed the fewest possible terms for the diagnostic to run. The interaction

term was the first term that was removed in these cases, as these are typically removed first when there is collinearity that can be observed, and because they are especially likely to be associated with other independent variables in the model (e.g., the individual factors in the product term) (88). Then, if collinearity diagnostics still did not run, we targeted variables that likely did not contribute to as much confounding (e.g., a more tenuous relationship in the literature) or showed little variation in the actual data (e.g., all or almost all of the participants have the same "former incarceration" status). When the diagnostic did run, we concluded that there was no longer collinearity in the model when the largest condition index (CI) was less than 30 or if there were fewer than two variance deviation proportions (VDPs) greater than 0.5 for CIs over 30, as recommended by Kleinbaum and Klein in their description of collinearity assessment (88). Covariates were removed if they were listed as having a VDP greater than 0.5 when the CI was greater than 30, and the interaction term in the model was prioritized for removal.

D.4: Interaction Assessment and Calculation of Adjusted Odds Ratios

Next, we determined whether there was statistically significant interaction between each of the main exposures and previous treatment for DR-TB, if the interaction term was not removed during the collinearity assessment. We included this interaction term because we predicted that the association between the main exposures and second-line drug resistance would differ for those who were and who were not previously treated for DR-TB. We predicted that this relationship may be modified by previous treatment because, for each exposure, we hypothesized that the effect of these exposures on resistance to second-line drugs was via acquisition of resistance, and acquiring resistance occurs during treatment for DR-TB (42; 43; 75). (We included these hypotheses in Section III of the Theoretical Framework for Research Question section.) Therefore, our proposed

pathway between these exposures and resistance to second-line drugs may not exist for individuals who were never treated for DR-TB.

For instance, increased alcohol consumption has been found to be a risk factor for loss to follow up during treatment for DR-TB, including in South Africa (77; 80). Treatment of DR-TB uses second-line drugs like those examined in this analysis, and loss to follow-up and failure during treatment of DR-TB are risk factors for the acquisition of resistance to second-line drugs (50). Because increased alcohol consumption is hypothesized to be related to second-line drug resistance through acquisition and has previously been found to be significantly associated with XDR-TB (50; 55) the association between these two variables may be different (and potentially stronger) for those who have previously undergone treatment for DR-TB compared to those who have not undergone treatment for DR-TB. After all, individuals who have not undergone treatment for DR-TB may not have had the same level of exposure to second-line drugs, and thus may not have the same chance to acquire resistance to second-line drugs. For this reason, the association between alcohol consumption and second-line drug resistance could be modified by previous treatment for DR-TB.

Assuming there was an interaction term in a model, we used the Likelihood Ratio Test and an alpha level of $\alpha = 0.05$ to assess whether the inclusion of the interaction term in the model was statistically significant. The interaction term was included only if the test statistic was statistically significant. This was done for each model with an interaction term. Finally, when the interaction assessment was completed, the adjusted odds ratio for each model was identified by exponentiating the coefficient and confidence interval of the main exposure. The Likelihood Ratio Test was used to identify p-values for the significance of the exposure's coefficient in the model, and an alpha level of $\alpha = 0.05$ was also used to determine statistical significance.

Appendix III: Example Thesis Code

This analysis was conducted in RStudio (version 2022.12.0, build 353), and this code was added to this document via R Markdown.

```
knitr::opts_chunk$set(echo = TRUE)
#Load necessary packages
packagelist <- c("readxl", "tidyverse", "janitor", "knitr", "GGally", "corrplot", "sc
ales", "data.table", "table1")
for (package in packagelist) {
  library(package, character.only = T)
}


#All analysis was conducted using the "R" programming language in RStudio 2022.12.0+3
53. First, a complete dataset from the CONTEXT study of data entered by Data Access G
roup 1 was downloaded from the study's online database management system, REDCap, on
November 22, 2022.

# Download the entire dataset from CONTEXT (downloaded
# November 22)
data <- read.csv("C:/Users/nickr/OneDrive - Emory University/Thesis/Analysis/CONTEXTD
ata_Copy.csv")

# lists of variables of interest I would be interested in
# including for analysis
varnames <- read_xlsx("C:/Users/nickr/OneDrive - Emory University/Thesis/Analysis/Var
Names.xlsx")
varfield <- as.vector(varnames$fieldvariable)
varlab <- as.vector(na.omit(varnames$labvariable))
```

The following is code used for data cleaning and descriptive statistics purposes.

```
#------------------EXTRACT LAB AND FIELD DATASETS---------------------------------
-----------------------------------------------
# create dataset for variables from field arm, excluding
# repeating instruments, individuals who were considered
# ineligible, and individuals who did not consent note that
# these datasets are being restricted to participants whose
# REDCap instruments for the F20, F40, L34, and L12 (which
# provide the needed information for analysis) are complete

field_data <- data %>%
    filter(redcap_event_name == "field_arm_2" & redcap_repeat_instrument ==
        "" & (is.na(s10_screenid) | s10_screenid == "") & f14_consent ==
        1 & f20_patient_interview_complete == 2 & f40_hiv_rapid_test_complete ==
        2) %>%
    select(varfield)

# This dataset below was created to identify flow of
# participants (e.g., initially referred to cohort,
# consented, complete interview data; table() function used
# to generate summaries of variables of interest) How many
# individuals eligible to be in the eThekwini cohort did
# not consent to participate, and were excluded from
```

```
# analysis?
inclusion_vars <- c("id", "f10_continue", "f12_teamscreen", "f12_successful",
    "f12_ptscreen", "f12_contactsuccess", "f14_indepconsent",
    "f14_inelig", "f14_consent", "f20_patient_interview_complete",
    "f20_studyid", "f40_hiv_rapid_test_complete", "s10_screenid")
field_data_consent <- data %>%
    filter(redcap_event_name == "field_arm_2" & redcap_repeat_instrument ==
        "") %>%
    select(inclusion_vars)

# create dataset for variables from lab arm, excluding
# repeating instruments and individuals considered
# ineligible
lab_dat_nonrepeating <- data %>%
    filter(redcap_event_name == "lab_arm_1" & redcap_repeat_instrument ==
        "" & is.na(redcap_repeat_instance) & (is.na(s10_screenid) |
        s10_screenid == "") & l34_phenotypic_dst_final_result_complete ==
        2 & l12_diagnostic_culture_result_complete == 2) %>%
    select(varlab)

# standardize study ID variable for field variable dataset
field_data$studyid <- strtrim(field_data$id, 5)
field_data$studyid <- as.integer(field_data$studyid)

#### Note: In the lab dataset, six participants were
#### originally missing sex, and a different eight
#### participants were missing DOB information. On 7 Feb
#### 2023, I went back into REDCap to determine if there
#### were any updates on sex/DOB info for these
#### participants, and manually changed their corresponding
#### values into the lab dataset. One participant had
#### missing DOB information in REDCap but completed a
#### field interview in which they stated their DOB, so I
#### entered that value as well. (There are still three
#### participants with missing DOB). This code is not being
#### shared as it includes potentially identifying
#### information.

#----------MERGE DATASETS AND RESTRICT TO PARTICIPANTS WITH PDST RESULTS-------------
-------------------------------------------------
# create merged dataset containing both field and L34/lab
# variables
fieldlab_combined <- merge(x = field_data, y = lab_dat_nonrepeating,
    by.x = "studyid", by.y = "l34_studyid")

# lab variable dataset, but only for participants with at
# least one pDST result
lab_dat_nonrepeating_pDST <- lab_dat_nonrepeating %>%
    filter(l34_dstavail == 1)
fieldlab_combined_pDST <- fieldlab_combined %>%
    filter(l34_dstavail == 1)

#-----------------------CREATE ADDITIONAL VARIABLES--------------------------------
-------------------------------------------------
```

```
# add variable indicating resistance to any fluoroquinolone
# provincial population
lab_dat_nonrepeating_pDST$flq <- ifelse((lab_dat_nonrepeating_pDST$l34_moxihighdst ==
    2 | lab_dat_nonrepeating_pDST$l34_moxilowdst == 2 | lab_dat_nonrepeating_pDST$l34
_levofloxdst ==
    2), 2, ifelse((lab_dat_nonrepeating_pDST$l34_moxihighdst ==
    1 | lab_dat_nonrepeating_pDST$l34_moxilowdst == 1 | lab_dat_nonrepeating_pDST$l34
_levofloxdst ==
    1), 1, 3))
# Add variable indicating cross-resistance ('2') to both
# bedaquiline and clofazimine. 1=sensitivity to both drugs,
# 4=discordant results (sensitivity to one, resistance to
# the other), 3=results are not available for both drugs
lab_dat_nonrepeating_pDST <- lab_dat_nonrepeating_pDST %>%
    mutate(bdq_cfz = ifelse(l34_bdqdst == 3 | l34_clofdst ==
        3, 3, ifelse(l34_bdqdst == 1 & l34_clofdst == 1, 1, ifelse(l34_bdqdst ==
        2 & l34_clofdst == 2, 2, 4))))
# eThekwini cohort
fieldlab_combined_pDST$flq <- ifelse((fieldlab_combined_pDST$l34_moxihighdst ==
    2 | fieldlab_combined_pDST$l34_moxilowdst == 2 | fieldlab_combined_pDST$l34_levof
loxdst ==
    2), 2, ifelse((fieldlab_combined_pDST$l34_moxihighdst ==
    1 | fieldlab_combined_pDST$l34_moxilowdst == 1 | fieldlab_combined_pDST$l34_levof
loxdst ==
    1), 1, 3))
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(bdq_cfz = ifelse(l34_bdqdst == 3 | l34_clofdst ==
        3, 3, ifelse(l34_bdqdst == 1 & l34_clofdst == 1, 1, ifelse(l34_bdqdst ==
        2 & l34_clofdst == 2, 2, 4))))

# transform date of collection of isolate with pDST results
# and date of birth into Date variable, and calculate age
# of participant at time of isolate collection, for lab and
# field datasets
lab_dat_nonrepeating_pDST$l12_ptdob <- as.Date(lab_dat_nonrepeating_pDST$l12_ptdob,
    "%m/%d/%Y")
lab_dat_nonrepeating_pDST$l34_datecollected <- as.Date(lab_dat_nonrepeating_pDST$l34_
datecollected,
    "%m/%d/%Y")
lab_dat_nonrepeating_pDST$age_at_doc <- as.integer((lab_dat_nonrepeating_pDST$l34_dat
ecollected -
    lab_dat_nonrepeating_pDST$l12_ptdob)/365.25)

fieldlab_combined_pDST$l34_datecollected <- as.Date(fieldlab_combined_pDST$l34_dateco
llected,
    "%m/%d/%Y")
fieldlab_combined_pDST$f20_dob <- as.Date(fieldlab_combined_pDST$f20_dob,
    "%m/%d/%Y")
fieldlab_combined_pDST$f40_datetest <- as.Date(fieldlab_combined_pDST$f40_datetest,
    "%m/%d/%Y")
fieldlab_combined_pDST$age_at_doc <- as.integer((fieldlab_combined_pDST$l34_datecolle
cted -
    fieldlab_combined_pDST$f20_dob)/365.25)

# This final step changes the participant age on the lab
```

```
# side to the participant age calculated using the field
# data. The participant's self-report of their DOB is
# viewed as more accurate than what was provided in lab
# reports at IALCH, so the age as calculated using the
# participant's self-reported DOB replaces the age
# calculated using lab reports from TrakCare where possible
lab_dat_nonrepeating_pDST$l12_ptdob <- ifelse(lab_dat_nonrepeating_pDST$l34_studyid %
in%
    fieldlab_combined_pDST$studyid, fieldlab_combined_pDST$f20_dob,
    lab_dat_nonrepeating_pDST$l12_ptdob)
lab_dat_nonrepeating_pDST$l12_ptdob <- as.Date.numeric(lab_dat_nonrepeating_pDST$l12_
ptdob,
    origin = "1970-01-01")
lab_dat_nonrepeating_pDST$age_at_doc <- ifelse(lab_dat_nonrepeating_pDST$l34_studyid
%in%
    fieldlab_combined_pDST$studyid, fieldlab_combined_pDST$age_at_doc,
    lab_dat_nonrepeating_pDST$age_at_doc)

# table() and summary() used to identify implausible values
# for each variable

# Note that repetitive values of '9' for continuous
# variables indicates that information was unknown (i.e.,
# by participant during interview); these will be replaced
# with 'NA' values for the following variables for which
# these values were identified
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(f20_totincome = ifelse(f20_totincome == 999999, NA,
        f20_totincome), f20_supportfully = ifelse(f20_supportfully ==
        99, NA, f20_supportfully), f20_supportpartially = ifelse(f20_supportpartially
==
        99, NA, f20_supportpartially), f20_hivnegyear = ifelse(f20_hivnegyear ==
        9999, NA, f20_hivnegyear), f20_mdrstartmonth = ifelse(f20_mdrstartmonth ==
        99, NA, f20_mdrstartmonth), f20_morethan1month = ifelse(f20_morethan1month ==
        99, NA, f20_morethan1month), f20_morethan2month = ifelse(f20_morethan2month =
=
        99, NA, f20_morethan2month), f20_xdrstartmonth = ifelse(f20_xdrstartmonth ==
        99, NA, f20_xdrstartmonth), f20_sltbstartmonth = ifelse(f20_sltbstartmonth ==
        99, NA, f20_sltbstartmonth))

# Create HIV variable
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(HIV = ifelse(f20_hivtest == 1, ifelse(f20_hivresult ==
        1, ifelse(f40_hivtestrec == 1, ifelse(f40_yeshivrec ==
        1, ifelse(f40_hivresult == 1, 1, 0), 0), 0), ifelse(f20_hivresult ==
        2, 1, ifelse(f40_hivtestrec == 1, ifelse(f40_yeshivrec ==
        1, ifelse(f40_hivresult == 1, 1, ifelse(f40_hivresult ==
        2, 0, 2)), 0), 2))), ifelse(f40_hivtestrec == 1, ifelse(f40_yeshivrec ==
        1, ifelse(f40_hivresult == 1, 1, ifelse(f40_hivresult ==
        2, 0, 2)), 2), 2)))

# Create previous DR-TB treatment variable: Has participant
# previously received treatment for MDR-TB and/or SLDR-TB
# at least once before date of collection of isolate with
# pDST results? First, transform start month and year of
```

```
# MDR-TB treatment into a date variable
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(f20_mdrstartmonth = as.character(f20_mdrstartmonth),
        f20_morethan1month = as.character(f20_morethan1month),
        f20_morethan2month = as.character(f20_morethan2month))

onenine <- as.character(c(1:9))

# some maneuvering to transform months into something that
# can be easily converted to a date variable--i.e., adding
# a '0' in front of a month (transforming 1 to 01 for
# January)
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(f20_mdrstartmonth = ifelse(f20_mdrstartmonth %in%
        onenine, paste("0", f20_mdrstartmonth), f20_mdrstartmonth),
        f20_morethan1month = ifelse(f20_morethan1month %in% onenine,
            paste("0", f20_morethan1month), f20_morethan1month),
        f20_morethan2month = ifelse(f20_morethan2month %in% onenine,
            paste("0", f20_morethan2month), f20_morethan2month))
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(f20_mdrstartmonth = gsub(" ", "", f20_mdrstartmonth),
        f20_morethan1month = gsub(" ", "", f20_morethan1month),
        f20_morethan2month = gsub(" ", "", f20_morethan2month))

# Concatenate month, a day (first of the month), and the
# year of treatment
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(mdrstart1 = paste(f20_mdrstartmonth, "01", as.character(f20_mdrstartyear),
        sep = "-"), mdrstart2 = paste(f20_morethan1month, "01",
        as.character(f20_morethan1year), sep = "-"), mdrstart3 = paste(f20_morethan2m
onth,
        "01", as.character(f20_morethan2year), sep = "-"))

# In special instances where the year is known, but not the
# month, set the date to December 02 [year treatment
# started]. This way, we know if treatment started in the
# year previous to the year of date of collection
# (12-02-year + 30 days < date of collection), even if we
# don't have specifics on whether it occurred in the month
# prior.
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(mdrstart1 = ifelse(!is.na(f20_mdrstartyear) & is.na(f20_mdrstartmonth),
        paste("12", "02", as.character(f20_mdrstartyear), sep = "-"),
        mdrstart1), mdrstart2 = ifelse(!is.na(f20_morethan1year) &
        is.na(f20_morethan1month), paste("12", "02", as.character(f20_morethan1year),
        sep = "-"), mdrstart2), mdrstart3 = ifelse(!is.na(f20_morethan2year) &
        is.na(f20_morethan2month), paste("12", "02", as.character(f20_morethan2year),
        sep = "-"), mdrstart3))


# Finally, convert this variable into a Date variable
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(mdrstart1 = as.Date(mdrstart1, "%m-%d-%Y"), mdrstart2 = as.Date(mdrstart2,
        "%m-%d-%Y"), mdrstart3 = as.Date(mdrstart3, "%m-%d-%Y"))
```

```
# create variable indicating if any treatment for MDR-TB
# startment before date of isolate collection.
# Specifically, did any treatment for MDR-TB (may be first
# round, or the second or third) occur at the very least in
# the month previous to collection of isolate with pDST
# results (or if month is not known, in at least the year
# prior to collection)? 0=no MDR treatment before pDST,
# 1=MDR treatment before pDST, 2=don't know

fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(MDRprepDST = ifelse(is.na(f20_txtb), 2, ifelse(f20_txtb ==
        1, ifelse(f20_txmdr == 1, ifelse(is.na(f20_mdrstartyear) &
        is.na(f20_morethan1year) & is.na(f20_morethan2year),
        2, ifelse(f20_mdrstartyear == 9999 & is.na(f20_morethan1year) &
            is.na(f20_morethan2year), 2, ifelse(f20_mdrstartyear ==
            9999 & f20_morethan1year == 9999 & is.na(f20_morethan2year),
            2, ifelse(f20_mdrstartyear == 9999 & f20_morethan1year ==
                9999 & f20_morethan2year == 9999, 2, ifelse((as.integer(mdrstart1) +
                30) < as.integer(l34_datecollected) & !is.na(mdrstart1),
                1, ifelse((as.integer(mdrstart2) + 30) < as.integer(l34_datecollected
) &
                  !is.na(mdrstart2), 1, ifelse((as.integer(mdrstart3) +
                  30) < as.integer(l34_datecollected) & !is.na(mdrstart3),
                  1, 0)))))), ifelse(f20_txmdr == 2, 0, 2)),
        ifelse(f20_txtb == 2, 0, 2))))


# create variable indicating if anyone received previous
# treatment for DR-TB in the past. 0=no previous treatment,
# 1=previous treatment, 2=don't know
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(prevtrt = ifelse(is.na(f20_txtb), 2, ifelse(f20_txtb ==
        1, ifelse(MDRprepDST == 1, 1, ifelse(MDRprepDST == 0,
        ifelse((f20_xdrdiag == 1 & is.na(f20_prevsltb)) | (f20_prevsltb ==
            1 & is.na(f20_xdrdiag)), ifelse((f20_txxdr == 1 &
            is.na(f20_txsltb)) | (f20_txsltb == 1 & is.na(f20_txxdr)),
            1, ifelse((f20_txxdr == 2 & is.na(f20_txsltb)) |
                (f20_txsltb == 2 & is.na(f20_txxdr)), 0, 2)),
            ifelse((f20_xdrdiag == 2 & is.na(f20_prevsltb)) |
                (f20_prevsltb == 2 & is.na(f20_xdrdiag)), 0,
                2)), ifelse(MDRprepDST == 2, ifelse((f20_xdrdiag ==
            1 & is.na(f20_prevsltb)) | (f20_prevsltb == 1 & is.na(f20_xdrdiag)),
            ifelse((f20_txxdr == 1 & is.na(f20_txsltb)) | (f20_txsltb ==
                1 & is.na(f20_txxdr)), 1, 2), 2), 2))), ifelse(f20_txtb ==
        2, 0, 2))))

# create income variable divide monthly hosuehold income by
# household members fully+partially supported by income
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(monthincpc = ifelse(is.na(f20_supportfully) | is.na(f20_supportpartially),
        ifelse(f20_support == 0, f20_totincome, f20_totincome/f20_support),
        ifelse(f20_supportfully == 0 & f20_supportpartially ==
            0, f20_totincome, f20_totincome/(f20_supportfully +
            f20_supportpartially))))
summary(fieldlab_combined_pDST$monthincpc)
```

```
# dichotimize per capita monthly income by per capita
# monthly food poverty line in South Africa in 2022, R663
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(income_cat = ifelse(is.na(monthincpc), NA, ifelse(monthincpc <
        663, 0, 1)))
table(fieldlab_combined_pDST$income_cat)

# categorize age variables, for both field and lab dataset.
# Age categories: 0-19, 20-34, 35-54, 55+

fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(age_cat = ifelse(is.na(age_at_doc), NA, ifelse(age_at_doc <
        20, 0, ifelse(age_at_doc >= 20 & age_at_doc < 35, 1,
        ifelse(age_at_doc >= 35 & age_at_doc < 55, 2, 3)))))

lab_dat_nonrepeating_pDST <- lab_dat_nonrepeating_pDST %>%
    mutate(age_cat = ifelse(is.na(age_at_doc), NA, ifelse(age_at_doc <
        20, 0, ifelse(age_at_doc >= 20 & age_at_doc < 35, 1,
        ifelse(age_at_doc >= 35 & age_at_doc < 55, 2, 3)))))

# Employment status: 1=current employment or employment in
# the previous two years, 0=no employment in previous two
# years, 2=unknown (i.e., they are currently unemployed but
# it is unknown if they were employed in the past two
# years)
fieldlab_combined_pDST <- fieldlab_combined_pDST %>%
    mutate(employment = ifelse(f20_currentemp == 1, 1, ifelse(is.na(f20_pastemploy),
        2, ifelse(f20_pastemploy == 1, 1, 0))))
table(fieldlab_combined_pDST$employment)

#-------------------------CREATE ANALYTIC DATASETS FOR REGRESSION-----------------
-----------------------------------------------
labfinalvars <- as.vector(na.omit(varnames$finallab))
fieldfinalvars <- as.vector(na.omit(varnames$finalfield))

# create final and formatted datasets, with variables coded
# in a way suitable for logistic regression
lab_final <- lab_dat_nonrepeating_pDST %>%
    select(labfinalvars)
field_final <- fieldlab_combined_pDST %>%
    select(fieldfinalvars)


# Provincial population recode variables for logistic
# regression (i.e., to 0/1 format for binary variables, or
# specifying '0' for referent categories)
lab_final <- lab_final %>%
    mutate(Isoniazid = ifelse(l34_isoniazid == 3, NA, l34_isoniazid -
        1), Rifampicin = ifelse(l34_rifampicin == 3, NA, l34_rifampicin -
        1), Capreomycin = ifelse(l34_capreodst == 3, NA, l34_capreodst -
        1), MoxifloxacinLow = ifelse(l34_moxilowdst == 3, NA,
        l34_moxilowdst - 1), MoxifloxacinHigh = ifelse(l34_moxihighdst ==
        3, NA, l34_moxihighdst - 1), Levofloxacin = ifelse(l34_levofloxdst ==
        3, NA, l34_levofloxdst - 1), Linezolid = ifelse(l34_linezoldst ==
```

```
        3, NA, l34_linezoldst - 1), Bedaquiline = ifelse(l34_bdqdst ==
        3, NA, l34_bdqdst - 1), Clofazimine = ifelse(l34_clofdst ==
        3, NA, l34_clofdst - 1), Fluoroquinolones = ifelse(flq ==
        3, NA, flq - 1), Sex = (l12_sex - 1), Age = factor(age_cat))

# Dataset that will be used for analysis of provincial
# population data
lab_final_noformat <- lab_final %>%
    select(-(l34_isoniazid:l34_clofdst), -flq, -l12_sex)
lab_final <- lab_final %>%
    mutate(bdq_cfz = ifelse(bdq_cfz == 3, NA, bdq_cfz))

# Now, recode and reformat variables in the field dataset.
# Recode the field dataset so it is suitable for logistic
# regression analysis (i.e., set referent category equal to
# 0.)
field_final <- field_final %>%
    mutate(Isoniazid = ifelse(l34_isoniazid == 3, NA, l34_isoniazid -
        1), Rifampicin = ifelse(l34_rifampicin == 3, NA, l34_rifampicin -
        1), Capreomycin = ifelse(l34_capreodst == 3, NA, l34_capreodst -
        1), MoxifloxacinLow = ifelse(l34_moxilowdst == 3, NA,
        l34_moxilowdst - 1), MoxifloxacinHigh = ifelse(l34_moxihighdst ==
        3, NA, l34_moxihighdst - 1), Levofloxacin = ifelse(l34_levofloxdst ==
        3, NA, l34_levofloxdst - 1), Linezolid = ifelse(l34_linezoldst ==
        3, NA, l34_linezoldst - 1), Bedaquiline = ifelse(l34_bdqdst ==
        3, NA, l34_bdqdst - 1), Clofazimine = ifelse(l34_clofdst ==
        3, NA, l34_clofdst - 1), Fluoroquinolones = ifelse(flq ==
        3, NA, flq - 1), Sex = (f20_sex - 1), MaritalStatus = factor(ifelse(f20_marit
al ==
        1 | f20_marital == 3 | f20_marital == 4, 0, ifelse(f20_marital ==
        2, 1, 2))), Education = factor(ifelse(f20_education ==
        1 | f20_education == 2, 0, ifelse(f20_education == 3,
        1, ifelse(f20_education == 4 | f20_education == 5, 2,
            3)))), AlcoholUse = factor(ifelse(f20_alcohol ==
        1, 0, ifelse(f20_alcohol == 2 | f20_alcohol == 3, 1,
        ifelse(f20_alcohol == 4 | f20_alcohol == 5, 2, 3)))),
        Incarceration = factor(ifelse(f20_prison == 2, 0, f20_prison)),
        HIV = factor(HIV), prevtrt = factor(prevtrt), employment = factor(employment)
,
        income_cat = factor(income_cat), age_cat = factor(age_cat))

# Final dataset for analysis of eThekwini cohort data
field_final_noformat <- field_final %>%
    select(-(l34_isoniazid:l34_clofdst), -flq, -f20_sex, -f20_marital,
        -f20_education, -f20_alcohol, -f20_prison)
field_final <- field_final %>%
    mutate(bdq_cfz = ifelse(bdq_cfz == 3, NA, bdq_cfz))

# Create final datasets, for provincial population and
# eThekwini cohort, for each second-line drug Provincial
# population Capreomycin
lab_cap <- lab_final_noformat %>%
    filter(!is.na(Capreomycin))
write.csv(lab_cap, file = "C:/Users/nickr/OneDrive - Emory University/Thesis/Analysis
/Analytic Datasets/lab_cap.csv")
```

```
# Fluoroquinolones
lab_flq <- lab_final_noformat %>%
    filter(!is.na(Fluoroquinolones))
write.csv(lab_flq, file = "C:/Users/nickr/OneDrive - Emory University/Thesis/Analysis
/Analytic Datasets/lab_flq.csv")

# Linezolid
lab_lzd <- lab_final_noformat %>%
    filter(!is.na(Linezolid))
write.csv(lab_lzd, file = "C:/Users/nickr/OneDrive - Emory University/Thesis/Analysis
/Analytic Datasets/lab_lzd.csv")

# Bedaquiline
lab_bdq <- lab_final_noformat %>%
    filter(!is.na(Bedaquiline))
write.csv(lab_bdq, file = "C:/Users/nickr/OneDrive - Emory University/Thesis/Analysis
/Analytic Datasets/lab_bdq.csv")

# Clofazimine
lab_cfz <- lab_final_noformat %>%
    filter(!is.na(Clofazimine))
write.csv(lab_cfz, file = "C:/Users/nickr/OneDrive - Emory University/Thesis/Analysis
/Analytic Datasets/lab_cfz.csv")

# eThekwini cohort Fluoroquinolones
field_flq <- field_final_noformat %>%
    filter(!is.na(Fluoroquinolones))
write.csv(field_flq, file = "C:/Users/nickr/OneDrive - Emory University/Thesis/Analys
is/Analytic Datasets/field_flq.csv")

# Linezolid
field_lzd <- field_final_noformat %>%
    filter(!is.na(Linezolid))
write.csv(field_lzd, file = "C:/Users/nickr/OneDrive - Emory University/Thesis/Analys
is/Analytic Datasets/field_lzd.csv")

# Bedaquiline
field_bdq <- field_final_noformat %>%
    filter(!is.na(Bedaquiline))
write.csv(field_bdq, file = "C:/Users/nickr/OneDrive - Emory University/Thesis/Analys
is/Analytic Datasets/field_bdq.csv")

# Clofazimine
field_cfz <- field_final_noformat %>%
    filter(!is.na(Clofazimine))
write.csv(field_cfz, file = "C:/Users/nickr/OneDrive - Emory University/Thesis/Analys
is/Analytic Datasets/field_cfz.csv")
```

Next is code specific to the unadjusted and adjusted regression analyses. The following code is specifically for the eThekwini cohort in which phenotypic resistance to fluoroquinolones is the outcome. Although not shown here, similar code was written for analyses involving the provincial population or for resistance to the other second-line drugs under consideration.

```
# Read in the dataset
df <- read.csv(file = "C:/Users/nickr/OneDrive - Emory University/Thesis/Analysis/Ana
lytic Datasets/field_flq.csv")
# Read in collinearity macro
source(file.path("C:/Users/nickr/OneDrive - Emory University/Thesis/Analysis/collinR.
R"))

# Unadjusted sex analysis

# create logistic regression model object
glm_flq_sex <- glm(Fluoroquinolones ~ Sex, family = binomial(link = "logit"),
    data = df)
# identify odds ratios and confindence intervals
cbind(exp(coef(glm_flq_sex)), exp(confint.default(glm_flq_sex)))
# Calculate p-value using LRT
drop1(glm_flq_sex, test = "Chisq")

df$age_cat <- factor(df$age_cat, levels = c(0:3))

glm_flq_age <- glm(Fluoroquinolones ~ age_cat, family = binomial(link = "logit"),
    data = df)
# identify odds ratios and confindence intervals
cbind(exp(coef(glm_flq_age)), exp(confint.default(glm_flq_age)))
# Calculate p-value using LRT
drop1(glm_flq_age, test = "Chisq")

glm_flq_income <- glm(Fluoroquinolones ~ income_cat, family = binomial(link = "logit"
),
    data = df)
# identify odds ratios and confindence intervals
cbind(exp(coef(glm_flq_income)), exp(confint.default(glm_flq_income)))
# Calculate p-value using LRT
drop1(glm_flq_income, test = "Chisq")

glm_flq_alc <- glm(Fluoroquinolones ~ factor(AlcoholUse), family = binomial(link = "l
ogit"),
    data = df)
# identify odds ratios and confindence intervals
cbind(exp(coef(glm_flq_alc)), exp(confint.default(glm_flq_alc)))
# Calculate p-value using LRT
drop1(glm_flq_alc, test = "Chisq")

glm_flq_hiv <- glm(Fluoroquinolones ~ factor(HIV), family = binomial(link = "logit"),
    data = df)
# identify odds ratios and confindence intervals
cbind(exp(coef(glm_flq_hiv)), exp(confint.default(glm_flq_hiv)))
# Calculate p-value using LRT
drop1(glm_flq_hiv, test = "Chisq")
```

```
#### Part 2: Adjusted Analysis

#### Adjusted analysis of HIV Note: needed to drop
#### factor(prevtrt)*factor(HIV) for collinearity
#### diagnostics to run. Dropped no variables to reduce
#### largest CI to under 30 and to ensure that <2 VDPs have
#### a value of 0.5 or greater. Creation of initial model
#### and collinearity assessment
glm_flq_hiv_adj <- glm(Fluoroquinolones ~ factor(HIV) + Sex +
    factor(Incarceration) + factor(employment) + factor(Education) +
    factor(AlcoholUse) + factor(age_cat) + factor(MaritalStatus) +
    factor(prevtrt), family = binomial(link = "logit"), data = df)
collinR(glm_flq_hiv_adj)
# Interaction assessment not possible.
cbind(exp(coef(glm_flq_hiv_adj)), exp(confint.default(glm_flq_hiv_adj)))
# Calculate p-value using LRT
drop1(glm_flq_hiv_adj, test = "Chisq")

#### Adjusted analysis of alcohol consumption Note: needed
#### to drop factor(prevtrt)*factor(AlcoholUse) for
#### collinearity diagnostics to run. Dropped no variables
#### to reduce largest CI to under 30 and to ensure that <2
#### VDPs have a value of 0.5 or greater. Creation of
#### initial model and collinearity assessment
glm_flq_alc_adj <- glm(Fluoroquinolones ~ factor(AlcoholUse) +
    factor(age_cat) + Sex + factor(Incarceration) + factor(Education) +
    income_cat + factor(prevtrt), family = binomial(link = "logit"),
    data = df)
collinR(glm_flq_alc_adj)
# Interaction assessment not possible.
cbind(exp(coef(glm_flq_alc_adj)), exp(confint.default(glm_flq_alc_adj)))
# Calculate p-value using LRT
drop1(glm_flq_alc_adj, test = "Chisq")

#### Adjusted analysis for income Note: needed to drop no
#### variables for collinearity diagnostics to run. Dropped
#### no variables to reduce largest CI to under 30 and to
#### ensure that <2 VDPs have a value of 0.5 or greater.
#### Creation of initial model and collinearity assessment
glm_flq_income_adj <- glm(Fluoroquinolones ~ income_cat + factor(employment) +
    Sex + factor(age_cat) + factor(Education) + factor(prevtrt) +
    factor(prevtrt) * income_cat, family = binomial(link = "logit"),
    data = df)
collinR(glm_flq_income_adj)
# Interaction assessment. We are interested in whether
# prevtrt*income is a significant interaction, and will use
# the Likelihood Ratio Test to assess whether this
# interaction is statistically significant. Create full and
# reduced models with and without the interaction term full
glm_flq_income_adj_full <- glm(Fluoroquinolones ~ income_cat +
    factor(employment) + Sex + factor(age_cat) + factor(Education) +
    factor(prevtrt) + factor(prevtrt) * income_cat, family = binomial(link = "logit")
,
    data = df)
# reduced
```

```
glm_flq_income_adj_reduced <- glm(Fluoroquinolones ~ income_cat +
    factor(employment) + Sex + factor(age_cat) + factor(Education) +
    factor(prevtrt), family = binomial(link = "logit"), data = df)
# LRT
anova(glm_flq_income_adj_reduced, glm_flq_income_adj_full, test = "Chisq")
# p=0.14; interaction is not significant. Gold standard
# model is reduced model.
cbind(exp(coef(glm_flq_income_adj_reduced)), exp(confint.default(glm_flq_income_adj_r
educed)))
# Calculate p-value using LRT
drop1(glm_flq_income_adj_reduced, test = "Chisq")

#### Adjusted analysis of age Note: needed to drop
#### factor(prevtrt)*factor(age_cat) for collinearity
#### diagnostics to run. Dropped no variables to reduce
#### largest CI to under 30 and to ensure that <2 VDPs have
#### a value of 0.5 or greater. Creation of initial model
#### and collinearity assessment
glm_flq_age_adj <- glm(Fluoroquinolones ~ factor(age_cat) + factor(prevtrt),
    family = binomial(link = "logit"), data = df)
collinR(glm_flq_age_adj)
# Interaction assessment not possible.
cbind(exp(coef(glm_flq_age_adj)), exp(confint.default(glm_flq_age_adj)))
# Calculate p-value using LRT
drop1(glm_flq_age_adj, test = "Chisq")

#### Adjusted analysis for sex Note: needed to drop no
#### variables for collinearity diagnostics to run. Dropped
#### no variables to reduce largest CI to under 30 and to
#### ensure that <2 VDPs have a value of 0.5 or greater.
#### Creation of initial model and collinearity assessment
glm_flq_sex_adj <- glm(Fluoroquinolones ~ Sex + factor(prevtrt) +
    factor(prevtrt) * Sex, family = binomial(link = "logit"),
    data = df)
collinR(glm_flq_sex_adj)
# Interaction assessment for sex Use Likelihood Ratio Test
# to assess significance of interaction between sex and
# previous treatment for DR-TB on the whole full model
glm_flq_sex_adj_full <- glm(Fluoroquinolones ~ Sex + factor(prevtrt) +
    factor(prevtrt) * Sex, family = binomial(link = "logit"),
    data = df)
# reduced model
glm_flq_sex_adj_reduced <- glm(Fluoroquinolones ~ Sex + factor(prevtrt),
    family = binomial(link = "logit"), data = df)
# LRT
anova(glm_flq_sex_adj_reduced, glm_flq_sex_adj_full, test = "Chisq")
# p=0.71; interaction is not significant. Use reduced model
# as gold standard
cbind(exp(coef(glm_flq_sex_adj_reduced)), exp(confint.default(glm_flq_sex_adj_reduced
)))
# Calculate p-value using LRT
drop1(glm_flq_sex_adj_reduced, test = "Chisq")
```

References

1) Bespiatykh, D., Bespyatykh, J., Mokrousov, I., & Shitikov, E. (2021). A Comprehensive Map of Mycobacterium tuberculosis Complex Regions of Difference. *mSphere*, *6*(4), e0053521. https://doi.org/10.1128/mSphere.00535-21

2) Peloquin, C. A., & Davies, G. R. (2021). The Treatment of Tuberculosis. *Clin Pharmacol Ther*, *110*(6), 1455-1466. https://doi.org/10.1002/cpt.2261

3) WHO. (2022). *Global tuberculosis report: 2022.* https://www.who.int/teams/global-tuberculosis-programme/tb-reports

4) Churchyard, G. J., Mametja, L. D., Mvusi, L., Ndjeka, N., Hesseling, A. C., Reid, A., Babatunde, S., & Pillay, Y. (2014). Tuberculosis control in South Africa: successes, challenges and recommendations. *S Afr Med J*, *104*(3 Suppl 1), 244-248. https://doi.org/10.7196/samj.7689

5) Brown, T. S., Challagundla, L., Baugh, E. H., Omar, S. V., Mustaev, A., Auld, S. C., Shah, N. S., Kreiswirth, B. N., Brust, J. C. M., Nelson, K. N., Narechania, A., Kurepina, N., Mlisana, K., Bonneau, R., Eldholm, V., Ismail, N., Kolokotronis, S. O., Robinson, D. A., Gandhi, N. R., & Mathema, B. (2019). Pre-detection history of extensively drug-resistant tuberculosis in KwaZulu-Natal, South Africa. *Proc Natl Acad Sci U S A*, *116*(46), 23284-23291. https://doi.org/10.1073/pnas.1906636116

6) STREPTOMYCIN treatment of pulmonary tuberculosis. (1948). *Br Med J*, *2*(4582), 769-782. https://www.ncbi.nlm.nih.gov/pubmed/18890300

7) Gandhi, N. R., Moll, A., Sturm, A. W., Pawinski, R., Govender, T., Lalloo, U., Zeller, K., Andrews, J., & Friedland, G. (2006). Extensively drug-resistant tuberculosis as a cause of

death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet*, *368*(9547), 1575-1580. https://doi.org/10.1016/S0140-6736(06)69573-1

8) Ndjeka, N., Hughes, J., Reuter, A., Conradie, F., Enwerem, M., Ferreira, H., Ismail, N., Kock, Y., Master, I., Meintjes, G., Padanilam, X., Romero, R., Schaaf, H. S., Riele, J. T., & Maartens, G. (2020). Implementing novel regimens for drug-resistant TB in South Africa: what can the world learn? *Int J Tuberc Lung Dis*, *24*(10), 1073-1080. https://doi.org/10.5588/ijtld.20.0174

9) WHO. (2022). *Rapid communication: Key changes to the treatment of drug-resistant tuberculosis.* https://apps.who.int/iris/rest/bitstreams/1420701/retrieve

10) CDC. (2006). Revised definition of extensively drug-resistant tuberculosis. *MMWR*, *55*, 1176. https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5543a4.htm

11) Flor de Lima, B., & Tavares, M. (2014). Risk factors for extensively drug-resistant tuberculosis: a review. *Clin Respir J*, *8*(1), 11-23. https://doi.org/10.1111/crj.12044

12) Roelens, M., Battista Migliori, G., Rozanova, L., Estill, J., Campbell, J. R., Cegielski, J. P., Tiberi, S., Palmero, D., Fox, G. J., Guglielmetti, L., Sotgiu, G., Brust, J. C. M., Bang, D., Lienhardt, C., Lange, C., Menzies, D., Keiser, O., & Raviglione, M. (2021). Evidence-based Definition for Extensively Drug-Resistant Tuberculosis. *Am J Respir Crit Care Med*, *204*(6), 713-722. https://doi.org/10.1164/rccm.202009-3527OC

13) WHO. (2021). *WHO announces updated definitions of extensively drug-resistant tuberculosis*. https://www.who.int/news/item/27-01-2021-who-announces-updated-definitions-of-extensively-drug-resistant-tuberculosis#:~:text=The%20definition%20of%20extensively%20drug,of%20these%20forms%20of%20TB.

14) Pontali, E., Raviglione, M. C., Migliori, G. B., & and the writing group members of the Global, T. B. N. C. T. C. (2019). Regimens to treat multidrug-resistant tuberculosis: past, present and future perspectives. *Eur Respir Rev*, *28*(152). https://doi.org/10.1183/16000617.0035-2019

15) Ndjeka, N., Campbell, J. R., Meintjes, G., Maartens, G., Schaaf, H. S., Hughes, J., Padanilam, X., Reuter, A., Romero, R., Ismail, F., Enwerem, M., Ferreira, H., Conradie, F., Naidoo, K., & Menzies, D. (2022). Treatment outcomes 24 months after initiating short, all-oral bedaquiline-containing or injectable-containing rifampicin-resistant tuberculosis treatment regimens in South Africa: a retrospective cohort study. *Lancet Infect Dis*, *22*(7), 1042-1051. https://doi.org/10.1016/S1473-3099(21)00811-2

16) Conradie, F., Enwerem, M., Ferreira, H., Ismail, N., Hughes, J., Maartens, G., Master, I., Mentjies, G., Ndjeka, N., Reuter, A., te Riele, J., Romero, R., Padanilam, X., Variava, E., Schaaf, S. (2018). *Interim clinical guidance for the implementation of injectable-free regimens for rifampicin-resistant tuberculosis in adults, adolescents, and children.*: TBOnline.info Retrieved from https://www.tbonline.info/media/uploads/documents/dr_tb_clinical_guidelines_for_rsa_september_2018.pdf

17) Conradie, F., Bagdasaryan, T. R., Borisov, S., Howell, P., Mikiashvili, L., Ngubane, N., Samoilova, A., Skornykova, S., Tudor, E., Variava, E., Yablonskiy, P., Everitt, D., Wills, G. H., Sun, E., Olugbosi, M., Egizi, E., Li, M., Holsta, A., Timm, J., . . . ZeNix Trial, T. (2022). Bedaquiline-Pretomanid-Linezolid Regimens for Drug-Resistant Tuberculosis. *N Engl J Med*, *387*(9), 810-823. https://doi.org/10.1056/NEJMoa2119430

18) Conradie, F., Diacon, A. H., Ngubane, N., Howell, P., Everitt, D., Crook, A. M., Mendel, C. M., Egizi, E., Moreira, J., Timm, J., McHugh, T. D., Wills, G. H., Bateson, A., Hunt, R., Van Niekerk, C., Li, M., Olugbosi, M., Spigelman, M., & Nix, T. B. T. T. (2020). Treatment of Highly Drug-Resistant Pulmonary Tuberculosis. *N Engl J Med*, *382*(10), 893-902. https://doi.org/10.1056/NEJMoa1901814

19) Nyang'wa, B. T., Berry, C., Kazounis, E., Motta, I., Parpieva, N., Tigay, Z., Solodovnikova, V., Liverko, I., Moodliar, R., Dodd, M., Ngubane, N., Rassool, M., McHugh, T. D., Spigelman, M., Moore, D. A. J., Ritmeijer, K., du Cros, P., Fielding, K., & Collaborators, T.-P. S. (2022). A 24-Week, All-Oral Regimen for Rifampin-Resistant Tuberculosis. *N Engl J Med*, *387*(25), 2331-2343. https://doi.org/10.1056/NEJMoa2117166

20) WHO. (2021). *Global tuberculosis report: 2021*. World Health Organization. Retrieved September 21 from https://www.who.int/teams/global-tuberculosis-programme/tb-reports

21) Soedarsono, S., Mertaniasih, N. M., Kusmiati, T., Permatasari, A., Juliasih, N. N., Hadi, C., & Alfian, I. N. (2021). Determinant factors for loss to follow-up in drug-resistant tuberculosis patients: the importance of psycho-social and economic aspects. *BMC Pulm Med*, *21*(1), 360. https://doi.org/10.1186/s12890-021-01735-9

22) Kim, D. H., Kim, H. J., Park, S. K., Kong, S. J., Kim, Y. S., Kim, T. H., Kim, E. K., Lee, K. M., Lee, S. S., Park, J. S., Koh, W. J., Lee, C. H., & Shim, T. S. (2010). Treatment outcomes and survival based on drug resistance patterns in multidrug-resistant tuberculosis. *Am J Respir Crit Care Med*, *182*(1), 113-119. https://doi.org/10.1164/rccm.200911-1656OC

23) Gandhi, N. R., Shah, N. S., Andrews, J. R., Vella, V., Moll, A. P., Scott, M., Weissman, D., Marra, C., Lalloo, U. G., Friedland, G. H., Tugela Ferry, C., & Research, C. (2010). HIV coinfection in multidrug- and extensively drug-resistant tuberculosis results in high early

mortality. *Am J Respir Crit Care Med*, *181*(1), 80-86. https://doi.org/10.1164/rccm.200907-0989OC

24) WHO. *Tuberculosis profile: South Africa*. https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&entity_type=%22country%22&lan=%22EN%22&iso2=%22ZA%22

25) WHO. *Tuberculosis profile: Global.* https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&lan=%22EN%22&entity_type=%22group%22&group_code=%22global%22

26) Eker, B., Ortmann, J., Migliori, G. B., Sotgiu, G., Muetterlein, R., Centis, R., Hoffmann, H., Kirsten, D., Schaberg, T., Ruesch-Gerdes, S., Lange, C., & German, T. G. (2008). Multidrug- and extensively drug-resistant tuberculosis, Germany. *Emerg Infect Dis*, *14*(11), 1700-1706. https://doi.org/10.3201/eid1411.080729

27) Magnet, S., & Blanchard, J. S. (2005). Molecular insights into aminoglycoside action and resistance. *Chem Rev*, *105*(2), 477-498. https://doi.org/10.1021/cr0301088

28) Johansen, S. K., Maus, C. E., Plikaytis, B. B., & Douthwaite, S. (2006). Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs. *Mol Cell*, *23*(2), 173-182. https://doi.org/10.1016/j.molcel.2006.05.044

29) Nimmo, C., Millard, J., van Dorp, L., Brien, K., Moodley, S., Wolf, A., Grant, A. D., Padayatchi, N., Pym, A. S., Balloux, F., & O'Donnell, M. (2020). Population-level emergence of bedaquiline and clofazimine resistance-associated variants among patients with drug-resistant tuberculosis in southern Africa: a phenotypic and phylogenetic analysis. *Lancet Microbe*, *1*(4), e165-e174. https://doi.org/10.1016/S2666-5247(20)30031-8

30) Andries, K., Verhasselt, P., Guillemont, J., Gohlmann, H. W., Neefs, J. M., Winkler, H., Van Gestel, J., Timmerman, P., Zhu, M., Lee, E., Williams, P., de Chaffoy, D., Huitric, E., Hoffner, S., Cambau, E., Truffot-Pernot, C., Lounis, N., & Jarlier, V. (2005). A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science*, *307*(5707), 223-227. https://doi.org/10.1126/science.1106753

31) Zhang, S., Chen, J., Cui, P., Shi, W., Shi, X., Niu, H., Chan, D., Yew, W. W., Zhang, W., & Zhang, Y. (2016). Mycobacterium tuberculosis Mutations Associated with Reduced Susceptibility to Linezolid. *Antimicrob Agents Chemother*, *60*(4), 2542-2544. https://doi.org/10.1128/AAC.02941-15

32) Kim, S. J. (2005). Drug-susceptibility testing in tuberculosis: methods and reliability of results. *Eur Respir J*, *25*(3), 564-569. https://doi.org/10.1183/09031936.05.00111304

33) Kadura, S., King, N., Nakhoul, M., Zhu, H., Theron, G., Koser, C. U., & Farhat, M. (2020). Systematic review of mutations associated with resistance to the new and repurposed Mycobacterium tuberculosis drugs bedaquiline, clofazimine, linezolid, delamanid and pretomanid. *J Antimicrob Chemother*, *75*(8), 2031-2043. https://doi.org/10.1093/jac/dkaa136

34) Maus, C. E., Plikaytis, B. B., & Shinnick, T. M. (2005). Molecular analysis of cross-resistance to capreomycin, kanamycin, amikacin, and viomycin in Mycobacterium tuberculosis. *Antimicrob Agents Chemother*, *49*(8), 3192-3197. https://doi.org/10.1128/AAC.49.8.3192-3197.2005

35) Lee, M., Lee, J., Carroll, M. W., Choi, H., Min, S., Song, T., Via, L. E., Goldfeder, L. C., Kang, E., Jin, B., Park, H., Kwak, H., Kim, H., Jeon, H. S., Jeong, I., Joh, J. S., Chen, R. Y., Olivier, K. N., Shaw, P. A., . . . Barry, C. E., 3rd. (2012). Linezolid for treatment of

chronic extensively drug-resistant tuberculosis. *N Engl J Med*, *367*(16), 1508-1518. https://doi.org/10.1056/NEJMoa1201964

36) Ismail, N. A., Omar, S. V., Moultrie, H., Bhyat, Z., Conradie, F., Enwerem, M., Ferreira, H., Hughes, J., Joseph, L., Kock, Y., Letsaolo, V., Maartens, G., Meintjes, G., Ngcamu, D., Okozi, N., Padanilam, X., Reuter, A., Romero, R., Schaaf, S., . . . Ndjeka, N. (2022). Assessment of epidemiological and genetic characteristics and clinical outcomes of resistance to bedaquiline in patients treated for rifampicin-resistant tuberculosis: a cross-sectional and longitudinal study. *Lancet Infect Dis*, *22*(4), 496-506. https://doi.org/10.1016/S1473-3099(21)00470-9

37) Petrella, S., Cambau, E., Chauffour, A., Andries, K., Jarlier, V., & Sougakoff, W. (2006). Genetic basis for natural and acquired resistance to the diarylquinoline R207910 in mycobacteria. *Antimicrob Agents Chemother*, *50*(8), 2853-2856. https://doi.org/10.1128/AAC.00244-06

38) Almeida, D., Ioerger, T., Tyagi, S., Li, S. Y., Mdluli, K., Andries, K., Grosset, J., Sacchettini, J., & Nuermberger, E. (2016). Mutations in pepQ Confer Low-Level Resistance to Bedaquiline and Clofazimine in Mycobacterium tuberculosis. *Antimicrob Agents Chemother*, *60*(8), 4590-4599. https://doi.org/10.1128/AAC.00753-16

39) Zhang, S., Chen, J., Cui, P., Shi, W., Zhang, W., & Zhang, Y. (2015). Identification of novel mutations associated with clofazimine resistance in Mycobacterium tuberculosis. *J Antimicrob Chemother*, *70*(9), 2507-2510. https://doi.org/10.1093/jac/dkv150

40) Liu, C. H., Yang, N., Wang, Q., Hu, Y. L., Li, L., Zhang, G. Y., & Zhu, B. (2011). Risk factors associated with fluoroquinolone-resistant tuberculosis in a Beijing tuberculosis referral hospital. *Respirology*, *16*(6), 918-925. https://doi.org/10.1111/j.1440-1843.2011.01990.x

41) Dalton, T., Cegielski, P., Akksilp, S., Asencios, L., Campos Caoili, J., Cho, S. N., Erokhin, V. V., Ershova, J., Gler, M. T., Kazennyy, B. Y., Kim, H. J., Kliiman, K., Kurbatova, E., Kvasnovsky, C., Leimane, V., van der Walt, M., Via, L. E., Volchenkov, G. V., Yagui, M. A., . . . Viiklepp, P. (2012). Prevalence of and risk factors for resistance to second-line drugs in people with multidrug-resistant tuberculosis in eight countries: a prospective cohort study. *Lancet*, *380*(9851), 1406-1417. https://doi.org/10.1016/S0140-6736(12)60734-X

42) Ershova, J. V., Kurbatova, E. V., Moonan, P. K., & Cegielski, J. P. (2012). Acquired resistance to second-line drugs among persons with tuberculosis in the United States. *Clin Infect Dis*, *55*(12), 1600-1607. https://doi.org/10.1093/cid/cis748

43) Smith, S. E., Ershova, J., Vlasova, N., Nikishova, E., Tarasova, I., Eliseev, P., Maryandyshev, A. O., Shemyakin, I. G., Kurbatova, E., & Cegielski, J. P. (2015). Risk factors for acquisition of drug resistance during multidrug-resistant tuberculosis treatment, Arkhangelsk Oblast, Russia, 2005-2010. *Emerg Infect Dis*, *21*(6), 1002-1011. https://doi.org/10.3201/eid2106.141907

44) Lai, C. C., Tan, C. K., Huang, Y. T., Chou, C. H., Hung, C. C., Yang, P. C., Luh, K. T., & Hsueh, P. R. (2008). Extensively drug-resistant Mycobacterium tuberculosis during a trend of decreasing drug resistance from 2000 through 2006 at a Medical Center in Taiwan. *Clin Infect Dis*, *47*(7), e57-63. https://doi.org/10.1086/591702

45) Lai, C. C., Tan, C. K., Lin, S. H., Liao, C. H., Huang, Y. T., Chou, C. H., Hsu, H. L., Wang, C. Y., Lin, H. I., & Hsueh, P. R. (2010). Clinical and genotypic characteristics of extensively drug-resistant and multidrug-resistant tuberculosis. *Eur J Clin Microbiol Infect Dis*, *29*(5), 597-600. https://doi.org/10.1007/s10096-010-0874-6

46) Shah, N. S., Pratt, R., Armstrong, L., Robison, V., Castro, K. G., & Cegielski, J. P. (2008). Extensively drug-resistant tuberculosis in the United States, 1993-2007. *JAMA*, *300*(18), 2153-2160. https://doi.org/10.1001/jama.300.18.2153

47) Shin, S. S., Keshavjee, S., Gelmanova, I. Y., Atwood, S., Franke, M. F., Mishustin, S. P., Strelis, A. K., Andreev, Y. G., Pasechnikov, A. D., Barnashov, A., Tonkel, T. P., & Cohen, T. (2010). Development of extensively drug-resistant tuberculosis during multidrug-resistant tuberculosis treatment. *Am J Respir Crit Care Med*, *182*(3), 426-432. https://doi.org/10.1164/rccm.200911-1768OC

48) Bonilla, C. A., Crossa, A., Jave, H. O., Mitnick, C. D., Jamanca, R. B., Herrera, C., Asencios, L., Mendoza, A., Bayona, J., Zignol, M., & Jaramillo, E. (2008). Management of extensively drug-resistant tuberculosis in Peru: cure is possible. *PLoS One*, *3*(8), e2957. https://doi.org/10.1371/journal.pone.0002957

49) Saifullah, A., Mallhi, T. H., Khan, Y. H., Iqbal, M. S., Alotaibi, N. H., Alzarea, A. I., & Rasheed, M. (2021). Evaluation of risk factors associated with the development of MDR- and XDR-TB in a tertiary care hospital: a retrospective cohort study. *PeerJ*, *9*, e10826. https://doi.org/10.7717/peerj.10826

50) Ignatyeva, O., Balabanova, Y., Nikolayevskyy, V., Koshkarova, E., Radiulyte, B., Davidaviciene, E., Riekstina, V., Jaama, K., Danilovits, M., Popa, C. M., & Drobniewski, F. A. (2015). Resistance profile and risk factors of drug resistant tuberculosis in the Baltic countries. *Tuberculosis (Edinb)*, *95*(5), 581-588. https://doi.org/10.1016/j.tube.2015.05.018

51) Murase, Y., Maeda, S., Yamada, H., Ohkado, A., Chikamatsu, K., Mizuno, K., Kato, S., & Mitarai, S. (2010). Clonal expansion of multidrug-resistant and extensively drug-resistant

tuberculosis, Japan. *Emerg Infect Dis*, *16*(6), 948-954. https://doi.org/10.3201/eid1606.091844

52) Vilarica, A. S., Gomes, C., & Pina, J. (2008). Comparative analysis of multidrug-resistant tuberculosis and extensively drug-resistant tuberculosis - epidemiology and predictive factors. *Rev Port Pneumol*, *14*(6), 829-842. https://doi.org/10.1016/s2173-5115(08)70310-5

53) Sun, Z., Chao, Y., Zhang, X., Zhang, J., Li, Y., Qiu, Y., Liu, Y., Nie, L., Guo, A., & Li, C. (2008). Characterization of extensively drug-resistant Mycobacterium tuberculosis clinical isolates in China. *J Clin Microbiol*, *46*(12), 4075-4077. https://doi.org/10.1128/JCM.00822-08

54) Peterson, M. L., Gandhi, N. R., Clennon, J., Nelson, K. N., Morris, N., Ismail, N., Allana, S., Campbell, A., Brust, J. C. M., Auld, S. C., Mathema, B., Mlisana, K., Moodley, P., & Shah, N. S. (2019). Extensively drug-resistant tuberculosis 'hotspots' and sociodemographic associations in Durban, South Africa. *Int J Tuberc Lung Dis*, *23*(6), 720-727. https://doi.org/10.5588/ijtld.18.0575

55) Kliiman, K., & Altraja, A. (2009). Predictors of extensively drug-resistant pulmonary tuberculosis. *Ann Intern Med*, *150*(11), 766-775. https://doi.org/10.7326/0003-4819-150-11-200906020-00004

56) Porwal, C., Kaushik, A., Makkar, N., Banavaliker, J. N., Hanif, M., Singla, R., Bhatnagar, A. K., Behera, D., Pande, J. N., & Singh, U. B. (2013). Incidence and risk factors for extensively drug-resistant tuberculosis in Delhi region. *PLoS One*, *8*(2), e55299. https://doi.org/10.1371/journal.pone.0055299

57) Banerjee, R., Allen, J., Westenhouse, J., Oh, P., Elms, W., Desmond, E., Nitta, A., Royce, S., & Flood, J. (2008). Extensively drug-resistant tuberculosis in california, 1993-2006. *Clin Infect Dis*, *47*(4), 450-457. https://doi.org/10.1086/590009

58) Jeon, C. Y., Hwang, S. H., Min, J. H., Prevots, D. R., Goldfeder, L. C., Lee, H., Eum, S. Y., Jeon, D. S., Kang, H. S., Kim, J. H., Kim, B. J., Kim, D. Y., Holland, S. M., Park, S. K., Cho, S. N., Barry, C. E., 3rd, & Via, L. E. (2008). Extensively drug-resistant tuberculosis in South Korea: risk factors and treatment outcomes among patients at a tertiary referral hospital. *Clin Infect Dis*, *46*(1), 42-49. https://doi.org/10.1086/524017

59) Velayati, A. A., Masjedi, M. R., Farnia, P., Tabarsi, P., Ghanavi, J., ZiaZarifi, A. H., & Hoffner, S. E. (2009). Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran. *Chest*, *136*(2), 420-425. https://doi.org/10.1378/chest.08-2427

60) Punga, V. V., Jakubowiak, W. M., Danilova, I. D., Somova, T. R., Volchenkov, G. V., Kazionnyy, B. Y., Nemtsova, E. S., Kiryanova, E. V., & Kourbatova, E. V. (2009). Prevalence of extensively drug-resistant tuberculosis in Vladimir and Orel regions, Russia. *Int J Tuberc Lung Dis*, *13*(10), 1309-1312. https://www.ncbi.nlm.nih.gov/pubmed/19793439

61) WHO. (2016). *The use of molecular line probe assays for the detection of resistance to second-line anti-tuberculosis drugs. Policy guidance.* https://apps.who.int/iris/handle/10665/246131

62) Helb, D., Jones, M., Story, E., Boehme, C., Wallace, E., Ho, K., Kop, J., Owens, M. R., Rodgers, R., Banada, P., Safi, H., Blakemore, R., Lan, N. T., Jones-Lopez, E. C., Levi, M., Burday, M., Ayakaka, I., Mugerwa, R. D., McMillan, B., . . . Alland, D. (2010). Rapid

detection of Mycobacterium tuberculosis and rifampin resistance by use of on-demand, near-patient technology. *J Clin Microbiol*, *48*(1), 229-237. https://doi.org/10.1128/JCM.01463-09

63) Migliori, G. B., Lange, C., Centis, R., Sotgiu, G., Mutterlein, R., Hoffmann, H., Kliiman, K., De Iaco, G., Lauria, F. N., Richardson, M. D., Spanevello, A., Cirillo, D. M., & Group, T. S. (2008). Resistance to second-line injectables and treatment outcomes in multidrug-resistant and extensively drug-resistant tuberculosis cases. *Eur Respir J*, *31*(6), 1155-1159. https://doi.org/10.1183/09031936.00028708

64) Gler, M. T., Skripconoka, V., Sanchez-Garavito, E., Xiao, H., Cabrera-Rivero, J. L., Vargas-Vasquez, D. E., Gao, M., Awad, M., Park, S. K., Shim, T. S., Suh, G. Y., Danilovits, M., Ogata, H., Kurve, A., Chang, J., Suzuki, K., Tupasi, T., Koh, W. J., Seaworth, B., . . . Wells, C. D. (2012). Delamanid for multidrug-resistant pulmonary tuberculosis. *N Engl J Med*, *366*(23), 2151-2160. https://doi.org/10.1056/NEJMoa1112433

65) WHO. (2022). *Case notifications.* [CSV File.]. https://www.who.int/teams/global-tuberculosis-programme/data

66) Mlambo, C. K., Warren, R. M., Poswa, X., Victor, T. C., Duse, A. G., & Marais, E. (2008). Genotypic diversity of extensively drug-resistant tuberculosis (XDR-TB) in South Africa. *Int J Tuberc Lung Dis*, *12*(1), 99-104. https://www.ncbi.nlm.nih.gov/pubmed/18173885

67) Auld, S. C., Shah, N. S., Mathema, B., Brown, T. S., Ismail, N., Omar, S. V., Brust, J. C. M., Nelson, K. N., Allana, S., Campbell, A., Mlisana, K., Moodley, P., & Gandhi, N. R. (2018). Extensively drug-resistant tuberculosis in South Africa: genomic evidence supporting transmission in communities. *Eur Respir J*, *52*(4). https://doi.org/10.1183/13993003.00246-2018

68) Nelson, K. N., Jenness, S. M., Mathema, B., Lopman, B. A., Auld, S. C., Shah, N. S., Brust, J. C. M., Ismail, N., Omar, S. V., Brown, T. S., Allana, S., Campbell, A., Moodley, P., Mlisana, K., & Gandhi, N. R. (2020). Social Mixing and Clinical Features Linked With Transmission in a Network of Extensively Drug-resistant Tuberculosis Cases in KwaZulu-Natal, South Africa. *Clin Infect Dis*, *70*(11), 2396-2402. https://doi.org/10.1093/cid/ciz636

69) Kapwata, T., Morris, N., Campbell, A., Mthiyane, T., Mpangase, P., Nelson, K. N., Allana, S., Brust, J. C. M., Moodley, P., Mlisana, K., Gandhi, N. R., & Shah, N. S. (2017). Spatial distribution of extensively drug-resistant tuberculosis (XDR TB) patients in KwaZulu-Natal, South Africa. *PLoS One*, *12*(10), e0181797. https://doi.org/10.1371/journal.pone.0181797

70) Nelson, K. N., Shah, N. S., Mathema, B., Ismail, N., Brust, J. C. M., Brown, T. S., Auld, S. C., Omar, S. V., Morris, N., Campbell, A., Allana, S., Moodley, P., Mlisana, K., & Gandhi, N. R. (2018). Spatial Patterns of Extensively Drug-Resistant Tuberculosis Transmission in KwaZulu-Natal, South Africa. *J Infect Dis*, *218*(12), 1964-1973. https://doi.org/10.1093/infdis/jiy394

71) Nelson, K. N., Gandhi, N. R., Mathema, B., Lopman, B. A., Brust, J. C. M., Auld, S. C., Ismail, N., Omar, S. V., Brown, T. S., Allana, S., Campbell, A., Moodley, P., Mlisana, K., Shah, N. S., & Jenness, S. M. (2020). Modeling Missing Cases and Transmission Links in Networks of Extensively Drug-Resistant Tuberculosis in KwaZulu-Natal, South Africa. *Am J Epidemiol*, *189*(7), 735-745. https://doi.org/10.1093/aje/kwaa028

72) Wasserman, S., Louw, G., Ramangoaela, L., Barber, G., Hayes, C., Omar, S. V., Maartens, G., Barry, C., Song, T., & Meintjes, G. (2019). Linezolid resistance in patients with drug-

resistant TB and treatment failure in South Africa. *J Antimicrob Chemother*, *74*(8), 2377-2384. https://doi.org/10.1093/jac/dkz206

73) Kaufman, J. S. (2008). Social epidemiology. In *Modern Epidemiology* (3rd ed., pp. 532-548). Lippincott Williams & Wilkins.

74) Africa, S. S. (31 August 2022). *National poverty lines: 2022.* (Department of Statistics South Africa, Issue. https://www.statssa.gov.za/?page_id=1854&PPN=P0310.1&SCH=73308

75) O'Toole, R. F. (2022). Antibiotic resistance acquisition versus primary transmission in the presentation of extensively drug-resistant tuberculosis. *Int J Mycobacteriol*, *11*(4), 343-348. https://doi.org/10.4103/ijmy.ijmy_187_22

76) Brust, J. C., Gandhi, N. R., Carrara, H., Osburn, G., & Padayatchi, N. (2010). High treatment failure and default rates for patients with multidrug-resistant tuberculosis in KwaZulu-Natal, South Africa, 2000-2003. *Int J Tuberc Lung Dis*, *14*(4), 413-419. https://www.ncbi.nlm.nih.gov/pubmed/20202298

77) McNabb, K. C., Bergman, A., & Farley, J. E. (2021). Risk factors for poor engagement in drug-resistant TB care in South Africa: a systematic review. *Public Health Action*, *11*(3), 139-145. https://doi.org/10.5588/pha.21.0007

78) Moyo, S., Cox, H. S., Hughes, J., Daniels, J., Synman, L., De Azevedo, V., Shroufi, A., Cox, V., & van Cutsem, G. (2015). Loss from treatment for drug resistant tuberculosis: risk factors and patient outcomes in a community-based program in Khayelitsha, South Africa. *PLoS One*, *10*(3), e0118919. https://doi.org/10.1371/journal.pone.0118919

79) Gajee, R., Schnippel, K., Mthupha, N., Muzah, B., & Berhanu, R. (2016). Missed appointments among rifampicin-resistant tuberculosis (RR-TB) patients at a decentralised RR-TB

outpatient clinic in Johannesburg, South Africa. *S Afr Med J*, *106*(9), 912-917. https://doi.org/10.7196/SAMJ.2016.v106i9.10570

80) Kendall, E. A., Theron, D., Franke, M. F., van Helden, P., Victor, T. C., Murray, M. B., Warren, R. M., & Jacobson, K. R. (2013). Alcohol, hospital discharge, and socioeconomic risk factors for default from multidrug resistant tuberculosis treatment in rural South Africa: a retrospective cohort study. *PLoS One*, *8*(12), e83480. https://doi.org/10.1371/journal.pone.0083480

81) Senthilingam, M., Pietersen, E., McNerney, R., Te Riele, J., Sedres, P., Wilson, R., & Dheda, K. (2015). Lifestyle, attitudes and needs of uncured XDR-TB patients living in the communities of South Africa: a qualitative study. *Trop Med Int Health*, *20*(9), 1155-1161. https://doi.org/10.1111/tmi.12532

82) Oladimeji, O., Oladimeji, K. E., Nanjoh, M., Banda, L., Adeleke, O. A., Apalata, T., Mbokazi, J., & Hyera, F. L. M. (2022). Contributory Factors to Successful Tuberculosis Treatment in Southwest Nigeria: A Cross-Sectional Study. *Trop Med Infect Dis*, *7*(8). https://doi.org/10.3390/tropicalmed7080194

83) Edessa, D., Adem, F., Hagos, B., & Sisay, M. (2021). Incidence and predictors of mortality among persons receiving second-line tuberculosis treatment in sub-Saharan Africa: A meta-analysis of 43 cohort studies. *PLoS One*, *16*(12), e0261149. https://doi.org/10.1371/journal.pone.0261149

84) Varshney, K., Anaele, B., Molaei, M., Frasso, R., & Maio, V. (2021). Risk Factors for Poor Outcomes Among Patients with Extensively Drug-Resistant Tuberculosis (XDR-TB): A Scoping Review. *Infect Drug Resist*, *14*, 5429-5448. https://doi.org/10.2147/IDR.S339972

85) Tiberi, S., Utjesanovic, N., Galvin, J., Centis, R., D'Ambrosio, L., van den Boom, M., Zumla, A., & Migliori, G. B. (2022). Drug resistant TB - latest developments in epidemiology, diagnostics and management. *Int J Infect Dis*, *124 Suppl 1*, S20-S25. https://doi.org/10.1016/j.ijid.2022.03.026

86) Rothman, K. J., Greenland, S., & Lash, T.L. (2008). *Modern epidemiology* (3rd ed.). Lippincott Williams & Wilkins.

87) McNamee, R. (2005). Regression modelling and other methods to control confounding. *Occup Environ Med*, *62*(7), 500-506, 472. https://doi.org/10.1136/oem.2002.001115

88) Kleinbaum, D. G., & Klein, M. (2010). *Logistic regression: A self-learning text.* (3rd ed.). Springer Science+Business Media.

89) NHLS. (2018). Standardisation of phenotypic drug susceptibility testing across NHLS TB culture laboratories. In N. H. L. Service (Ed.), (pp. 3). Sandringham, South Africa: Centre for Tuberculosis.

90) Davis, C. G., Thake, J., & Vilhena, N. (2010). Social desirability biases in self-reported alcohol consumption and harms. *Addict Behav*, *35*(4), 302-311. https://doi.org/10.1016/j.addbeh.2009.11.001

91) Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality and Quantity*, *2013*(47), 2025-2047. https://doi.org/10.1007/s11135-011-9640-9

92) Mvelase, N. R., Balakrishna, Y., Lutchminarain, K., & Mlisana, K. (2019). Evolving rifampicin and isoniazid mono-resistance in a high multidrug-resistant and extensively drug-resistant tuberculosis region: a retrospective data analysis. *BMJ Open*, *9*(11), e031663. https://doi.org/10.1136/bmjopen-2019-031663

93) Guthrie, T., Chaitkin, M., Khoza, N., Zulu, N., Madisha, V., Ndlovu, N., Shezi, S., Karume, J., Motsoeneng, P., Simelane, S., Meyer-Rath, G., Masuku, S., Jaimeson, L., & Ghai, K. (2018). *Consolidated spending on HIV and TB in South Africa (2014/15-2016/17)*. R. f. D. I. National Department of Health; Health Finance & Governance Project. https://www.r4d.org/wp-content/uploads/South-Africa-HIV-and-TB-Expenditure-Review-2014-15-2016-17-Full-Report_vf.pdf?_ga=2.231321887.1619095278.1680407571-1576026351.1680407571

94) Hernan, M. A. (2022). Causal analyses of existing databases: no power calculations required. *J Clin Epidemiol*, *144*, 203-205. https://doi.org/10.1016/j.jclinepi.2021.08.028

95) Fontes Marx, M., London, L., Harker, N., & Ataguba, J. E. (2021). Assessing Intertemporal Socioeconomic Inequalities in Alcohol Consumption in South Africa. *Front Public Health*, *9*, 606050. https://doi.org/10.3389/fpubh.2021.606050

96) Sreeramareddy, C. T., & Acharya, K. (2021). Trends in Prevalence of Tobacco Use by Sex and Socioeconomic Status in 22 Sub-Saharan African Countries, 2003-2019. *JAMA Netw Open*, *4*(12), e2137820. https://doi.org/10.1001/jamanetworkopen.2021.37820

97) Sreeramareddy, C. T., Pradhan, P. M., & Sin, S. (2014). Prevalence, distribution, and social determinants of tobacco use in 30 sub-Saharan African countries. *BMC Med*, *12*, 243. https://doi.org/10.1186/s12916-014-0243-x

98) Ugwu, C. L. J., & Ncayiyana, J. R. (2022). Spatial disparities of HIV prevalence in South Africa. Do sociodemographic, behavioral, and biological factors explain this spatial variability? *Front Public Health*, *10*, 994277. https://doi.org/10.3389/fpubh.2022.994277

99) van de Water, B. J., Silva, S. G., Prvu Bettger, J., Humphreys, J., Cunningham, C. K., & Farley, J. E. (2018). Provision of guideline-based care for drug-resistant tuberculosis in South

Africa: Level of concordance between prescribing practices and guidelines. *PLoS One*, *13*(11), e0203749. https://doi.org/10.1371/journal.pone.0203749

100) Townsend, L., Flisher, A. J., Gilreath, T., & King, G. (2006). A systematic literature review of tobacco use among adults 15 years and older in sub-Saharan Africa. *Drug Alcohol Depend*, *84*(1), 14-27. https://doi.org/10.1016/j.drugalcdep.2005.12.008

101) Galal, S. (8 December 2022). *Unemployment rate in South Africa 2019-2022, by age group*. https://www.statista.com/statistics/1129482/unemployment-rate-by-age-group-in-south-africa/#:~:text=In%20the%20third%20quarter%20of,force%20participation%20in%20the%20country.

102) Africa, S. S. (1 June 2022). *South Africa's youth continues to bear the burden of unemployment.* https://www.statssa.gov.za/?p=15407

103) Necho, M., Belete, A., & Getachew, Y. (2020). The prevalence and factors associated with alcohol use disorder among people living with HIV/AIDS in Africa: a systematic review and meta-analysis. *Subst Abuse Treat Prev Policy*, *15*(1), 63. https://doi.org/10.1186/s13011-020-00301-6

104) Chersich, M. F., & Rees, H. V. (2010). Causal links between binge drinking patterns, unsafe sex and HIV in South Africa: its time to intervene. *Int J STD AIDS*, *21*(1), 2-7. https://doi.org/10.1258/ijsa.2000.009432

105) Schneider, M., Chersich, M., Temmerman, M., Degomme, O., & Parry, C. D. (2014). The impact of alcohol on HIV prevention and treatment for South Africans in primary healthcare. *Curationis*, *37*(1), 1137. https://doi.org/10.4102/curationis.v37i1.1137

106) Matzopoulos, R., Cois, A., Probst, C., Parry, C. D. H., Vellios, N., Sorsdahl, K., Joubert, J. D., Pillay-van Wyk, V., Bradshaw, D., & Pacella, R. (2022). Estimating the changing

burden of disease attributable to alcohol use in South Africa for 2000, 2006 and 2012. *S Afr Med J*, *112*(8b), 662-675. https://doi.org/10.7196/SAMJ.2022.v112i8b.16487

107) OECD. (2019). *Education at a glance: OECD indicators: South Africa.* https://www.oecd.org/education/education-at-a-glance/EAG2019_CN_ZAF.pdf

108) AIDSInfo. *Country factsheets: South Africa: 2021.* UNAIDS. https://aidsinfo.unaids.org/

109) McKinnon, L. R., & Karim, Q. A. (2016). Factors Driving the HIV Epidemic in Southern Africa. *Curr HIV/AIDS Rep*, *13*(3), 158-169. https://doi.org/10.1007/s11904-016-0314-z

110) Peltzer, K., Davids, A., & Njuho, P. (2011). Alcohol use and problem drinking in South Africa: findings from a national population-based survey. *Afr J Psychiatry (Johannesbg)*, *14*(1), 30-37. https://doi.org/10.4314/ajpsy.v14i1.65466

111) Groenewald, P., Pacella, R., Sitas, F., Awotiwon, O. F., Vellios, N., Van Rensburg, C. J., Manda, S., Laubscher, R., Nojilana, B., Joubert, J. D., Labadarios, D., Ayo-Yusuf, L., Roomaney, R. A., Turawa, E. B., Neethling, I., Abdelatif, N., Pillay-van Wyk, V., & Bradshaw, D. (2022). Estimating the changing disease burden attributable to smoking in South Africa for 2000, 2006 and 2012. *S Afr Med J*, *112*(8b), 649-661. https://doi.org/10.7196/SAMJ.2022.v112i8b.16492

112) Lewis, L., Kharsany, A. B. M., Humphries, H., Maughan-Brown, B., Beckett, S., Govender, K., Cawood, C., Khanyile, D., & George, G. (2022). HIV incidence and associated risk factors in adolescent girls and young women in South Africa: A population-based cohort study. *PLoS One*, *17*(12), e0279289. https://doi.org/10.1371/journal.pone.0279289

113) Africa, S. S. (23 August 2022). *Nearly half of SA women are out of the labour force in Q2:2022*. https://www.statssa.gov.za/?p=15668

114) Maloma, I. (2016). The socioeconomic determinants of household poverty status in a low-income settlement in South Africa. *International Journal of Social Sciences and Humanity Studies*, *8*, 122-131.

115) Soepnel, L. M., Kolkenbeck-Ruh, A., Crouch, S. H., Draper, C. E., Ware, L. J., Lye, S. J., & Norris, S. A. (2022). Prevalence and socio-structural determinants of tobacco exposure in young women: Data from the Healthy Trajectories Initiative (HeLTI) study in urban Soweto, South Africa. *Drug Alcohol Depend*, *232*, 109300. https://doi.org/10.1016/j.drugalcdep.2022.109300

116) Sauer, C. M., Sasson, D., Paik, K. E., McCague, N., Celi, L. A., Sanchez Fernandez, I., & Illigens, B. M. W. (2018). Feature selection and prediction of treatment failure in tuberculosis. *PLoS One*, *13*(11), e0207491. https://doi.org/10.1371/journal.pone.0207491

117) Magee, M. J., Kempker, R. R., Kipiani, M., Tukvadze, N., Howards, P. P., Narayan, K. M., & Blumberg, H. M. (2014). Diabetes mellitus, smoking status, and rate of sputum culture conversion in patients with multidrug-resistant tuberculosis: a cohort study from the country of Georgia. *PLoS One*, *9*(4), e94890. https://doi.org/10.1371/journal.pone.0094890

118) Megbowon, E. T., David, O. O., & Makhalima, J. L. (2022). Behavioral Risk Factor and Primary Healthcare Utilization in South Africa. *Healthcare (Basel)*, *10*(11). https://doi.org/10.3390/healthcare10112186

119) Welekidan, L. N., Skjerve, E., Dejene, T. A., Gebremichael, M. W., Brynildsrud, O., Agdestein, A., Tessema, G. T., Tonjum, T., & Yimer, S. A. (2020). Characteristics of pulmonary multidrug-resistant tuberculosis patients in Tigray Region, Ethiopia: A cross-

sectional study. *PLoS One*, *15*(8), e0236362. https://doi.org/10.1371/journal.pone.0236362

120) Hermans, S. M., Zinyakatira, N., Caldwell, J., Cobelens, F. G. J., Boulle, A., & Wood, R. (2021). High Rates of Recurrent Tuberculosis Disease: A Population-level Cohort Study. *Clin Infect Dis*, *72*(11), 1919-1926. https://doi.org/10.1093/cid/ciaa470

121) Johnson, J. E., Carney, T., Kline, T., Browne, F. A., & Wechsberg, W. M. (2012). Incarceration history relative to health, substance use, and violence in a sample of vulnerable South African women: implications for health services in criminal justice settings. *Subst Abuse Rehabil*, *3*(Suppl 1), 59-69. https://doi.org/10.2147/SAR.S21351

122) UNODC. *HIV/AIDS Prevention, Care Treatment and Support in Prison Settings in Southern Africa.* United Nations Office on Drugs and Crime. https://www.unodc.org/southernafrica/en/hiv/prison-settings_index.html

123) Christodoulou, J., Stokes, L.R., Bantjes, J., Tomlinson, M., Stewart, J., Rabie, S., Gordon, S., Mayekiso, A., & Rotheram-Borus, M.J. (2019). Community context and individual factors associated with arrests among young men in a South African township. . *PLoS One*, *14*(1), e0209073. https://doi.org/10.1371/journal.pone.0209073

124) Jonck, P., Goujon, A., Testa, M.R., Kandala, J. (2015). Education and crime engagement in South Africa: A national and provincial perspective. *International Journal of Educational Development*, *45*(2015), 141-151. https://doi.org/https://doi.org/10.1016/j.ijedudev.2015.10.002

125) Andrews, J. R., Shah, N. S., Weissman, D., Moll, A. P., Friedland, G., & Gandhi, N. R. (2010). Predictors of multidrug- and extensively drug-resistant tuberculosis in a high HIV

prevalence community. *PLoS One*, *5*(12), e15735. https://doi.org/10.1371/journal.pone.0015735

126) Denman, D. C., Baldwin, A. S., Betts, A. C., McQueen, A., & Tiro, J. A. (2018). Reducing "I Don't Know" Responses and Missing Survey Data: Implications for Measurement. *Med Decis Making*, *38*(6), 673-682. https://doi.org/10.1177/0272989X18785159