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:

_____ _____

Amy Whitesell Date

Multisite Evaluation of Phenotypic *Mycobacterium tuberculosis* Drug Susceptibility Testing Methods

By

Amy Whitesell
Master of Public Health

Epidemiology

_____

Kenneth G. Castro, MD
Committee Chair

_____

Angela M. Starks, PhD
Committee Member

Multisite Evaluation of Phenotypic *Mycobacterium tuberculosis* Drug Susceptibility Testing Methods

By

Amy Whitesell

B.A.
University of North Carolina at Chapel Hill
2016

Thesis Committee Chair: Kenneth G. Castro, MD, FIDSA

Thesis Committee Member: Angela M. Starks, PhD

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 Public Health
in Epidemiology
2018

Abstract

Multisite Evaluation of Phenotypic *Mycobacterium tuberculosis* Drug Susceptibility Testing Methods
By Amy Whitesell

Drug susceptibility testing (DST) of *Mycobacterium tuberculosis* (MTB) is crucial in rapidly detecting and eliminating drug resistance. However, phenotypic DST methods have proven difficult to standardize, and previous studies have shown that discordant DST results may occur between and within laboratories based on multiple factors including method, drug, concentration, and genetic mutation(s) present in the isolate. To further investigate this discordance, an expanded Model Performance Evaluation Program (MPEP) study was conducted by the Centers for Disease Control and Prevention (CDC) Division of Tuberculosis Elimination, Laboratory Branch and the Association of Public Health Laboratories (APHL). Thirty MTB isolates previously characterized by CDC DST methods were available, including 8 sets of duplicates. These were sent to a total of 12 public health and clinical laboratories where phenotypic DST was performed by three different methods: indirect agar proportion method, BACTEC™ mycobacterial growth indicator tubes (MGIT) 960™, and Sensititre™ *Mycobacterium tuberculosis* MIC plate. Molecular detection of drug resistance (MDDR) was also performed by sequencing target loci to detect mutations associated with drug resistance. Site-specific agreement with a growth-based, expected result as well as agreement with a composite result incorporating resistance-conferring mutations was assessed. A Fleiss kappa estimate was calculated in order to measure interlaboratory agreement independent of previous characterization. Agreement was found to vary by site, method, drug, and concentration. Interlaboratory agreement and agreement to the expected, growth-based result was highest for amikacin and kanamycin (range: 0.90 [95% CI: 0.71, 1.00], 1.00 [95% CI: 1.00, 1.00]) and lowest for ethambutol and para-aminosalicylic acid. Agreement with the composite result was worse for rifampin and rifabutin compared to agreement with the expected results, and improved for capreomycin. Many findings were consistent with previously described patterns of discordance, and further supported evidence that discordant results can be observed across multiple methods, drugs, and strains.

Multisite Evaluation of Phenotypic *Mycobacterium tuberculosis* Drug Susceptibility Testing Methods

By

Amy Whitesell

B.A.
University of North Carolina at Chapel Hill
2016

Thesis Committee Chair: Kenneth G. Castro, MD, FIDSA

Thesis Committee Member: Angela M. Starks, PhD

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 Public Health
in Epidemiology
2018

TABLE OF CONTENTS

CHAPTER I: BACKGROUND

I. TUBERCULOSIS & DRUG RESISTANCE

Tuberculosis (TB) is a major public health concern and one of the leading causes of death worldwide, despite the existence of effective treatment capable of curing the disease with timely diagnosis and proper administration. In 2016, there were an estimated 10.4 million cases and 1.3 million deaths attributable to TB, making it the leading cause of death due to a single infectious agent (1). Currently, the standard treatment for drug-susceptible (DS) TB is a six-month regimen consisting of the four first-line antituberculosis drugs: rifampin (RMP), isoniazid (INH), ethambutol (EMB), and pyrazinamide (PZA). First-line treatment typically requires a two-month intensive phase involving a daily dose of all four drugs, followed by a four-month continuation phase that includes only RMP and INH; however, the length of treatment ultimately depends on patient response as well as associated risk factors present (1, 2). With proper and timely administration, treatment of DS-TB has a reported >90% cure rate, and in 2016, had an 85% global completion rate (1, 3).

Although global cases of TB have been in decline, the emergence of drug resistance jeopardizes progress made by TB control and elimination programs and threatens to become a greater issue as resistant strains become a larger proportion of TB cases in the future (1, 4). While any degree of drug resistance increases the likelihood of poor treatment outcomes compared to DS-TB, there are two definitions of particular clinical significance regarding drug resistant (DR) TB (5). Multidrug-resistant TB (MDR-TB) refers to TB that is resistant to at least the two most effective first-line drugs, RMP and INH, and extensively drug-resistant TB (XDR-TB) refers to MDR-TB that has

additional resistance to at least one fluoroquinolone (FQ) and at least one second-line injectable (SLI) (1). RMP-resistant TB (RR-TB) is also of particular clinical significance, as RMP resistance alone is an important indicator of treatment failure (1, 6).

In 2016, there were an estimated 600,000 cases of RR/MDR-TB worldwide, including 490,000 estimated MDR-TB cases, and an estimated 240,000 deaths attributable to RR/MDR-TB (1). Reports of RR-TB have increased in recent years due to the increased global use of the Xpert MTB/RIF® assay, which allows for rapid and simultaneous detection of TB and RMP resistance (1). Notably, RR-TB incidence may be indicative of undetected MDR-TB, as further INH susceptibility results may not be available in certain settings (1). Additionally, by the end of the year, 123 countries had reported at least one case of XDR-TB (1). MDR-TB and XDR-TB are associated with higher morbidity and mortality, and require a lengthy treatment regimen utilizing expensive second-line drugs with higher toxicity, lower efficacy, and more adverse side-effects (1, 7–9). Compared to patients with drug-susceptible strains, MDR-TB patients have a significantly higher all-cause mortality rate and a lower rate of recurrence-free survival (10, 11). Additional resistance to second-line drugs has been found to significantly increase mortality risk; therefore, patients with XDR-TB have even lower rates of survival and worse long-term outcomes compared to MDR-TB (7, 12–15).

Treatment of DR-TB ideally consists of a combination of effective second-line drugs and any remaining first-line drugs to which the strain is susceptible. A regimen typically consists of at least 5 drugs, and should always include at least 4 that are known to be effective (8, 16). However, this individualized treatment requires knowledge of drug susceptibility, which is not always available, especially for second-line drugs. In

areas without regular second-line drug susceptibility testing (DST), a standardized MDR-TB treatment may be utilized based on the estimated prevalence of resistance in a population (17). The majority of second-line treatment regimens were designed to last 20 months or longer, with many lasting 2 years or more, depending on the patient (8,16). Recently, a shorter MDR treatment course lasting 9–12 months has been recommended for patients who have not been previously treated with second-line drugs and for whom second-line resistance is not expected (1, 16). These longer regimens involving a larger number of drugs are associated with lower adherence and treatment success rates. Compared to an overall treatment success rate of 83% for global TB cases, the success rate for RR/MDR-TB treatment is 54% and is even lower for XDR-TB, with less than a third of XDR patients (30%) reported to have completed treatment in 2016 (1).

Additionally, the financial burden of DR-TB is significantly higher compared to that of DS-TB. A study conducted between 2005 and 2007 by the Centers for Disease Control and Prevention (CDC) abstracted patient data from three high-burden states in the United States and found the average direct cost of MDR-treatment was $134,000, which is nearly an 8-fold increase compared to the average DS-TB treatment cost of $17,000. Furthermore, the average cost of XDR-TB treatment was $430,000, over three times higher than MDR-TB treatment (18).

Errors in the prescription of drug regimens, management of drug supply, patient management, and treatment delivery processes can all contribute to the evolution of DR-TB strains (19). Therefore, resistance has emerged on multiple occasions worldwide, frequently originating in areas of both high TB incidence and improper treatment practices (5, 17, 20). Differing from other bacteria in that horizontal gene transfer of

resistance plasmids does not contribute to the acquisition and spread of drug resistance, DR-TB strains are caused by the slow, adaptive process of these errors enacting selective pressures that allow strains with resistance conferring mutations to survive and multiply (5, 21). These mutations are primarily found in loci associated with inhibiting direct interaction between the drug and its target, preventing prodrug conversion, or affecting other mechanisms that diminish the impact of the drug, such as overexpression of the drug target (5, 21). Resistance to multiple drugs, therefore, requires the sequential acquisition of multiple resistance conferring mutations, usually at multiple target loci (5). In addition to DR-TB developing in patients due to inappropriate previous treatment, referred to as acquired resistance, drug resistant strains can also be spread via person-to-person transmission from someone with TB disease due to drug resistant strains, referred to as primary resistance (19).

To effectively combat both the emergence of resistance and transmission of resistant strains, it is necessary to ensure the rapid detection and elimination of both DS and DR-TB (5). However, major gaps still exist in the detection of DR-TB (1, 22). Delays in diagnosis are common, and in 2016, only an estimated 26% of global RR/MDR-TB cases were detected and reported to national surveillance systems (1, 22). These gaps are likely wider for detection of XDR-TB, as it requires second-line DST in order to diagnose, which is not readily available in many low-resource and high-burden settings. Failure to detect drug-resistance can lead to the utilization of ineffective treatment regimens that can cause further acquisition of resistance and allow for ongoing transmission of the strain. This has been documented in South Africa, where the use of a standardized MDR treatment regimen, rather than routine susceptibility testing, led to

improper treatment of undetected pre-XDR TB, causing the emergence and transmission of XDR-TB as well as cases of "totally drug resistant" (TDR) TB, which is sometimes used to describe strains resistant to all first- and second-line drugs tested (17, 23). This exacerbation of drug resistance due to the lack of susceptibility results emphasizes the importance of accurate and rapid detection of resistance in TB control and elimination.

II. DRUG SUSCEPTIBILITY TESTING OF TB

DST is essential in helping clinicians determine optimal drug regimens for patients, in turn preventing transmission of disease and acquisition of additional resistance, as well as accurately assessing the burden of drug-resistant strains in a population (12, 24, 25). DST is recommended for all TB patients, and universal coverage is a major component of the World Health Organization (WHO) End TB Strategy, as it is an important tool in detecting and eradicating resistance as it emerges (1, 5, 16).

While coverage of DST, particularly for RMP, has been increasing globally since 2009, coverage varies from country to country, and there is still a great need for strengthening laboratory capacity in resource-limited settings to guarantee timely and accurate results (1, 12, 22). Second-line DST is necessary for the detection of XDR-TB and individualized second-line treatment. However, standardization of second-line DST has been a slow process, and there are still gaps in knowledge concerning the correlation between second-line DST results and clinical outcomes, leading to poor clinical predictive values for some drugs (25, 26).

DST can be conducted utilizing phenotypic or molecular methods (1). Molecular DST methods typically involve the detection of mutations associated with the resistant

phenotype at specific loci in the *Mycobacterium tuberculosis* (MTB) genome (25, 27). These methods have the potential for shorter turnaround times (TATs) as well as accurate results (5, 27). Molecular methods may, in some cases, offer better predictability of resistance due to the association of certain genetic loci and mutations with variations in phenotypic expression of resistance that may not be detected by phenotypic methods, such as low-level resistance (5, 28). However, there are several limitations associated with these methods, including insufficient coverage of the MTB genome by some currently available assays, lack of knowledge concerning resistance conferring loci and mutations, and the lack of an easily deployable point of care test especially those associated with second-line drugs. For these reasons, molecular methods have often been recommended as an adjunct to phenotypic methods (25, 27, 29, 30). Despite these potential barriers, some molecular methods, such as the Xpert MTB/RIF® assay, may be easier to implement as they may not require the complex biosafety infrastructure needed for phenotypic methods and may involve limited training by comparison (1, 31). Recently, methods such as whole genome sequencing allow for a thorough analysis of known and unknown mutations associated with drug resistance and have the potential to detect the emergence of resistance to newer drugs as it occurs, allowing new treatment regimens to remain effective for a longer period of time (5).

Most phenotypic DST methods, also referred to as growth-based methods, determine an isolate's susceptibility by observing growth of MTB isolates in the presence of a specified concentration of drug (25, 28).

Indirect Agar Proportion

The indirect agar proportion method (APM) is considered the gold standard method and is typically used as a reference when evaluating new methods. This method involves inoculating solid agar containing a critical concentration (CC) of drug with MTB and calculating the proportion of growth seen on the drug-containing agar compared to the growth seen on an inoculated drug-free agar. An isolate is considered resistant if there is observed growth greater than or equal to 1% of the growth seen on the drug-free agar (25, 32). This is based on the theory that all wild-type (WT) strains contain some mutants with antimicrobial resistance, and the difference between clinically resistant and susceptible strains is the proportion of these mutants present (32).

The use of CCs differs from many other microbiology procedures that test a series of drug dilutions in order to determine the minimum inhibitory concentration (MIC), defined as the lowest concentration to inhibit visible growth of MTB *in vitro* (25, 28). In contrast, the CC for a given drug is defined as the concentration that inhibits ≥95% of WT strains that have never been exposed to antituberculosis medication while allowing growth of resistant strains (28). This concentration can be determined by the epidemiological cut-off (ECOFF) value, which is defined as the highest MIC value within a WT distribution and, hypothetically, corresponds to the concentration that best discriminates between resistant and susceptible isolates (25, 28, 33, 34). Therefore, if the recommended CC for a given drug significantly differs from the corresponding ECOFF value, it can lead to poor correlation between DST results and clinical outcomes (28, 34). Originally, CCs were established using Löwenstein–Jensen (LJ) media. Equivalent

critical concentrations were later established for other media, such as Middlebrook 7H10 and 7H11, as well as liquid media and are expected to provide equivalent results to those obtained using the LJ proportion method (25).

While this method is typically thought to produce results that correlate well with clinical outcomes, TATs associated with APM are long, usually 1–2 months from time of MTB identification (25, 29). Additionally, solid media is not recommended in testing the first-line drug PZA due to the acidic conditions required for optimal drug performance (25). Some discrepancies have been reported for APM DST, including some possible differences between 7H10, 7H11, and LJ media, which could be affected by site-specific differences in laboratory procedures as well as failure to detect low-levels of resistance that may be clinically significant (35–37).

BACTEC MGIT 960

Automated broth systems, such as the BACTEC™ mycobacterial growth indicator tubes (MGIT) 960™, allow for shorter TATs of about one week, but are not FDA approved for testing second-line agents (25, 29, 33). The MGIT system was originally evaluated using the radiometric BACTEC™ 460 TB system (BACTEC 460) as a reference, rather than the proportion method, and therefore, may not be well calibrated to clinical results (29, 38). Concerns related to the radioactivity associated with the BACTEC 460 system eventually led to its replacement with the MGIT system, which had overall high agreement with BACTEC 460 despite some discrepancies reported between the two methods (38–40). Additionally, both systems have a low associated sensitivity in detecting EMB resistance as well as borderline RMP resistance compared to APM, which

has led to the questioning of MGIT results and their reliability (6, 29, 38). This method has also been associated with higher rates of contamination compared to other methods, leading to repeat testing and delayed results (38). MGIT may not be practical in all settings, particularly those with fewer resources, due to the high cost, and need for technical support, stable electricity, and trained personnel (29, 30).

Sensititre MIC Plate

The Sensititre™ *Mycobacterium tuberculosis* MIC plate (Sensititre) allows the simultaneous testing of 12 first- and second-line antituberculosis drugs included on the plate, determining the MIC values associated with each (41). Both the APM and MGIT methods return qualitative results, classifying an isolate as either "resistant" or "susceptible," which may not capture the true range of resistance that can be seen clinically (42). The quantitative result returned by Sensititre could potentially better inform clinicians of the degree of resistance associated with a particular strain (41–43). In some cases, the dose of a drug may be increased in the presence of a MIC value indicating borderline resistance, allowing the inclusion of more effective drugs in a given regimen. Alternatively, if an MIC value indicates a high level of susceptibility, the dose of a drug may be reduced in order to avoid adverse side-effects (42).

Studies have shown high agreement between APM and Sensititre, with a few exceptions regarding certain drugs (21, 26, 41, 44, 45). Additionally, the MIC plate has a 2-year shelf-life, is quality controlled by the manufacturer, and can utilize the same inoculum as rapid broth testing methods (41). However, this methodology is not yet FDA approved and standardized MIC breakpoints have not been established for this method,

making it difficult to determine whether an isolate should be reported as resistant or susceptible (41, 42, 45). Previous studies have used recommended CCs for Middlebrook 7H10 APM as breakpoints to determine susceptibility and resistance. However, some CCs, such as those for INH and EMB, are not included on the MIC plate, which may explain some discordance observed between methods for these drugs (44, 45).

III. DISCORDANCE

These phenotypic methods have proven difficult to standardize, and previous studies have shown that discordance between DST results can occur between and within laboratories as well as methods (21, 26, 35–37, 41, 43–46). Discordant results not only cause confusion for a clinician, but can delay proper treatment in individuals, increasing the likelihood of transmission, additional resistance, and treatment failure (7, 23, 37, 47). While differences in laboratory procedures may cause some level of disagreement observed between sites, previous analyses of DST data have indicated an overall high-level of agreement among proficient laboratories, with errors typically pertaining to certain combinations of methodology, drug, and strain (29, 35, 45).

Certain drugs are thought to have higher rates of these errors due to their established CCs failing to accurately discriminate between the MIC distributions of WT and resistant strains (48). This would lead to oscillation between resistant and susceptible results and poor reproducibility, even within a laboratory (26, 45). Additionally, qualitative DST results fail to capture the heterogeneity of resistance phenotypes, with low-level but still clinically relevant resistance difficult to detect by growth-based methods (6, 28, 36, 37). In these cases, isolates tend to have MIC values close to the CC;

therefore, DST results may not correspond well with the clinical outcomes, as isolates characterized as susceptible may or may not respond to treatment (28). The issue of correlating DST results with clinical outcomes becomes more complicated when accounting for the fact that CCs utilized by DST methods do not necessarily relate to drug concentrations *in vivo*, and that treatment consists of combination therapy, leading to a possible disconnect between singular *in vitro* DST results and combined *in vivo* clinical outcomes (28).

Rifampin & Rifabutin

RMP is the most effective first-line drug and a key determinant of treatment success (6). Acting as a bactericidal agent, RMP binds to the β-subunit of RNA polymerase and disrupts RNA synthesis within MTB cells (49). Overall, RMP DST is considered the most reliable among all antituberculosis drugs, with typically high concordance and reproducibility observed between and within laboratories across multiple methods (29, 30, 43, 45, 46). The majority of RMP-resistant strains, up to 96%, have a mutation within the rifampin resistance-determining region (RRDR) of *rpoB* (50). Mutations in this region are typically associated with high-level resistance, but certain mutations have been found to confer clinically significant, low-level resistance and are highly associated with discordant results (6, 27, 28). While this failure to detect low-level resistance has been observed across multiple phenotypic methods, the discordance has been particularly evident in systems utilizing liquid culture, such as MGIT (21, 35, 37). One study noted the mutations missed by MGIT were observed to typically fall at the end or outside of the *rpoB* core region, at positions 511, 533, and 572 (37). It has been

hypothesized that these mutations likely hinder the fitness of the isolate to the point that growth may be undetected under certain circumstances (37).

Rifabutin (RBT) is another rifamycin, belonging to the same group of antibiotics as RMP, and is of particular importance in treating TB in HIV-infected individuals due to fewer associated drug-drug interactions (49, 51). Since both drugs target the same mechanism, mutations in the *rpoB* locus can also confer resistance to RBT (49). Therefore, cross-resistance of RBT and RMP is common, and correlated MIC values have been reported within strains (21, 49, 52). However, not all *rpoB* mutations confer resistance to both drugs, as some RMP resistant strains are not resistant to RBT (52). Discordance has been previously observed among RBT DST results, with one study reporting susceptible MGIT results that were clinically resistant (21, 44). Some of this discordance may be explained by the recommended Middlebrook 7H10 CC for RBT (0.5 μg/ml) being higher than the ECOFF, leading to missed resistance (34).

Isoniazid

INH is a pro-drug and another bactericidal agent that plays an important role in both TB treatment and preventive therapy for latent TB infection (35, 49). The catalase-peroxidase enzyme, KatG, activates INH which then targets multiple mechanisms, mainly the InhA enzyme, which in turn inhibits the synthesis of mycolic acid, an important component of the MTB cell wall (49). Therefore, mutations in the *katG* gene, which encodes the activating enzyme, typically confer high levels of resistance, as they affect the pro-drug conversion, and mutations within the *inhA* gene are typically associated with low-level resistance (28). When these loci are considered together,

molecular detection of resistance improves greatly compared to when considered separately, although there are likely other loci that contribute to INH resistance in the MTB genome (27).

Utilizing phenotypic methods, INH can be tested at two different CCs, corresponding to low-level and high-level resistance. A comparison of DST methods across multiple proficient laboratories reported more errors associated with low-level INH resistance, varying by method, and several studies have reported lower levels of agreement associated with the MGIT method in particular (35, 38, 40). However, other studies have shown overall high levels of agreement and reproducibility for INH at both concentrations of interest, implying there is good discrimination between resistant and susceptible phenotypes (29, 43, 45, 46).

Ethambutol

EMB is a bacteriostatic first-line drug that inhibits MTB cell wall arabinogalactan synthesis (49). Mutations at the *embB* locus are usually associated with low- or moderate-level EMB resistance, although this locus does not account for all EMB resistance conferring mutations in the MTB genome (28, 40). Low agreement and high variability of EMB DST results has been reported frequently across multiple laboratories and methods, particularly for MGIT which commonly fails to detect EMB resistance (35, 36, 38, 40, 43, 45, 46). One reason for this observed discordance is the presence of resistance conferring *embB* mutations that are associated with MIC values only slightly higher than the CC, causing susceptible phenotypic characterization when resistance is present (27). Additionally, the recommended CC for EMB splits the upper limit of the WT MIC

distribution, causing poor discrimination between resistant and susceptible phenotypes, which can further explain the common discordance (34, 48).

Pyrazinamide

PZA is a pro-drug activated by the pyrazinamidase/nicotinamidase enzyme and targets multiple cellular mechanisms (49). The inclusion of PZA in the first-line treatment regimen played an important role in shortening the treatment length to the six-month short course therapy currently used by killing persistent strains of MTB not targeted by other drugs (53). The activating enzyme is encoded by the *pncA* gene, in which mutations are typically associated with PZA resistance; however, the large diversity of mutations within this locus makes molecular PZA DST challenging (27, 28). Additionally, in order to optimally perform testing, PZA requires acidic conditions that are not conducive with MTB growth, making many growth-based methods unreliable with regards to determining PZA susceptibility (53). For this reason, broth-based methods that allow for an acidic environment and MTB growth are recommended for PZA susceptibility testing (25, 35). Due to the limited number of methods capable of reliably performing PZA DST, data regarding concordance is limited, but overall good agreement has been reported across proficient laboratories (35).

Fluoroquinolones

FQs are a class of antibiotics that inhibit DNA synthesis within MTB cells and include ofloxacin (OFL), moxifloxacin (MFX), ciprofloxacin (CIP), and levofloxacin (LEV) (49). These drugs are important when resistance to first-line drugs is detected and

are typically preferred for treatment of MDR-TB over other second-line agents (16). Mutations in the *gyrA* gene, which encodes subunit A of DNA gyrase, confer FQ resistance by preventing the drug from binding to its target. Mutations in the *gyrB* gene, encoding DNA gyrase subunit B, have also been shown to confer resistance, though not to the extent of *gyrA* (49). Mutations in *gyrA* are typically associated with low- to moderate-levels of FQ resistance, but some mutations, such as those at codons 94, 90, and 88, have been associated with high levels of resistance (28, 49, 54).

Cross-resistance is common among FQs, and correlated MIC values have been reported within strains (21, 55). However, some studies have shown that certain mutations in *gyrA* confer clinically significant resistance to OFL but not MFX, especially at the higher MFX CC of 2 μg/ml, indicating that certain doses of MFX may remain effective for treatment in some cases, even when there is resistance to other FQs (28, 55). This has led to the recommendation of reflex testing for MFX DST, meaning isolates that are resistant at the lower MFX CC will be subsequently tested for resistance at the higher CC in order to better determine the level of resistance present (55). Overall, high agreement has been reported regarding FQ DST, with some variability across method and drug, though fewer studies have evaluated these second-line drugs to the extent of first-line drugs (45, 46).

Second-Line Injectables

SLIs belong to a class of second-line drugs that inhibit protein synthesis and include the aminoglycosides amikacin (AMK) and kanamycin (KAN), as well as the cyclic peptide capreomycin (CAP) (49, 56). Mutations in the *rrs* gene, which encodes

16S rRNA, are associated with resistance in all three drugs, causing high levels of cross resistance among SLIs (27, 49). Additionally, mutations in the promoter region of the *eis* gene, encoding aminoglycoside acetyltransferase, have been shown to confer mostly low-level resistance for KAN (27, 49). For CAP, *tlyA* mutations have also been shown to confer low-level resistance, although these mutations are not as common as those at the other loci (27, 28).

Both *rrs* and *eis* have been shown to reliably detect resistance regarding AMK and KAN, although there is some evidence of low-level resistance in WT strains that may cause discordance (27). Missed CAP resistance may occur utilizing APM due to the Clinical and Laboratory Standards Institute (CLSI) recommended 7H10 CC of 10 μg/ml, which is higher than the ECOFF for CAP (34, 35). Certain *rrs* mutations that confer CAP resistance are associated with MIC values close to this concentration, which has led to discordance among APM results (27, 56). Notably, the WHO recommends a lower 7H10 CC for CAP of 4 μg/ml, which is expected to be closer to the ECOFF value associated with the WT MIC distribution (57).

Streptomycin

Streptomycin (STR) is an aminoglycoside, like AMK and KAN, that inhibits protein synthesis. Mutations found in the *rrs* gene and the *rpsL* gene, which encodes the S12 ribosomal protein, are thought to confer moderate drug resistance, but there are likely other loci responsible for STR resistance elsewhere in the MTB genome (28, 40, 49). Across multiple studies evaluating STR DST, performance has varied, with high

levels of discordance reported across multiple proficient laboratories as well as across multiple methodologies within a single laboratory (36, 43, 46).

Ethionamide

Ethionamide (ETO) is a second-line pro-drug that inhibits mycolic acid synthesis. Activated by EtaA/EthA, mutations in the genes responsible for these mono-oxygenases confers ETO resistance. Additionally, the same *inhA* locus that confers low-level INH resistance also confers low or moderate ETO resistance (28, 49). Overall, few evaluations of ETO DST reliability have been performed, and performance has varied across studies, many of which include the Sensititre™ method since ETO is one of the second-line drugs included on the MIC plate (21, 26, 33, 41, 42, 44, 45).

Para-aminosalicylic acid

Para-aminosalicylic acid (PAS) is a bacteriostatic agent that inhibits folic acid and thymine nucleotide metabolism in MTB cells (49). The mechanisms of PAS resistance are not well understood, and there is little data evaluating PAS DST concordance across laboratories and methods (49). Performance of PAS DST varies across studies, with evaluations reporting low agreement and high agreement regarding the same methods (21, 26, 43–45).

Despite these findings, the understanding of discordant DST results is limited, especially concerning agreement across multiple phenotypic DST methods, second-line drugs, and genotypic mechanisms. Typically, evaluations of DST agreement concern

only two methods, including the proportion method as a reference (43). However, the use of APM as a reference may be flawed, in that APM results are not guaranteed to be accurate and are not guaranteed to correlate with clinical outcomes (35–37). The use of molecular methods to confirm or contradict susceptibility results could be useful in evaluating phenotypic DST methods by accounting for possible errors in the growth-based reference results (5, 21, 27). Additionally, data regarding second-line DST is limited. Not only have fewer studies assessed second-line performance, but studies that include second-line DST may lack a variety of second-line resistance phenotypes to meaningfully evaluate performance (25, 43, 46). The inclusion of a variety of resistant isolates with various resistance patterns, including rare mutations, is important in first- and second-line DST evaluation, because certain patterns of discordance may not be evident using a representative population of TB strains (43).

These patterns of discordance, as well as their causes, are important to identify for a full understanding of the limitations of DST methods moving forward. Further research is required in these areas to better evaluate the true performance of current methods, as well as accurately assess new methods. Reliable DST results are essential in the detection and eradication of DR-TB; therefore, steps to improve DST performance are necessary for effective TB control and elimination programs.

CHAPTER II: MANUSCRIPT

INTRODUCTION

TB remains a major public health concern as one of the leading causes of death worldwide. Although global cases have been in decline, the emergence of drug resistance jeopardizes progress made by TB control and elimination programs and threatens to become a greater issue as resistant strains become a larger proportion of TB cases in the future (1, 4). While any degree of drug resistance increases the likelihood of poor treatment outcomes compared to DS-TB, there are two definitions of particular clinical significance regarding DR-TB (5). MDR-TB refers to TB that is resistant to at least the two most effective first-line drugs, RMP and INH, and XDR-TB refers to MDR-TB that has additional resistance to at least one FQ and at least one SLI (1). MDR-TB and XDR-TB are not only associated with higher morbidity and mortality, but also require a lengthy treatment regimen utilizing expensive second-line drugs with higher toxicity, lower efficacy, and more adverse side-effects, and are significantly costlier to treat than drug susceptible TB (1, 7–10, 13, 14, 18).

DST is essential in helping clinicians determine optimal drug regimens for patients, in turn preventing transmission of disease and acquisition of additional resistance, as well as accurately assessing the burden of drug-resistant strains in a population (12, 24, 25). DST can be conducted utilizing phenotypic or molecular methods. Molecular DST methods typically involve the detection of mutations associated with the resistant phenotype at specific loci in the MTB genome (25, 27). Despite the potential for shorter TATs and accurate results, insufficient coverage of the MTB genome by some currently available assays, lack of knowledge concerning resistance conferring

loci and mutations, and the lack of an easily deployable point of care test prevent most programs from relying solely on these methods (25, 27, 29, 30).

Most phenotypic DST methods determine susceptibility by observing growth of MTB isolates in the presence of a CC of drug (25, 28). APM is considered the gold standard method and involves inoculating solid agar containing a CC of drug with MTB and comparing the proportion of growth seen on the drug-containing agar to the growth seen on an inoculated drug-free agar. While this method is typically thought to produce results that correlate well with clinical outcomes, TATs associated with APM are long, usually 1–2 months from time of identification (25, 29). Automated broth systems, such as MGIT, allow for shorter TATs, but are not FDA approved to evaluate most second-line drugs (25, 29, 33). Sensititre allows the simultaneous testing of 12 first- and second-line antituberculosis drugs included on the plate, determining the MIC values associated with each (41). Additionally, this quantitative result could potentially better inform clinicians of the degree of resistance associated with a particular strain (41–43). However, this methodology is not yet FDA approved, and standardized MIC breakpoints have not been established for this method, making it difficult to determine whether a given isolate should be reported as resistant or susceptible (41, 42, 45).

These phenotypic methods have proven difficult to standardize, and previous studies have shown that discordance between DST results can occur between and within laboratories as well as methods (21, 26, 35–37, 41, 43–46). Discordant results not only cause confusion for a clinician, but can delay proper treatment in individuals, increasing the likelihood of transmission, additional resistance, and treatment failure (7, 23, 37, 47). Some have speculated differences between laboratory DST implementation could

influence discrepancies seen between sites; however, others have noted that the distribution of errors across proficient laboratories indicate an overall high-level of agreement, with discordance typically pertaining to certain combinations of methodology, drug, and strain (29, 35, 45). Certain drugs are thought to have higher rates of discordance due to their established CCs failing to accurately discriminate between the MIC distributions of WT and resistant strains (48). This would lead to oscillation between resistant and susceptible results and poor reproducibility, even within a laboratory (26, 45). Additionally, qualitative DST results fail to capture the heterogeneity of resistance phenotypes, with low-level but still clinically relevant resistance difficult to detect by growth-based methods (6, 28, 36, 37). Despite these findings, the understanding of discordant DST results is limited, especially concerning agreement across multiple phenotypic DST methods, second-line drugs, and genotypic mechanisms.

This study aimed to investigate discordance and evaluate agreement within and between laboratories based on method, drug, concentration, and genotype. This was done by conducting a multisite expanded Model Performance and Evaluation Program (MPEP) study in which 30 MTB isolates, including 8 sets of duplicates, previously characterized by CDC DST methods were sent to 12 public health and clinical laboratories where phenotypic DST was performed by three methods: APM, MGIT, and Sensititre.

MATERIALS AND METHODS

Site Selection

This multisite evaluation was funded by the CDC through a cooperative agreement (U60HM000803) with the Association of Public Health Laboratories (APHL).

This one-time funding opportunity was available to up to 12 public health or clinical laboratories to evaluate first- and second-line DST methods for MTB. Laboratories could submit applications to perform at least one, but had the option to apply for up to three, of the following methods as part of the evaluation, MGIT 960 Mycobacterial Detection System (MGIT™, Becton Dickinson), TREK Sensititre MYCOTB MIC (Sensititre™, Thermo Fisher Scientific) and indirect agar proportion (APM). Eligibility criteria included that awardees must be an APHL member state or local public health laboratory or a clinical laboratory that is currently enrolled in CDC's MPEP for DST of MTB (58); must be willing to sign a biosafety compliance letter; must be willing to accept drug resistant isolates of MTB; must have established first and second-line DST methods to detect drug resistance in MTB for at least one of the methods included in the evaluation; must provide DST results for at least RMP, INH, EMB, PZA, at least 1 FQ (e.g., OFL, CIP, MFX, or LEV), and at least 2 SLIs (e.g., CAP, AMK, or KAN).

Eleven applications were received for MGIT, eight for APM, and six for Sensititre. A team of subject matter experts from CDC/Division of Tuberculosis Elimination (DTBE) and a panel of APHL members selected from non-applicant public health laboratories reviewed applications. Proposals were evaluated and scored based on responses to four questions provided in the application; these were related to laboratory experience with the test methods (length of time performing method and test volume) and number of second-line drugs included in their testing panel. Applications were compared to each other by method. If applicants applied to participate for more than one method, their applications were evaluated separately, by method. The four highest scoring proposals were selected as the awardees for MGIT and APM. All six applicants

submitting proposals for the Sensititre evaluation were selected. Including all methods, the sites selected represented 11 unique laboratories (8 public health and 3 clinical laboratories); three of which were awarded for two methods. In addition, CDC DTBE Laboratory Branch (LB) participated in the evaluation for all methods. The project period was Dec 1, 2013 – June 30, 2014.

Strain Selection

Thirty strains of MTB, including eight sets of duplicates, were used in this study. All were selected from a culture collection maintained by the LB at the CDC. To effectively evaluate multiple DST methods for concordance, strain selection was biased to include a wide variety of drug resistance patterns to both first and second-line antituberculosis drugs as well as strains known to demonstrate low-level resistance and potentially have variable results among the methods (e.g., isolates with disputed *rpoB* mutations) (59, 60). In addition, one pan-susceptible strain, H37Rv, was included in the panel. Each strain had been previously well characterized both phenotypically and genotypically at CDC. APM was used to determine susceptibility results for all drugs except for PZA, for which MGIT was used. These phenotypic results were considered the expected results for this evaluation. Molecular detection of drug resistance (MDDR) was performed by sequencing targeted loci associated with resistance: *rpoB* (RMP), *inhA* and *katG* (INH), *embB* (EMB), *pncA* (PZA), *gyrA* (FQs), and *rrs*, *eis*, and *tlyA* (SLIs) (27, 61). To determine composite results for additional comparisons, isolates were considered resistant to a particular drug if a mutation associated with resistance was present,

regardless of phenotypic results. Mutations were considered to be resistance conferring based on previous literature (60).

Strain Preparation

The 22 strains selected for the panel were each cultured in 5 ml 7H9 broth by inoculating approximately 100 µl of well-mixed, thawed freezer stock of the isolate and incubating at 37°C until turbid. Once turbid, 100 µl was used to inoculate LJ slants. Eight of the 22 strains were inoculated in duplicate for a total panel of 30 isolates per site. Duplicates were chosen at random. LJs were incubated at 37°C for at least 3 weeks until confluent growth was observed. Panels of LJ slants were then packaged and shipped to participating laboratories.

Laboratory Methods

Awardees were required to submit their laboratory specific DST protocols to CDC and APHL. For this study, no attempts were made to standardize methodologies, drug panels, or drug concentrations across sites. Antituberculosis drugs and concentration(s) tested by each site for APM and MGIT are included in Table 2a-b. For Sensititre, the standard MYCOTB MIC plate was used with no variations to the drug concentrations. For APM, the majority of drugs were tested using CLSI or WHO recommended CCs (25, 62). Variations in media and drug concentrations across sites are indicated in Table 2a. For MGIT, first-line drugs were tested according to the package insert, while second-line drugs were tested using concentrations validated within each laboratory (Table 2b). For Sensititre, all sites tested according to the package insert to

determine the MIC for each drug. Site A performed APM, site B performed all three methods, sites C and E performed APM and MGIT, site D performed APM and Sensititre, sites F and G performed MGIT, sites H – L performed Sensititre.

Data Collection

Following receipt of the isolates, laboratories had 6 months to complete the evaluation with data due to CDC and APHL June 30, 2014. Data collection instruments were designed for each method and provided to sites. Data collection forms were site-specific, in that they only included fields for the first and second-line drugs in their routine panels. Additional fields were included on the data collection forms to capture turnaround times and any comments related to any issues with the testing such as problems with specific drugs or isolates, repeats, or inoculum issues. The data collection form for Sensititre was specifically designed for awardees to indicate growth or no growth in each well and to indicate the reported MIC. Data was returned to CDC and entered into Excel spreadsheets. Double data entry was performed and site-specific data was provided back to sites for confirmation. Discrepancies were reconciled with reporting sites, and in some cases, missing information was added during data cleaning if discrepancies were not resolved.

Data Analysis

The collected data were imported into datasets for analysis utilizing SAS® 9.4 (SAS Institute Inc., Cary, NC). Results were considered missing if a site reported insufficient growth, contamination, or a corresponding quantitative results that was

discordant with the qualitative result (i.e. percent resistance, MGIT growth units). Cycloserine (CYC) results were excluded from analysis due to the lack of a recommended CC and previously noted low reliability of DST results (25). “Borderline resistant” results were considered resistant for the purposes of these analyses due to the corresponding quantitative results indicating resistance by CLSI standards (25). To determine Sensititre susceptibility results, MIC breakpoints were established for each drug based on the dilution of drug included on the MIC plate that was closest to the recommended Middlebrook 7H10 APM CC (25, 62). Additional breakpoints were established based on the concentration utilized by CDC DST to determine the expected, growth-based results if they differed from the recommended concentrations (Table 2c). Results were considered susceptible if the reported MIC was less than or equal to the established breakpoint and resistant if the reported MIC was greater than the established breakpoint.

Site-specific agreement with previous characterization was assessed utilizing SAS. Sensitivities, specificities, and Cohen’s Kappa estimates, as well as their 95% confidence intervals were calculated using the expected results as a reference. These estimates were also calculated using the composite results as a reference for additional comparison. Both categorical and conditional agreement was considered for Sensititre. Categorical agreement was defined as both Sensititre and the reference result categorizing an isolate as susceptible, or both categorizing an isolate as resistant for a given drug. Conditional agreement allowed for concordance when the reported MIC was within one doubling dilution of a given MIC breakpoint, as previously described (45). For an isolate with a susceptible reference result, there was conditional agreement if the

MIC was less than or equal to the established MIC breakpoint plus one doubling dilution. For an isolate with a resistant reference result, there was conditional agreement if the MIC was greater than or equal to the established MIC breakpoint. An interpretation of Cohen's kappa values suggested by McHugh was used to determine whether estimates indicated an acceptable level of agreement: 0-0.20 no agreement, 0.21-0.39 minimal agreement, 0.40-0.59 weak agreement, 0.60-0.79 moderate agreement, 0.80-0.90 strong agreement, >0.90 almost perfect agreement (63).

Interlaboratory concordance was assessed using R software (R Development Core Team, 2010). For each of the three methods, a Fleiss kappa estimate for multiple raters was used to determine agreement between laboratories for a particular drug (64). To calculate Fleiss kappa estimates and corresponding 95% confidence intervals, an R function was used which allowed for the inclusion of subjects with missing raters. At least two laboratories for each drug were required for inclusion in the Fleiss kappa estimates (65,66). Comparable to the Cohen's kappa, the Fleiss kappa reports the likelihood of observing a certain level of agreement across multiple reporting sites, given the likelihood of agreement by chance. The interpretation of the Fleiss kappa is as follows: <0 poor agreement, 0-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, 0.81-1.0 almost perfect agreement (67). This interpretation differs from that of the Cohen's kappa due to the lower expected agreement involving a larger number of raters. While any kappa value greater than zero indicates agreement better than the expected agreement by chance, a value of 0.60 was selected as a cutoff for acceptable interlaboratory concordance.

Site-specific duplicate agreement was also assessed for each method, though due to a small sample size, only descriptive analyses were performed using SAS.

The Emory Institutional Review Board determined this study to be exempt from IRB review. This study was also determined to not be human subjects research by the U.S. CDC, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, as defined by 45 CFR 46.

RESULTS

Twelve laboratories were competitively selected to participate in this study. Across all sites and methods, DST was performed for 16 different first- and second-line drugs by at least one site, though not all laboratories performing the same method tested the same drugs. The same panel of 30 MTB isolates, including 8 sets of duplicates, was distributed to each participating laboratory, although the number of isolates tested for a given drug and concentration varied across method, drug, and laboratory. The panel consisted of isolates were specifically selected due to different drug resistance profiles as determined by both growth-based and molecular methods. Some of the isolates had mutations known to exhibit variability with some DST methods (e.g., *rpoB* Leu511Pro, *embB* Met306Ile) while others were reported as highly associated with resistance (e.g., *rpoB* Ser531Leu, *gyrA* Ala90Val).

Expected and Composite Results

Table 1a describes the expected, growth-based results from APM for 22 unique isolates and eight duplicates. Results were available for 12 first- and second-line

antituberculosis drugs, including a low and high INH concentration, and MDDR results were available for 9 loci associated with resistance, described in Table 1b. In constructing the composite results, 12 results reported as susceptible by growth-based methods were considered resistant due to the presence of a resistance conferring mutation. Four results, including three sets of duplicates, were modified for RMP due to mutations found in *rpoB* (Asp516Tyr; Leu571Leu/ Ile572Phe; Ser531Cys; Leu511Pro). Reports in the literature indicate these *rpoB* mutations are clinically significant (6, 37, 59, 60, 68). The same three sets of duplicates were also determined to be resistant to RBT, in addition to three other isolates with Ser531Leu *rpoB* mutations. EMB, PZA, and CAP each had one susceptible phenotypic result changed due to a Met306Ile *embB* mutation, a Tyr103Cys *pncA* mutation, and a A1401G *rrs* mutation, respectively. Each of these three results corresponded to a duplicate pair. Analyses were conducted for both expected and composite results for comparison.

Indirect Agar Proportion

Participants.

Five laboratories performed DST by APM, all of which tested the first-line drugs, as well as the second-line drugs ETO, CAP, KAN, OFL, and STR. Additional second-line drugs were tested by at least two sites (Table 2a). All but two laboratories performed APM DST on the entire panel of 30 isolates across all drugs tested. For Site D, three isolates failed to grow, resulting in a sample size of 27 for all APM DST. Site A reported seven indeterminate results, six for ETO susceptibility testing and one for OFL susceptibility testing. This site also reported six isolates as having “borderline

resistance," which were considered resistant for the purpose of these analyses due to the corresponding percent resistant values being greater than 1% of growth on no-drug controls.

For the purposes of assessing agreement across sites by drug, the two concentrations of INH (0.2 and 1 μg/ml) and STR (2 and 10 μg/ml) were each evaluated separately. Otherwise, susceptibility results for each drug were combined across sites, regardless of concentrations utilized.

Site vs. Expected Agreement.

Figure 1 shows the site-specific Cohen's kappa estimates for agreement between the APM and the expected reference results by drug. Agreement to the expected result was highest for AMK, for which all sites reported perfect agreement (1.00 [95% CI: 1.00, 1.00]), and KAN, for which all sites reported perfect or strong agreement (range: 0.90 [95% CI: 0.71, 1.00], 1.00 [95% CI: 1.00, 1.00]). Additionally, all sites reported kappa estimates ≥ 0.80 for RMP, INH (0.2 μg/ml), and RBT, indicating strong agreement.

Cohen's kappa estimates ≥ 0.60 across all sites indicated moderate agreement to the expected result for all drugs, except EMB, STR (2 μg/ml), ETO, and PAS. Additionally, the high variability and wide confidence intervals associated with CAP estimates suggests a high level of uncertainty pertaining to the quality of agreement observed. Agreement with the expected result was lowest for the drugs PAS (range: 0.20 [95% CI: -0.14, 55], 0.90 [95% CI: 0.71, 1.00] and ETO (range: 0.16 [95% CI: -0.02, 0.33], 0.78 [95% CI: 0.55, 1.00]), for which at least one site had a corresponding kappa estimate that was not statistically significant, indicating the agreement observed was no

better than the agreement expected by chance. These second-line drugs as well as EMB were the only drugs with at least one kappa estimate <0.40, indicating minimal agreement, and had the largest range of kappa estimates across reporting sites.

For the majority of drugs, lower sensitivity was observed compared to higher specificity, which was >80% for all drugs except STR (2 μg/ml). Perfect sensitivity was observed for RMP, RBT, AMK, and STR (2 μg/ml), meaning the imperfect agreement with the expected result observed for RMP, RBT, and STR (2 μg/ml) was solely driven by false-resistant results. The lowest estimates of sensitivity were observed for the drugs with the lowest agreement and largest variation between sites (EMB, ETO, CAP, and PAS) (Table 3a).

Interlaboratory Agreement.

The Fleiss kappa statistic for multiple raters was calculated for all 14 drugs tested by APM, since all drugs were tested by at least two sites. Figure 4 shows the estimated measures of interlaboratory concordance by drug for APM DST. Overall, agreement was good for most drugs, with statistically significant concordance measured for all drugs except PAS, for which the corresponding confidence interval contained the null value of zero. Agreement between sites was highest for AMK (1.00 [95% CI: 1.00, 1.00]), KAN (0.96 [0.89, 1.00]), RBT (0.96 [0.86, 1.00]), and INH (0.2 μg/ml) (0.93 [0.84, 1.00]). All first-line drugs, except EMB had high Fleiss kappa estimates >0.80, indicating almost perfect agreement.

Agreement between sites was lowest for EMB (0.47 [0.27, 0.67]), ETO (0.47 [0.28, 0.65]), and PAS (0.14 [-0.09, 0.36]). Additionally, the second-line drugs STR (10

μg/ml), CAP, MFX, and LEV had estimates >0.60, indicating substantial agreement, but low precision. Due to this high level of uncertainty associated with the magnitude of agreement, it could not be concluded with confidence that these drugs had acceptable levels of interlaboratory agreement.

BACTEC MGIT 960

Participants.

Five laboratories performed DST utilizing MGIT, and 13 first- and second-line drugs were tested by at least one site. All reporting laboratories tested the four first-line drugs (RMP, INH, EMB, and PZA) and all but one laboratory tested STR. Table 2b summarizes the number of isolates tested by each site for each drug. At site G, an isolate had inconsistent reported growth units and susceptibility result for ETO DST, so it was dropped from analysis. Site C reported a "borderline" result for PZA DST, which was considered a resistant result for the purpose of the analyses because the reported MGIT growth units were ≥ 100.

All laboratories utilized MGIT broth and followed the CC as defined in the package insert for RMP, INH, EMB, and PZA as well as the CLSI recommended CC for ETO (25). Drug concentrations used for the remaining second-line drugs varied depending on site (Table 2b). Site F determined susceptibility for MFX at a higher concentration using reflex testing when resistance was detected at a lower MFX concentration. These differences in testing algorithms resulted in 25 isolates tested at the CLSI recommended CC of 0.25 μg/ml and five isolates tested at a higher concentration of 0.5 μg/ml.

For the purposes of assessing site-specific agreement and interlaboratory concordance, the two concentrations of INH (0.1 and 0.4 μg/ml) were evaluated separately; otherwise, susceptibility results for each drug were combined across sites, regardless of concentrations utilized.

Site vs. Expected Agreement.

Figure 2 shows the site-specific Cohen's kappa estimates for agreement between MGIT susceptibility results and the expected reference results shown by drug. Agreement to the expected result was highest for RMP, AMK, and KAN, for which all sites reported perfect agreement, though only one laboratory conducted MGIT DST for KAN. Additionally, agreement to the expected result was strong for STR (range: 0.86 [95% CI: 0.69, 1.00], 0.93 [95% CI: 0.80, 1.00]) and ETO (range: 0.93 [95% CI: 0.80, 1.00], 0.93 [95% CI: 0.78, 1.00]). Cohen's kappa estimates ≥0.60 for all reporting sites indicated moderate agreement for all drugs except EMB (range: 0.14 [95% CI: -0.02, -.29], 0.39 [95% CI: 0.15, 0.63]) and CAP (range: 0.44 [95% CI: 0.04, 0.85], 0.89 [95% CI: 0.67, 1.00]). EMB had one site-specific kappa estimate that was not statistically significant, indicating agreement no better than chance, and all kappa estimates were <0.40 indicating weak agreement across sites.

Sensitivity was lowest for EMB and CAP and contributed to imperfect agreement for all drugs except RMP, AMK, and KAN. Specificity was perfect for all drugs except INH (0.4 μg/ml), EMB, PZA, STR, and CAP (Table 3b).

Interlaboratory Agreement.

The Fleiss kappa statistic for multiple raters was calculated for 10 drugs tested by MGIT, since these drugs were tested by two or more sites. Figure 4 shows the estimated measures of interlaboratory concordance by drug for MGIT DST. Agreement between sites was highest for RMP and AMK, which both had perfect agreement (1.00 [95% CI: 1.00, 1.00]), while agreement between sites was lowest for EMB (0.56 [95% CI: 0.33, 0.80]) and MFX (0.56 [95% CI: 0.19, 0.92]). The drugs EMB and MFX, with the addition of CAP (0.73 [95% CI: 0.49, 0.97]), had the lowest precision associated with their estimates and contained the Fleiss kappa cutoff of 0.60 within their 95% confidence intervals.

All drugs were significantly concordant, although there was less precision associated with the Fleiss kappa estimates, most likely due to fewer laboratories performing DST, particularly for second-line drugs.

Sensititre MIC Plate

Participants.

Seven laboratories reported MIC values corresponding to the 12 first- and second-line drugs included on the Sensititre MYCOTB MIC plate. CYC results were excluded from analysis due to the lack of a recommended CC and known challenges associated with CYC DST (25). After exclusion, 11 drugs remained for analysis. Two isolates failed to grow at Site D, resulting in a reduced panel of 28 isolates; otherwise, the remaining six sites performed Sensititre DST on the full panel of 30 isolates.

Sensititre susceptibility results were determined using the MIC value closest to the Middlebrook 7H10 CLSI recommended CC. Additional breakpoints were selected for RBT and ETO that were equivalent to APM concentration used to determine expected, growth-based results (2 μg/ml and 10 μg/ml, respectively), and for MFX that was equivalent to the higher WHO recommended CC (2 μg/ml) (62). Table 2c summarizes the breakpoints used for each drug. Inoculum for all isolates were made from growth on solid media (Middlebrook 7H10, 7H11, or LJ media).

Site vs. Expected Agreement.

Corresponding expected results were available for 10 drugs included on the Sensititre MYCOTB MIC plate, as MFX was the only drug included on the MIC plate that lacked a corresponding expected result. Figure 3a shows the site-specific Cohen's kappa estimates for categorical agreement (i.e., resistant or susceptible) between Sensititre susceptibility results and the expected reference results by drug.

Categorical agreement to the expected result was highest for AMK and KAN, for which all sites had perfect agreement. RMP, INH (0.25 and 1 μg/ml), STR, and OFL had moderate or strong agreement with kappa estimates ≥0.60 for all reporting sites. Categorical agreement to the expected result was lowest for EMB, ETO (5 and 10 μg/ml), and PAS, for which all had at least one site with a corresponding kappa estimate <0.40, indicating minimal agreement. PAS was the only drug to have a site-specific kappa estimate that indicated no statistically significant agreement (0.20 [95% CI: -0.14, 0.55]).

Figure 3b shows the site-specific Cohen's kappa estimates for conditional agreement, which allows for agreement when a MIC is one doubling dilution from the

established breakpoint, between Sensititre susceptibility results and the expected reference results shown by drug. Conditional agreement improved concordance to the expected result for the majority of sites for all drugs except AMK and KAN, which already had perfect categorical agreement. All estimates of conditional agreement were statistically significant.

Using conditional agreement, kappa estimates increased at all sites for all drugs except, RMP, INH (0.25 μg/ml), RBT (0.5 μg/ml) and PAS, compared to categorical results. Conditional agreement was at least moderate for all sites, with kappa estimates ≥0.60, for all drugs except ETO (10 μg/ml) and PAS. Conditional kappa estimates were ≥0.80, indicating strong agreement, at all sites for RMP, INH (0.25 and 1 μg/ml), STR, and OFL. Conditional agreement greatly improved agreement between EMB and ETO results and expected results. All sites reporting EMB and ETO (5μg/ml) results had corresponding kappa estimates ≥0.60. Additionally, all but one site reporting results for EMB had corresponding kappa estimates ≥0.80, indicating strong agreement compared to weak and minimal categorical agreement.

Categorical sensitivity was highest for RMP, INH (0.25 μg/ml), RBT (0.5 μg/ml), AMK, KAN, and OFL, for which all sensitivity estimates were >80%. Sensitivity was lowest for the drugs with low categorical agreement, EMB, ETO (10 μg/ml), and PAS (Table 3c). Conditional agreement increased sensitivity for at least one site for all drugs except RMP, AMK and KAN, which already had perfect sensitivity. All drugs had conditional sensitivity estimates >80% for all sites except EMB, ETO (5 and 10 μg/ml), and PAS.

Overall, specificity was high for all drugs, with estimates >80% for all sites, except for RMP, RBT (0.5 μg/ml), and STR (Table 3c). Conditional agreement improved specificity so that estimates for all sites and all drugs were >80%. RMP, INH (1 μg/ml), STR, RBT (0.5 and 2 μg/ml), and PAS were the only drugs without perfect conditional specificity.

Interlaboratory Agreement.

The Fleiss kappa statistic for multiple raters was calculated for all 11 drugs tested by Sensititre, excluding CYC, since all seven participating laboratories tested each drug. Figure 4 shows the estimated measures of categorical interlaboratory concordance by drug for the Sensititre susceptibility results. All drugs were significantly concordant between laboratories. Agreement between sites was highest for AMK and KAN, for which all sites perfectly agreed on DST results. Agreement was almost perfect for RMP (0.94 [95% CI: 0.87, 1.00]), INH (0.25 μg/ml) (0.95 [95% CI: 0.86, 1.00]), and OFL (0.92 [95% CI: 0.83, 1.00]).

Agreement between sites was lowest for EMB, ETO (10 μg/ml), and PAS. PAS was the only drug to have a Fleiss kappa estimate < 0.60, indicating less than substantial agreement.

Composite Agreement

Compared to agreement with the expected, growth-based reference results, agreement with the composite reference results was lower for the majority of reporting sites regarding RMP, EMB, and RBT and higher for the majority of reporting sites

regarding CAP and PZA. Cohen's kappa estimates differed the most for RMP and RBT, with all sites reporting lower agreement compared to the expected results across all methods. The difference in agreement for RMP was greatest for MGIT with all sites reporting a Cohen's kappa estimate <0.50, compared to perfect agreement with the expected, growth-based results. Estimates for CAP agreement improved the most among the drugs with composite results, with higher agreement across all reporting sites for MGIT and at 3 sites for APM compared to agreement with the expected result.

These differences in agreement seen between the two reference results were primarily due to higher specificity and lower sensitivity associated with the composite reference results. Therefore, the imperfect agreement to the composite result observed for RMP, RBT, and CAP was driven solely by false-susceptible results.

Duplicate Agreement

Overall, duplicate agreement varied by method, site, and drug. All sites reported at least one discordant duplicate pair across all methods, with the exception of Site F which reported perfect duplicate agreement across all drugs tested for MGIT. When discordance was reported, the majority of sites reported only one discordant pair of isolates for a given drug. The only drugs to have multiple discordant pairs reported by a site were EMB, PAS, ETO, and RBT (0.5). Across all methods, perfect duplicate agreement was reported for AMK, KAN, LEV, and CIP across all reporting sites, although only one site reported results for KAN, CIP, and LEV using MGIT DST.

For APM, site C reported the most discordance, reporting multiple discordant duplicate pairs for EMB and ETO. The majority of sites testing EMB and PAS reported

at least one discordant duplicate pair, and both sites testing MFX reported 1 discordant duplicate. In addition to the aforementioned drugs with perfect agreement across all methods, perfect duplicate agreement was reported across all sites for the drugs INH (0.2 μg/ml and 1 μg/ml), RBT, and STR (2 μg/ml).

For MGIT, additional perfect duplicate agreement was reported across all sites for RMP and OFL. The most duplicate discordance was observed for INH (0.4 μg/ml), for which four of the five sites reported a discordant pair.

For Sensititre, site L reported the most duplicate discordance, and was the only site across all methods to report three discordant duplicate pairs for EMB, RBT (0.5 μg/ml), and PAS. Additional perfect duplicate agreement was reported for INH (0.25 μg/ml) and OFL across all sites. The majority of the seven sites reported at least one discordant duplicate for the drugs EMB, STR, RBT (0.5 μg/ml), MFX (0.5 μg/ml), and PAS.

DISCUSSION

In this study, the agreement of phenotypic DST results among sites varied by drug, method, concentration, and genotype. Discordance was observed across all sites, methods, and drugs, with the exception of AMK. Observed agreement to an expected, growth-based result was typically lower than agreement described by previous studies, but interlaboratory agreement was substantial for the majority of evaluated drugs. Additionally, DST results regarding the pan-susceptible H37Rv strain were concordant across all three methods, with all sites reporting the isolate (EM027) as susceptible. The

drugs with the lowest agreement typically had worse duplicate agreement, although analysis was limited to descriptive methods due to inadequate duplicate sample size.

The analysis utilized multiple measures of concordance, allowing for a multi-faceted evaluation of DST agreement among and between individual sites. Site-specific agreement to an expected, growth-based reference was considered, as well as agreement to a composed composite reference that combined phenotypic and genotypic DST information for comparison purposes. The incorporation of molecular results potentially allowed for a more accurate determination of resistance by accounting for mutations associated with low-level resistance or that were otherwise challenging to detect by growth-based methods. By comparing agreement with expected results to agreement with composite results, the analysis was able to account for possible errors in the determination of the reference and elucidate discrepancies between phenotypic and molecular methods. Since the Sensititre MYCOTB MIC plate does not have standardized MIC breakpoints to determine susceptibility, breakpoints were established for this study following methods described by previous literature, where the concentration included on the plate that was closest to the recommended APM CC was considered the breakpoint (41). An additional breakpoint was considered for certain drugs if the concentration used to determine the expected results differed from the recommended CC. Conditional agreement was evaluated in addition to categorical agreement for MIC testing by Sensititre to account for the recommended APM CC for EMB, which was not included on the MIC plate, as well as to provide information of how close reported MIC values were to the susceptibility breakpoint for a given drug. Interlaboratory agreement,

measured by the Fleiss kappa statistic, allowed for the evaluation of agreement between sites independent of previous characterization.

The observed differences in agreement to the expected, growth-based results for some drugs, while differing from agreement reported by previous evaluations of TB DST, was not completely unexpected in this study. The panel of isolates utilized in this study was composed of strains that were intentionally selected to display a wide variety of both first and second-line resistances with some isolates encoding mutations known to result in variable DST results. For example, isolate EM026 contains an *rpoB* Leu511Pro mutation which has been previously associated with resistant and susceptible growth-based DST results but reported to have clinical significance in that patients maintained on rifampin-based therapies can have poor outcomes (6, 37, 59, 60). By including isolates with patterns of resistance that are likely to yield discordance, such as low-level resistance, the study was able to highlight challenges and limitations associated with phenotypic DST that may not have been apparent when considering a more representative sample of MTB strains. Additionally, site-specific procedures and methodologies were not standardized in order to best reflect conditions that would be present for a given site when reporting DST results for patient care. Concentrations utilized for a specific drug may have differed by site, but were combined for the purpose of these analyses. While this may have led to additional discordance between sites, the disagreement found in this study is likely to resemble disagreement that would be observed when reporting for patient care.

Similarly, differences in the methodology used to determine the expected results may have led to lower site-specific agreement with the reference. For example, the CCs used to determine expected ETO and RBT results (10 and 2 μg/ml, respectively) were

higher than the CLSI recommended CC utilized by some participating sites (25). Due to these differences in methodologies, lower site-specific agreement with sites that utilized lower concentrations may have been observed.

Despite lower agreement, this study was able to validate results that highlight patterns of discordance previously described, as well as provide new insight into agreement across phenotypic methods regarding first- and second-line antituberculosis drugs. Quantitative results, such as MIC values, as well as genotypic results, can be used to help explain the observed discordance. Figure 5a and Figure 5b show the distribution of MIC values for isolates determined to be discordant to the expected, growth-based result for the relevant drugs with reported MIC values.

Rifampin and Rifabutin

Overall, RMP and RBT DST results had high agreement with the expected results, as well as substantial interlaboratory concordance. However, these drugs had the largest difference in agreement between expected and composite reference results across all methods, indicating agreement was lower when considering disputed *rpoB* mutations (59).

For RMP, the majority of observed discordance appeared to be caused by low-level resistance, which is notoriously difficult to detect via some growth-based methods, especially liquid culture systems (6, 28, 37). The disagreement with the expected results was due solely to what would be characterized as false resistant results, although all discordant isolates had reported MIC values that indicated possible low-level resistance and reports in the literature indicate that isolates with the *rpoB* mutations included in this

study (Asp516Tyr, Leu511Pro, Leu571Leu/Ile572Phe, Ser531Cys) would be considered as resistant (59, 60). To adjust for this resistance undetected by the growth-based reference, these *rpoB* mutations were incorporated into composite reference results. While all false-resistant results were considered concordant with the composite reference, incorporation of these mutations increased the observed discordance, with sites failing to detect low-level resistance across all three methods. This was partially due to changing the reference result for additional isolates (EM028, EM014, EM018) that did not originally have false resistant results. Corresponding MIC values show all but two discordant isolates had potentially borderline or low-level RMP resistance (Figure 5a). The two isolates with MIC values well below the susceptibility breakpoint (≤0.12) were duplicates that contained a Ser531Cys mutation. While it is possible these isolates have undetected low-level resistance by all sites and methods, the clinical significance for this mutation has not be determined, and these findings support previous evidence that this mutation may not be associated with RMP resistance (69).

High interlaboratory concordance independent of agreement with reference results indicated that the majority of sites agreed for isolates considered resistant by the composite results but were tested as susceptible. For MGIT, all isolates with disputed mutations were considered susceptible across all sites, supporting previous literature indicating challenges associated with automated systems detecting low-level resistance (21, 35, 37). By incorporating these molecular results, performance of RMP DST significantly changed, and the cause of DST discordance was found to be due to failure in detecting low-level resistance, rather than false-resistant results. These results demonstrate the importance of molecular methods and suggest they may be more

accurate than phenotypic methods in detecting low-level but clinically significant resistance.

For RBT, expected DST results were determined using a CC of 2 μg/ml, which differs from the recommended Middlebrook 7H10 CC of 0.5 μg/ml, and may have contributed to the observed discordance. For APM, all discordant results were due to sites reporting resistant results for isolates with Ser531Leu *rpoB* mutations that were determined susceptible by the expected, growth-based results, indicating the concentration used to determine the expected result missed clinically significant RBT resistance (69, 70). Both the recommended 7H10 CC, 0.5 μg/ml, and the CC used to determine the expected results, 2 μg/ml, were considered as susceptibility breakpoints for the Sensititre method. Isolates that were discordant at both concentrations had MIC values, as determined by Sensititre, that indicated low-level or borderline resistance at the lower concentration, had MIC values below the 0.5 μg/ml breakpoint, or had high variability of reported MIC values, which were distributed across both breakpoints (Figure 5b). After incorporating *rpoB* mutations expected to confer RBT resistance as part of the composite result, isolates with Ser531Leu *rpoB* mutations were concordant at the lower concentration across all sites and methods, but remained discordant at the higher concentration of RBT, with MIC values equivalent to or greater than the 2 μg/ml breakpoint, suggesting the higher RBT concentration may not reliably detect resistance in these isolates due to the associated MIC values being close to the CC, which would cause variability between resistant and susceptibility results as well as poor reproducibility. Additional isolates with His526Tyr *rpoB* mutations that were expected to confer resistance were discordant at the higher concentration as well, with the majority having at

least one reported MIC at the breakpoint and the remaining MICs above the breakpoint, potentially indicating a similar issue (69). The observed discordance associated with RBT DST increased with the incorporation of *rpoB* mutations as part of the composite result due to the incorporation of the mutations Asp516Tyr, Leu571Leu/Ile572Phe, Ser531Cys, which had corresponding MIC values clustered around the 0.5 µg/ml breakpoint or below it. It is possible these mutations confer RMP resistance but not RBT resistance, as noted in previous studies, or perhaps, similar to RMP, these mutations confer low-level resistance that are not detected reliably via the phenotypic methods employed (69,70).

These results show the recommended Middlebrook 7H10 APM CC of 0.5 µg/ml was associated with higher concordance and fewer discordant isolates, with the higher concentration of 2 µg/ml failing to detect resistance at a higher rate. These results also demonstrate that resistance can be missed using only growth-based methods for RBT DST, especially when using the higher concentration, but molecular methods can be employed in order to reliably detect resistance that may be missed otherwise.

Ethambutol

EMB had low agreement with the expected results as well as poor interlaboratory concordance across all methods. EMB DST has been known to produce unreliable results and is notoriously problematic with issues of false susceptibility in some methods (35, 36, 38, 40, 43, 45, 46). Previous studies have speculated that high rates of discordance are due to the recommended APM CC splitting the upper end of the WT MIC distribution, thus failing to accurately discriminate between WT and resistant strains (34, 48). The findings in this study support this theory, as the majority of reported MIC values for

discordant isolates are equivalent to or within one doubling dilution of the breakpoint (4 µg/ml) (Figure 5a). This caused conditional agreement to greatly improve agreement estimates for Sensititre, which also may have accounted for isolates that were in agreement but were not considered concordant due to the recommended APM CC not being included on the MIC plate. Agreement with the expected result was highest for sites utilizing APM, although this method also had the lowest interlaboratory agreement for the drug implying high variability of results between sites. Additionally, these results contribute to the body of evidence indicating poor reliability of EMB DST utilizing MGIT due to a failure to detect resistance. Previous studies have suggested changing EMB's CC to overcome this limitation, and recommend using methods other than MGIT for determining EMB susceptibility (40, 48). Our results support these recommendations and further imply that EMB DST reliability is questionable for all methods, not just MGIT.

Isoniazid

Discordance for the lower concentration of INH was due to false susceptible results. All methods failed to detect low level INH resistance for one duplicate pair (EM012 and EM017) with a C(-15)T *inhA* mutation. Other isolates with this mutation were considered resistant by all sites across all methods, but the majority of reported corresponding MIC values were below the susceptibility breakpoint of 0.25 µg/ml, implying there may be another factor influencing discordance (Figure 5a). Additionally, one site did not detect INH resistance at the low and high concentration for an isolate with a Ser315Thr *katG* mutation using MGIT. The high reported MIC values as well as

the presence of a mutation conferring a high level of resistance indicates possible site-specific factors may have contributed to this disagreement. Notably, the higher concentration of INH had lower interlaboratory concordance across all methods compared to the lower concentration. Discordance at the high concentration was due to both false susceptibility and false resistance, with the reported MICs typically falling on the breakpoint or within one doubling dilution of the breakpoint. Conditional concordance greatly improved agreement for INH for this reason. All isolates discordant at the higher concentration contained either an *inhA* or *katG* mutation, with the majority containing a C(-15)T *inhA* mutation. These *inhA* mutations, which are expected to confer low-level INH resistance, may have contributed to discordance at the higher concentration due to corresponding MIC values close to 1 µg/ml, leading to oscillation between resistant and susceptible results across laboratories (28). This could also explain the discordance observed between duplicates using the MGIT method. These results conflict with previous studies that found DST results for the higher concentration of INH to be more reliable than the lower concentration and demonstrate the possibility of varying DST results at both concentrations (35).

Pyrazinamide

PZA was only tested by MGIT but had substantial interlaboratory concordance and good agreement with the expected result, although estimates of agreement varied by site. Interlaboratory concordance remained high due to a single site in disagreement for the majority of discordant isolates. One duplicate pair was susceptible by growth-based results, but had a Tyr103Cys *pncA* mutation expected to confer resistance (60).

Susceptibility results for these isolates were evenly split between resistant and susceptible, indicating potential borderline resistance or some other factor that led to oscillation of susceptibility. False resistance has previously been reported for PZA DST in liquid broth systems. After considering the isolates with Tyr103Cys *pncA* mutations as resistant, two isolates with WT *pncA* had false-resistant results reported by at least one site. This is likely due to the weaker PZA activity observed *in vitro* compared to *in vivo,* allowing isolates that would be clinically susceptible to grow and therefore be reported as resistant (71–73).

The high observed agreement in this assessment may have been due to a biased sample of isolates, for which PZA-resistant isolates included in the panel consistently tested as resistant, and, therefore, had lower variability in DST results. Overall, previous data describing PZA DST agreement is limited due to the required acidic conditions for optimal drug activity limiting the methods that can be performed, therefore, further research using a more diverse sample of isolates expressing various levels of PZA resistance may be necessary (35).

Second-Line Injectables

AMK and KAN had the highest agreement, with perfect or near perfect site-specific and interlaboratory agreement across all sites and all methods. For KAN, one site failed to detect resistance regarding an isolate with a C(-14)T *eis* mutation using Middlebrook 7H11 agar and the recommended CC of 6 µg/ml. The reported MIC values for this isolate were all well above the established breakpoint, indicating an error

potentially related to the site-specific factors rather than the level of resistance conferred by the *eis* mutation (Figure 5b).

For CAP, agreement with the expected results varied greatly by site, and was associated with high levels of uncertainty. This was potentially due to the high CC utilized for DST in determining the expected results. The APM concentration of 10 µg/ml utilized by CDC is higher than the ECOFF separating the WT and resistant MIC distributions, which may have contributed to undetected resistance for reference results. The WHO now recommends a lower CC of 4 µg/ml for Middlebrook 7H10 agar (57). The use of this higher concentration by sites performing APM could also explain some observed discordance, due to isolates with certain *rrs* mutations that are associated with an MIC close to 10 µg/ml, causing high variability for APM susceptibility results. One such mutation is the A1401G *rrs* mutation, which confers CAP resistance and was present in the majority of isolates discordant by APM (27, 56). A duplicate pair with this mutation was susceptible by the expected, growth-based reference but resistant by composite. When considering this isolate as resistant, agreement was higher across all sites for MGIT and the majority of sites for APM, indicating MGIT discordance was due to an inaccurate reference result, while discordance associated with this isolate remained between sites for APM, likely due to the poor reproducibility associated with this mutation when utilizing the higher CC. Additionally, a duplicate pair with frameshift *tlyA* mutations were considered discordant by both APM and MGIT, which may be due to low-level resistance that is thought to be conferred by mutations at this locus (28). These results suggest that molecular methods may be beneficial, particularly in conjunction with APM in order to avoid false susceptible results. Further studies may be needed to

evaluate reliability of CAP DST but the lower recommended concentration should aid performance (56).

Fluoroquinolones

OFL had high agreement across all sites as well as substantial interlaboratory agreement. Discordance varied across method, and resulted from false resistant and false susceptible results. All discordant isolates with reported false-resistance were WT and all isolates with false-susceptible results had a *gyrA* mutation. All WT isolates had reported MIC values at or within one doubling dilution of the susceptibility breakpoint indicating possible low-level or borderline resistance (Figure 5b). This suggests there may be another mechanism of resistance not accounted for by the measurements in this study, or possible heteroresistance, which has been previously reported in a high proportion of FQ resistant isolates (74). Heteroresistance describes the presence of multiple MTB populations within a sample, containing different nucleotide sequences at resistant-conferring loci. In this case, a population of organisms with resistance conferring *gyrA* mutations may be present at a level that is not detectable by molecular methods; therefore, these isolates would be phenotypically resistant, but may have corresponding WT molecular results (74, 75). Additionally, missed resistance in isolates harboring *gyrA* mutations may be due to varying degrees of resistance conferred by the mutation, which may cause variability in reported DST results (55).

Among the remaining FQs, only interlaboratory concordance was calculated. CIP and LEV were only tested by APM, for which CIP had substantial agreement and better precision than LEV and MFX. LEV had a lower estimate and lower precision,

comparable to that of MFX. MFX was tested by all three methods, but precision was poor for all Fleiss kappa estimates, particularly for APM and MGIT, likely due to the small number of sites performing MFX DST using these methods. While the estimate of agreement was substantial for APM and Sensititre, low precision indicated high levels of uncertainty associated with the measure, making it difficult to determine with confidence whether interlaboratory agreement for MFX using these methods was adequate. MFX had the worst interlaboratory concordance for MGIT, in addition to low precision. Differences in concordance between laboratories using this method may have been affected by the reflex testing utilized by Site F that was not utilized by the other site conducting MFX DST. Previous studies have shown that while MFX resistance may be present at 0.5 µg/ml, an isolate may be susceptible at 2 µg/ml, allowing for a higher dose to be included in a treatment regimen (28, 55). Therefore, isolates only tested at 2 µg/ml may be expected to differ from results obtained from testing at 0.5 µg/ml. Further evaluations may be necessary with larger number of strains to determine the reliability of MFX DST with greater certainty.

Limitations

There were several limitations to this study. We evaluated a relatively small sample size of 30 MTB isolates, and a large amount of random error was associated with many of the estimates of agreement. The low precision of these estimates contributed to many confidence intervals to contain kappa values associated with inadequate agreement and led to lower overall confidence in the interpretation of the observed measures. The sample included eight duplicates, which only allowed for descriptive analyses, making it

difficult to determine if there were any significant patterns of duplicate agreement across drug, site, and method. Additionally, not every laboratory performing the same method tested the same drugs. This resulted in less precision for many interlaboratory estimates, especially for second-line drugs, and interlaboratory concordance could not be calculated for drugs with only one reporting site. Expected, growth-based results were not provided for every drug tested, limiting the evaluation of site-specific agreement to drugs previously tested by the CDC. Across all three methods, not all sites tested the full panel of isolates, causing a smaller sample to be analyzed resulting in more random error associated with calculated estimates. The Cohen's kappa estimate is not recommended for samples smaller than 30, implying some of the calculated kappa estimates for sites with missing isolates may lack the power to detect meaningful differences. Missing data may have also caused inaccurate estimates via selection bias if isolates were not missing at random. However, the multisite nature of the study allowed for a different site to report results for isolates that were missing at another site for a particular drug. Across all methods, all isolates were tested by at least one site for each drug. A limitation of many DST agreement studies is the lack of corresponding clinical information, which limits the interpretation of results beyond the laboratory setting. Without a "gold standard" measure, it is unclear how these DST results would correlate with clinical outcomes.

Conclusions

This assessment found that agreement of DST results varied by method, drug, concentration, and genotype, with discordance observed across all methods, sites, and drugs, excluding AMK. Little of the observed disagreement appears due to site-specific

differences alone, and the understanding of the majority of the discordance was enhanced by the incorporation of molecular and quantitative information, such as MIC values. Many of the discordant isolates had corresponding MIC values equivalent to or within one doubling dilution of the recommended APM CC, which suggests limitations regarding the detection of low-level resistance, an observation which has been well-documented for RMP DST. Similar limitations in the detection of low-level resistance may exist for other first- and second-line drugs as well. The use of molecular methods in this study provided useful information regarding the determination of low-level resistance, and for some drugs, proved more accurate than phenotypic methods.

While the level of discordance observed in this study is not expected to be observed in a representative sample of MTB strains, the results emphasize some of the challenges and limitations associated with DST, calling into question the reliability of certain phenotypic methods that may have been taken for granted over time. Furthermore, the results demonstrate that discordance is not limited to one method, drug, or concentration. Even quantitative results, which are considered ideal for clinical reporting, may have limitations, illustrated by isolates with highly variable MIC results distributed across six or seven doubling dilutions. Ultimately, further research is necessary regarding the reasons for discordant DST results, including further research on the mechanisms of resistance, particularly for second-line antituberculosis drugs. These limitations need to be addressed by currently available methods in combination with new DST methods in order to provide accurate and timely susceptibility results. Rapid DST methods remain necessary to inform the optimal selection of treatment regimens and improve patient outcomes.

REFERENCES

1. Global tuberculosis report 2017. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.
2. Nahid P, Dorman SE, Alipanah N, et al. Executive Summary: Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. *Clin. Infect. Dis.* 2016;63(7):853–867.
3. Glaziou P, Sismanidis C, Floyd K, et al. Global epidemiology of tuberculosis. *Cold Spring Harb. Perspect. Med.* 2014;5(2):a017798.
4. Gandhi NR, Nunn P, Dheda K, et al. Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *Lancet*. 2010;375(9728):1830–1843.
5. Trauner A, Borrell S, Reither K, et al. Evolution of drug resistance in tuberculosis: recent progress and implications for diagnosis and therapy. *Drugs*. 2014;74(10):1063–1072.
6. Van Deun A, Barrera L, Bastian I, et al. Mycobacterium tuberculosis strains with highly discordant rifampin susceptibility test results. *J. Clin. Microbiol.* 2009;47(11):3501–3506.
7. Kvasnovsky CL, Cegielski JP, Erasmus R, et al. Extensively drug-resistant TB in Eastern Cape, South Africa: high mortality in HIV-negative and HIV-positive patients. *J. Acquir. Immune Defic. Syndr.* 2011;57(2):146–152.

8. Falzon D, Jaramillo E, Schünemann HJ, et al. WHO guidelines for the programmatic management of drug-resistant tuberculosis: 2011 update. *Eur. Respir. J.* 2011;38(3):516–528.
9. Winston CA, Mitruka K. Treatment duration for patients with drug-resistant tuberculosis, United States. *Emerg. Infect. Dis.* 2012;18(7):1201–1202.
10. Jeon DS, Shin DO, Park SK, et al. Treatment outcome and mortality among patients with multidrug-resistant tuberculosis in tuberculosis hospitals of the public sector. *J. Korean Med. Sci.* 2011;26(1):33–41.
11. Chung-Delgado K, Guillen-Bravo S, Revilla-Montag A, et al. Mortality among MDR-TB cases: comparison with drug-susceptible tuberculosis and associated factors. *PLoS One*. 2015;10(3):e0119332.
12. Gandhi NR, Moll A, Sturm AW, et al. 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*. 2006;368(9547):1575–1580.
13. Kim DH, Kim HJ, Park S-K, et al. Treatment outcomes and long-term survival in patients with extensively drug-resistant tuberculosis. *Am. J. Respir. Crit. Care Med.* 2008;178(10):1075–1082.
14. Chan ED, Strand MJ, Iseman MD. Treatment outcomes in extensively resistant tuberculosis. *N. Engl. J. Med.* 2008;359(6):657–659.
15. Falzon D, Gandhi N, Migliori GB, et al. Resistance to fluoroquinolones and second-line injectable drugs: impact on MDR-TB outcomes. *Eur. Respir. J.* 2012;erj01347–2012.

16. Falzon D, Schünemann HJ, Harausz E, et al. World Health Organization treatment guidelines for drug-resistant tuberculosis, 2016 update. *Eur. Respir. J.* [electronic article]. 2017;49(3). (http://dx.doi.org/10.1183/13993003.02308-2016)
17. Müller B, Chihota VN, Pillay M, et al. Programmatically selected multidrug-resistant strains drive the emergence of extensively drug-resistant tuberculosis in South Africa. *PLoS One*. 2013;8(8):e70919.
18. Marks SM, Flood J, Seaworth B, et al. Treatment practices, outcomes, and costs of multidrug-resistant and extensively drug-resistant tuberculosis, United States, 2005--2007. *Emerg. Infect. Dis.* 2014;20(5):812.
19. Crofton SJ, Chaulet P, Maher D, et al. Guidelines for the management of drug-resistant tuberculosis. Geneva: World Health Organization; 1997. (http://apps.who.int/iris/bitstream/10665/63465/1/WHO_TB_96.210_(Rev.1).pdf)
20. Udwadia ZF, Amale RA, Ajbani KK, et al. Totally drug-resistant tuberculosis in India. *Clin. Infect. Dis.* 2012;54(4):579–581.
21. Nosova EY, Zimenkov DV, Khakhalina AA, et al. A Comparison of the Sensititre MycoTB Plate, the Bactec MGIT 960, and a Microarray-Based Molecular Assay for the Detection of Drug Resistance in Clinical Mycobacterium tuberculosis Isolates in Moscow, Russia. *PLoS One*. 2016;11(11):e0167093.
22. Cazabon D, Alsdurf H, Satyanarayana S, et al. Quality of tuberculosis care in high burden countries: the urgent need to address gaps in the care cascade. *Int. J. Infect. Dis.* 2017;56:111–116.

23. Klopper M, Warren RM, Hayes C, et al. Emergence and spread of extensively and totally drug-resistant tuberculosis, South Africa. *Emerg. Infect. Dis.* 2013;19(3):449–455.
24. Guidelines for treatment of drug-susceptible tuberculosis and patient care, 2017 update. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.
25. CLSI. Susceptibility Testing of Mycobacteria, Nocardiae, and Other Aerobic Actinomycetes; Approved Standard - Second Edition. CLSI document M24-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2011.
26. Yu X, Ma Y-F, Jiang G-L, et al. Sensititre® MYCOTB MIC plate for drug susceptibility testing of Mycobacterium tuberculosis complex isolates. *Int. J. Tuberc. Lung Dis.* 2016;20(3):329–334.
27. Campbell PJ, Morlock GP, Sikes RD, et al. Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of Mycobacterium tuberculosis. *Antimicrob. Agents Chemother.* 2011;55(5):2032–2041.
28. Böttger EC. The ins and outs of Mycobacterium tuberculosis drug susceptibility testing. *Clin. Microbiol. Infect.* 2011;17(8):1128–1134.
29. Kim SJ. Drug-susceptibility testing in tuberculosis: methods and reliability of results. *Eur. Respir. J.* 2005;25(3):564–569.
30. Van Deun A, Martin A, Palomino JC. Diagnosis of drug-resistant tuberculosis: reliability and rapidity of detection. *Int. J. Tuberc. Lung Dis.* 2010;14(2):131–140.

31. Boehme CC, Nabeta P, Hillemann D, et al. Rapid molecular detection of tuberculosis and rifampin resistance. *N. Engl. J. Med.* 2010;363(11):1005–1015.
32. Canetti G, Froman S, Grosset J, et al. MYCOBACTERIA: LABORATORY METHODS FOR TESTING DRUG SENSITIVITY AND RESISTANCE. *Bull. World Health Organ.* 1963;29:565–578.
33. Lin S-YG, Desmond E, Bonato D, et al. Multicenter evaluation of Bactec MGIT 960 system for second-line drug susceptibility testing of Mycobacterium tuberculosis complex. *J. Clin. Microbiol.* 2009;47(11):3630–3634.
34. Ängeby K, Juréen P, Kahlmeter G, et al. Challenging a dogma: antimicrobial susceptibility testing breakpoints for Mycobacterium tuberculosis. *Bull. World Health Organ.* 2012;90(9):693–698.
35. Angra PK, Taylor TH, Iademarco MF, et al. Performance of tuberculosis drug susceptibility testing in U.S. laboratories from 1994 to 2008. *J. Clin. Microbiol.* 2012;50(4):1233–1239.
36. Van Deun A, Wright A, Zignol M, et al. Drug susceptibility testing proficiency in the network of supranational tuberculosis reference laboratories. *Int. J. Tuberc. Lung Dis.* 2011;15(1):116–124.
37. Rigouts L, Gumusboga M, de Rijk WB, et al. Rifampin resistance missed in automated liquid culture system for Mycobacterium tuberculosis isolates with specific rpoB mutations. *J. Clin. Microbiol.* 2013;51(8):2641–2645.
38. Tortoli E, Benedetti M, Fontanelli A, et al. Evaluation of automated BACTEC MGIT 960 system for testing susceptibility of Mycobacterium tuberculosis to four major antituberculous drugs: comparison with the radiometric BACTEC 460TB

method and the agar plate method of proportion. *J. Clin. Microbiol.* 2002;40(2):607–610.

39. Scarparo C, Ricordi P, Ruggiero G, et al. Evaluation of the fully automated BACTEC MGIT 960 system for testing susceptibility of Mycobacterium tuberculosis to pyrazinamide, streptomycin, isoniazid, rifampin, and ethambutol and comparison with the radiometric BACTEC 460TB method. *J. Clin. Microbiol.* 2004;42(3):1109–1114.
40. Ahmad S, Mokaddas E, Al-Mutairi N, et al. Discordance across Phenotypic and Molecular Methods for Drug Susceptibility Testing of Drug-Resistant Mycobacterium tuberculosis Isolates in a Low TB Incidence Country. *PLoS One*. 2016;11(4):e0153563.
41. Hall L, Jude KP, Clark SL, et al. Evaluation of the Sensititre MycoTB plate for susceptibility testing of the Mycobacterium tuberculosis complex against first- and second-line agents. *J. Clin. Microbiol.* 2012;50(11):3732–3734.
42. Heysell SK, Pholwat S, Mpagama SG, et al. Sensititre MycoTB plate compared to Bactec MGIT 960 for first- and second-line antituberculosis drug susceptibility testing in Tanzania: a call to operationalize MICs. *Antimicrob. Agents Chemother*. 2015;59(11):7104–7108.
43. Banu S, Rahman SMM, Khan MSR, et al. Discordance across several methods for drug susceptibility testing of drug-resistant Mycobacterium tuberculosis isolates in a single laboratory. *J. Clin. Microbiol.* 2014;52(1):156–163.

44. Xia H, Zheng Y, Zhao B, et al. Assessment of a 96-Well Plate Assay of Quantitative Drug Susceptibility Testing for Mycobacterium Tuberculosis Complex in China. *PLoS One*. 2017;12(1):e0169413.

45. Lee J, Armstrong DT, Ssengooba W, et al. Sensititre MYCOTB MIC plate for testing Mycobacterium tuberculosis susceptibility to first- and second-line drugs. *Antimicrob. Agents Chemother.* 2014;58(1):11–18.

46. Horne DJ, Pinto LM, Arentz M, et al. Diagnostic accuracy and reproducibility of WHO-endorsed phenotypic drug susceptibility testing methods for first-line and second-line antituberculosis drugs. *J. Clin. Microbiol.* 2013;51(2):393–401.

47. Williamson DA, Roberts SA, Bower JE, et al. Clinical failures associated with rpoB mutations in phenotypically occult multidrug-resistant Mycobacterium tuberculosis. *Int. J. Tuberc. Lung Dis.* 2012;16(2):216–220.

48. Schön T, Juréen P, Giske CG, et al. Evaluation of wild-type MIC distributions as a tool for determination of clinical breakpoints for Mycobacterium tuberculosis. *J. Antimicrob. Chemother.* 2009;64(4):786–793.

49. Zhang Y, Yew W-W. Mechanisms of drug resistance in Mycobacterium tuberculosis: update 2015. *Int. J. Tuberc. Lung Dis.* 2015;19(11):1276–1289.

50. Telenti A, Imboden P, Marchesi F, et al. Detection of rifampicin-resistance mutations in Mycobacterium tuberculosis. *Lancet*. 1993;341(8846):647–650.

51. Lee SS, Meintjes G, Kamarulzaman A, et al. Management of tuberculosis and latent tuberculosis infection in human immunodeficiency virus-infected persons. *Respirology*. 2013;18(6):912–922.

52. Jamieson FB, Guthrie JL, Neemuchwala A, et al. Profiling of rpoB mutations and MICs for rifampin and rifabutin in Mycobacterium tuberculosis. *J. Clin. Microbiol.* 2014;52(6):2157–2162.

53. Zhang Y, Shi W, Zhang W, et al. Mechanisms of Pyrazinamide Action and Resistance. *Microbiol Spectr*. 2014;2(4):MGM2–0023–2013.

54. Li J, Gao X, Luo T, et al. Association of gyrA/B mutations and resistance levels to fluoroquinolones in clinical isolates of Mycobacterium tuberculosis. *Emerg. Microbes Infect.* 2014;3(3):e19.

55. Willby M, Sikes RD, Malik S, et al. Correlation between GyrA substitutions and ofloxacin, levofloxacin, and moxifloxacin cross-resistance in Mycobacterium tuberculosis. *Antimicrob. Agents Chemother.* 2015;59(9):5427–5434.

56. Reeves AZ, Campbell PJ, Willby MJ, et al. Disparities in capreomycin resistance levels associated with the rrs A1401G mutation in clinical isolates of Mycobacterium tuberculosis. *Antimicrob. Agents Chemother.* 2015;59(1):444–449.

57. Technical Report on critical concentrations for drug susceptibility testing of medicines used in the treatment of drug-resistant tuberculosis. Geneva: World Health Organization; 2018 (WHO/CDS/TB/2018.5). Licence: CC BY-NC-SA 3.0 IGO.

58. Centers for Disease Control and Prevention. Tuberculosis (TB): Model Performance Evaluation Program (MPEP). https://www.cdc.gov/tb/topic/laboratory/mpep/default.htm. Published May 13, 2013. Accessed April 21, 2018.

59. Miotto P, Cabibbe AM, Borroni E, et al. Role of disputed mutations in the rpoB gene in the interpretation of automated liquid MGIT culture results for rifampicin susceptibility testing of Mycobacterium tuberculosis. *J. Clin. Microbiol.* 2018;JCM–01599.

60. Miotto P, Tessema B, Tagliani E, et al. A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. *Eur. Respir. J.* [electronic article]. 2017;50(6). (http://dx.doi.org/10.1183/13993003.01354-2017)

61. Centers for Disease Control and Prevention. *Laboratory user guide for U.S. public health laboratories: molecular detection of drug resistance (MDDR) in* Mycobacterium tuberculosis *complex by DNA sequencing (version 2.0)*. Atlanta, GA: CDC, Division of Tuberculosis Elimination Laboratory Branch; 2012. (http://www.cdc.gov/tb/topic/laboratory/mddrusersguide.pdf).

62. World Health Organization. Updated interim critical concentrations for first-line and second-line DST (as of May 2012). *World Health Organization, Geneva, Switzerland*; 2012.

63. McHugh ML. Interrater reliability: the kappa statistic. *Biochem. Med. .* 2012;22(3):276–282.

64. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol. Bull.* 1971;76(5):378.

65. Gwet KL. Handbook of Inter-Rater Reliability, 4th Edition: The Definitive Guide to Measuring The Extent of Agreement Among Raters. Advanced Analytics, LLC; 2014.

66. Advanced Analytics, LLC. Inter-Rater Reliability Publications: Inter-Rater Reliability With R. Gaithersburg, MD: Advanced Analytics, LLC. http://www.agreestat.com/r_functions.html. Published 2010. Accessed March 11, 2018.
67. Hartling L, Hamm M, Milne A, et al. Validity and Inter-Rater Reliability Testing of Quality Assessment Instruments. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012.
68. Van Deun A, Aung KJM, Bola V, et al. Rifampin drug resistance tests for tuberculosis: challenging the gold standard. *J. Clin. Microbiol.* 2013;51(8):2633–2640.
69. Berrada ZL, Lin S-YG, Rodwell TC, et al. Rifabutin and rifampin resistance levels and associated rpoB mutations in clinical isolates of Mycobacterium tuberculosis complex. *Diagn. Microbiol. Infect. Dis.* 2016;85(2):177–181.
70. Jing W, Pang Y, Zong Z, et al. Rifabutin Resistance Associated with Double Mutations in rpoB Gene in Mycobacterium tuberculosis Isolates. *Front. Microbiol.* 2017;8:1768.
71. Morlock GP, Tyrrell FC, Baynham D, et al. Using Reduced Inoculum Densities of Mycobacterium tuberculosis in MGIT Pyrazinamide Susceptibility Testing to Prevent False-Resistant Results and Improve Accuracy: A Multicenter Evaluation. *Tuberc. Res. Treat.* 2017;2017:3748163.
72. Chedore P, Bertucci L, Wolfe J, et al. Potential for erroneous results indicating resistance when using the Bactec MGIT 960 system for testing susceptibility of

Mycobacterium tuberculosis to pyrazinamide. *J. Clin. Microbiol.* 2010;48(1):300–301.

73. Zhang Y, Mitchison D. The curious characteristics of pyrazinamide: a review. *Int. J. Tuberc. Lung Dis.* 2003;7(1):6–21.
74. Eilertson B, Maruri F, Blackman A, et al. High proportion of heteroresistance in gyrA and gyrB in fluoroquinolone-resistant Mycobacterium tuberculosis clinical isolates. *Antimicrob. Agents Chemother.* 2014;58(6):3270–3275.
75. Coeck N, de Jong BC, Diels M, et al. Correlation of different phenotypic drug susceptibility testing methods for four fluoroquinolones in Mycobacterium tuberculosis. *J. Antimicrob. Chemother.* 2016;71(5):1233–1240.

Table 1a. CDC expected susceptibility results by isolate and drug[1]

Isolate	Duplicate	RMP	INH (0.2)	INH (1.0)	EMB	PZA	ETO	STR	RBT	AMK	KAN	CAP	OFL	PAS
EM001		R	R	R	R	R	R	S	S*	S	S	S	R	S
EM002	EM010	R	R	R	R	R	R	R	S	R	R	R	S	R
EM003		S	S	S	S	S	S	S	S	S	S	S	R	S
EM004		R	R	R	R	R	R	S	R	S	R	S	S	S
EM005		S	R	S	S	S	R	S	S	S	S	S	S	S
EM006		R	S	S	S	S	S	S	R	S	S	S	S	S
EM007	EM016	R	R	R	R	S*	S	R	R	S	S	S	S	R
EM008		R	S	S	S	S	S	S	R	S	S	S	R	S
EM009	EM011	S*	R	R	R	S	R	S	S*	S	S	S	S	S
EM012	EM017	R	R	S	S	S	R	S	R	S	S	S	S	S
EM013	EM028	S*	R	R	R	R	R	R	S*	S	S	S	R	S
EM014	EM018	S*	R	R	R	R	S	R	S*	R	R	S*	S	S
EM015	EM029	R	R	R	R	R	R	R	R	S	S	R	S	R
EM019		S	R	R	S	S	S	R	S	S	S	S	R	S
EM020		S	R	R	S	S	S	S	S	S	S	S	R	S
EM021		S	S	S	S	S	S	R	S	R	R	R	S	S
EM022		R	S	S	R	R	R	R	R	S	R	S	R	S
EM023	EM030	R	R	S	S*	R	R	S	R	S	S	S	S	S
EM024		R	R	R	R	R	S	R	S*	S	S	S	S	S
EM025		R	R	R	R	R	R	R	S*	S	S	S	S	R
EM026		S*	R	R	R	R	S	R	S	S	S	S	R	S
EM027		S	S	S	S	S	S	S	S	S	S	S	S	S

* phenotypic DST indicated isolate was susceptible; however, molecular results indicated a mutation consistent with resistance. These isolates were considered resistant with regards to the composite result

[1] Results were determined by the Division of Tuberculosis Elimination Laboratory Branch at the CDC. Indirect agar proportion was used to determine susceptibility results for all drugs except for pyrazinamide (PZA), for which Bactec MGIT 960 was used. Molecular detection of drug resistance (MDDR) was performed by sequencing targeted loci associated with resistance. For composite results, isolates were considered resistant to a particular drug if a mutation associated with resistance was present, regardless of phenotypic results.

Abbreviations: rifampin (RMP), isonizaid (INH), ethambutol (EMB), pyrzainamide (PZA), streptomycin (STR), kanamycin (KAN), capreomycin (CAP), amikacin (AMK), ofloxacin (OFL), para-aminosalicylic acid (PAS), ethionamide (ETO), rifabutin (RBT)

Table 1b. CDC molecular detection of drug resistance (MDDR) results by isolate and drug[1]

Isolate	Duplicate	*rpoB*	*inhA*	*katG*	*embB*	*rrs*	*eis*	*tlyA*	*gyrA*	*pncA*
EM001		**Ser531Leu**	WT	Ser315Thr	Met306Val	T1239C	WT	WT	Asp94Asn	A(−11)G
EM002	EM010	Asp516Val	WT	Ser315Thr	WT	A1401G	WT	WT	WT	A(−11)G
EM003		Phe514Phe	WT	WT	WT	WT	WT	WT	Asp94Asn	WT
EM004		Ser531Leu	WT	Ser315Thr	Met306Ile	WT	C(−14)T	WT	WT	Ala146Glu
EM005		WT	C(−15)T	WT	Glu378Ala	WT	WT	WT	WT	WT
EM006		His526Tyr	WT	WT	WT	T1239C	WT	WT	WT	WT
EM007	EM016	Ser531Leu	WT	Ser315Thr	WT	WT	WT	WT	WT	**Tyr103Cys**
EM008		His526Asp	WT	WT	WT	WT	WT	WT	Asp94Gly	WT
EM009	EM011	**Asp516Tyr**	C(−15)T	WT	Leu355Leu, Glu378Ala	WT	WT	WT	WT	WT
EM012	EM017	His526Tyr	C(−15)T	WT	Leu355Leu, Glu378Ala	WT	WT	WT	WT	WT
EM013	EM028	**Leu571Leu, Ile572Phe**	C(−15)T	Ser315Thr	Met306Ile, Asp328Gly	WT	WT	WT	Asp94Gly	A(−11)G
EM014	EM018	**Ser531Cys**	WT	Ser315Asn	Met306Val	**A1401G**	WT	WT	WT	Trp119Arg
EM015	EM029	Ser531Trp	C(−15)T	Ser315Thr	Met306Ile	WT	WT	Frameshift	WT	Trp68Arg
EM019		WT	WT	Ser315Thr	Glu378Ala	WT	WT	WT	Ala90Val	WT
EM020		WT	WT	Thr394Pro	Glu378Ala	WT	WT	WT	Ala90Val	WT
EM021		WT	WT	WT	WT	A1401G	WT	WT	WT	WT
EM022		Ser531Leu	WT	WT	Met306Ile	WT	G(−37)T	WT	Asp94Gly	Frameshift
EM023	EM030	Ser531Leu	C(−15)T	WT	**Met306Ile**	WT	WT	WT	WT	Frameshift
EM024		**Ser531Leu**	WT	Ser315Thr	Met306Leu	WT	WT	WT	WT	Asp12Ala
EM025		**Ser531Leu**	WT	Ser315Thr	Met306Val	WT	WT	WT	WT	Leu151Ser
EM026		**Leu511Pro**[†]	WT	Ser315Thr	Met306Val	WT	WT	WT	WT (Asp94Gly)^	Ala102Val
EM027		WT	WT	WT	WT	WT	WT	WT	WT	WT

^ Initial sequencing results indicate no mutation; however, examination of growth in the presence of ofloxacin revealed a mutation consistent with resistance

[1] Molecular results in **red** were identified as resistant conferring mutations in isolates with susceptible growth-based results. These isolates were considered resistant with regards to the composite results

[†]disputed *rpoB* mutation likely to confer resistance to RMP but not RBT

Abbreviations: rifampin (RMP), isonizaid (INH), ethambutol (EMB), pyrzainamide (PZA), streptomycin (STR), kanamycin (KAN), capreomycin (CAP), amikacin (AMK), ofloxacin (OFL), para-aminosalicylic acid (PAS), ethionamide (ETO), rifabutin (RBT)

Table 2a. Summary of participating laboratories performing APM

Site	Media[1]	Concentrations (ug/ml)[2]																											
		RMP		INH		EMB		PZA		ETO		STR		RBT		AMK		KAN		CAP		MFX		OFL		CIP		LEV		PAS	
		conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*
Site A	7H10	**1**	*30*	**0.2**	*30*	**5**	*30*	N/A	-	**5**	*24*	**2**	*30*	N/A	-	**4**	*30*	**5**	*30*	**10**	*30*	N/A	-	**2**	*29*	N/A	-	N/A	-	N/A	-
				1	*30*							**10**	*30*																		
Site B	7H10	**1**	*30*	**0.2**	*30*	**5**	*30*	N/A	-	**10**	*30*	**2**	*30*	**2**	*30*	**4**	*30*	**5**	*30*	**10**	*30*	N/A	-	**2**	*30*	**2**	*30*	N/A	-	**2**	*30*
				1	*30*							**10**	*30*																		
Site C	7H11	**1**	*30*	**0.2**	*30*	**7.5**	*30*	N/A	-	**10**	*30*	**2**	*30*	N/A	-	**6**	*30*	**6**	*30*	**10**	*30*	**0.5**	*30*	**2**	*30*	**2**	*30*	**1**	*30*	**8**	*30*
				1	*30*																										
Site D	7H10	**1**	*27*	**0.2**	*27*	**5**	*27*	N/A	-	**5**	*27*	**2**	*27*	**2**	*27*	N/A	-	**5**	*27*	**10**	*27*	N/A	-	**2**	*27*	N/A	-	N/A	-	N/A	-
				1	*27*																										
Site E	7H10	**1**	*30*	**0.2**	*30*	**5**	*30*	N/A	-	**5**	*30*	**2**	*30*	**0.5**	*30*	**4**	*30*	**5**	*30*	**10**	*30*	**0.5**	*30*	**2**	*30*	N/A	-	**1**	*30*	**2**	*30*
				1	*30*							**10**	*30*																		

[1] Solid Media: Middlebrook 7H10 (7H10) agar, Middlebrook 7H11 agar (7H11)

[2] Concentration (conc.) refers to the critical/equivalent concentrations used to determine susceptibility. The number of isolates tested for each drug/concentration combination (n) is also reported.

Abbreviations: rifampin (RMP), isoniazid (INH), ethambutol (EMB), pyrazinamide (PZA), ethionamide (ETO), streptomycin (STR), rifabutin (RBT), amikacin (AMK), kanamycin (KAN), capreomycin (CAP), moxifloxacin (MFX), ofloxacin (OFL), ciprofloxacin (CIP), levofloxacin (LEV), para-aminosalicylic acid (PAS)

Table 2b. Summary of participating laboratories performing MGIT

Site	Concentrations (ug/ml)[1]																									
	RMP		INH		EMB		PZA		ETO		STR		RBT		AMK		KAN		CAP		MFX		OFL		CIP		LEV		PAS	
	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*	conc.	*n*
Site B	**1**	*27*	**0.1**	*27*	**5**	*27*	**100**	*30*	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-
			0.4	*27*																										
Site C	**1**	*30*	**0.1**	*30*	**5**	*30*	**100**	*28*	N/A	-	**2**	*30*	N/A	-	**2**	*30*	**2**	*30*	**2**	*30*	**0.5**	*30*	**2**	*27*	**1**	*30*	**1**	*29*	N/A	-
			0.4	*30*																										
Site E	**1**	*30*	**0.1**	*30*	**5**	*30*	**100**	*30*	N/A	-	**1**	*30*	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-
			0.4	*28*																										
Site F	**1**	*30*	**0.1**	*30*	**5**	*30*	**100**	*30*	**5**	*30*	**1**	*30*	N/A	-	**1.5**	*30*	N/A	-	**3**	*30*	**0.25**	*25*	N/A	-	N/A	-	N/A	-	N/A	-
			0.4	*30*																	**0.5**	*5*								
Site G	**1**	*29*	**0.1**	*29*	**5**	*29*	**100**	*29*	**5**	*28*	**1**	*29*	N/A	-	N/A	-	N/A	-	**3**	*29*	N/A	-	**1.5**	*29*	N/A	-	N/A	-	N/A	-
			0.4	*29*																										

[1] Concentration (conc.) refers to the equivalent concentrations used to determine susceptibility. The number of isolates tested for each drug/concentration combination (n) is also reported.

Abbreviations: rifampin (RMP), isoniazid (INH), ethambutol (EMB), pyrazinamide (PZA), ethionamide (ETO), streptomycin (STR), rifabutin (RBT), amikacin (AMK), kanamycin (KAN), capreomycin (CAP), moxifloxacin (MFX), ofloxacin (OFL), ciprofloxacin (CIP), levofloxacin (LEV), para-aminosalicylic acid (PAS)

Table 2c. Summary of participating laboratories performing Sensititre and MIC breakpoints utilized

Site[1]	MIC Breakpoints (ug/ml)[2]														
	RMP conc.	**INH conc.**	**EMB conc.**	**PZA conc.**	**ETO conc.**	**STR conc.**	**RBT conc.**	**AMK conc.**	**KAN conc.**	**CAP conc.**	**MFX conc.**	**OFL conc.**	**CIP conc.**	**LEV conc.**	**PAS conc.**
Site B, D, H-L	**1**	**0.25** **1**	**4**	N/A	**5** **10**	**2**	**0.5** **2**	**4**	**5**	N/A	**0.5** **2**	**2**	N/A	N/A	**2**

[1]The same MIC breakpoints were utilized for all sites performing this method

[2]Concentration (conc.) refers to the MIC breakpoint used to classify isolates as susceptible or resistant based on Middlebrook 7H10 recommended critical concentrations and additional APM concentrations used to establish the CDC expected results.

Abbreviations: rifampin (RMP), isoniazid (INH), ethambutol (EMB), pyrazinamide (PZA), ethionamide (ETO), streptomycin (STR), rifabutin (RBT), amikacin (AMK), kanamycin (KAN), capreomycin (CAP), moxifloxacin (MFX), ofloxacin (OFL), ciprofloxacin (CIP), levofloxacin (LEV), para-aminosalicylic acid (PAS)

Table 3a. Sensitivity and Specificity estimates for sites performing APM[1]

Drug	Site A		Site B		Site C		Site D		Site E	
	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)
RMP	100 (80, 100)	85 (55, 98)	100 (80, 100)	85 (55, 98)	100 (80, 100)	92 (64, 100)	100 (77, 100)	92 (64, 100)	100 (80, 100)	100 (75, 100)
INH (0.2)	92 (70, 99)	100 (54, 100)	100 (86, 100)	100 (54, 100)	92 (73, 99)	100 (54, 100)	90 (70, 99)	100 (54, 100)	92 (73, 99)	100 (54, 100)
INH(1.0)	89 (67, 99)	100 (72, 100)	100 (82, 100)	82 (48, 98)	89 (67, 99)	100 (72, 100)	89 (65, 99)	100 (66, 100)	89 (67, 99)	82 (48, 98)
EMB	44 (22, 69)	100 (74, 100)	83 (59, 96)	100 (74, 100)	72 (47, 90)	100 (74, 100)	47 (23, 72)	100 (69, 100)	94 (73, 100)	83 (52, 98)
STR (2.0)	100 (79, 100)	57 (29, 82)	100 (79, 100)	79 (49, 95)	100 (79, 100)	79 (49, 95)	100 (78, 100)	83 (52, 98)	100 (79, 100)	93 (66, 100)
STR (10)	94 (70, 100)	86 (57, 98)	88 (62, 98)	100 (77, 100)					75 (48, 93)	100 (77, 100)
RBT			100 (74, 100)	83 (59, 96)			100 (66, 100)	89 (65, 99)	100 (74, 100)	83 (59, 96)
ETO	55 (23, 83)	100 (75, 100)	88 (64, 99)	100 (75, 100)	18 (4, 43)	100 (75, 100)	79 (49, 95)	100 (75, 100)	65 (38, 86)	100 (75, 100)
AMK	100 (48, 100)	100 (86, 100)	100 (48, 100)	100 (86, 100)	100 (48, 100)	100 (86, 100)			100 (48, 100)	100 (86, 100)
CAP	100 (48, 100)	92 (74, 99)	100 (48, 100)	96 (80, 100)	40 (5, 85)	100 (86, 100)	75 (19, 99)	91 (72, 99)	60 (15, 95)	100 (86, 100)
KAN	100 (59, 100)	100 (85, 100)	100 (59, 100)	100 (85, 100)	86 (42, 100)	100 (85, 100)	100 (59, 100)	100 (83, 100)	100 (59, 100)	100 (85, 100)
OFL	88 (47, 100)	100 (84, 100)	100 (66, 100)	95 (76, 100)	89 (52, 100)	90 (70, 99)	89 (52, 100)	100 (81, 100)	78 (40, 97)	100 (84, 100)
PAS			86 (42, 100)	100 (85, 100)	14 (0, 58)	100 (85, 100)			29 (4, 71)	100 (85, 100)

[1]Sensitivities (Se) and Specificities (Sp) were calculated using the expected, growth-based results as a reference

Table 3b. Sensitivity and Specificity estimates for sites performing MGIT[1]

Drug	Site B		Site C		Site E		Site F		Site G	
	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)
RMP	100 (79, 100)	100 (72, 100)	100 (80, 100)	100 (75, 100)	100 (80, 100)	100 (75, 100)	100 (80, 100)	100 (75, 100)	100 (80, 100)	100 (74, 100)
INH (0.1)	91 (71, 99)	100 (48, 100)	88 (68, 97)	100 (54, 100)	96 (79, 100)	100 (54, 100)	92 (73, 99)	100 (54, 100)	91 (72, 99)	100 (54, 100)
INH(0.4)	94 (71, 100)	80 (44, 97)	84 (60, 97)	91 (59, 100)	95 (74, 100)	78 (40, 97)	100 (82, 100)	82 (48, 98)	94 (73, 100)	73 (39, 94)
EMB	41 (18, 67)	100 (69, 100)	17 (4, 41)	100 (74, 100)	44 (22, 69)	92 (62, 100)	44 (22, 69)	100 (74, 100)	28 (10, 53)	100 (72, 100)
PZA	100 (79, 100)	71 (42, 92)	100 (77, 100)	100 (77, 100)	94 (70, 100)	100 (77, 100)	100 (79, 100)	86 (57, 98)	94 (70, 100)	92 (64, 100)
STR			94 (70, 100)	93 (66, 100)	100 (79, 100)	86 (57, 98)	100 (79, 100)	86 (57, 98)	100 (78, 100)	93 (66, 100)
ETO							94 (71, 100)	100 (75, 100)	94 (71, 100)	100 (74, 100)
AMK			100 (48, 100)	100 (86, 100)			100 (48, 100)	100 (86, 100)		
CAP			60 (15, 95)	88 (69, 97)			100 (48, 100)	92 (74, 99)	100 (48, 100)	96 (79, 100)
KAN			100 (59, 100)	100 (85, 100)						
OFL			78 (40, 97)	100 (81, 100)					88 (47, 100)	100 (84, 100)

[1]Sensitivities (Se) and Specificities (Sp) were calculated using the expected, growth-based results as a reference

Table 3c. Sensitivity and Specificity estimates for sites performing Sensititre[1]

Drug	Site B		Site D		Site H		Site I		Site J		Site K		Site L	
	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)
RMP	100 (80, 100)	85 (55, 98)	100 (78, 100)	92 (64, 100)	100 (80, 100)	85 (55, 98)	100 (80, 100)	85 (55, 98)	100 (80, 100)	77 (46, 95)	100 (80, 100)	77 (46, 95)	100 (80, 100)	85 (55, 98)
INH (0.25)	92 (73, 99)	100 (54, 100)	86 (65, 97)	100 (54, 100)	88 (68, 97)	100 (54, 100)	88 (68, 97)	100 (54, 100)	92 (73, 99)	100 (54, 100)	92 (73, 99)	100 (54, 100)	92 (73, 99)	100 (54, 100)
INH (1)	95 (74, 100)	100 (72, 100)	84 (60, 97)	100 (66, 100)	84 (60, 97)	82 (48, 98)	89 (67, 99)	100 (72, 100)	89 (67, 99)	100 (72, 100)	79 (54, 94)	91 (74, 100)	89 (67, 99)	100 (72, 100)
EMB	44 (22, 69)	100 (74, 100)	33 (13, 59)	100 (69, 100)	50 (26, 74)	100 (74, 100)	22 (6, 48)	100 (74, 100)	61 (36, 83)	100 (74, 100)	33 (13, 59)	100 (74, 100)	44 (22, 69)	92 (62, 100)
STR	100 (79, 100)	79 (49, 95)	88 (62, 98)	100 (74, 100)	69 (41, 89)	93 (66, 100)	88 (62, 98)	93 (66, 100)	94 (70, 100)	93 (66, 100)	100 (79, 100)	79 (49, 95)	100 (79, 100)	100 (77, 100)
RBT (0.5)	100 (74, 100)	78 (52, 94)	80 (44, 97)	83 (59, 96)	92 (62, 100)	83 (59, 96)	100 (74, 100)	83 (59, 96)	92 (62, 100)	83 (59, 96)	100 (74, 100)	83 (59, 96)	83 (52, 98)	78 (52, 94)
RBT (2)	100 (74, 100)	89 (65, 99)	70 (35, 93)	83 (59, 96)	75 (43, 95)	89 (65, 99)	75 (43, 95)	89 (65, 99)	83 (52, 98)	89 (65, 99)	83 (52, 98)	89 (65, 99)	67 (35, 90)	83 (59, 96)
ETO (5)	76 (50, 93)	100 (75, 100)	60 (32, 84)	92 (64, 100)	47 (23, 72)	100 (75, 100)	53 (28, 77)	100 (75, 100)	41 (18, 67)	100 (75, 100)	53 (28, 77)	100 (75, 100)	65 (38, 86)	100 (75, 100)
ETO (10)	47 (23, 72)	100 (75, 100)	53 (27, 79)	100 (75, 100)	24 (7, 50)	100 (75, 100)	41 (18, 67)	100 (75, 100)	24 (7, 50)	100 (75, 100)	35 (14, 62)	100 (75, 100)	47 (23, 72)	100 (75, 100)
AMK	100 (48, 100)	100 (86, 100)	100 (48, 100)	100 (85, 100)	100 (48, 100)	100 (86, 100)	100 (48, 100)	100 (86, 100)	100 (48, 100)	100 (86, 100)	100 (48, 100)	100 (86, 100)	100 (48, 100)	100 (86, 100)
KAN	100 (59, 100)	100 (85, 100)	100 (59, 100)	100 (84, 100)	100 (59, 100)	100 (85, 100)	100 (59, 100)	100 (85, 100)	100 (59, 100)	100 (85, 100)	100 (59, 100)	100 (85, 100)	100 (59, 100)	100 (85, 100)
OFL	89 (52, 100)	90 (70, 99)	89 (52, 100)	89 (67, 99)	89 (52, 100)	90 (70, 99)	89 (52, 100)	90 (70, 99)	89 (52, 100)	86 (64, 97)	100 (66, 100)	81 (58, 95)	89 (52, 100)	90 (70, 99)
PAS	100 (59, 100)	96 (78, 100)	43 (10, 82)	100 (84, 100)	57 (18, 90)	100 (85, 100)	43 (10, 82)	100 (85, 100)	14 (0, 58)	100 (85, 100)	71 (29, 96)	91 (72, 99)	57 (18, 90)	96 (78, 100)

[1]Categorical Sensitivities (Se) and Specificities (Sp) were calculated using the expected, growth-based results as a reference

FIGURE LEGENDS

Figure 1.

Concordance of site-specific indirect agar proportion (APM) DST results with expected DST results provided by the CDC. Cohen's Kappa estimates and corresponding 95% confidence intervals were calculated by drug using the expected results as the reference. The majority of sites utilized CLSI or WHO recommended critical concentrations, although some sites tested RBT and ETO at higher 7H10 concentrations (25,62). Apart from isoniazid and streptomycin, sites are grouped by drug regardless of the concentration used for DST.

Figure 2.

Concordance of site-specific MGIT 960 Mycobacterial Detection System (MGIT) DST results with expected DST results provided by the CDC. Cohen's Kappa estimates and corresponding 95% confidence intervals were calculated by drug using the expected results as the reference. Equivalent concentrations recommended by the CLSI were used for the first-line drugs (rifampin, isoniazid, ethambutol, and pyrazinamide) and ethionamide (25). All other concentrations were determined by the sites reporting DST results. Apart from isoniazid, sites are grouped by drug regardless of the concentration used for DST.

Figure 3a.

Categorical concordance of site-specific TREK Sensititre MYCOTB MIC plate (Sensititre) DST results with expected DST results provided by the CDC. Cohen's Kappa

estimates and corresponding 95% confidence intervals were calculated by drug using the expected results as the reference. The breakpoints for susceptibility were determined using the concentration included on the plate that was closest to the recommended Middlebrook 7H10 indirect agar proportion critical concentration and any additional concentrations utilized to determine expected results (25,62).

Figure 3b.

Conditional concordance of site-specific TREK Sensititre MYCOTB MIC plate (Sensititre) DST results with expected DST results provided by the CDC. Cohen's Kappa estimates and corresponding 95% confidence intervals were calculated by drug using the expected results as the reference. The breakpoints for susceptibility were determined using the concentration included on the plate that was closest to the Middlebrook 7H10 indirect agar proportion critical concentration and any additional concentrations utilized to determine expected results (25,62).

Figure 4.

Fleiss Kappa estimates and 95% confidence intervals measuring interlaboratory concordance by drug for each method: indirect agar proportion (APM), MGIT 960 Mycobacterial Detection System (MGIT), and TREK Sensititre MYCOTB MIC plate (Sensititre). A Fleiss Kappa estimate of 0.6 and above was considered substantial agreement and an acceptable magnitude of agreement between sites.

Figure 5a.

Distribution of minimum inhibitory concentrations (MICs) for discordant isolates corresponding to the drugs rifampin (RMP), streptomycin (STR), isoniazid (INH), and ethambutol (EMB). Isolates were considered discordant if at least one site reported a result that was not in agreement with the corresponding expected, growth-based result for any method. Discordant isolates are grouped by mutations present at loci associated with resistance: *rpoB* (RMP), *inhA* (INH 0.2 μg/ml), *katG* (INH 1 μg/ml), *embB* (EMB) (27,60). STR was not grouped by *rrs* mutations. Dashed lines indicate the susceptibility breakpoint at the concentration closest to the recommended 7H10 APM critical concentration.

Figure 5b.

Distribution of minimum inhibitory concentrations (MICs) for discordant isolates corresponding to the drugs rifabutin (RBT), para-aminosalicylic acid (PAS), ofloxacin (OFL), kanamycin (KAN), and ethionamide (ETO). Isolates were considered discordant if at least one site reported a result that was not in agreement with the corresponding expected, growth-based result for any method. Discordant isolates are grouped by mutations present at loci associated with resistance: *rpoB* (RBT), *inhA* (ETO), *gyrA* (OFL), and *rrs*/*eis* (KAN) (27,60). PAS did not have a corresponding locus associated with drug resistance. Grey, dashed lines indicate the susceptibility breakpoint at the recommended 7H10 APM critical concentration (CC). Red dashed lines indicate higher susceptibility breakpoints established in order to account for concentrations used to determine the expected, growth-based result that differed from the recommended CC.

Figure 1.

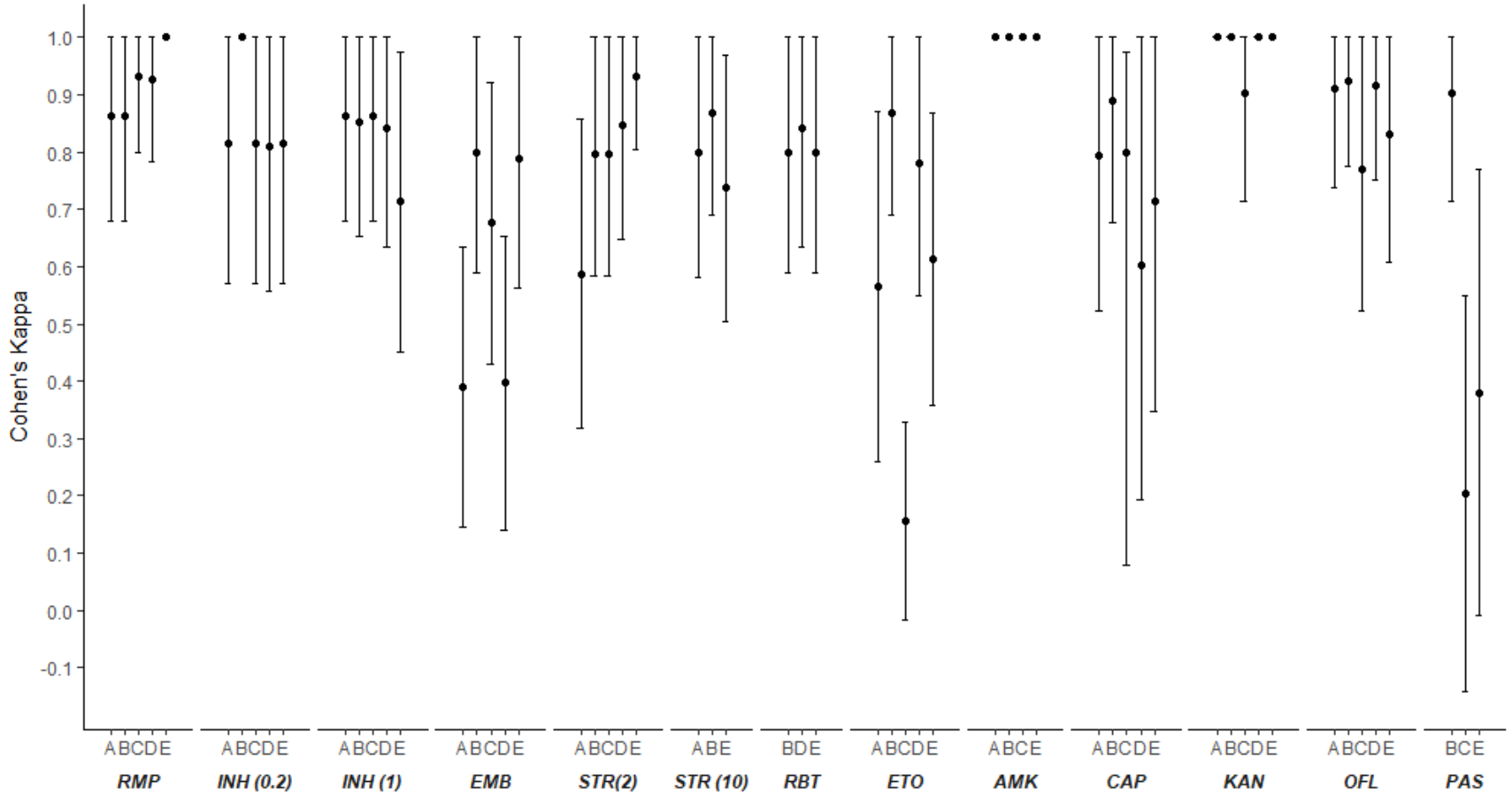

Cohen's Kappa
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
ABCDE
ABCDE
ABCDE
ABCDE
ABCDE
ABE
BDE
ABCDE
ABCE
ABCDE
ABCDE
ABCDE
BCE
RMP
INH (0.2)
INH (1)
EMB
STR(2)
STR (10)
RBT
ETO
AMK
CAP
KAN
OFL
PAS

Figure 2.

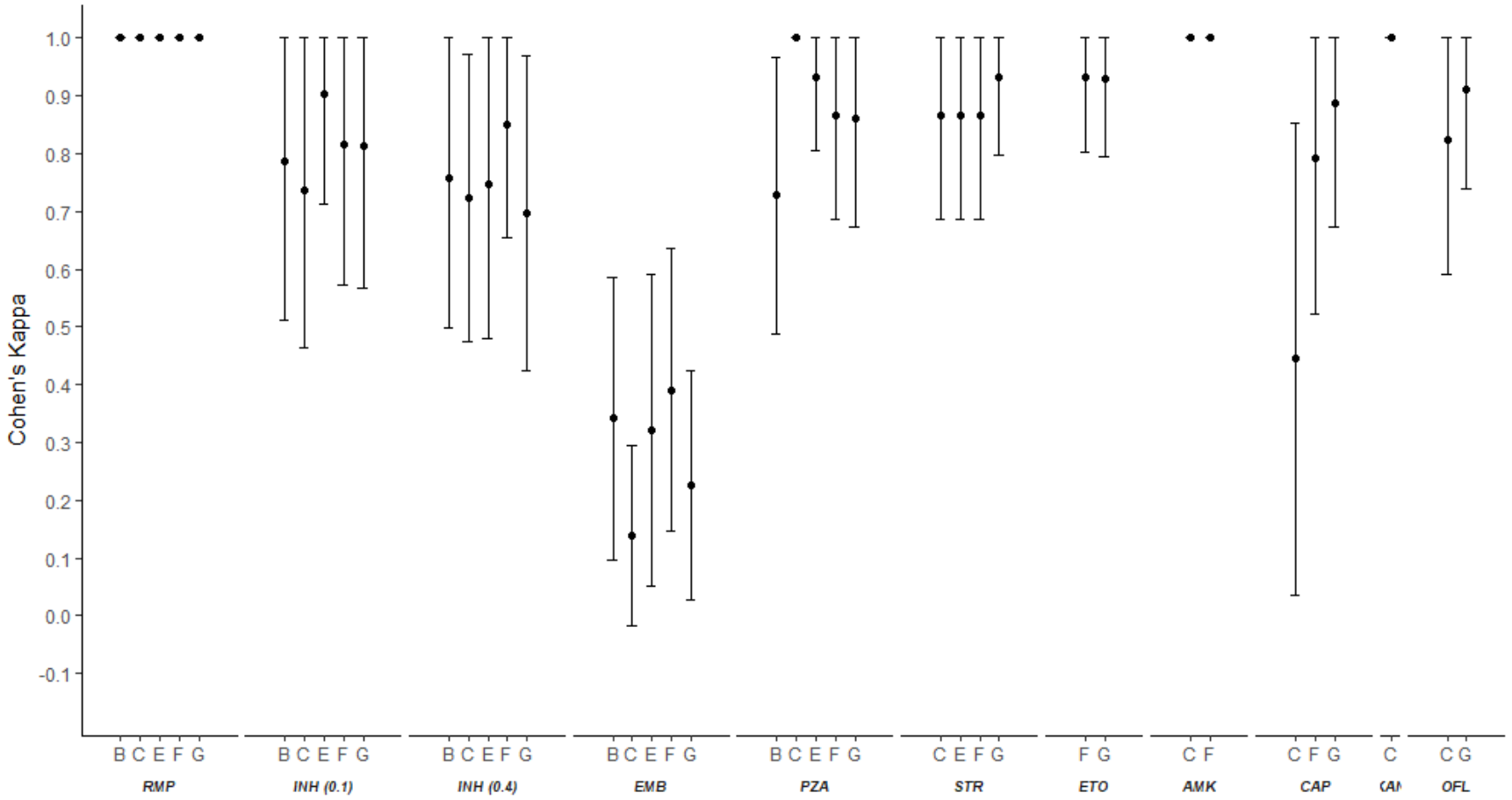

Cohen's Kappa
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
B C E F G
RMP
B C E F G
INH (0.1)
B C E F G
INH (0.4)
B C E F G
EMB
B C E F G
PZA
C E F G
STR
F G
ETO
C F
AMK
C F G
CAP
C
C G
OFL

Figure 3a.

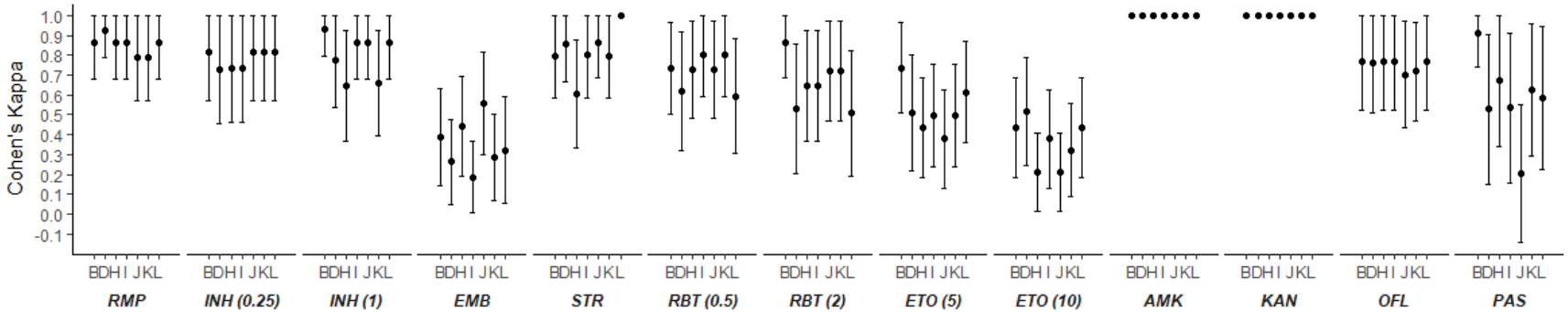


Figure 3b.

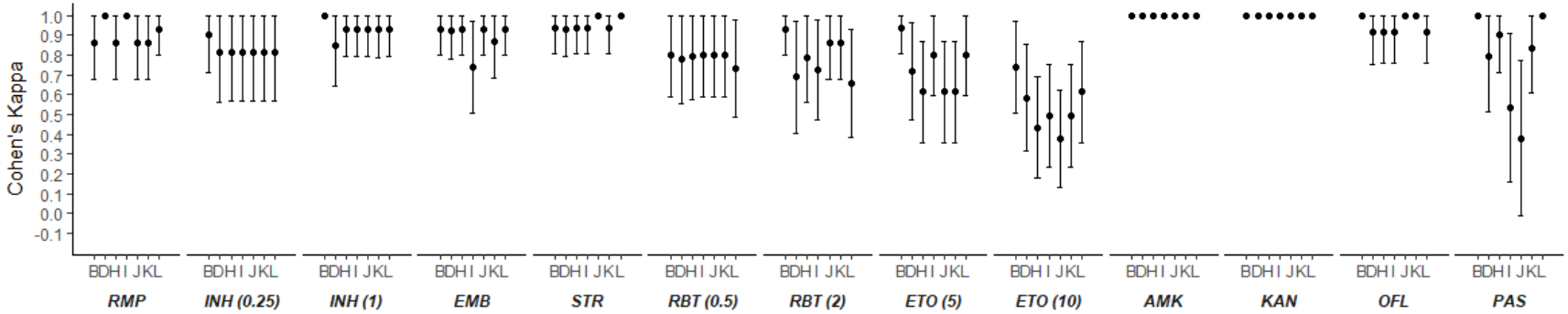


Figure 4.

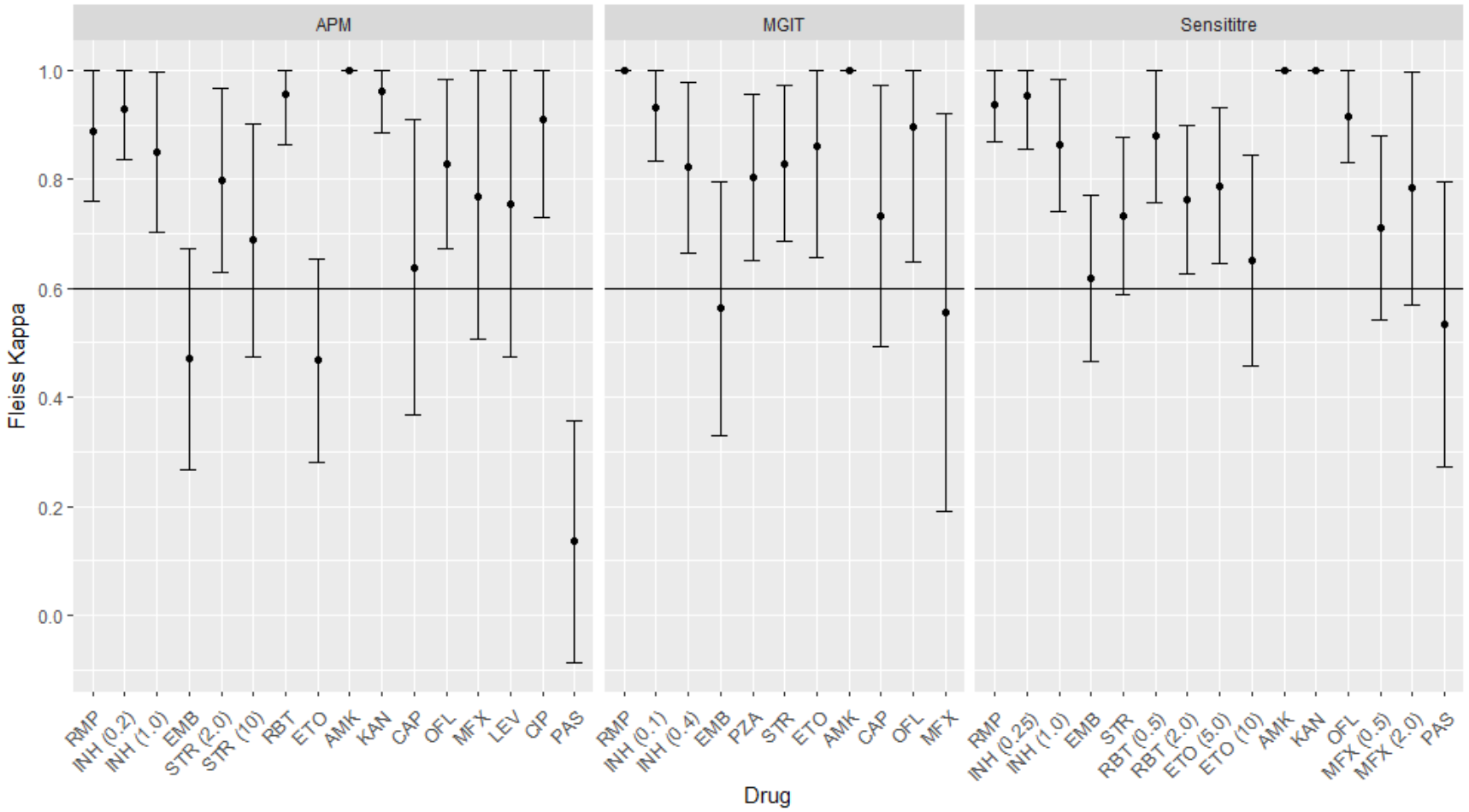

APM
MGIT
Sensititre
Fleiss Kappa
1.0
0.8
0.6
0.4
0.2
0.0
RMP
INH (0.2)
INH (1.0)
EMB
STR (2.0)
STR (10)
RBT
ETO
AMK
KAN
CAP
OFL
MFX
LEV
CIP
PAS
RMP
INH (0.1)
INH (0.4)
EMB
PZA
STR
ETO
AMK
CAP
OFL
MFX
RMP
INH (0.25)
INH (1.0)
EMB
STR
RBT (0.5)
RBT (2.0)
ETO (5.0)
ETO (10)
AMK
KAN
OFL
MFX (0.5)
MFX (2.0)
PAS
Drug

Figure 5a.

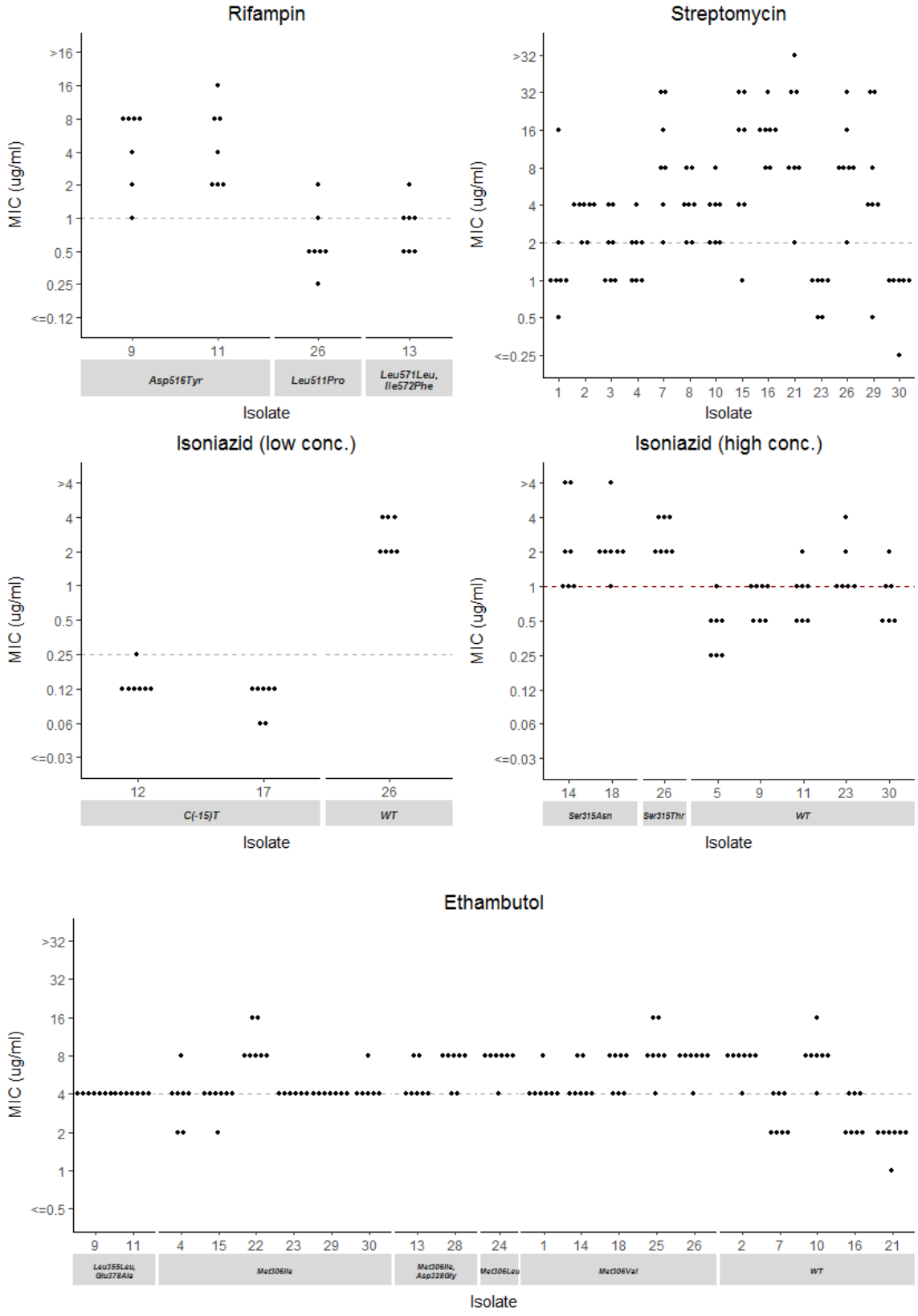
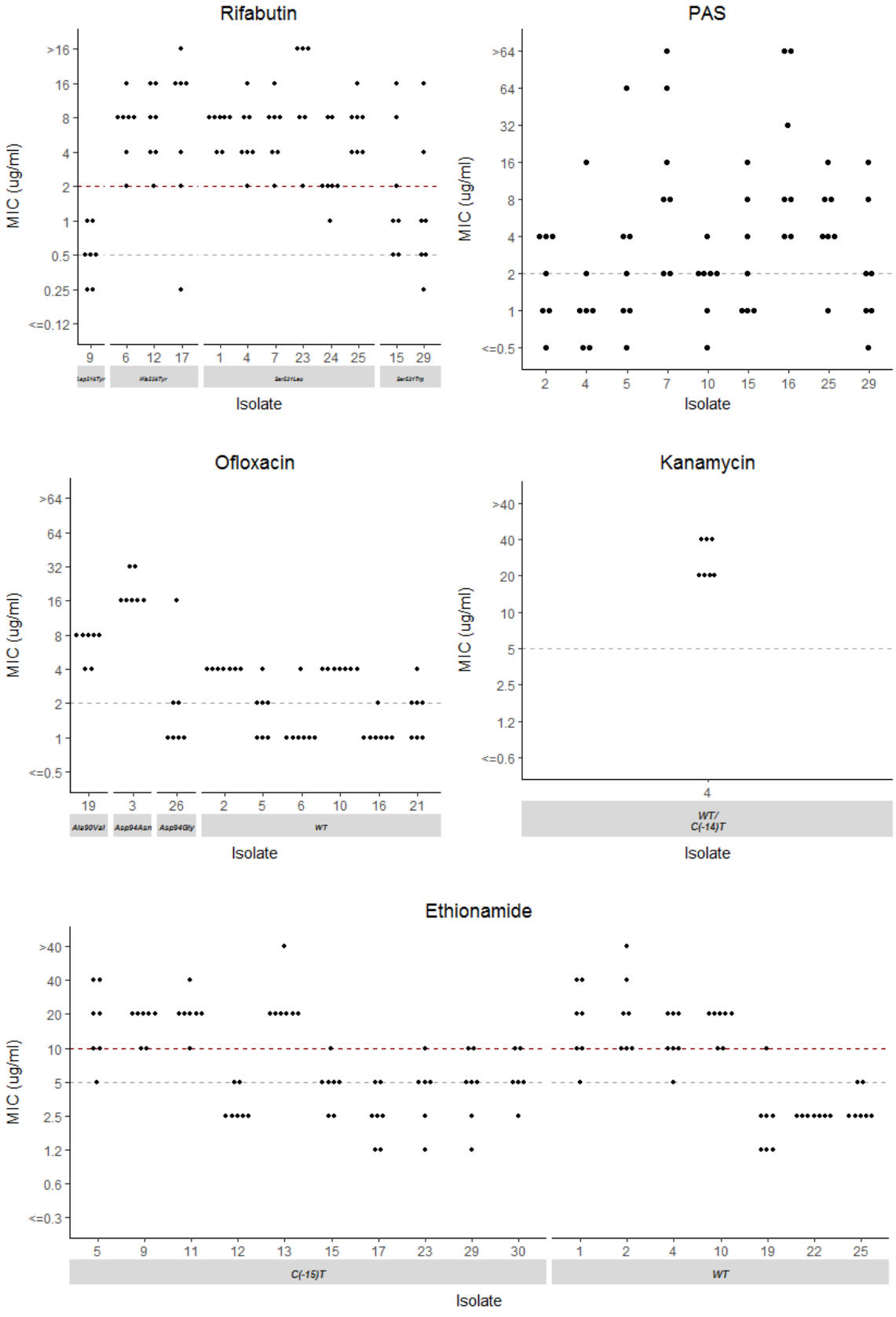

Rifampin
MIC (ug/ml)
>16
16
8
4
2
1
0.5
0.25
<=0.12
9
11
26
13
Asp516Tyr
Leu511Pro
Leu571Leu, Ile572Phe
Isolate
Streptomycin
MIC (ug/ml)
>32
32
16
8
4
2
1
0.5
<=0.25
1 2 3 4 7 8 10 15 16 21 23 26 29 30
Isolate
Isoniazid (low conc.)
MIC (ug/ml)
>4
4
2
1
0.5
0.25
0.12
0.06
<=0.03
12
17
26
C(-15)T
WT
Isolate
Isoniazid (high conc.)
MIC (ug/ml)
>4
4
2
1
0.5
0.25
0.12
0.06
<=0.03
14 18
26
5 9 11 23 30
Ser315Asn
Ser315Thr
WT
Isolate
Ethambutol
MIC (ug/ml)
>32
32
16
8
4
2
1
<=0.5
9 11
4 15 22 23 29 30
13 28
24
1 14 18 25 26
2 7 10 16 21
Leu355Leu, Glu378Ala
Met306Ile
Met306Ile, Asp328Gly
Met306Leu
Met306Val
WT
Isolate

Figure 5b.


Rifabutin
MIC (ug/ml)
>16
16
8
4
2
1
0.5
0.25
<=0.12
9 6 12 17 1 4 7 23 24 25 15 29
Asp516Tyr
His526Tyr
Ser531Leu
Ser531Trp
Isolate
PAS
MIC (ug/ml)
>64
64
32
16
8
4
2
1
<=0.5
2 4 5 7 10 15 16 25 29
Isolate
Ofloxacin
MIC (ug/ml)
>64
64
32
16
8
4
2
1
<=0.5
19 3 26 2 5 6 10 16 21
Ala90Val
Asp94Asn
Asp94Gly
WT
Isolate
Kanamycin
MIC (ug/ml)
>40
40
20
10
5
2.5
1.2
<=0.6
4
WT/
C(-14)T
Isolate
Ethionamide
MIC (ug/ml)
>40
40
20
10
5
2.5
1.2
0.6
<=0.3
5 9 11 12 13 15 17 23 29 30 1 2 4 10 19 22 25
C(-15)T
WT
Isolate