The reliability and reproducibility of WHO-endorsed phenotypic drug susceptibility testing methods for first- and second-line anti-tuberculosis drugs: A systematic review and meta-analysis

A case study

Karen R Steingart, MD, MPH
Montreal, July 2012
karenst@uw.edu

Background

- An unprecedented effort to improve and expand TB laboratory capacity is currently under-way
- WHO has issued numerous guidelines on drug susceptibility testing, including the 2008 policy guidance on drug susceptibility testing of second-line anti-TB drugs
- WHO commissioned a systematic review to inform an update of the interim guidelines

World Health Organization

2007

20, AVENUE APPIA – CH-1211 GENEVA 27 – SWITZERLAND – TEL CENTRAL +41 22 791 2111 – FAX CENTRAL +41 22 791 3111 – WWW.WHO.INT

Use of Liquid TB Culture and Drug Susceptibility Testing (DST) in Low and Medium Income Settings

Summary report of the

Expert Group Meeting on the use of liquid culture media,

Geneva, 26 March 2007

World Health Organization

2008

MOLECULAR LINE PROBE ASSAYS FOR RAPID SCREENING OF PATIENTS AT RISK OF MULTIDRUG-RESISTANT TUBERCULOSIS (MDR-TB)

WHO THE STOP TB DEPARTMENT

2008

Policy guidance on drug-susceptibility testing (DST) of second-line antituberculosis drugs

World Health Organization

2010

World Health Organization

2011

Noncommercial culture and drug-susceptibility testing methods for screening patients at risk for multidrug-resistant tuberculosis

AUTOMATED REAL-TIME NUCLEIC ACID AMPLIFICATION TECHNOLOGY FOR RAPID AND SIMULTANEOUS DETECTION OF TUBERCULOSIS AND RIFAMPICIN RESISTANCE: Xpert MTB/RIF SYSTEM

Institute of Medicine, standards for systematic reviews

http://www.iom.edu/Reports/2011/Finding-What-Works-in-Health-Care-Standards-for-Systematic-Reviews.aspx

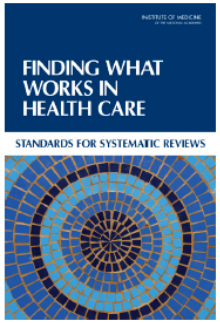
REPORT BRIEF MARCH 2011

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES
Advising the nation • Improving health

For more information visit www.iom.edu/srstandards

Finding What Works in Health Care

Standards for Systematic Reviews



Healthcare decision makers in search of reliable information comparing health interventions increasingly turn to systematic reviews for the best summary of the evidence. Systematic reviews identify, select, assess, and synthesize the findings of similar but separate studies and can help clarify what is known and not known about the potential benefits and harms of drugs, devices, and other healthcare services. Systematic reviews can be helpful for clinicians who want to integrate research findings into their daily practices, for patients to make well-informed choices about their own care, and for professional medical societies and other organizations that develop clinical practice guidelines.

In the *Medicare Improvement for Patients and Providers Act of 2008*, Congress directed the Institute of Medicine (IOM) to develop standards for conducting systematic reviews and to develop standards for clinical practice guidelines, which are evidence-based recommendations for clinicians to use when treating patients. The IOM formed two distinct committees to respond to this charge, and each committee assessed the relevant evidence and considered expert guidance to develop the standards. This report, *Finding What Works in Health Care: Standards for Systematic Reviews*, recommends standards for systematic reviews of the comparative effectiveness of medical or surgical interventions (see the insert for a list of the standards).

Systematic reviews ... can help clarify what is known and not known about the potential benefits and harms of drugs, devices, and other healthcare services.

Standard 2.1 Establish a team with appropriate expertise and experience to conduct the systematic review

- 2.1.1 Include expertise in the pertinent clinical content areas
- 2.1.2 Include expertise in systematic review methods
- 2.1.3 Include expertise in searching for relevant evidence
- 2.1.4 Include expertise in quantitative methods

Standard 2.2 Manage bias and conflict of interest (COI) of the team conducting the systematic review

- 2.2.1 Require each team member to disclose potential COI and professional or intellectual bias
- 2.2.2 Exclude individuals with a clear financial conflict
- 2.2.3 Exclude individuals whose professional or intellectual bias would diminish the credibility of the review in the eyes of the intended users

Standard 2.5 Formulate the topic for the systematic review

- **2.5.1 Confirm the need for a new review**

  - We identified prior systematic reviews on specific index tests
- **2.5.2 Develop an analytic framework that clearly lays out the chain of logic that links the health intervention to the outcomes of interest and defines the key clinical questions to be addressed by the systematic review**
- 2.5.3 Use a standard format to articulate each clinical question of interest
- 2.5.4 State the rationale for each clinical question
- **2.5.5 Refine each question based on user and stakeholder input**

The review question

- What is the reliability and reproducibility of WHO-endorsed phenotypic DST methods for first- and second-line anti-TB drugs?
  - Line probe assays were included per WHO request
  - Xpert was not included because a systematic review is currently underway

- Small is beautiful

The review question - questions

- What type of systematic review is this?
- What is the scope?
- What is the definition of reliability?
- What is the definition of reproducibility?

Standard 2.6 Develop a systematic review protocol

- 2.6.1 Describe the context and rationale for the review from both a decision making and research perspective
- **2.6.2 Describe the study screening and selection criteria (inclusion/exclusion criteria)**
- **2.6.3 Describe precisely which outcome measures, time points, interventions, and comparison groups will be addressed**

PICO or PPPICPTR for systematic review of diagnostic test accuracy?

- **P**atients, **P**resentation, **P**rior tests
- **I**ndex test, **C**omparator tests
- **P**urpose: comparative question, role of test
- **T**arget condition, **R**eference standard

Inclusion/exclusion criteria

- Participants?
- Index tests?
- Comparator tests?
- Outcomes?

- Types of studies?
- Target condition?
- Reference standard?

Inclusion/exclusion criteria

Participants/samples

- all patients suspected or confirmed as having TB
- all direct clinical specimens and culture isolates
- all settings (clinical and laboratory) and countries

Types of studies

- all study designs for which we could extract true positive (TP), false positive (FP), false negative (FN), and true negative (TN values)
- excluded letters and abstracts

First-line drugs, phenotypic index tests

A. Commercial
- MGIT Manual (Becton Dickinson)
- MGIT 960 (Becton Dickinson)
- VersaTREK (TREK, Trek Diagnostic Systems, USA)

B. Noncommercial newer tests
- Microscopic Observation Drug Susceptibility (MODS) assay
- Nitrate reductase assay (NRA)
- Colorimetric redox indicator (CRI) methods
    1. Alamar blue
    2. Resazurin
    3. Tetrazolium bromide

First-line drugs, genotypic index tests

- GenoType® MTBDR assay (MTBDR, Hain LifeScience GmbH, Nehren, Germany)
- GenoType® MTBDR plus assay (MTBDR plus, Hain LifeScience GmbH, Nehren, Germany)
- GenoType® MTBDRsl (MTBDRsl, Hain LifeScience GmbH, Nehren, Germany)
- INNO-LiPA Rif.TB (Innogenetics, Ghent, Belgium)

Second-line drugs

Phenotypic index tests

- Commercial (MGIT 960, etc)
- Noncommercial newer tests (MODS, NRA, CRI)
- Noncommercial solid media
  - Löwenstein-Jensen
  - Middlebrook 7H10
  - Middlebrook 7H11

Genotypic index tests

- GenoType® MTBDRsl assay

Reference standards

First-line drugs

- Löwenstein-Jensen, 7H10, and 7H11 medium (all by proportion, absolute concentration, or resistance ratio method)
- BACTEC 460

Second-line drugs

- MGIT 960
- BACTEC 460

- Caution: The same test can appear as both an index and a reference standard

Some clear definitions

Drug susceptibility testing refers to tests that classify TB isolates as drug resistant or drug susceptible, based on the ability of the isolate to grow in the presence of the test drug at a "critical concentration"

Type of testing

- Direct testing: a set of drug-containing and drug-free media is inoculated directly with a concentrated specimen

- Indirect testing: involves inoculation of drug-containing media with a pure culture grown from the original specimen

Some more clear definitions

Critical concentration (CC) is the lowest concentration of a drug that inhibits 95% of "wild-type" strains of *M. tuberculosis* that have not been exposed to the drug, but that simultaneously does not inhibit strains of *M. tuberculosis* considered resistant that are isolated from patients who are not responding to therapy

Clinical and Laboratory Standards Institute.2003. Susceptibility testing of mycobacteria, nocardiae, and other aerobic actinomycetes; approved standard. Clinical and Laboratory Standards Institute, Wayne, PA)

Group	Drug	DST Critical Concentrations (µg/ml)				
		LJ	7H10	7H11	BACTEC	MGIT 960
1	Isoniazid	0.2	0.2	0.2	0.1	0.1
	Rifampicin	40.0	1.0	1.0	2.0	1.0
	Ethambutol	2.0	5.0	7.5	2.5	5.0
	Pyrazinamide	-	-	-	100.0	100.0
2	Streptomycin	4.0	2.0	2.0	2.0	1.0
	Kanamycin	30.0	5.0	6.0	4.0	-
	Amikacin	-			1.0	1.0
	Capreomycin	40.0	10.0	10.0	1.25	2.5
	Viomycin	-	-	-	-	-
3	Ciprofloxacin	2.0	2.0	2.0	2.0	1.0
	Ofloxacin	2.0	2.0	2.0	2.0	2.0
	Levofloxacin	-	2.0	-	-	2.0
	Moxifloxacin	-	-	-	0.5	0.25
	Gatifloxacin	-	1.0	-	-	-
4	Ethionamide	40.0	5.0	10.0	2.5	5.0
	Prothionamide	40.0	-	-	1.25	2.5
	Cycloserine	40.0	-	-	-	-
	Terizidone	-	-	-	-	-
	PAS	1.0	2.0	8.0	2.0	-
	Thioacetazone	-	-	-	-	-

How to define reliability?

Reliability is the comparison of the results of the index test with those of a reference standard

Sensitivity = proportion of resistant TB samples correctly identified [TP/(FN + TP)]

Specificity = proportion of susceptible TB samples correctly identified [(TN/(FP + TN)]

Agreement = (R-R + S-S)/(R-R + S-S + S-R + R-S)

R, resistant; S, susceptible; TP, true positives, FP, false positives, FN, false negatives; TN, true negatives

How to define reproducibilty?

Reproducibility is agreement when DST by a given index test is repeated on the same M. tb isolate

1. Presented as agreement between index test results regardless of what the drug susceptibility of isolates was known to be

2. Presented as agreement with a reference method: "reproducibility of expected result”

Which definition, 1? 2? Both 1 and 2?

Example 1. Reproducibility

- **During the initial phase a panel of 10 strains among 100 clinical isolates was tested in triplicate at each of the three sites to establish the reproducibility of the MGIT 960 testing.**

Giampaglia et al. Multicentre evaluation of an automated BACTEC 960 system for susceptibility testing of Mycobacterium tuberculosis. IJTLD 2007

Example 2. Reproducibility of expected results

- **A judicial strain susceptibility profile was obtained by the majority concordant results of the three reference methods: BACTEC 460, proportion method, and resistance ratio method. MGIT 960 results were obtained at the three individual sites and compared with the judicial susceptibility profiles.**

Giampaglia et al. Multicentre evaluation of an automated BACTEC 960 system for susceptibility testing of Mycobacterium tuberculosis. IJTLD 2007

Standard 4.4 If conducting a meta-analysis, then do the following:

- 4.4.1 Use expert methodologists to develop, execute, and peer review the meta-analyses
- 4.4.2 Address the heterogeneity among study effects
- 4.4.3 Accompany all estimates with measures of statistical uncertainty
- 4.4.4 Assess the sensitivity of conclusions to changes in the protocol, assumptions, and study selection (sensitivity analysis)

Meta-analysis

- Pooled sensitivity and specificity estimates using a hierarchical random effects regression model
- Pooled agreement estimates using a random effects model
- Stata/IC, version 11.0

Splitting studies into subgroups

What drug?

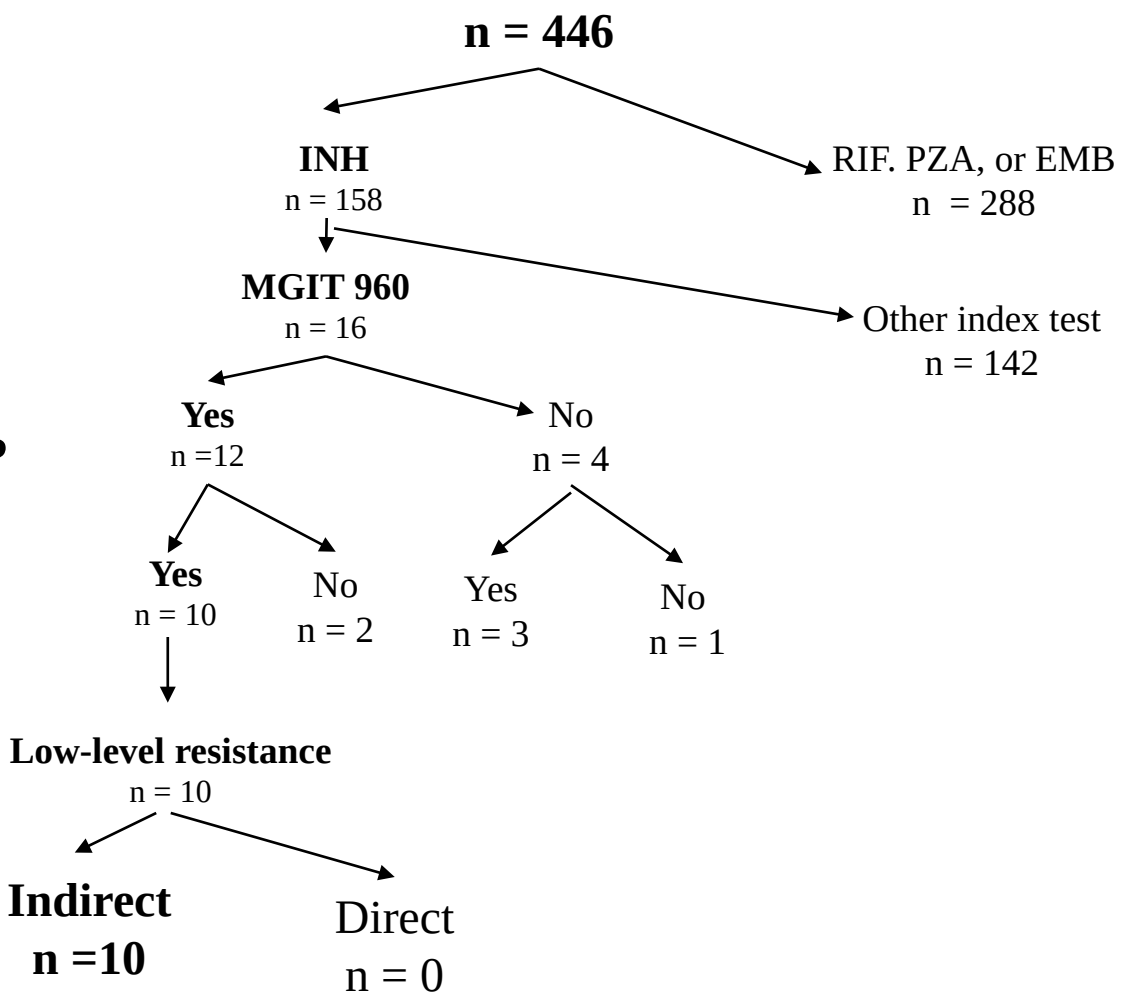
What index test?

Currently recommended CC for reference standard?

Currently recommended CC for index test?

What level resistance?

Direct or indirect?



Index Tests Included

MGIT 960	Tetrazolium
MGIT Manual	Resazurin
MODS	**T**REK
NRA Solid	MTBDR
NRA liquid	MTBDRplus
Alamar blue	INNO-LiPA

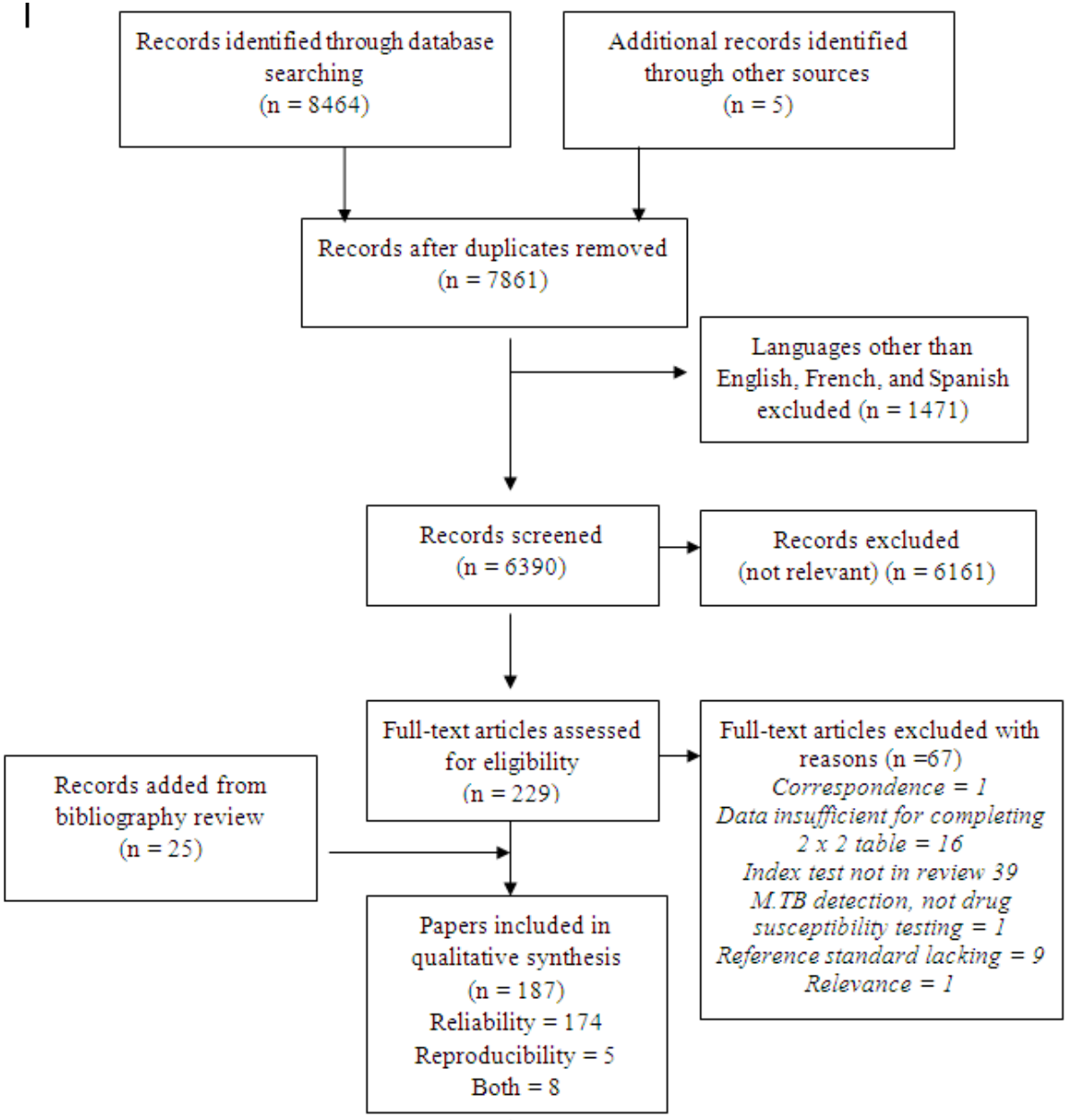
CC, critical concentration

RESULTS

Flow of studies

8464 citations
- 229 full-texts
- 25 papers added from bibliography review

187 papers
600 studies reliability
93 studies reproducibility



Index Test	Number	Percent %
Nitrate Reductase Assay (NRA) solid	88	14.7
MGIT 960	84	14.0
MGIT Manual	73	12.2
Genotype® MTBDR Plus	56	9.3
Resazurin	50	8.3
Tetrazolium	43	7.2
INNO-LiPA Rif.TB	42	7.0
Alamar Blue	35	5.8
Genotype® MTBDRsl	35	5.8
Microscopic Observation Drug Susceptibility (MODS) assay	35	5.8
Genotype® MTBDR	25	4.2
VersaTREK	7	1.2
7H10 Proportion method	6	1.0
Nitrate Reductase Assay (NRA) liquid	6	1.0
LJ Resistance ratio method	5	0.8
LJ Absolute concentration method	4	0.7
7H11 Proportion method	3	0.5
BACTEC 460	2	0.3
LJ Proportion method	1	0.2

Drug	Number	Percent %
Rifampicin	197	32.8
Isoniazid	158	26.3
Ethambutol	80	13.3
Streptomycin	67	11.2
Ofloxacin	27	4.5
Kanamycin	15	2.5
Capreomycin	14	2.3
Pyrazinamide	11	1.8
Amikacin	7	1.2
Ethionamide	7	1.2
Moxifloxacin	6	1.0
P-aminosalicylic acid	4	0.7
Linezolid	2	0.3
Rifabutin	2	0.3
Gatifloxacin	1	0.2
Cycloserine	1	0.2
Prothionamide	1	0.2

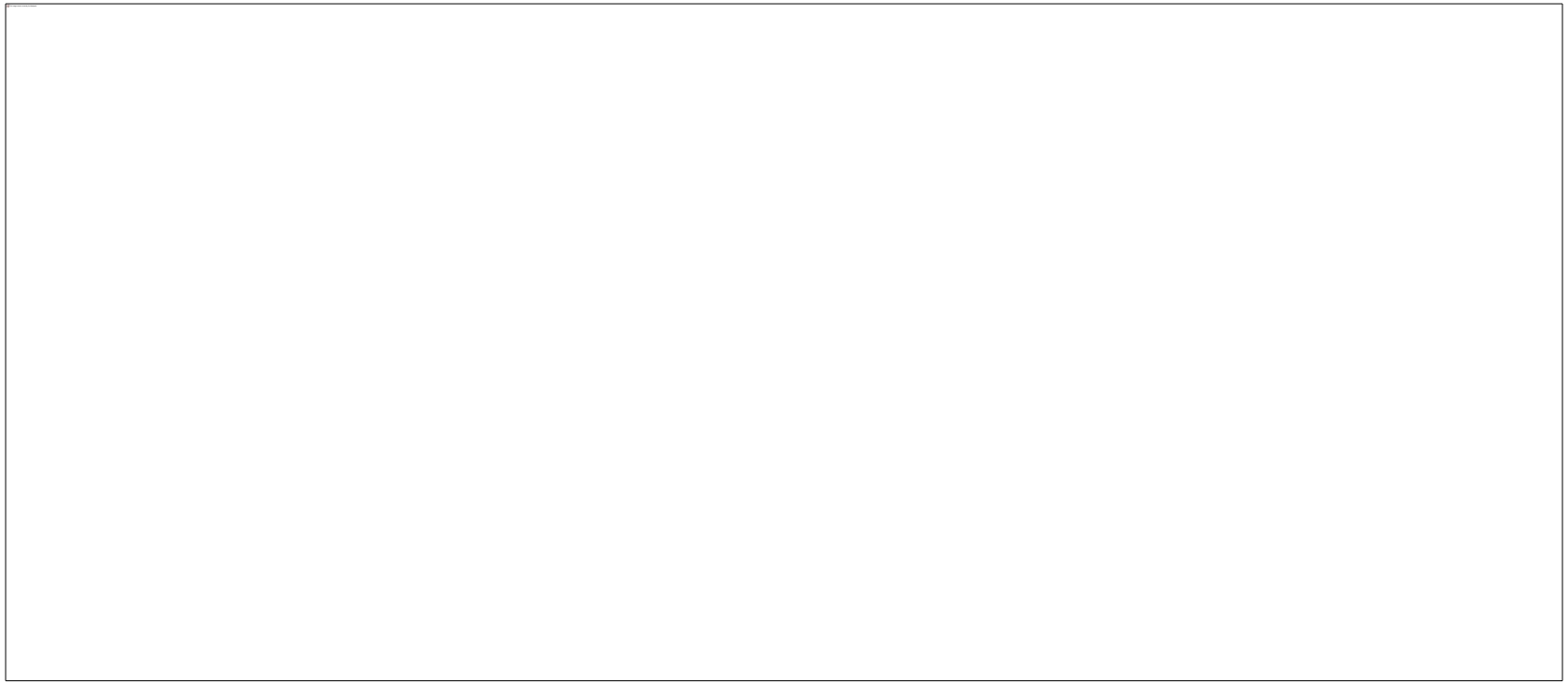
Reliability results from the Meta-analysis

Reliability of DST by MGIT 960

Drug	# Studies (total n)	Sensitivity (95% CI)	Specificity (95% CI)	Agreement (95% CI)
Isoniazid (CC=0.1ug/ml)	10 (811)	98.9% (94.4-99.8)	98.2% (95.4-99.3)	98.7% (97.7-99.7)
Rifampicin (CC= 1ug/ml)	10 (800)	98.2% (92.8-99.6)	99.6% (98.5-99.9)	99.5% (98.6-100)
Ethambutol (CC=5ug/ml)	7 (647)	83.9% (72.8-91.1)	95.8% (81.0-99.2)	95.3% (92.5-98.0)
Streptomycin (CC=1ug/ml)	6 (607)	99.6% (73.6-100)	95.3% (73.4-99.3)	97.1% (95.0-99.2)
Ofloxacin (CC=2ug/ml)	4 (1106)	99.2% (76.4-100)	99.9% (76.4-100)	100% (99.8-100)

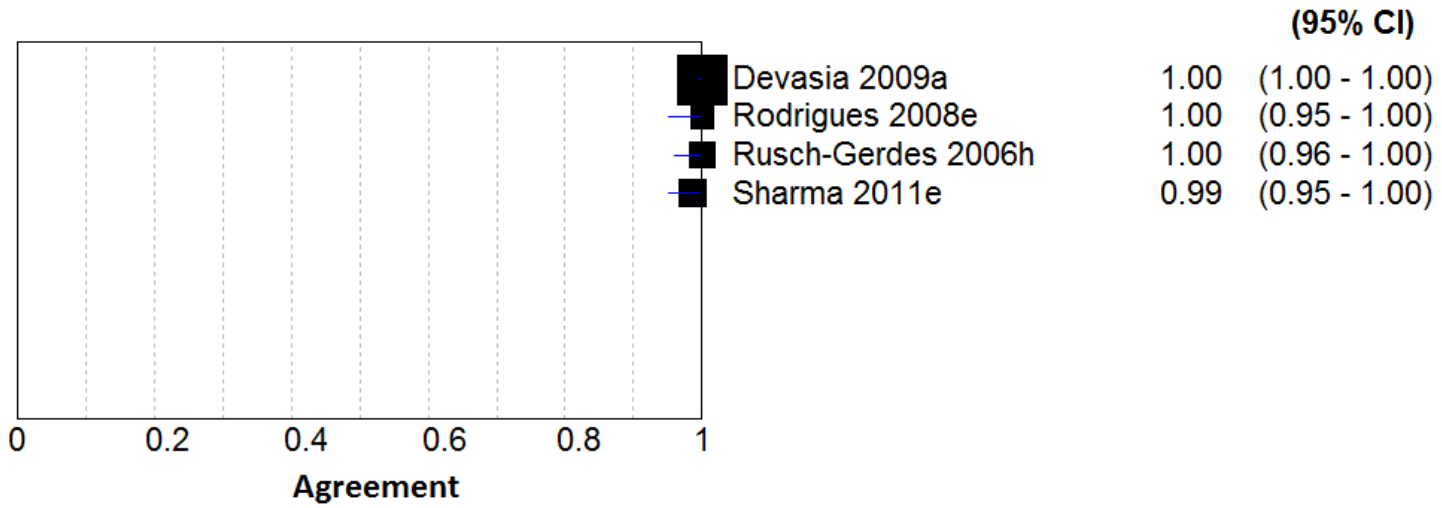
Pooled estimates are from the meta-analysis

Forest plot comparing agreement of MGIT 960 for ethambutol susceptibility testing with a reference standard, CC=5ug/ml , indirect testing



Sensitivity (95% CI)	Specificity (95% CI)	Agreement (95% CI)
83.9% (72.8-91.1)	95.8% (81.0-99.2)	95.3% (92.5-98.0)

Forest plot comparing agreement of MGIT 960 for ofloxacin susceptibility testing with a reference standard, CC=2ug/ml



Sensitivity (95% CI)	Specificity (95% CI)	Agreement (95% CI)
99.2% (76.4-100)	99.9% (76.4-100)	100% (99.8-100)

Reliability of DST by MODS, Direct Testing

Drug	# Studies (total n)	Sensitivity (95% CI)	Specificity (95% CI)	Agreement (95% CI)
Isoniazid (CC=0.1ug/ml)	4 (691)	94.4% (90.1-96.9)	91.8% (82.9-96.2)	92.9% (88.9-96.8)
Rifampicin (CC= 1ug/ml)	5 (823)	97.9% (85.3-99.7)	98.8% (90.8-99.8)	97.5% (94.9-100)

Pooled estimates from the meta-analysis

Reliability of DST by NRA Solid Media, Indirect and Direct

Drug	# Studies (total n)	Sensitivity (95% CI)	Specificity (95% CI)	Agreement (95% CI)
Isoniazid (Indirect) (CC=0.2ug/ml)	14 (1704)	96.5% (94.5-97.8)	100% (93.6-100)	99.1% (98.5-99.7)
Isoniazid (Direct) (CC=0.1ug/ml)	8 (934)	97.2% (94.1-98.7)	98.5% (96.2-99.4)	98.4% (97.3-99.5)
Rifampicin (Indirect) (CC=40ug/ml)	16 (1782)	97.8%* (96.3-98.8)	99.7%* (99.1-99.9)	99.4% (98.9-100)
Rifampicin (Direct) (CC=40ug/ml)	9 (1200)	96.3%* (93.6-98.1)	99.5%* (98.8-99.9)	99.4% (98.7-100)
Ethambutol (Indirect) (CC=2ug/ml)	11 (1333)	94.4% (89.9-97.0)	98.8% (93.6-99.8)	96.8% (94.9-98.7)
Streptomycin (Indirect) (CC=4ug/ml)	11 (1333)	92.4% (84.2-96.5)	96.8% (90.1-99.0)	94.6% (92.1-97.1)

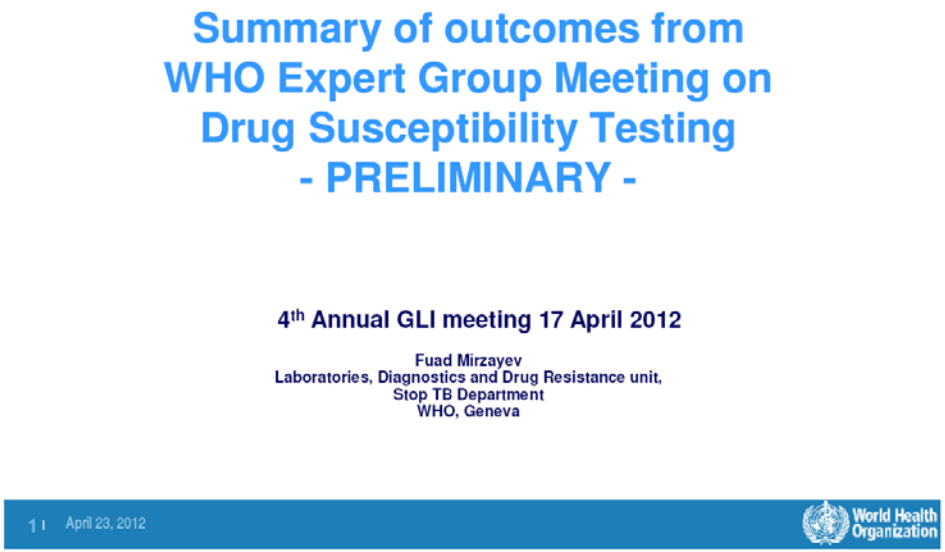
Pooled estimates from the meta-analysis

Intralaboratory and Interlaboratory reproducibility of first-line drugs

Author Year	Method	CC	Comparison (Sites)	Tests	Agreement %
Ethambutol					
Giampaglia 2007	MGIT 960	5.0	Intralaboratory (3)	90	88.9
Giampaglia 2007	MGIT 960	5.0	Interlaboratory (3)	90	88.9
Laszlo 1987	7H10 Proportion	5.0	Intralaboratory (4)	240	97.5
Laszlo 1987	7H10 Proportion	10.0	Intralaboratory (3)	180	98.0
Laszlo 1987	LJ Resistance ratio		Intralaboratory (1)	60	100.0

Intralaboratory and interlaboratory reproducibility of second-line drugs

Author Year	Method	CC	Comparison (Sites)	Tests	Agreement %
Streptomycin					
Giampaglia 2007	MGIT 960	1.0	Intralaboratory (3)	90	97.8
Giampaglia 2007	MGIT 960	1.0	Interlaboratory (3)	90	95.6
Laszlo 1987	7H10 Proportion	2.0	Intralaboratory (5)	300	93.4
Laszlo 1987	7H10 Proportion	10.0	Intralaboratory (5)	300	92.8
Laszlo 1987	LJ Resistance ratio		Intralaboratory (1)	60	100.0
Amikacin					
Lin 2009	MGIT 960	1.5	Interlaboratory (2)	96	100.0
Rusch-Gerdes 2006	MGIT 960	1.0	Interlaboratory (3)	93	100.0
Capreomycin					
Lin 2009	MGIT 960	3.0	Interlaboratory (2)	96	98.0
Rusch-Gerdes 2006	MGIT 960	2.5	Interlaboratory (3)	93	100.0



http://www.stoptb.org/wg/gli/assets/html/day%201/Mirzayev%20-%20Outcomes%20of%20DST%20EGM.pdf

2. DST method accuracy and reproducibility

- Expert Group agreed that DST for **Isoniazid**, **Rifampicin**, **SL injectables** and **fluoroquinolones** are accurate and reproducible across various settings.

It was therefore concluded that testing for these drugs be recommended. All FQs should be tested to guide the choice of the most appropriate agent.

- Expert Group agreed that reaching accuracy and reproducibility for most of Group 4 and 5 drugs remain technically challenging or problematic.

It was therefore concluded that country investment in developing such capacity cannot be recommended until more research has been done.

Additional conclusions

- We found wide ranges of agreement when testing for ethambutol susceptibility for all tests including MGIT 960 at currently recommended critical concentration of 5 ug/ml; **hence a re-evaluation of the currently recommended ethambutol critical concentrations for the index tests studied may be warranted**

Lessons Learned

"What's in a name? that which we call a rose
By any other name would smell as sweet" William Shakespeare

- Identify the review as a diagnostic test accuracy review
- Ensure that the scope is reasonable
- If needed, redefine the review question(s)

- **Diagnostic accuracy and reproducibility of WHO-endorsed phenotypic drug susceptibility testing methods for first-line anti-TB drugs: A systematic review and meta-analysis, manuscript in preparation**

Acknowledgments

- Karin Weyer, WHO Stop TB Department
- Wayne Van Gemert, WHO Stop TB Department
- Christopher Gilpin, WHO Stop TB Department
- Vittoria Lutje, Cochrane Infectious Diseases Group and Liverpool School of Medicine
- Members of the WHO Expert Group on Drug Susceptibility Testing