INNOVATIONS IN TB DATA QUALITY:

An M&E Workshop Facilitators Guide

USAID
FROM THE AMERICAN PEOPLE

TB CARE I

Contents

Authorship

These sessions were compiled by the persons listed below. We developed original material where necessary, but often extracted (with permission) from existing training materials to ensure consistency. Special efforts have been made to acknowledge the authors however authorship should extend to include the authors of the original content on which these training materials are based. The many contributors to each session are noted in the slide set.

Session	Compiled by
- 1: Elements of Data Quality: A Game	Ellen M. H. Mitchell, Rachel Ochola
- 2: How to produce quality data prospectively	Rachel Ochola
- 3: Introduction to Form Design and Quality-Assured Data Entry	Navindra Persaud, Rachel Ochola
- 4: Hands-On Skills Practice in Quality-Assured Data Entry	Rachel Ochola, Navindra Persaud
- 5: Country Experiences in Developing an Electronic R&R System	Claire Moodie
- 6: How to Develop Electronic Recording & Reporting Systems	Claire Moodie
- 7: How to Ensure Quality of Qualitative Data	Suzanne Cloutier, Navindra Persaud
- 8: Data Management of Large Electronic Datasets	Rachel Ochola
- 9: Quality-Assured Joining of TB Data Sets with a Unique ID	Navindra Persaud
- 10: Theory, Structure, and Process of Data Quality Assessment	Suzanne Cloutier
-11: Sampling and Logic Checks of Paper Surveillance Data	Ellen M. H. Mitchell
- 12: Exploring the Quality of Your Electronic TB Data	Ellen M. H. Mitchell, Suzanne Cloutier
- 13: Introduction to Routine Data Quality Assessment	Suzanne Cloutier
- 14: Practice - with Audit Case Study	Suzanne Cloutier
- 15: Use of WHO Checklist of Standards and Benchmarks for TB Surveillance and Vital Registration Systems	Emily Bloss, Rachel Ochola
- 16: M&E of TB Mortality	Ellen M. H. Mitchell
- 17: Revised WHO Case Definitions and Reporting Forms for 2013	Claire Moodie
- 18: M&E of Contact Investigations & Screening Programs	Ellen M. H. Mitchell, Abbas Zezai, Knut Lönnroth
- 19: M&E of TB in Health Care Workers and other Occupational Groups	Suzanne Verver, Max Meis, Ellen M. H. Mitchell
- 20: Screening M&E Skills Practice with Electronic Data	Ellen M. H. Mitchell
- 21: M&E of Programmatic Management of MDR (PMDT)	Claire Moodie
- 22: How to Cope with Poor Quality Data	Ellen M. H. Mitchell
- 23: How to Link Datasets When There are No Unique IDs	Suzanne Cloutier, Rachel Ochola
- 24: Skills Practice with Link Plus Software	Suzanne Cloutier
- 25: Data Are Human – The Politics and Practice of TB Data Exchange	Ellen M. H. Mitchell
- 26: M&E for TB/HIV Data Integration	Ellen M. H. Mitchell
- 27: Assessing Under-Reporting through Inventory Studies	Ellen M. H. Mitchell, Emily Bloss
TB M&E Data Quality Glossary	Rachel Ochola

Acknowledgments

This Facilitator's guide benefitted from the contributions of many technical experts in the M& E and data quality fields. Ineke Huitema designed the template for this course and provided valuable input on its design. Nico Kalisvaart and Navindra Persaud reviewed the document for technical content.

Colleen Mazin edited the entire document. Colleagues at WHO and US CDC provided many of the slides and assisted with the interpretation of dense technical content for a wider audience.

The Global Health Bureau, Office of Health, Infectious Disease and Nutrition (HIDN), US Agency for International Development, financially supports this publication through TB CARE I under the terms of Agreement No. AID-OAA-A-10-00020. Suggested citation:

Mitchell, E., Cloutier, S., Moodie, C. Ochola, R., Persaud, N., Bloss, E. Huitema, I. Innovations in TB Data Quality: An M&E Workshop Facilitators guide. 2014 TB CARE I www.tbcare1.org

Acronym List

ART	Antiretroviral treatment
BMU	basic management unit
CBOs	Community-based organizations
CDC	Centers for Disease Control
CI	Contact investigation
CPT	cotrimoxizole
DLTLD	Kenyan Division of Leprosy, tuberculosis, and Lung Disease
DQ	Data quality
ERR	Electronic Recording & Reporting
H_0	Null hypothesis
HCW	Health care workers
HIV	Human immunodeficiency virus
IPT	Isoniazid preventive therapy
M&E	Monitoring and evaluation
NGOs	Nongovernmental organizations
NTPs	National TB Programs
NNS	Number needed to screen
PMDT	Programmatic Management of MDR
PTB	Pulmonary Tuberculosis
PRJ	Project file
QA	Quality assurance
RDQA	Routine data quality assessment
R&R	Recording & reporting
SOPs	Standard operating procedures
TB	Tuberculosis
TIBU	Kenya's TB surveillance system
WHO	World Health Organization
XDR-TB	Multidrug resistant TB

Background on the Need for and Purpose of the Training

Monitoring and evaluation (M&E) are critical to measuring and reporting on the success of National TB (Tuberculosis) Programs (NTPs) and the TB CARE I/II projects. While governments and donors are placing greater emphasis on results, at the country level, greater attention is being paid to the use of data for improving patient care and enhancing program management. In order to ensure that adequate capacity exists to meet the increasingly stringent M&E requirements, this course was designed to build the capacity of M&E Officers of NTPs and technical partners.

This course has three over-arching themes. They are to **avoid, detect**, and **fix** data quality problems. These three themes seamlessly map onto the three tracks of our TB work, which is to **prevent, diagnose**, and **treat** TB. They also permeate the monitoring and evaluation sessions, including those that focus on the measurement of challenges posed by the rollout of:

- The 2013 Revised Case Definitions & Reporting Forms
- The 2012 Programmatic Management of MDR (PMDT) Guidelines & Companion Handbook
- The 2011 World Health Organization (WHO) Contact Investigation Guidelines
- The 2013 WHO Screening Guidelines
- The new Stop TB Targets and Indicators Post-2015

The 5 modules include:

1. The Prevention Module (Avoid Problems), which includes sessions on database design, quality-assured TB data entry and management, and standardization, including the development of standard operating procedures (SOPs).
2. The Detection Module (Find Problems), which includes both basic and advanced skills in TB data appending, merging, and linking, using both unique identifiers as well as probabilistic and deterministic algorithms.
3. The Treatment Module (Fix Problems), which covers management of missing, duplicative, and inaccurate TB data.
4. New Challenges in TB M&E, which offers some new areas and emerging challenges in measurement of active case finding, prevention, and infection control
5. The Humanity Module (M&E as Collaboration), which covers the often man-made obstacles to good M&E (e.g., lack of trust) and how to use persuasion and collaboration to resolve them.

Increasingly, the successful avoidance of TB data quality problems requires both technical skills as well as "soft skills," i.e., the ability to encourage safe and ethical data

exchange by often reluctant stakeholders. This course includes both core computational skills and a negotiation skills component.

The identification of quality problems in TB data has recently advanced significantly with the publication of a suite of five audit tools (see Core Texts below). This course offers learners the chance to contrast the various approaches to routine monitoring versus periodic audit. We cover both more participatory and less collaborative methods of assessing data quality, and the relative merits of both. The curriculum includes methods for scrutinizing data quality at all levels of a health system.

Moreover, M&E officers are introduced to techniques to assess the validity of both quantitative and qualitative data, using state-of-the-art techniques. Learners gain skills in discerning how and when to apply different tools and, more importantly, how to translate findings into an action plan.

Unlike other courses, which emphasize a strictly preventative approach to ensuring data quality *a priori*, this course empowers participants to triage data quality problems and fix them, if possible. The authors openly acknowledge that M&E officers often confront situations in which low quality data are all that is available. The curriculum covers both ideal as well as more pragmatic methods of "fixing" data quality problems on a tight timeline. This course helps learners to distinguish when data can be salvaged and when they cannot.

This training is part of a capacity building effort that consists of virtual and in-person mentoring of country-level staff, short term technical assistance to selected countries, and development of in-person training on TB data management at the country level. Additional complementary training materials can be found at www.tbcare1.org.

Goal

The overall goal is to strengthen the M&E capacity among staff and partners of the national TB program.

Specific Objectives

By the end of the course, participants will be able to:

- Apply best practices to produce quality TB data prospectively.
- Assess the quality of TB data using new tools.
- Differentiate between valid and invalid methods of handling poor quality data.
- Develop and implement appropriate strategies to deal with low quality TB data in the short term (remedial) and the long term (prevention).

Target Group(s)

This training will build upon the foundation established during a three-day, in-person M&E training for NTP M&E staff from 15 countries and TB CARE I and II M&E counterparts held in 2011. The previous training provided a forum to share experiences on common challenges and approaches.

This guide is intended to support M&E planning and practice for a wide audience of stakeholders, including:

- M&E officers from projects and NTP (central level) involved in either routine TB data management or other TB program M&E efforts.
- M&E officers working in TB programs at all levels of the health system.
- Technical partners who design, implement, and evaluate TB activities.
- Global Fund recipients and consultants who provide technical assistance for Global Fund projects and applications.
- Civil society organizations working at all levels to improve TB services. These include community-based organizations (CBOs), faith-based organizations, and other nongovernmental organizations (NGOs) implementing TB activities.

Course Philosophy

The course is guided by three maxims:

1. An ounce of prevention is worth a pound of cure.
2. Learn by doing and through reflection.
3. Strive to improve M&E in areas you can influence, and learn to deal with data from sources that you could not influence.

Pre-requisites and Inclusion Criteria for the Training

This course is designed for individuals who are:

- Working in the field of TB M&E;
- have basic computational skills (arithmetic);
- have a basic knowledge of TB care and control; and
- Have completed prior basic (TB) M&E training.

Computing Requirements

As this course is data-intensive, participants must have the following:

- A functional lap top with Windows or Mac OS.
- A functioning version of CDC (Centers for Disease Control) Link Plus or other linking software.
- A functioning version of Epi-Info 3.5.5, Epi-Info 7, Epidata 3.1, and MS excel 2007.

If participants do not have administrator privileges on their computers, facilitators should encourage them to work with IT administrators before the course to ensure they are able to save executable files and install these software tools.

Link Plus 2.0[1]

Link Plus is a small piece of a bigger family of cancer surveillance software, but we only want to utilize this small part of the program. Link Plus 2.0 requires administrator privileges to install, so it is essential that you work with your systems administrator to install it prior to leaving your workplace.

It can be downloaded for free here:
http://www.cdc.gov/cancer/npcr/tools/registryplus/lp_tech_info.htm

1. Download Link Plus, RPLinkPLus_2.0.exe (executable file, 20.7 MB, June 29, 2007) to your computer.
2. Open the downloaded file. The installation program will direct you through the steps for installing Link Plus. If you are a first-time user, we recommend you select the defaults.
3. Once installed, click on the Windows Start button and select Programs, then Registry Plus, then Link Plus.

Epi-Info

Please go to http://wwwn.cdc.gov/epiinfo/ to download your free version of Epi-Info. Please install and configure versions 3 and 7 and open them to make sure they are functional. Note that the two versions can co-exist on your laptop without any problems. During the course, participants will use Epi-Info versions 3 and 7 to undertake different tasks. While the intention was to use only version 7 for the training, this is not possible, as a number of the key functions in version 7 are still not yet operational.

Self-Study Preparation for the Course

The course facilitators assume that learners have completed the following pre-requisites:

1. Completed the MEASURE evaluation online course on human immunodeficiency virus (HIV) data quality (1.5 hrs):
https://training.measureevaluation.org/related-online-courses/data-quality

2. Viewed two of the following YouTube videos:
    - Getting Started with Epi-Info 7 (5 min)

[1] A Portuguese video tutorial on Link Plus is available from Dr. Ana Luiza Bierrenbach albierrenbach@yahoo.com.br

- http://www.youtube.com/watch?v=FYrWiLGO7ZE
- Epi-Info7: An Overview by Erik Knudsen (42:53min)
- http://www.youtube.com/watch?v=-uTHl9E6NK8

Core Texts for this Course

These core texts are all provided electronically and can also be downloaded before the course.

Understanding and Using Tuberculosis Data. WHO 2014
www.who.int.

Companion Handbook to the WHO Guidelines for the Programmatic Management of Drug-Resistant Tuberculosis. 2014 www.who.int.

Definitions and Reporting Framework for Tuberculosis – 2013 Revision:
http://apps.who.int/iris/bitstream/10665/79199/1/9789241505345_eng.pdf

Checklist of Standards and Benchmarks for TB Surveillance and Vital Registration Systems. WHO 2013.

Assessing tuberculosis under-reporting through inventory studies. WHO 2012.

Manual on use of routine data quality assessment (RDQA) tool for TB monitoring. WHO 2011.

Data Quality Audit Tool. Guidelines for implementation. MEASURE Evaluation.

Routine Data Quality Assessment: Standard Operating Procedure. Botswana MEASURE Evaluation.

Electronic recording and reporting for tuberculosis care and control. WHO 2010:
whqlibdoc.who.int/publications/2012/9789241564465_eng.pdf

Systematic screening for active tuberculosis: principles and recommendations:
http://www.who.int/tb/tbscreening/en/

Recommendations for investigating contacts of persons with infectious tuberculosis in low- and middle-income Countries. WHO 2012:
http://www.tbcare1.org/publications/toolbox/tools/hss/Guidelines_on_TB_Contact_Investigation.pdf

The Lessons from Loss Tool: www.tbcare1.org/access

Tuberculosis prevalence surveys: A handbook. WHO 2011 (aka The Lime Book).

Guide on the Monitoring of TB Disease Incidence Among Health Care Workers: http://www.tbcare1.org/publications/toolbox/tools/hss/HCW_TB_Incidence_Measuring_Guide.pdf

Guide to Measure the Prevalence of Active TB Disease Among Health Care Workers: http://www.tbcare1.org/publications/toolbox/tools/hss/HCW_TB_Prevalence_Measuring_Guidelines.pdf

A Guide to Monitoring and Evaluation for Collaborative TB/HIV Activities. Geneva, WHO 2009: http://www.who.int/hiv/pub/tb/hiv_tb_monitoring_guide.pdf.

Compendium of Indicators for Monitoring and Evaluating National Tuberculosis Programs. Geneva, WHO 2004: www.who.int/tb/publications/tb_compendium_of_indicators/en/index.html.

Conducting Quality Impact Evaluations under Budget, Time and Resource Constraints. Washington, D.C., The World Bank 2006: http://siteresources.worldbank.org/EXTEVACAPDEV/Resources/4585672-1251461875432/conduct_qual_impact.pdf.

Useful videos for self-study and reinforcement of key skills

For teaching in areas without internet access it is recommended that facilitators contact CDC to obtain digital copies of these files. It is helpful for learners to have their own copies to refresh their memories.

1. 8:30 min Epi-Info 7 overview
   http://www.youtube.com/watch?v=tnWiGyIgnV4

2. Epi Info 7 Create a project template
   http://www.youtube.com/watch?v=uvPpKW1LnmY

3. 2:36 min how to import Excel files into Epi_Info 7:
   http://www.youtube.com/watch?v=CgRCBord-YA

4. 3:31 Epi Info 7 Skip pattern using an "if, then" statement
   http://www.youtube.com/watch?v=Ww4dAplEnTI

5. 12:03 min Creating legal values
   http://www.youtube.com/watch?v=Lv6_v-pGeRA

6. CDC Epi Info 7 How-to: Moving Files
   http://www.youtube.com/watch?v=VxMVCdYUzVo

7. Epi Info 7 Create a form from an Excel spreadsheet 3:30
   http://www.youtube.com/watch?v=dWQLozkWOiE&feature=c4-overview&list=UUGQtxBzqAAQErVOikz2JTpQ

8. 7:37 min Methods and hazards of deduplication in Excel
   http://www.youtube.com/watch?v=6HNX_tk2VxU

9. How To Merge / Join Data From Tables In Excel Using *vLookup*
   http://www.youtube.com/watch?v=3tk_Mif7040

10. How to use Epi-Info Compare Utility **v. 3.4.3**
    http://www.youtube.com/watch?v=24gBswHdZ8U

11. 8:17 Epi Info 7 Aberration Detection with visit level data
    http://www.youtube.com/watch?v=kmCtyX5OlcU

12. 3:10 Epi Info 7 Aberration detection analysis with aggregated data
    http://www.youtube.com/watch?v=kmCtyX5OlcU

13. Epi Info 7 Packaging data for sharing
    http://www.youtube.com/watch?v=1rKLAPRp71g

14. 3:55 Epi Info 7 2x2 Table Analysis
    http://www.youtube.com/watch?v=8BZZvU4zy_c

15. Epi Info 7 Linear regression analysis
    http://www.youtube.com/watch?v=A7R1rjjHb98

16. Epi Info 7 Create a column chart
    http://www.youtube.com/watch?v=Fr5hR01GV8Q&list=UUGQtxBzqAAQErVOikz2JTpQ

Methodology

The methods of the course reflect a range of individual learning styles and the principles of adult learning. It is a highly interactive training using actual electronic and paper-based data. The methodology includes:

- Debates and triangulation
- Field visits
- Games and friendly competition
- Group exercises using actual data from the countries
- Home work, including problem sets and individual exercises
- Pre- and post-tests on knowledge and application of skills
- Self-study, including a review of various texts
- Sharing of country experiences
- YouTube videos

Guidelines on the Organization, Logistics, and Preparations for the Training

Course organization will depend on resources and group size. We recommend no more than 30 people be enrolled in the course, and a student to facilitator ratio no greater than 4:1. Important points to remember are:

- Leave sufficient time for the students to grapple with the tools. Do not lecture the whole time, as this will not properly convey the material or help build skills in the learners.
- Make sure learners are given sufficient advance notice of the course to allow them to complete all of the pre-requisites and readings.
- . A needs assessment before the course can give you a sense of the skill range among the students (see sample in appendix). This will allow you to identify those who should be given co-facilitation and teaching roles.

Facilitators

Facilitators for this course should be a mix of experience and expertise. All facilitators need a strong understanding of the basic principles of tuberculosis and monitoring and evaluation. At least one facilitator should be a data manager or someone who has extensive experience managing datasets. Facilitators who are very senior researchers are unlikely to have the data quality trouble shooting skills that learners may require. At least one facilitator should be very conversant in the use of all of the software. At least one facilitator should be highly competent in training dynamics and pedagogy.

Overview of the Training Modules

Module 1: Avoid Problems

1. A game to explore elements of data quality.
2. How to produce quality data prospectively (i.e., best practices).
3. Introduction to quality assured data entry in Epi-Info.
4. Hands-on skills practice in quality assured data entry in Epi-Info.
5. Short film on Kenya's TB Surveillance system called TIBU - health care delivery innovation for tuberculosis (Kenya DLTLD) (8 min).
6. How to develop electronic recording & reporting (ERR) systems.
7. How to ensure quality of qualitative data.
8. Back up and security procedures for large electronic datasets with emphasis on good practices.
9. Introduction to quality-assured matching/joining/appending of TB data sets with a unique identifiers (IDs).

Module 2: Find Problems

10. Theory, structure, and process of data quality assessment.
11. No time for a proper audit? How to do quick and simple logic checks of paper-based TB surveillance data.
12. How to explore quality of electronic data – Epi-Info skills practice session and aberration detection.
13. Introduction to routine data quality assessment (RDQA).
14. Practice session with audit case study.
15. Quality of TB surveillance systems and use of WHO checklist of standards and benchmarks for TB surveillance and vital registration systems discussion.
16. M&E of TB mortality – lessons from loss tools and death audits.

Module 3: New Challenges in TB M&E

17. Revised WHO case definitions & reporting forms for 2013.
18. Monitoring & Evaluation of contact investigations and other screening/active case finding programs.
19. How to measure TB prevalence among health care workers (HCW) as a part of M&E of infection control.
20. How to monitor, evaluate, and report on TB/HIV efforts – Implications of the new case definitions.
21. Introduction to PMDT efforts – Implications of the new case definitions.

Module 4: Fix Problems

22. Working with data using revised WHO case definitions & reporting forms for 2013.

23. How to cope with poor quality TB data – Guidance for triage of imprecise, incomplete, inaccurate, and >5% missing data (from inference and imputation) and exercises in working with poor quality data.
24. Screening M&E skills practice with electronic data to calculate the yield of TB screening, the scope of a contact investigation, and to evaluate isoniazid preventive therapy (IPT) uptake among eligible individuals.
25. M&E Skills practice calculating PMDT treatment outcomes and core indicators.

Module 5: M&E as Collaboration

26. Data are human: The politics and practice of TB data exchange for safe and ethical sharing of data and role playing.
27. How to link datasets when there are no unique IDs, and an introduction to CDC's Link Plus Software.
28. Overview of methods of detecting under-reporting (i.e., low level of completeness): The data verification simulation game.

Sample Timetable of an Innovations in Data Quality M&E Course

Below is a sample agenda for a six day course. This content could best be taught over 8 days, but that length is rarely feasible.

Agenda for Innovations in Data Quality: An M&E Workshop

Day 1 Theme: AVOID PROBLEMS

Time	No.	Description
8:30 – 9:00		Registration of participants
9:00 – 9:30		Opening Session - Inspiration
9:30 – 10:00		Introductions and& icebreaker
10:00-10:30		Course orientation, ground rules, self-study
10:30 – 10:45		Break
11:00 – 11:30	1	A game to explore the 16 elements of data quality
11:30 – 12:30	2	How to produce quality data prospectively (i.e., best practices) and SOPs data dictionaries
12:30 – 13:30		Lunch
13:30 – 14:30	3	Introduction to quality assured data entry in Epi-Info 7 • create a form - walk them through it • skip patterns - video in class • legal values - homework
14:30 – 15:30	4	Hands-on skills practice in quality assured data entry in Epi-Info 7
15:30 – 15:45		Break
15:45 – 15:55	5	Short film on TIBU - Health Care Delivery Innovation for Tuberculosis (DLTLD) (8 min)
16:00-16;45	6	How to develop electronic recording & reporting (ERR) systems
16:45 – 17:00		Wrap-up

Day 2 Theme: PLAN AHEAD

Time	No.	Description
9:00 – 9:30		Homework highlights
9:30 – 10:30	7	How to ensure quality of qualitative data?
10:30 – 10:45		Break
10:45 – 12:30	8	Back up and security procedures for large electronic datasets, with emphasis on good practices
12:30 – 13:30		Lunch
13:30 – 15:30	9	Introduction to quality-assured matching/joining/appending of TB data sets with a unique ID
15:30 – 15:45		Break
15:45-16:45	9	Practice session joining data bases
16:45 – 17:00		Wrap-up

Day 3 Theme: FIND PROBLEMS

Time	No.	Description
9:00 – 9:30		Homework highlights
9:30 – 10:30	10	Theory, structure and process of data quality assessment
10:30-10:45		Break
10:45 – 12:30	11	Sampling and Logic checks of Paper Surveillance Data
12:30 – 13:30		Lunch
13:30 – 14:15	13	How to conduct routine data quality audits
14:15 – 15:15	15	Use of WHO *Checklist of Standards and Benchmarks for TB Surveillance*
15:15 – 15:30		Break
15:30 – 17:00	14	Practice session with audit case study
17:00 – 17:15		Wrap-up with feedback and explanation of homework

Day 4 Theme: NEW CHALLENGES IN M&E FOR TB

Time	No.	Description
9:00 – 9:30		Homework highlights
9:30 – 10:15	17	*Revised WHO case definitions and reporting forms for 2013*
10:15-10:30		Break
10:30 – 11:45	17	Review of new TB Registers and reporting forms
11:45-12:30	18	M&E of contact investigations and other screening/active case finding programs
12:30 – 13:30		Lunch
13:30 – 14:30	19	How to measure TB prevalence among HCW as a part of M&E of infection control
15:00 – 15:15		Break
15:15 – 16:45	20	Screening M&E skills practice with electronic data, to calculate the yield and scope of a CI and to calculate the percentage of IPT uptake among eligible persons.
16:45 – 17:00		Wrap-up
TBD		Group dinner/social event

Day 5 Theme: FIX PROBLEMS

Time	No.	Description
9:00 – 9:15		Highlights from homework, warm-up
9:15-10:30	21b	Interpreting paper data based upon *Revised WHO case definitions and reporting forms for 2013*
10:30-10:45		Break
10:45 – 12:00	12	Exploring the quality of your electronic data
12:00-12:30	22	How to cope with poor quality TB data - Guidance for triage
12:30 – 13:30		Lunch
13:30-14:00	22	Exercises in working with poor quality and missing data- Imputation
14:00 – 15:00	21	Introduction to Programmatic Management of MDR (PMDT) efforts and implications of the new case definitions
15:00 – 15:30	21	Review of the new PMDT Registers
15:30 – 15:45		Break
15:30 – 16:45	21	M&E skills practice calculating PMDT treatment outcomes and core indicators
16:45 – 17:15		Wrap-up

Day 6 Theme: TB M&E AS COLLABORATION

Time	No.	Description
9.00 – 9:30		Highlights from homework, warm-up
9:30-10:30	25	Data are human – the politics and practice of data exchange, covering the safe and ethical sharing of data with role plays
10:30-10:45		Break
10:45 – 12:30	23	How to link datasets when there are no unique IDs, and an introduction to Link Plus Software
12:30 – 13:30		Lunch
13:30-15:00	24	Skills Practice with Link Plus software
15:00 – 15:15		Break
15:15 – 16:45	26	How to monitor, evaluate, and report on TB/HIV efforts, and implications of the new case definitions
16:45 – 17:00		Wrap Up

Day 7 Theme: TB M&E AS INNOVATION

Time	No.	Description
9.00 – 9:30		Highlights from homework, warm-up
9:30-10:30	16	M&E of TB Mortality
10:30-10:45		Break
10:45 – 12:30	1	Part II Data Quality Game- TB Examples
12:30 – 13:30		Lunch
13:30-14:30	27	Overview of methods of detecting under-reporting (low level of completeness)
14:30 – 15:00	27	Data verification simulation game
15:00 – 15:15		Break
15:15 – 16:15		Course exam (Post test)
16:15 – 16:45		Course evaluation
16:45 – 17:00		Course closure/group photo

Sample Ground Rules (adapt as desired)

1. Hold yourself accountable for achieving the outcomes of this training.
2. Take risks, such as telling the truth or questioning conventional thinking.
3. Make maximum use of our multi-disciplinarity by asking each other questions. We all know many things.
4. Try to practice *creative abrasion* (Hirschberg et al). In other words, challenge each other and debate the options.
5. Distinguish between processes that require us to be creative vs. processes for which we need to be systematic.
6. Resist the urge to take the easy way out or to do something superficial.
7. Define the jargon and acronyms you use so that others can understand you.
8. Be as flexible as your mind will allow. Think outside the box. Be open to new ideas and methods.
9. Try to reframe the problems we will encounter as *challenges*.
10. If what we are doing seems boring or unnecessary, gently suggest a new direction.

Comprehension Exercise

This can serve as a pre-test and a post-test if facilitators desire to measure short-term knowledge gain. Of course we are more concerned with competency, but these tests can be useful to highlight who may need additional support in the field.

Name: _____ Title: _____ ____Country:_____

Innovations in Data Quality: an M&E workshop
Learning Summary Exercise

The purpose of this exercise is to assess the effectiveness of our teaching
Complete the answers for the sessions you attended

Data quality

1. **Accurate data** have (Choose one)

    a. Minimal errors
    b. Minimal bias
    c. Validity
    d. All of the above

2. True or False: Data are **reliable** when they are measured and collected *consistently over time*. (Choose one)

    a. True
    b. False

3. **Completeness** means that an information system (Choose one)

    a. Captures *all* of the eligible persons, services, sites, or other units that it is supposed to measure.
    b. Is unaffected by timeliness of data.
    c. Is free of bias.

4. The information system lacks _____ if it is not designed to record the exact age of individuals diagnosed with TB. (Choose one*)*

    a. Robustness
    b. Precision
    c. Replicability
    d. Fairness

5. Data have _____ when the information system is *protected from deliberate bias or manipulation* for political or personal reasons. (Choose one) *(*

    a. Equality
    b. Sustainability
    c. Integrity

d. Elasticity

6. **Confidentiality** means that (Choose one)

a. Clients are assured that their data will be maintained in a way that does not cause them social or physical harm.
b. Privacy and security measures are in place.
c. No deductive disclosure occurs.
d. all of the above

7. When some reports are received late, that information is not available for the aggregate report, which affects the **accuracy, reliability, completeness,** and **timeliness** of the data. The possible causes of the late data reports are: (Choose one)

a. M&E staff at all reporting sites have not been informed in writing the date reporting is required.
b. M&E staff turnover has resulted in new staff members without written documentation of reporting requirements.
c. M&E staff has not demanded accurate, reliable, complete, and timely data reports.
d. All of the above

8. Parallel data collection systems can also lead to _____of the services provided. (Choose one)

e. Miscounting
f. Under-counting
g. Double counting

9. If a dataset records the same patient more than once, what must be done?

a. Verification
b. De-duplication
c. Regionalization

10. True or False: Reliability and Validity mean the same thing.

d. True
e. False

11. If a program decides to redefine an indicator it is collecting from month to month, which dimension of data quality is most directly affected? (Choose one)

f. Accuracy
g. Reliability
h. Precision
i. Completeness
j. Integrity

12. Which of the following is a strategy to maintain the reliability of the data? (Choose one)
a. Develop standardized, written instructions for data collection.
b. Keep records in a locked cabinet.

c. Protect electronic files with a password.
d. All of the above

13. What is meant by the term accuracy? (Choose one)

    a. The level of detail at which data is stored.
    b. The lack of bias in the data.
    c. The extent to which a value approaches its true value.
    d. The overall quality of the data.

14. What is meant by the term data quality? (Choose one)

    e. The lineage of the data.
    f. The resolution of the data.
    g. The generalization present in the source data.
    h. The inherent quality of the data as characterized by its accuracy, precision, bias, level of error, etc.

<u>TB-HIV Session</u>

15. Name three types of double counting:

    a.
    b.
    c.

16. What is benchmarking? (Choose one)

    a. A means of identifying outliers through ecological comparison.
    b. A method for triangulation of qualitative data.
    c. A method of verifying accuracy.

<u>Coping with Bad Data Session</u>

17. Missing data on sexual behavior due to participant non-response is very unlikely to be which type of missing data: (Choose one)

    a. Missing completely at random
    b. Missing at random
    c. Missing not at random

18. To determine what to do about missing data, the following information is needed: (Choose one)

    a) When is the report due?
    b) Who will be reading the report?
    c) How much data are missing?
    d) Is the data paper or electronic?

<u>Screening Sessions</u>

19. Why do we consider HCW a priority population for TB screening? (Choose one)

a. They are a vital resource for every country.
b. They work in a congregate setting.
c. They work with vulnerable populations.
d. They have a right to work in a safe environment.
e. All of the above

20. If there are 40,000 health workers in Maravilha, and 10,000 are screened each year, and 200 cases of TB are found, what is the number needed to screen to find 1 case of active TB among the health care workers? (Choose one)

    a. 100
    b. 50
    c. 200
    d. 1000

21. What is the key impact indicator for M&E of contact investigation? (Choose one)

    a. Percentage of eligible child contacts under 5 year placed on IPT.
    b. Percentage of people with HIV put on IPT.
    c. Percentage of smear positive TB patients who report names of contacts.
    d. Percentage of TB patients who complete treatment.
    e. Percentage of screened contacts that are diagnosed with TB.
    f. All of the above

22. What are the main differences between a screening test and a diagnostic test? (Choose one)

    a. A screening test is for ruling-out disease, but a diagnostic test is for ruling it in.
    b. A screening test should always be low tech, and a diagnostic test is high tech.
    c. A screening test needs to be inexpensive, but a diagnostic test can be expensive.
    d. A screening test should be sensitive, and a diagnostic test should be both sensitive and highly specific.

<u>Routine Data Quality Assessment Session</u>

23. In general, how long should you allow per site for an RDQA? (Choose one)

    a. 2- 4 hours
    b. ½ -1 day
    c. 1-1½ days
    d. 1-2 days

24. Which is not part of the RDQA implementation process? (Choose one)

    a. Interpret results
    b. Indicator selection
    c. M&E framework development
    d. Action plan development
    e. Site selection

<u>Quality Data Management</u>

25. Accidental or malicious loss of data can be due to: (Choose one)

a. Hardware faults or failure
b. Software or media faults
c. Virus infection or malicious hacking
d. Power failure
e. Human errors by changing or deleting files
f. All of the above

26. A plausible order for Data Life is: (Choose one)

a. Plan, Process, Acquisition, Analyze, Preserve, Publish/Share
b. Plan, Acquisition, Analyze, Process, Preserve, Publish/Share
c. Plan, Acquisition, Process, Analyze, Preserve, Publish/Share
d. Plan, Publish/Share, Preserve, Analyze, Process, Acquisition

Case Definitions Session

27. True or False: One of the main reasons WHO has revised reporting forms is to permit the inclusion of TB cases detected using WHO-approved rapid diagnostics.

a. True
b. False

28. TB terminology was changed to be less judgmental. Select the correct change(s): (Choose one)

a. *MDR-TB* is now known as *RR-TB*
b. *Defaulter* is now known as *Lost to follow-up*
c. *TB suspect* is now known as *presumptive TB*
d. b and c
e. all of the above

Electronic Recording and Reporting Sessions

29. Which of the following is NOT an advantage to a well-functioning electronic recording and reporting system? (Choose one)

a) Data quality
b) Timeliness
c) Managing complex data
d) Upfront costs

30. True or False: You need to have a strong paper-based system in place before implementing an ERR system effectively. (Choose one)

a. True
b. False

Exploring Your Data Session

31. Which software cannot currently be used for double data entry? (Choose one)

a. Excel 2007
b. Epi-Info v.3.1
c. Epi-Data v. 3.1
d. Epi-Info v. 7

32. What are exploratory frequencies good for? (Choose one)

    a. To calculate the error rate
    b. To look for missing data
    c. To do multiple imputation
    d. To do double data entry

33. What are exploratory cross tabulations (2x2 tables) good for? (Choose one)

    a. To derive standard deviations
    b. To do logic checks
    c. To calculate the means
    d. To apply weights

Qualitative Data Quality Session

34. Inter-rater reliability is used in what types of M&E? (Choose one)

    a. Qualitative
    b. Quantitative
    c. Neither
    d. Both

35. Select the output-oriented methods used to assess quality in qualitative research: (Choose one)

    a. Triangulation
    b. Comprehensiveness
    c. Deviant/negative case analysis
    d. All of the above
    e. a and c only

36. The fundamental principles of qualitative research include: (Choose one)

    a. Reflexivity, transparency, comprehensiveness, responsibility, ethical practice, systematic approach
    b. Validity, rigor, confirmability, credibility, trustworthiness
    c. Triangulation, respondent validation, reflexivity, attention to negative cases, transparency, relevance
    d. Rigor, objectivity, representativeness, comprehensiveness, context sensitivity
    e. None of the above

Standards and Benchmarks Session

37. True or False: The who standards & benchmarks checklist consists of a set of 13 standards and associated benchmarks with nine standards related to TB case measurements, and one standard related to TB deaths measurement. (Choose one*)*

a. True
b. False

43. The lab-confirmed cases vs. clinically diagnosed cases is an example of :

    a. Standard
    b. Benchmark
    c. Both
    d. None of the above

Epi-Info 7 Session

44. True or false: A view is a data entry interface for an individual table.

    a. True
    b. False

45. True or false: It is advisable to use both Microsoft Access and Epi-Info to enter data into the same data table?

    a. True
    b. False

PMDT Session

46. Tracking interim results of MDR-TB patients is important for:

    a. Preventing the development of XDR-TB.
    b. Giving you a chance to improve your PMDT care services.
    c. Increasing the workload of your M&E team.
    d. a and b
    e. All of the above

Paper-Check Session

47. If the data in a particular area are of low quality, do we need a big sample or a little sample to measure the proportion of incomplete records?

    a. Big sample
    b. Little sample

48. In Link Plus, what is a blocking variable?

    a. A means of pre-selecting key variables to reduce the number of comparisons.
    b. A confounder.
    c. An obstacle to assessing data quality.
    d. A sampling strategy.

Under-Reporting Session

49. What are the pre-requisites for a capture–recapture study?

a. Ethical permission to interview TB patients.
b. Unique identifiers.
c. Three independent sources of TB notification data.

50. What are the potential benefits of measuring underreporting? (Check all that apply)

a. Can identify gaps to target available resources.
b. Can help improve estimates of TB incidence.
c. Can identify countries where TB data are so good that incidence can be measured directly from surveillance data.

Data Sharing Session

51. When sharing TB data, what ethical principles should be kept in mind?

a. Justice
b. Non-Malfeasance
c. Respect for Persons
d. Generalizability

52. How can deductive disclosure be prevented?

a. Remove all identifying information from a data set.
b. Keep linking tables under lock and key.
c. Use password protected data bases and computers.
d. All of the above

TB/HIV Session

53. What is benchmarking? (Choose one)

a. A means of identifying outliers through ecological comparison.
b. A method for triangulation of qualitative data.
c. A method of verifying the accuracy of data.

54. What are some potential pitfalls of de-duplifying TB data using Excel?

a. No syntax to re-run if needed.
b. No audit trail.
c. Duplicates are removed in order.
d. All of the above

All Sessions

55. I found this exercise to be: (choose one)

a. Easy
b. Difficult
c. Neither easy or difficult

Theme 1: Avoid Problems

Session 1: Elements of Data Quality: A Game

Objectives:

1. Reinforce the self-study and online components of the data quality (DQ) course.
2. Set a playful, collaborative, and dynamic tone for the course.

Background Preparation to be completed before the session:

1. Complete an online data quality course (1.5 hrs): https://training.measureevaluation.org/related-online-courses/data-quality

Time	Content	Methodology	Materials
40 min	A game to work out 16 data quality concepts involving two or more teams competing to develop these concepts and to group them by themes.	Slide presentation with instructions & basic background	• Big blank cards • Cards with data elements on them • Sticky tack or tape • Markers

Homework assignment: None

Session Notes

High quality data have a positive cascading affect on surveillance program's outcome measures (accuracy, completeness, timeliness)
Basis of M & E as means of improving program performance

Take Home Message

There is no single definitive list of Data Quality elements
There are lots of different ways to look at Data Quality, but the common features should be clear
We already know the basics of Data Quality if you completed the on-line pre-requisite course

Discussion points for debriefing:

1. Data Quality is a multi-dimensional concept. What aspects are most important to you? What aspects are most important for TB control?

Multiple choice questions for assessing comprehension:

1. What is meant by the term "accuracy"?
   a. The level of detail at which data is stored
   b. The lack of bias in the data
   c. The overall quality of the data
   d. The extent to which a value approaches its true value

2. What is meant by the term "data quality"?
   a. The lineage of the data
   b. The generalization present in the source data
   c. The inherent quality of the data as characterized by its accuracy, precision, bias, level of error, etc.
   d. The resolution of the data

TB/HIV Data Quality Game
(60 minutes required)

Introduction

The ostensible objective of the game is for the teams to thoughtfully arrange these pieces of data quality on the wall as a means of better understanding the meaning of the rather abstract 16 dimensions of data quality in practical terms for TB.

In addition, this game is intended to get participants up and moving and talking to each other about technical issues. It is intentionally left a bit vague how teams should function and the fact that there is not a 1:1 relationship between the 3 elements makes it additionally challenging. However this is fine for adult learners who are professional problem solvers. The fewer instructions you can give the better. Teams will no doubt begin to debate and disagree about how data quality definitions, data quality indicators, and data quality elements all fit together or cluster.

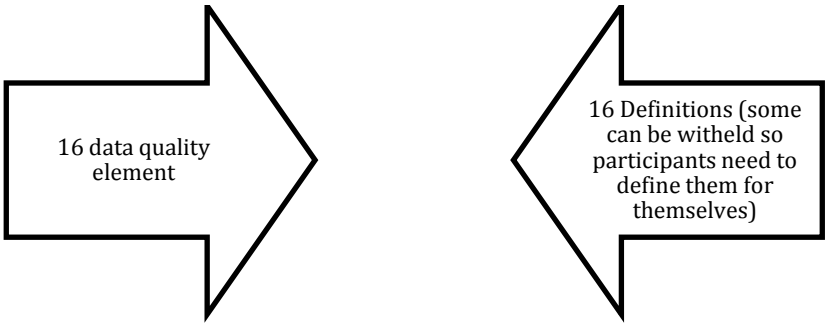
Finally, this exercise serves as a check on whether learners have in fact completed the on-line Data Quality pre-requisite and rewards those who have made that early investment in learning.

Materials:

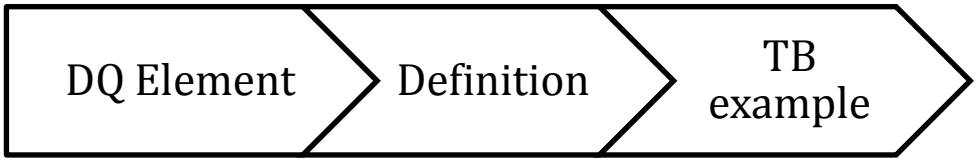
- Big blank cards
- Cards with data elements on them
- Sticky tack or tape
- Markers

This game has three stages: matching, clustering, discussion

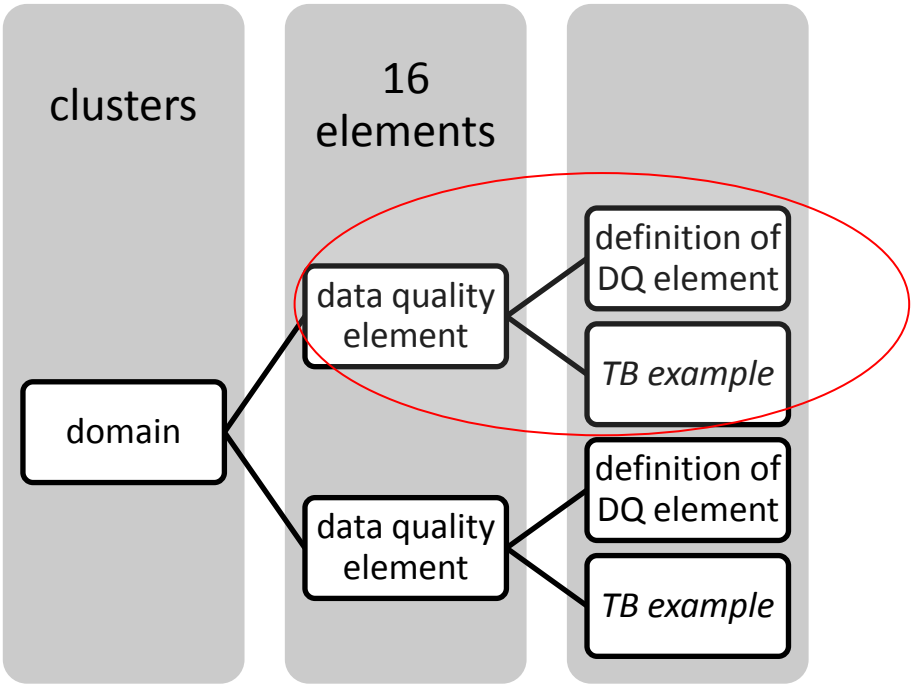
1. Form two or three teams.
2. Have each team move to one large blank wall.
3. Give each team (see Elements below):
    a. eight Elements cards with an element of data quality written on it
    b. eight blank Element cards
    c. eight long cards containing definitions (untitled)
    d. eight blank long cards.
    Try to ensure that there is not a1:1 relationship between elements and definitions, so that the team has to brainstorm definitions and elements.
4. **First, each team should try to match up their DQ definitions and their DQ Elements.** (15 min) Let them know that if they cannot find an appropriate term to match the definition they have, they should invent an appropriate DQ terms for the given definitions and write them on the blank title cards. If they have Element cards without definitions, they should draft definitions.



5. **Second**, each team should match a TB example to their DQ-Element+Definition pairings.(10min) to create trios of DQ element with accompanying definition and a TB example. The ability to link the DQ concept to specific TB examples may seem "advanced" at this point in the training. It may be OK to expose the learners to things they do not yet understand. This will get them thinking and talking and encourage those who do understand to share their knowledge with fellow learners. This sets the stage for more horizontal learning and sharing in the course. If you think this part of the game is too tricky for learners on the first day, you can shift this part of the game to the middle or end of the course. Because it involves moving around, it can be a good energizer.



6. **Third,** each team should group the 16 trios into clusters of concepts that seem to go together on the wall (10 min). This process will involve debate among the members about why things go together.



7. **Fourth** : Have the two teams switch sides of the room to examine the other team's wall (15 min).
    a. Discuss the matching. Do you agree that all the Elements have coherent definitions? Discuss diverging opinions.
    b. Look at the groupings by domain and consider the pros and cons of different ways of organizing these DQ Elements. Is there only one way to do it? Or are their multiple ways to organize the DQ elements?

8. **Wrap Up: Return to Plenary. Ask learners**

    a. What DQ elements were familiar? Which were new? Does everyone understand what is meant by each DQ element? Clarify any misunderstandings.
    b. Which domains were generated by each group to cluster the trios? What was the underlying logic? Were the groups the same or distinct?
    c. What was hard about this exercise?

d. What do you think its purpose was?

Cards to prepare:

The 16 Elements of Data Quality

1. Accuracy
2. Completeness
3. Validity
4. Consistency
5. Appropriateness
6. Coverage
7. Ease of use
8. Relevance
9. Understandability/ Interpretability
10. Confirmability
11. Value Added
12. Accessibility
13. Timeliness
14. Confidentiality
15. Security
16. Reputation
17. Objectivity
18. Reliability

Cards to prepare :

These can be laminated for easy re-use.

Data quality Element Cards

Accuracy

Completeness

Validity

Consistency

Appropriateness

Coverage

Ease of use

Relevance

Understandability/ Interpretability

Confirmability

Value Added

Accessibility

Timeliness

Confidentiality

Security

Reputation

Objectivity

Reliability

CARDS WITH DEFINITIONS OF ELEMENTS OF DATA QUALITY

Data quality Element Cards

Data Quality Definitions Cards

Accuracy	Data are considered correct: the data measure what they are intended to measure.
Completeness	Means that the information system from which the results are derived is appropriately inclusive. It represents the complete list of eligible persons or units, and not just a fraction of the list.

Validity	Requires minimizing interference, e.g., social desirability, interviewer bias, political pressures to show improvements, progress, and performance.
Consistency	Refers to data that are collected by adherence to protocols and SOPs that do not change according to who is using them and when or how often they are used
Appropriateness	Implies data without transcription errors or sampling errors.

Coverage	The geographic extent of the information system
Ease of use	The simplicity and user-friendliness of the data
Relevance	The salience and importance of the data for the program
Understandability/ Interpretability	The degree to which the data can be readily grasped and comprehended
Confirmability	The ability to verify or validate the data
Value Added	The contribution that the data makes to program performance
Accessibility	Refers to whether the data are obtainable or not

Timeliness	Refers to the strategic frequency of data collection, data reporting, and feedback loops that facilitate the governance of health programs
Confidentiality	Means that clients are assured that their personal data are not disclosed inappropriately, and that data in hard copy and electronic form are treated with appropriate levels of security (e.g.. kept in locked cabinets and in password protected files.)
Security	Implies frequent data protection, back up, and building in redundancies so nothing is lost.

Reputation	The public perception of the quality of the data and those who generate it
Objectivity	The absence of favoritism or bias
Reliability	Implies systematic data collection through the repeated use of a scientific instrument or a data collection procedures used under the same conditions

Cards to prepare of TB-specific examples of individual Data elements

Accuracy	The smear results in the laboratory register match the smear results in the TB treatment register.
Completeness	Less than 2% missing treatment outcome data

Validity	Over 90% of childhood TB cases diagnosed using a validated scoring system.
Consistency	Using the same case definitions in all the public and private facilities.
Appropriateness	TB among children disaggregated into 0-4 and 5-14 years of age.

Coverage

Percentage of districts (or basic management units BMU) reporting every quarter.

Ease of use

TB data and HIV data both use the same unique identifier

Relevance

All the donors ask for the same set of TB/HIV indicators

Understandability / Interpretability

Bacteriologically confirmed TB is widely understood to mean either smear positive or GeneXpert positive TB

Confirmability

The smear results in the laboratory register can be matched to the smear results in the TB treatment register using a unique identified.

Value Added

Knowing not only how many people have TB, but also how many were screened and then tested so program performance can be assessed.

Accessibility

All aggregated quarterly reports posted on the NTP website.

Timeliness	Annual reports are ready by the first quarter of the following year
Confidentiality	Electronic patient-based records that can be anonymized, encrypted, and emailed
Security	No names on the laboratory forms or samples. Each TB patient and presumptive TB client has a unique ID.

Reputation

Policy makers quote the data and researchers seek permission to use it

Objectivity

External data audits by independent experts

Reliability

Clear protocols about handling missing data re followed.

All areas of the country define loss to follow up in the same way.

Here is an example of 4 potential clusters, but there are many possibilities. The goal is not to find the "right" clusters, but to catalyze a discussion of when and where data quality issues become important and to whom are they important. There is no right answer.

Cluster 1: Technical Content of Data

- Accuracy
- Completeness
- Validity
- Consistency
- Appropriateness
- Coverage
- Ease of use

Cluster 2: The use of TB/HIV data

- Relevance
- Understandability/ Interpretability
-
- Confirmability

Value Added

Cluster 3: Exchange of TB/HIV data

- Accessibility
- Timeliness
- Confidentiality
- Security

Cluster 4: The image of TB/HIV data:

- Reputation
- Objectivity
- Believability/reliability

Another potential way of clustering them would be by junctures when data quality issues are discovered or averted:

- Data collection
- Data entry
- Data synthesis
- Data cleaning
- Data quality check
- Data analysis

Session 2: How to produce quality data prospectively

Objectives:

1. Enhance understanding of the best practices for quality data collection, including planning and the data collection life cycle.
2. Understand the importance of TB data dictionaries, SOPs for TB data collection, and choice of software in quality data collection.

Background Preparation to be completed before the session:

1. Introductory EPI-Info 7video (8:30 minutes) located at: http://www.youtube.com/watch?v=tnWiGyIgnV4
2. Chapter 15, "Data management," *Tuberculosis prevalence surveys: a handbook:* http://www.who.int/tb/advisory_bodies/impact_measurement_taskforce/resources_documents/thelimebook/en/

Time	Content	Methodology	Materials
40 min	Overview of data quality in practice –SOPS, data dictionaries	guided presentation	Powerpoint presentation on USB stick
20 min	• Basic background to quality data and Epi-Info 7demonstration on how to make data entry questionnaires and properly enter data •	Live demonstration	EPI-Info7 project

Homework assignment:

1. Read: Epi-Info User Guide on making forms (version 7) and views (version 3)

Take home message: Data management should be planned well in advance of data collection & should continue throughout study/survey to lead to high quality data for analysis

Overview

Quality data leads to confidence in

- Data source &
- Data quality assessments
- Data linkages

- Complete & reliable data sets i.e. integrity of the study
- Unbiased data

FLOW of TB Data

Data Management

Integrated system that allows for:

- Collection
- Cleaning
- Storing
- Monitoring
- Reviewing
- Reporting

Data Collection

Data collection should be carried out in compliance with regulatory standards
Aim to keep minimal errors & missing data whilst gathering maximum data for analysis
Adoption of best practices to meet objectives

Unique Identifiers

TB and HIV programs should consider the use of a unified unique identifier that combines patient and facility characteristics such as a 17 digit alphanumeric code

Facility code(3 digits)+ BMU code(3 digits)+date of birth(8 digits)+3 letters of name

The IMPORTANT steps

1. PLAN
2. Ensure existence of unique data & audit trail
3. Existence of
4. SOPs (standard operating procedures)
5. Data dictionary
6. Capacity for Data linking
7. Choice of software

PIN or UPIC or UNIQUE ID

The importance of the PIN: example

How link the laboratory file to the case record from?

The importance of the PIN: example

How to link the laboratory file to the case record from?

This is only possible when you use a PIN!

Some definitions

Data dictionary **(Metadata)** =Contains data about data
Describes nature of database/system catalogue holds info on

- Name
- Type of unique identifiers
- Format
- Coding
- Range of values
- Source
- Access authorization

Indicates which application programs use data so if a change in data structure is contemplated, a list of possibly affected programs can be anticipated
Example: data dictionary

Importance of data dictionary

Facilitates:

- Communication
- Collaboration
- Analysis

Some definitions

SOPs

- Gives full details of all procedures
- Standardizes process by providing step-by-step guide; anyone should be able to perform task in consistent manner
- Describes qualifications roles & responsibilities of all team members, ensuring standardization of performed tasks
- Allows for accountability

Purpose of SOP

- Serve as framework for organizational policy to provide direction and structure
- Written documentation of best practice
- Tells what, how, when, why, and who
- Provide foundation for:
- job descriptions
- employee training
- corrective action and discipline
- performance review

Elements of a SOP

- Rationale for SOP
- Detailed description of procedure – based on best practice/standards
- Monitoring actions
- Accountability
- Corrective Actions
- Date of last review or revision date

Example of SOP

Putting it all together ……..

Prevention of errors

Prevention of errors is better than correction afterwards!
Takes less time
Therefore, think about the design of your data-entry file as a vaccine against data diseases
Check dataset after pilot data-entry
first questionnaires or fake data
Feedback data-entry errors and agreements made to those who enter data
Change data-entry program where necessary

Discussion points for debriefing:

1. Whose responsibility is data quality?
2. Why is there so much emphasis on SOPs?
3. In your opinion, what is the most important safeguard to ensure data quality?
4. All stakeholders must be involved at all levels.

Multiple choice questions for assessing comprehension:

1. SOPs are:

    a. An international ethical and scientific quality standard for designing, conducting, recording, and reporting research.
    b. <u>A set of written instructions that document routine activity in a step by step guide, and hence standardize procedures.</u>
    c. A set of documents put together to satisfy regulators.
    d. A step by step guide of how to create questionnaires and care report forms for use in the field.

2. The main purpose of a data dictionary is to provide a source of reference in which the _____can look up content and any other relevant information.

a. analyst
b. user
c. designer
d. all of the above

Session 3: Introduction to Form Design and Quality-Assured Data Entry

Objectives:

1. To learn how to create well-designed questionnaires.
2. To appreciate the relative merits of open and closed questions.
3. To master the theory of double data entry.

Background Preparation to be completed before the session:

1. Epi-Info User Guide on making forms (version 7) and views (version 3)
2. Preview the Epi-Info video called "Legal Values" (12 minutes) in the videos folder on USB stick.

Time	Content	Methodology	Materials
30 min	Learn to use CREATE FORMS and ENTER DATA commands in Epi-Info	Powerpoint demo of Epi-Info 7 commands	Powerpoint slides
45 min	Skills practice in-class assignment	individual skills practice	DemonstrationMDB (Access file and PRJ- Epi- Info file)

EPI INFO 7

Introduction to the Form Designer

Opening the Form Designer

Form Designer Work Areas

The Form Designer has several "work areas":

The **Menu**

The **toolbar** contains buttons for creating projects, editing the form's check code, going to the data entry module, and undo/redo.
The **Project Explorer** is where you can add and remove forms from your project, add, edit, and remove pages from individual forms, and work with templates.
The **Canvas** is where fields are placed, moved, and edited.

Form Designer Areas

The **Menu**

The **Project Explorer**

The **Canvas**

The Menu and Toolbar

The Form Designer main menu provides an easy way to access your projects and gives you tools to edit your forms..

The Menu and Toolbar

The Form Designer main menu provides additional tools to help you manage your project and customize your Canvas.

The Toolbar Buttons

The toolbar contains buttons for directly entering a function without going through the main menu.

Creating Projects

New
Open (existing)
Undo
Redo
Check Code editing
Enter Data (module)

The Project Explorer

The Project Explorer is where you can add and remove forms from your project, add, edit, and remove pages from individual forms, and work with templates.
Usually, items in the Project Explorer have a right-click context menu. The Project Explorer also has a list of "open fields" that can be dragged directly on to the canvas.

The Canvas

The canvas is where fields are placed, moved, and edited. Fields can be dragged around the canvas by left-clicking to hold them and then moving the mouse.
The canvas has a right-click context menu that allows users to add new fields, set the tab order for the current page, and more.

Creating Projects

Module 2 – Form Designer

Form Design

Introduction

Uses Microsoft Access database or SQL server database format
Creates **projects** (analogy: filing cabinet)
Contains 1 or more **forms** (analogy: folders)
Each form may have 1 or more **data tables** (analogy: questionnaire)
Fields (variables) on forms designed to hold data

Field or Variable Types

Each field/variable has its own properties when selected
At least 20 types exist

A **Required** field is mandatory
A **Read Only** field does not allow the placement of the cursor in the field or data entry.
The **Range** property can be applied to Number or Date field types
Field Types and Creating Fields

Module 2 – Form Designer
Field Types

A variety of field types exist in Epi Info™ 7 to help customize the data entry experience.

Choosing the right field for the type of data being collected:

- Reduces data entry errors
- Ensures the data collected can be analyzed (meaningful results)
- Allows faster data entry
- Improves user satisfaction with the data entry process

Further manipulations

Try the following:
DELETE a field
Right-click on the **field**. The pop-up menu opens. Click Delete (NOTE: any data previously collected will be deleted & is not recoverable)
EDIT a field
Right-click on the **field**. The pop-up menu opens. Click CHANGE TO
Set Tab Order

Tab Order

Manually Change Tab Order

Manually Change Tab Order

Homework assignment:

1. Watch this short film: 3:31 Epi Info 7 Skip pattern using an "if, then" statement http://www.youtube.com/watch?v=Ww4dAplEnTI
2. Further Refine the Project you created in class and email the file(hint: *see video on packaging data for sharing)

Homework hand out:
Try this practice session

1. Add your name and mine into data table, and add fake data.
2. Edit form so that the cough label is the same font as the rest of the data, and move to page 2.
3. Familiarize yourself with the navigation menu.
4. Label Page 2 Symptoms, and add the following labels:

    a. Wheezing (Yes/No)
    b. HIV results (Yes/No)
    c. Sex (option button)
    d. PTB (Check box)
    e. ETB (Check box)
    f. Phone number (choosing an appropriate pattern)
5. Follow instructions from Epi-Info to try and create:
    a. Legal values
    b. Comment legal
    c. Codes

Discussion points for debriefing:

1. What is the difference between a variable and a label?
2. What are some of the advantages of using pre-defined fields or variables?
3. Why are some questions better if left open ended? Give an example.
4. Which elements of data quality might be enhanced by open-ended questions?
5. Which elements of data quality are likely to be enhanced by closed-ended questions?

Multiple choice questions for assessing comprehension:

1. At least 20 types of variables types exist and can have either of the following properties:
    a. Required field
    b. A read only field
    c. Range property
    d. <u>All of the above</u>
    e. None of the above

2. Projects in Epi-Info 7 must contain
    a. One or more forms
    b. One or more data tables
    c. Fields/variables
    d. None of the above
    e. <u>All of the above</u>

Session 4: Hands-On Skills Practice in Quality-Assured Data Entry

Objectives:

1. To practice data entry skills using the ENTER DATA command in Epi-Info 7.

Pre-requisite: Session 3**:** Introduction to Form Design and Quality-Assured Data Entry

Background Preparation to be completed before the session:

1. Watch this video on Epi-Info 7 file management: How-to (3 of 3): Moving Files http://www.youtube.com/watch?v=VxMVCdYUzVo

2. Watch the Epi-Info 7 video: Create a project template. http://www.youtube.com/watch?v=uvPpKW1LnmY

Time	Content	Methodology	Materials
15 min	Overview of ENTER DATA command in Epi-Info 7	Live demonstration	Demonstration MDB (Access file and PRJ- epi info file)
45 min	Skills practice (Class/homework) Enter data into the questionnaire	Individual hands-on practice	Demonstration MDB (Access file and PRJ- epi info file)

Homework assignment: See next page.

Session Notes

In Epidata, Epi Info, SPSS, Access, ….
Where possible, build in checks to prevent typing errors
In case of sex, only '1' and '2' or 'M' and 'F' can be entered (and '9' for missing)
Double data entry to reveal typing errors

Which software for data entry?

EPI-DATA or EPI-Info (older version) are best for double data entry at the present time.
See: www.tbrieder.org for detailed instructions on double data entry using EPI-DATA
As of July 2013, EPI-INFO 7 does not have a functional double data entry module

EPI-Info 3

EPI-Info 3.02 does have a robust data comparison feature and can be used for double data entry.

Excerpt from H.Rieder Epi Data course:

Data entry errors will occur, and worse, *to an unknown extent.*
The only way, and the only acceptable one, is to enter the data twice into two different files, and then to compare the two files for discordances.
Any discordance uncovered will then be corrected against the original paper record.

Rationale for double data entry (Rieder, 2013)

the probability of committing the same error in the same field twice when data entry is done independently by two persons is very small.
You MUST have Unique Identifiers for each record.

Example double data check in Excel

Proposed Data Life Cycle

Acquisition

Methods
New data collection (SOPs)
Converting/transforming legacy data
Sharing/exchanging data
Purchasing data
Security Requirements
MOUs
Data Sharing Agreements

Data Management Software

Depends on complexity of study
Depends on different types of observational units at different hierarchical levels
Simple data structures at one level can be handled by:
Statistical packages, e.g. SAS, SPSS, Status, Epi-info, CS-Pro, and R.

Data Management Software

- Spreadsheet packages, e.g. Excel.
- Relational database management systems, e.g. Access, dBase
- Geographic information systems are also available for storing spatial data.
- Statistical, spreadsheet and database management packages have overlapping facilities for data management, and all can now 'talk' to each other.

Data Management Software

MS Excel….

- Cannot handle longitudinal data properly

- Cannot responsibility handle anything more than simple edit checks require programming
- Limited ability to define data types
- Can't select subsets of data
- No audit trail

MS Access…

- Can easily handle longitudinal data
- Can do edit checks and validations
- Easy to import and merge data files
- Can easily define data types
- Can select subsets of data
- No audit trail
- Is being phased out by Microsoft

Epi-info 7 Handout

Follow these instructions in Epi-Info and try to create a data entry file:

If using version 3, open the program, click on Make View, highlight the File Tab and select NEW. This should lead to the popup window requesting the name for a new project. Name it Training by typing TRAINING in the file name box. Click open, and then type the name of the view (questionnaire) in the new pop up window. This is your actual view file. Name it <Train1>.

Click on the INSERT Tab and follow the instructions to add the following variables:

1. LABEL: Name of questionnaire
2. TEXT: Clinic name
3. Numerical: TB_ID
4. Date <dd/mm/yyyy>: Date of lab test
5. Yes/No: Wheezing
6. Check box: PTB
7. Text: First name
8. Text: Last name
9. Numerical: Cough (Insert legal values)
10. Numerical: Fever (Insert legal value)
11. Text: Night sweats (Limit the length of the text field)

If using version 7, open the program, then click on Create Forms, and select new project on the top left. This should lead to the popup window asking the name for a new project. Name it Training by typing TRAINING in the file name box. Then go to the bottom of the

popup window and give the form a name. This is your actual view file. Name it <Train1>. End by clicking OK.

Using the relevant command from the field command on the left window, follow the instructions to add the following variables:

1. LABEL: Name of questionnaire
2. TEXT: Clinic name
3. Numerical: TB_ID
4. Date <dd/mm/yyyy>: Date of lab test
5. Yes/No: Wheezing
6. Check box: PTB
7. Text: First name
8. Text: Last name
9. Numerical: Cough (Insert legal values)
10. Numerical: Fever (Insert legal value)
11. Text: Night sweats (Limit the length of the text field)

Discussion points for debriefing:

1. Do you understand how to CREATE forms, with the necessary variables and checks, skip patterns, and Tab orders?
2. Do you understand how to ENTER DATA the commands and how to switch back to CREATE Form format to update your questionnaires?

Data Entry Homework Instructions:
Enter the data below as it is (small & big cap for names) into your Epi-Info form. Ensure all variables are present and in the correct format.

Record No.		Clinic Name	TB ID	Date of Lab Test	First Name	Last Name	Cough	Fever	Night Sweats
1		A	1	09/09/2011	Thomas	JEFFERSON	0	1	Over 2 weeks ago
2		A	2	31/08/2012	George	WASHINGTON	1	1	Last 2 weeks
3		B	1	01/06/2012	Ben	FRANKLIN	0	1	Over 6 months ago
4		B	2	22/06/2012	Abraham	LINCOLN	1	1	Over 6 months ago
5		A	4	31/05/2013		SOCRATES	1	1	Last 2 weeks
6		B	3	23/02/2012	Napoleon	BONAPARTE	1	1	Over 6 months ago
7		A	3	24/02/2013	Marie	ANTOINETTE	1	0	Over 2 weeks ago
8		A	2	01/01/2013	George	WASHINGTON	1	0	Last 2 weeks
9		B	1	31/03/2012	Ben	FRANKLIN	1	1	Over 6 months ago

Session 5: Country Experiences in Developing an Electronic R&R System

Objectives:

1. To explore different ERR approaches.
2. To share lessons learned in ERR system development.

Background Preparation to be completed before the session:

1. *Electronic Recording and Reporting for Tuberculosis Care and Control* (WHO): whqlibdoc.who.int/publications/2012/9789241564465_eng.pdf
2. Fraser, H., D. Thomas, et al. (2012). *Adaptation of a web-based, open source electronic medical record system platform to support a large study of tuberculosis epidemiology*. BMC Medical Informatics and Decision Making 12(1): 125.

Time	Content	Methodology	Materials
8 min	TIBU movie on Kenya ERR system	Video stimulus	Video
60 min	Panel discussion with countries on their experiences with ERR planning/implementation (Indonesia, Kenya, Namibia/Vietnam?)	Prepared questions posed to panelists (30 minutes) in addition to questions directly from the audience (30 minutes)	Prepared questions

Homework assignment: none

Panel discussion - Experiences with ERR planning/implementation

Panel questions

1. Can you briefly describe what ERR system(s) you have in place in country and what the scope of the system is? (i.e. case-based? TB or MDR TB only? # sites/level of use? How long it's been in place?)
2. What was the process your NTP went through to select the system(s) currently being used?
3. How long did the process take from envisioning an ERR system to having it 'up and running'?
4. What would you do differently in the planning/roll-out/implementation process?

5. If you were advising a neighboring country on how to implement an ERR system effectively, what top three tips on implementing an ERR would you give them?

Additional questions for the group

1. What are your country's experiences with linking TB ERRs with other systems (i.e. HIV, HMIS)? What has worked well? What hasn't?
2. How has your country dealt with the issue of changing dx and tx approaches and TB case definitions? How well/not well can ERR systems respond to changing technology on the diagnostics side (i.e. Xpert)?
3. How have your ERRs improved or complicated data quality?

Discussion points for debriefing:

1. What was/has been the biggest challenge for your country in the establishment of an ERR system?
2. How long did the development process take, from system selection to maintenance?

Session 6: How to Develop Electronic Recording & Reporting Systems

Objectives:

1. Identify the advantages & pitfalls of ERRs.
2. Recognize questions to ask when developing an electronic system.
3. Discuss general steps to follow in the ERR system development process.

Background Preparation to be completed before the session:

1. *Electronic Recording and Reporting for Tuberculosis Care and Control* (WHO): whqlibdoc.who.int/publications/2012/9789241564465_eng.pdf
2. *Definitions and reporting framework for tuberculosis – 2013 revision*. WHO 2013. Available at: http://apps.who.int/iris/bitstream/10665/79199/1/9789241505345_eng.pdf
3. Blaya, J. A., S. S. Shin, et al. (2007). *A web-based laboratory information system to improve quality of care of tuberculosis patients in Peru: design, implementation and evaluation methodology*. BMC Med Inform DecisMak 7(1): 33. http://www.biomedcentral.com/1472-6947/7/33
   *See also: www.bio**med**central.com/1472-6947/12/125*

Time	Content	Methodology	Materials
5 min	Discuss the advantages and potential pitfalls of implementing an ERR system.	Presentation/ discussion	PowerPoint presentation
15 min	Introduce participants to key questions to ask when defining the scope of an ERR system.		
10 min	Highlight questions that should be asked when defining the detailed system requirements.		
30 min	Discuss the general steps for selecting and implementing a system.		

Homework assignment:

1. Draw a map of the data flow in your country's surveillance system and identify which (if any) points could or should be electronic.

Session Notes

Electronic Recording & Reporting

Data quality perks of ERRs

Data collection (range checks, legal values, automated unique IDs – no duplicates)
Data monitoring
Data transfer (de-identification, packing, e-mailing)
Data sorting and filing
Data entry (possibilities of double entry)
Data validation (automated queries)
Data cleaning

You want an electronic system?

Cyclical implementation process (MSH)

Envisioning your ERR system

Looking at the organization

Is there a functioning TB recording and reporting system in place?

An ERR cannot fix an already broken system!

Who needs to provide overall oversight and participate in decision-making related to the adoption, design and implementation of an electronic recording and reporting system for TB?

Define the scope

What do you want your system to look like?

Defining the scope

What are the primary objectives of building an electronic recording and reporting system for TB care and control?
Who are the users and beneficiaries of the system?

Which patients will the system cover?

All diagnosed TB patients;
Only a subset (i.e. MDR-TB patients);
Initially only a subset, but will expand later
or
Links to different systems so all patients are covered for national surveillance.

Defining the scope

What are the primary objectives of building an electronic recording and reporting system for TB care and control?
Who are the users and beneficiaries of the system?
Which patients will the system cover?

Which locations will the system cover?

Will the system be a stand-alone system or will it be integrated with other electronic systems?

Linking and 'talking' with other systems

ERRs rarely exist in isolation;
Other systems may already have element of TB R&R (i.e. vital registration, HIV, lab, pharmacy);
3 options:
Interoperatibility – ability to send/receive data b/w systems – possible if you plan for it, but very difficult to do after the fact!
Module/extension of existing system (i.e. Nat. patient record system)
No exchange/interaction w/ other system;

Defining the scope

What are the primary objectives of building an electronic recording and reporting system for TB care and control?
Who are the users and beneficiaries of the system?
Which patients will the system cover?
Which locations will the system cover?
Will the system be a stand-alone system or will it be integrated with other electronic systems?
What elements of paper-based recording and reporting should be maintained?

What data items need to be captured?

What data items need to be captured?

What is essential vs. useful vs. unnecessary?
Develop and update a data dictionary to clearly outline the data held in the system (variable names, definitions, codes, relationships b/w data items, etc.)

Defining the scope

What are the primary objectives of building an electronic recording and reporting system for TB care and control?
Who are the users and beneficiaries of the system?
Which patients will the system cover?
Which locations will the system cover?
Will the system be a stand-alone system or will it be integrated with other electronic systems?
What elements of paper-based recording and reporting should be maintained?
What data items need to be captured?

Is the basic unit of recording clinical data a patient, a case or a group of cases?

Is the basic unit of recording clinical data a patient, a case or a group of cases?

Individual data
Patient-based records
(ideal, using unique national personal identifier)
Case-based records
Aggregated totals only
Doesn't offer full benefits of an ERR system; not encouraged.

Defining the scope – lessons learned

ID which items are 'non-negotiable' and which are desirable (but optional);
Unrealistic to expect a completely electronic system – paper will still likely play an important role;
Think about interoperability BEFORE creating the system – nearly impossible to make systems talk after the fact;
Parallel runs of paper and electronic systems may be needed initially, but limit to test period to avoid chronic 'pilot project' – data may suffer;

Identifying detailed requirements

Capabilities

Who enters data, where and when will data be entered, and how do data flow within the system?
What data quality assurance processes are required?
How is feedback provided to system users?
What standard outputs, reports and other analyses are required?
What are the data entry screen or interface requirements?
How will data confidentiality and security be ensured?

Resources

What staffing is required?
What user support is needed?
What technical support is needed?
What level of service availability, response times and contingency planning is required?
What funding is required for both start-up and routine operations?
How long will electronic data be retained and will they be archived?

Infrastructure

- How is the electronic recording and reporting software made available to users?
- What devices will users need to use the system?
- What database software is required?
- Where will the servers be located?
- What communications networks are needed?
- What are the electrical power needs?

How to get there?

Cyclical implementation process (MSH)

Action Plan

Purpose: Define & organize the implementation process;
Critical since several stakeholders may be involved – keep everyone on the same page;
Builds on and incorporates requirements identified in needs assessment process.

Action Plan components

- System description
- Roles & responsibilities
- Human resources
- Funding
- Project plan & timeline
- Expected outcomes

Action plan – lessons learned

1. Clear strategy for transitioning from paper-based to ERR (concurrent use or transfer paper to ERR?)
2. Engage stakeholders early (ex. action planning workshop);
3. Engage users of other systems in country (HIV, drug, etc.) – leverage resources/learn from their experience;
4. Experienced project manager/team is priceless!

Action plan – lessons learned (2)

- Up-to-date SOPs for current R&R system;
- Plan to develop/disseminate training materials (for sustainability);
- Exit strategy and hand over plan if external agencies involved;
- Learn from others.

Cyclical implementation process (MSH)

Development/adaptation – lessons learned

- Pilot training courses, not just ERR system – hard to change training during roll-out;
- Freeze versions (training materials & system) during scale-up; fix bugs in systematic way for version control/efficiency;
- Staff challenges during roll-out…

Human resource troubleshooting during roll-out

- Reluctance to change way of working (hierarchies, roles, etc.);
- Redundancies → lower morale & poor performance;
- Reconcile expectations of users and managers;
- High staff turnover demands regular training;
- Misaligned job roles (under/over-qualified for new work).

Cyclical implementation process (MSH)

Maintenance

Maintenance – lessons learned

- Long-term maintenance/evolution plan (ground rules for ongoing implementation; funding);
- System to record/track requested & implemented changes – increase transparency & efficiency;
- Disaster recovery or business continuity plan (worst case scenario troubleshooting)
- Server crash
- NTP building burns down
- Funds cut

Monitoring of new system

Needed to show progress in roll-out, ID weaknesses & demonstrate impact of system.

Sample Indicators

SYSTEM PERFORMANCE

● % of ERR system users trained and active out of target number of users

● Cumulative annual or monthly downtime of central server (for web-based systems)

SYSTEM COVERAGE

● # of TB units where the system has been rolled out and is in use (and % out of all units in the country)

DETECTION, REGISTRATION AND TREATMENT

● # of TB cases diagnosed and registered in the system (and % of all cases)
● % of MDR-TB suspects screened and registered in the system out of all MDR-TB suspects screened and registered.

Sample Indicators

TB PROGRAM ACTIVITY
● # of TB or DR-TB patients registered in the system who started treatment during a given period.
● # and % of TB patients registered in the system during the previous calendar year for whom treatment outcomes have been recorded.
DRUG MANAGEMENT (IF APPLICABLE)
● Number of stock outs at sites using the drug management component of the system

Remember! Major upgrades will likely need a new implementation cycle

A few approaches

e-TB manager (https://www.etbmanager.org)
Open MRS (http://openmrs.org/demo/)
TIBU - Kenya's system (http://www.youtube.com/watch?v=zEXjm51o_64)
- Open-source, web-based tool for managing NTP information
- Integrates data across all aspects of TB control, including information on suspects, patients, medicines, laboratory testing, diagnosis, treatment, and outcome.

Open MRS

Developed by Regenstrief Institute & Partners in Health
Software platform & reference application that enables design of a customized medical records system with no programming knowledge (although medical and systems analysis knowledge is required).
Open-source
For more information: http://openmrs.org/

Discussion points for debriefing:

1. What are some of the big picture questions you want to ask when you are first defining the scope of your ERR?
2. What are the four general phases of implementation of an ERR?

Multiple choice test questions to assess comprehension:

1. Which of the following is NOT an advantage to a well-functioning electronic recording and reporting system?
   a. Data quality

b. Timeliness
c. Managing complex data
d. Upfront costs

ANSWER: D. ERR systems often require considerable upfront costs. There are also costs necessary to maintain the system, but these aren't always significantly more expensive than paper-based system costs.

2. True/False: You need to have a strong paper-based system in place before implementing an ERR system effectively.

ANSWER: True. ERR systems cannot fix a broken paper-based system. A strong paper-based system is needed to build an electronic system.

Session 7: How to Ensure Quality of Qualitative Data

Objectives:

1. To distinguish between output-oriented and process-oriented quality assurance (QA)
2. To describe key QA concepts: transparency and reflexivity.
3. To practice coding narrative responses to open-ended questions.

Back ground Reading:

1. *Qualitative research review guidelines – RATS*: http://www.biomedcentral.com/info/ifora/rats?layout=printer

Time	Content	Methodology	Materials
10 min	Overview of TB qualitative research.	Guided discussion/interactive	Power point
30 min	Overview of output-oriented quality assurance methods.		
20 min	Skills practice in critique of quality of qualitative data.	Group exercise	Interview transcript
15 min	Overview of process-oriented quality assurance methods.	Guided discussion/interactive	Power point
15 min	Skills practice	Inductive coding exercise	Transcript paragraph

Homework: None

Session Notes:

Why do Qualitative Research for TB?
•Behavior
•Social & political organization of people, groups, & organizations
TB deals with
•TB prevention
•Access to services
•Diagnosis
•Adherence
Qualitative research is essential to improving
•Processes and practices
•How/why people relate to programs & interventions
•The meanings it generates
•The effects this has on treatment

Need to understand
Qualitative Research
.Must be suitable for the research
question, e.g. not good for hypothesis testing
.Not ad hoc – follows a methodology
.Requires technical expertise in qualitative methods
Assessing Quality in Qualitative Research
* Reynolds et al. Quality Assurance of Qualitative Research: A Review of the Discourse. Health Research Policy and System 2011,9:43

- Output-oriented
- Process-oriented

Output-oriented Quality Assurance

- .Post-hoc – external process
- .Objectivity
- .Establish distance from the data

.Quality is conceptualized in relation to theoretical constructs like validity, rigor, credibility, relevance, etc. (evidence-based model)
.Demonstrate quality in research outputs
.Use of methods/techniques deemed to be indicators of quality, e.g. checklists (RATS)
C:\Program Files (x86)\Microsoft Office\MEDIA\CAGCAT10\j0292020.wmf
Output-oriented QA: Design Phase
.Study methods
.Research aims/objectives were clear
.Appropriate method was selected based
on research question

- .Observations
- .Interviews
- .Case studies

.Determined etic codes if appropriate
Output-oriented QA: Data Collection
.Used an explicit analytic framework,
e.g. .Demonstrated how the analytic procedure was consistently applied

Output-oriented QA: Analysis

.Transcription and translations are validated
.Multiple independent coders
.Ensure strong intercoder agreement
.Compare sets of codes assigned to specific text passages by each coder
.Compare how each code was used by the coders
.Measure agreement by using the Kappa statistic

.Establishes a replicable analysis

Inductive Coding Exercise*

Q: What do you think causes tuberculosis?
A1: TB sometimes caused by working too much and too hard (overworking). For example, some Vietnamese that work beyond their strength like farmers, factory workers, are working all day (some work 12, 14 hrs/day, and 7 days/wk) and when home and very little and also no nutritional food come. When you overwork and don't have enough calories in your body, it is very easy to get all kinds of germs to enter your body because your body's immune system is not strong enough. Also, they may acquire the TB germ from someone that has TB. Some cause by smoking too much, or drinking. The amount of nicotine and alcohol that you take into your body will be very harmful for lungs (if you have been smoking and drink a lot).
A2: I was told that everyone does have the Koch virus in his/her body, and if one is overworked without proper nutrition can get Tuberculosis.

* Carey et al. Intercoder Agreement in Analysis of Responses to Open-ended Interview Questions:
Examples from Tuberculosis Research. Cultural Anthropology Methods 8(3):1-5.

Output-oriented QA: Analysis

- .Ensure that all cases were included and reported, not just those that support conclusions
- .Deviant/negative case analysis
- .Search for and discuss elements in the data that seemingly contradict emerging patterns
- .Refine the analysis until it explains all or most of the cases (theory building)

Output-oriented QA: Analysis

.Triangulation: cross-check information and conclusions via use of multiple procedures or sources
.Common types of triangulation

- .Method – compare data that comes from different methods, e.g. interviews and observations
- .Data – compare data from different sources,
- e.g. interviews with different interest groups
- .Agreement may confirm the interpretation
- .Assumes that any weaknesses in one method will be compensated by strengths in another

Output-oriented QA: Conclusions

- .Member/participant validation of findings
- .Compare researchers' account to participants to determine level of agreement
- .Use this method as part of error reduction rather than a indicator of credibility
- .Researchers and participants have different roles/ perspectives, i.e. researchers' account is designed for a wide audience
- .Peer review of findings

Process-oriented Quality Assurance

- .Internal, on-going
- .Consider quality throughout the research process
- .Fundamental, internal set of values/principles indicative of the qualitative approach
- .Principles must be understood and upheld by the M&E team

Principles of Process-oriented QA

- Reflexivity of the researcher's position, assumptions and practice
- Transparency of decisions made and assumptions held
- Comprehensiveness of approach to the research question
- Responsibility towards decision-making acknowledged by the researcher
- Upholding good ethical practice throughout the research
- Systematic approach to designing, conducting and analyzing a study

1
2
3
4
5
6

Reflexivity =.Sensitivity to the ways in which the researcher and the research process have affected the data, including

- .Role of prior assumptions and experience
- .Personal and intellectual biases
- .Effects of personal characteristics, e.g. age, sex, social class, professional status, etc.
- .Researchers must document their beliefs, attitudes, values and reactions to the object of the study
- .Active, iterative process

Reflexivity
Personal

Reflect on how values, experiences, interests, beliefs, social identities, etc. affect the research
Epistemological
Reflect on how the research question, design, and analysis define or limit the results and conclusions
Could the research question have been investigated differently and would this have resulted in a different conclusion?

Facilitators of Reflexivity

.Use field diaries to explore and capture assumptions and biases
.Hold on-going dynamic discussions of quality issues among the research team
.Ensure researchers' comprehension of and engagement with their role in assuring quality

Transparency

Document all decisions and interpretations made at each stage of the research (audit trail)
Design
•Sampling techniques: rationale and theory behind them

Data Collection

•Description of context/setting
•How and why the techniques/focus were changed in response to data
Analysis
•**Internal validity:** sufficient information about pathway from data to conclusions
Conclusion

- .Qualitative research requires theoretical expertise
- .Several ways to assess the quality of qualitative research
- .Basic strategy: systematic, self-conscious research design, data collection, analysis/interpretation, and communication

Inductive coding exercise-
1) independently pick the key themes in the following response (5 min)
2) share with the group your "codes" –i.e. concept + definitions (10 min)
3) note the subtle and not so subtle differences in definition, scoping of the codes to emphasize the importance of code books for quality analysis

Inductive Coding Exercise*
Q: What do you think causes tuberculosis?
A1: TB sometimes caused by working too much and too hard (overworking). For example, some Vietnamese that work beyond their strength like farmers, factory workers, are working all day (some work 12, 14 hrs/day, 7 days/wk) and when come home and very little and also no nutritional food. When you overwork and don't have

enough calories in your body, it is very easy to get all kinds of germs to enter your body because your body's immune system is not strong enough. Also, they may acquire the TB germ from someone that has TB. Some cause by smoking too much, or drinking. The amount of nicotine and alcohol that you take into your body will be very harmful for lungs (if you have been smoking and drink a lot).
A2: I was told that everyone does have the Koch virus in his/her body, and if one is overworked without proper nutrition can get
tuberculosis.

Homework assignment: Review the interview video on the USB stick and identify data quality issues from the data collection phase.

Discussion points for debriefing:

1. Explain reflexivity in qualitative research.
2. Compare and contrast output-oriented and process-oriented quality assurance.

Multiple choice test questions to assess comprehension:

1. Select the output-oriented methods used to assess quality in qualitative research:
   a. Triangulation
   b. Comprehensiveness
   c. Deviant/negative case analysis
   d. All of the above
   e. a and c only

2. The fundamental principles of qualitative research include:
   a. Reflexivity, transparency, comprehensiveness, responsibility, ethical practice, and systematic approach.
   b. Validity, rigor, confirmability, credibility, trustworthiness.
   c. Triangulation, respondent validation, reflexivity, attention to negative cases, transparency, relevance.
   d. Rigor, objectivity, representativeness, comprehensiveness, context sensitivity.
   e. None of the above.

INTERVIEW TRANSCRIPT Handout
KEY INFORMANT INTERVIEW WITH A TB PATIENT

Interviewer: Do you consent to have this conversation recorded?
Respondent; Yeah, It's ok.

QN.1 Interviewer: Think about your community, what are the main health problems in this community?
Respondent: I am just a visitor, I was brought here when I became sick, was living in QQQQQ.

Interviewer: Probe: That means you don't actually know what the health problems are in this community??!??
Respondent: No.

Interviewer: OK, well let's continue..

Qn.2:
Interviewer: In your opinion, what is TB?
I was told that TB is airborne.

Interviewer (Probing): Ok, yes, that may be how one gets it. But what do *you* think TB *is?*

Respondent: I am not sure what you ask… For my case, I had boils and cough. I was taken to ZZZZ and got treatment.

Interviewer (Probing): yes that is where I am heading. I am wondering what you feel TB is about. What it looks and feels like.

Respondent: At times it feels tiring, like wasting away …but then it can feel like just a nagging cold sometimes. It depends..

Probing: Depends on what (patient's name)?
Respondent: Who you are… what else is..there—happening—going on in your life.

Qn.3: What are the different TB names in this community?
Respondent: Am not very sure… --as I told you before

Probing: just try!
Respondent.. Besides TB…, I don't know…. Perhaps tuberculosis… some people may call it "consumption"… or "the white plague" but those are old names…I don't know if they are used here.

Qn. 4
Interviewer: How does one get TB?
Respondent: It's airborne.

Probe: Is that all you know?
Respondent: Yes

QN.5
Interviewer: How did you know that you had TB?
Respondent: I went to ZZZZ and was checked with an x-ray'

QN6.
Interviewer: In your opinion, which part of the body is mainly affected by TB?
Respondent: It's the chest.

Probe: Anything else?
Respondent: NO.

QN.7
Interviewer: What diseases are associated with TB?
Respondent: I also had ulcers, even fever.

Interviewer: Is there any other disease which you think is associated with TB?
Respondent: Even stomach ache.

QN.8
Interviewer: Who in your opinion is commonly affected by TB?
Respondent: It affects both men and women.

QN.9
Interviewer: Why do you think it affects both men and women?
Respondent: Because when your chance has come, you will get it.

QN.10

Interviewer: What of young people?
Respondent: It affects all, even the newborns.

Interviewer: Why do you think it affects both young people and newborns?
Respondent: Sometimes, they say if the family once had TB, then the children will also develop it.

QN.11
Interviewer: What made you decide seek health care?
Respondent: I wasn't eating, I was vomiting, I had lost weight and had cough. People were saying its HIV/AIDS.

QN.12
Interviewer: Where did you seek health care?
Respondent: I went to Buluba, they took my blood sample, checked it and found it wasn't HIV/AIDS. They also performed a sputum test and x-ray and discovered it was TB. I used to overwork until late in the night, around 2.00 am. I think that's how I got it from the coldness.

Interviewer: What are the reasons for your choice?
Respondent: From VVVV, I was referred to MMMM and then from MMMM, I was told that there is also treatment in BBBB H/C and I was referred to there.

Interviewer: Now, there are people who decide to go to traditional healers or health facilities. What made you decide to go to a health facility?
Respondent: Because this kind of illness isn't treated by traditional medicine.

Interviewer: How long is the treatment period?
Respondent: I was told 8 months.

QN. 13
Interviewer: What problems do you encounter in following up treatment?
Respondent: Transport. <PPPPP> is far, yet I can't walk. I spent two months admitted in the hospital, couldn't turn on my own, they were only helping me.

QN.14
Interviewer: In your opinion, do you think TB cures?
Respondent: I was told it cures.
Probe: But in your opinion, do you think it cures?
Respondent: Yes, because now, there is a very big improvement from the situation I was in.
Probe: How has your situation improved, exactly?
Respondent: I can eat now and my back is aching less. When I move I do not feel that I am carrying a heavy weight.
QN.16
Interviewer: Please, describe the attitude of health workers towards TB patients? When you are a TB patient, how do you they treat you?
Respondent: They treat you as if you were like any other patient.
Probe: Yes, what does that mean exactly?
Respondent: They make you wait for them in the early morning and when they see you, they are rushing through to get to the next one. Some have the friendly ways, but some are cold like stones.

QN.17

Interviewer: What of the community members, how do you they treat you?
Respondent: They are good, they don't treat me badly.

QN.18

Interviewer: What is your opinion about TB health services?
Respondent: They give you tablets or injections.
Probe: I mean, how do you find it? Do you think it's good?
Respondent: The treatment is good because they don't ask for money.

QN.19

Interviewer: What are the different ways used to prevent diseases among children in this community?
Respondent: They are immunized against measles, polio, TB

QN.20

Interviewer: What is your opinion about immunization?
Respondent: When a child is immunized, he/she isn't prone to diseases like one who was not immunized. So, it's good.

QN.22
Interviewer: Why do we immunize against TB?
Respondent: To ensure they don't spread the infection to other people.

QN.23
Interviewer: What is your opinion about TB immunization?
Respondent: It avoids spreading diseases.

Probe: Anything else? Is there any fear you have towards TB immunization?
Respondent: No.

QN.24
Interviewer: How does it protect us?
Respondent: Don't know. But when one is immunized, he or she doesn't contract diseases. Blood is heavy; when the vaccines power is in your body, the disease will meet the drug. I think that when a child is still young; he is immunized so that when there is a disease outbreak, it will find that he is already immunized.

Probe: Any other ideas?
Respondent: germs of that disease are taken and used as "Askaris" that weaken that disease in case it infects. One who is immunized is different from one who is not. When there is measles outbreak, a child who was not immunized against it will be seriously infected, but one who was immunized has a mild attack.

I think they should be giving us support, us people in the rural area, we are helpless, they should give us things like ground nuts, mosquito nets, seeds for planting, etc.

QN.25
Interviewer: Does it protect us 100%?
Respondent: I don't know.

Session 8: Data Management of Large Electronic Datasets

Objectives:

1. To understand the basis of continuous data management throughout the lifecycle of project and the involvement of every stake holder at every step.

Background Preparation to be completed before the session:

1. Understanding and Using Tuberculosis Data. WHO 2014 Chapter 2 Analysis of case-based TB notification data
2. Watch this video on confidential data sharing protocols: Epi Info 7 Packaging data for sharing
   http://www.youtube.com/watch?v=1rKLAPRp71g
3. How to use Epi-Info Compare Utility v. 3.4.3 (**NOTE for OLD VERSION OF EPI_INFO) *new Epi-info 7 has no way to check double entered data yet.*
   http://www.youtube.com/watch?v=24gBswHdZ8U

Time	Content	Methodology	Materials
30 min	Re-iterate good data management best practices, especially in light of BIG DATA (electronic).	Guided discussion demonstration	Powerpoint slides
30 min	Show audit trails, security features, how to do a query and logic checks.	demonstration	Big TB surveillance database in EPI-Info7

Homework:

Using the sample TB patient database on the USB stick, identify all the security features.

Topics to be covered

Change theory
Data quality activities within the data management system
Data quality responsibilities by level
Data quality assessment tools

Connecting Data & Health Outcomes

Who/What Needs to Change?

Overview of Change Theory & Basic Principles

Despite people's conviction about a course of action, they often need prompts and triggers (cues to action) which move them forward
Behaviors may resist change if they have significant social costs or reinforcements in institutional cultures

Factors that Influence Behavior Change

- Expected Outcomes
- Intention
- Self-Image
- Skills
- Self-Efficacy
- Emotions
- Perceived social norms

Effective Change Process
Principles of Effective Change
Data Management System Components
M&E Structures, Functions, and Capabilities
Indicator Definitions & Reporting Guidelines
Data Collection & Reporting Forms/Tools
Data Management Processes
Training
Evidence-based Decision Making
Data Quality Responsibilities
Each level of the health system has its own responsibilities for maintaining high quality data
DQ Responsibilities – Health Facility
DQ Responsibilities – Intermediate Level
DQ Responsibilities – NTP/M&E Unit
DQ Assessment Tools

Discussion points for debriefing:

1. Whose responsibility is data management?
2. Describe steps in data management, starting with the earliest.

Multiple choice questions for assessing comprehension:

2. Accidental or malicious loss of data can be due to:

    a. Hardware faults or failure
    b. Software or media faults
    c. Virus infection or malicious hacking
    d. Power failure
    e. Human errors by changing or deleting files
    f. <u>All of the above</u>

3. A plausible order for Data Life is:
    a. Plan, Process, Acquisition, Analyze, Preserve, Publish/Share
    b. Plan, Acquisition, Analyze, Process, Preserve, Publish/Share

c. Plan, Acquisition, Process, Analyze, Preserve, Publish/Share
d. Plan, Publish/Share, Preserve, Analyze, Process, Acquisition
e. None of the above

Session 9: Quality-Assured Joining of TB Data Sets with a Unique ID

Objectives:

1. To learn the preparatory steps required for linking.
2. Introduce learners to performing de-duplification procedures on datasets
3. To demonstrate the use of the **join** and **concatenate** functions in Epi-Info for joining two data sets**.**
4. To reiterate the importance of a unique ID.

Background reading:

1. 7:37 min **Methods and hazards of deduplication in Excel** http://www.youtube.com/watch?v=6HNX_tk2VxU
2. **How To Merge / Join Data From Tables In Excel Using *vLookup***
   http://www.youtube.com/watch?v=3tk_Mif7040

Time	Content	Methodology	Materials
30 min	Principles of merging	Presentation	Powerpoint
20 min	Importance of preparation of datasets and deduplication	Presentation	• Powerpoint • Deduplication video
20 min	Merging datasets with a unique ID using the **join** and **concatenate** functions in EPI-Info 3	Live demonstration	2 datasets for joining

Homework: None
Session Notes:

Record Linkage with Unique IDs

Questions for adult learners

1. How do you usually link your data?
2. What do you do with the missing identifiers?

Why link?
Record linkage carried out to
Accurately identify 2+ records from same entity (person, hospital, community, geographical area etc)
Consolidation of different databases into 1 central database (looks for duplicates-skews data)
Introduction
Record-linkage highly sensitive to quality of data

The potential for linkage varies greatly between countries according to how info collected & identified

Importance of record linkage

Creates data required for examining health of the public & health care system itself.
Improve data holdings, data collection, quality assessment, and the dissemination of information.
Data sources can be examined to eliminate duplicate records, identify underreporting & missing cases

Importance of record linkage tool

Creates data required for:

- Create person-oriented health statistics,
- Generate disease registries & health surveillance systems

Create TB indicators e.g.
% of HIV + clients diagnosed with TB
% of persons screened for TB, who are diagnosed with TB

Introduction

In 1946, H. L. Dunn of the US National Bureau of Statistics introduced the term in this way:
"Each person in the world creates a Book of Life. This Book starts with birth and ends with death. Record linkage is the name of the process of assembling the pages of this Book into a volume" *(Dunn, 1946)*

Introduction

Computerized record linkage was first undertaken by the Canadian geneticist Howard Newcombe and his associates in 1959

Record Linkage Process

5 main phases in linkage

1. Pre-processing
2. Cleaning data
3. Decision-making
4. Selecting matching variables
5. Grouping Blocking or indexing

6. Searching/scoring

Reviewing results manually

Data quality assessment pre-linkage

Essential 1st step

Especially KEY IDENTIFIER FIELDS

Unique Identifiers

Programs should consider the use of a unified unique identifier that combines patient and facility characteristics such as a 17 digit alphanumeric code that makes duplication unlikely and deductive disclosure very difficult.

Facility code(3 digits)+ BMU code(3 digits)+date of birth(8 digits)+years of education (2 digits)

A randomly generated Unique identifier is preferable in cases where stigma is a major issue and people are well known to each other. However, an advantage of a non-computer generated ID is that patients who access multiple health facilities will be assigned the same code if the Unique Identifier is based upon immutable patient characteristics.

Format Key identifiers (unique) to minimize complexity

Data cleaning prior to linking

Pre-requisites

- Removal of commas & punctuation marks, unnecessary blanks, accent marks, invalid values
- Upper/lower case variation
- String vs. numeric variables
- Missing values

Standardization

Many software programs require you to sort your datasets before merging

Standardization

Review Process

Intuition & intrinsic knowledge of data needed

Access to additional variables not used in search necessary

Do not delete duplicate record, should be marked and kept on file for further re-assessment

Whole process of linkage exercise should be accounted by full documentation

Good software for linking with Unique IDs

- SPSS
- STATA
- MS Access

Reminders for good linking
Data must be unique, but also *ordered (sort both file by PIN)*

DOUBLE COUNTING

What are the 3 types of double counting?

3 types of double counting
***Type I*:** *Within* Partner Double Counting of *Individuals*
Type II: *Between Partner* Double Counting of *Individuals*
Type III: Double Counting of *Sites*

Check for Duplicate Records in Epi-Info 7

Using Frequency, check for duplicate records.
How to:
READ C:\
FREQ TBpatientnumber
Delete Duplicate Records

How To:
READ {C:\
FREQ TBpatientnumber
SELECT TBpatientnumber "118080"
LIST * GRIDTABLE
DELETE UNIQUEKEY = 1 PERMANENT

Can I de-duplify in Excel?
What are the pros and cons of de-duplifying in Excel?
http://www.youtube.com/watch?v=6HNX_tk2VxU

It is not recommended.

Demonstrate the use of the **join** and **concatenate** functions in Epi-Info for joining two data sets**. Please note that these functions do not work properly yet in the current versions of Epi 7 and so we strongly suggest that the participants practice doing this in EPI-Info version 3.**

Handout Exercise 1:

Step 1: Open the TB screening register data and the TB patient register data.

Step 2: Clean and prepare the data for joining in the following ways:

Clean/prepare ScreeningRegister.xlsx

1. Open worksheet in Excel

2. Sort worksheet by Date of TB Screening (oldest to newest)
    a. Delete first record due to invalid and missing data

3. Sort worksheet by Region (largest to smallest)
    a. Re-enter Region in rows two through four

4. Sort worksheet by ScreeningNumber (largest to smallest)
    a. Delete row with ScreeningNumber equal to all 9s, i.e., 9999999999

5. Save cleaned file, as ScreeningRegister-clean.txt (n=17,668)

Step 3: Merge the presumptive TB and TB patient data registers using Excel vlook up or EPI-Info 3, matching by UNIQUE ID.

Discussion points for debriefing:

1. What are the main phases in the data linkage process?
2. How might you handle missing unique identifiers?

Theme 2: Find Problems

Session 10: Theory, Structure, and Process of Data Quality Assessment

Objectives:

1. Describe how change theory can be applied to data quality improvement.
2. List activities within each component of a data management system that help ensure data quality.
3. Describe the main roles and responsibilities for data quality at each level of the health system.
4. Compare and contrast different tools for assessing data quality.

Time	Content	Methodology	Materials
15 min	Overview of change theory	Presentation	Powerpoint
10 min	Data quality activities within the data management system		
15 min	Data quality responsibilities by level		
15 min	Data quality assessment tools		

Homework assignment:

1. Map the levels of responsibility for Data quality in your own TB program or organization.
2. Where and when does data quality get assessed?

Discussion points for debriefing:

1. Choose two factors that influence behavior change at the individual level and discuss how each could be addressed to improve data quality.
2. Describe two key differences between data quality audits and routine data quality assessments.

Multiple choice questions for assessing comprehension:

2. Which component of a data management system would include providing regular feedback to lower levels as a way to improve data quality?
   a. Data management processes
   b. Evidence-based decision making
   c. Data collection and reporting forms/tools
   d. M&E structures, functions, and capabilities

3. Which is not a factor that influences behavior change at the individual level?
    a. Social status
    b. Self-image
    c. Skills
    d. Emotions
    e. Perceived social norms

Session11: Sampling and Logic Checks of Paper Surveillance Data

Objectives:

1. To learn basic principles of sampling.
2. To practice using a random numbers table.
3. To practice deriving a systematic random sample of TB cases from a paper-based TB register.

Time	Content	Methodology	Materials
10 min	Overview of current practices for assessing the quality of paper register data.	Brainstorm	Flip chart
30 min	Overview of sampling of paper-based surveillance data.	dialogue	Power point
30 min	Systematic random sampling exercise for paper data.	Exercise- hands on practice	Paper TB registers

Homework assignment: Consider the following

1. If we scan a TB register and find that over 50% of the TB patients under five are smear positive, what types of data quality issues might be occurring? Describe two approaches you might use to trying to unravel this mystery.
2. If a culture laboratory has a cross-contamination problem in their Bactec MGIT machine, how could we discover it in the laboratory register?
3. What data quality clues might we expect to find in this situation?

Session Notes

Objectives

Share knowledge on our existing practices

Explore the issue of sampling from paper lists

What are your current practices?

What do you do to check data quality on supervision visits?

Short exercise in groups of 3 – 4 questions

What is sampling?

A process by which we study a small part of a population (sample) to make judgments about the entire population

Sampling involves selecting a number of units from a defined population

Definitions

Study <u>population</u>

All the sampling units or individuals which could possibly be included in the sample
Sampling frame
A list of all the available sampling units in the study population
Sampling unit
The item which is sampled
Sampling interval
The proportion of a study population sampled

Representative Sample

A representative sample has all the important characteristics of the study population from which it is drawn
Example:

Population	Sample
50% Male	50% Male
50% Female	50% Female

Random Sampling Strategies
Simple random sampling
Systematic sampling
Stratified sampling
Cluster sampling
Multistage Sampling

1. Simple random sampling

Used in situations where the number of sampling units is relatively small
Determine units available for sampling: i.e. Study population of 100 individuals
Decide on sample size: i.e. Sample 10 individuals
Lottery method (random number table, computer program
Advantages and Disadvantages
Advantages

- Easy to understand
- Easy to analyze

Disadvantages

- Requires a list of the population
- Cost may be prohibitive

May miss or undersample key subsets

2. Systematic sampling

Individuals are chosen at regular intervals.
Determine units available for sampling: i.e. Study population of 100 individuals
Decide on sample size: i.e. Sample 25 individuals
Calculate sampling interval 100/25=4
Start at random student between 1 and 4, i.e. 2.

Advantages and Disadvantages

Advantages

Less time consuming than simple random sampling
Easy to perform
Disadvantages
Risk of bias (i.e. sampling days of the week with sampling interval of 7 – will always select Tuesday which may be a market day).

Discussion

If the data are low quality, do we need a big sample or a little sample to measure the % of incomplete records?

Why?

Sampling practice with paper TB register

Draw 5 10% samples
Draw 5 20% sample
Compare
Thank you
Acknowledgements: Eveline Klinkenberg

Discussion points for debriefing:

1. If the data are very low quality, what kind of sample size would be needed and why?
2. What are some of the potential risks with systematic random sampling? When would it be a mistake to use it?

Multiple choice questions to assess comprehension:

1. If the TB data are very low quality, what size sample is needed to measure the proportion of incomplete data?
   a. a very big sample
   b. a very small sample
   c. it cannot be measured by sampling

2. If all the data are in a chronological list on paper, what sampling methods might be easy to use?
   a. Sampling proportional to size
   b. Lot quality assurance sampling
   c. Systematic random sampling

Systematic Random Sampling(SRS) Exercise
Sampling with Paper-Based TB Data Team Handout

This activity is based on an adaptation by Joan Garfield and Dani Ben-Zvi of an exercise from Rossman and Chance (2000), *Workshop Statistics: Discovery with Data*, second edition.

Materials Needed:

1. Sets of paper TB registers, each containing about 100 TB patients.
2. Handout (see below).
3. Random numbers table.

Systematic Random Sampling Exercise
Sampling with paper-based TB data Team Handout

Individual TB patients have two types of entries: complete and incomplete

Without counting, but just scanning quickly, guess the proportion of complete entries in Ellenville:

Complete _____%
Incomplete _____%

1. If each M&E Officer takes a sample of six TB patients from 60 TB patients(10%), would you expect every person to have the number of complete entries in their sample? Explain.

2. Randomly select five systematic random samples of six TB patients (10%). Write down the number of each color for these 5samples:

Five 10% Samples of the Ellenville TB Register data

Sample Number	Number complete	Percent complete	Number incomplete	Percent incomplete
A.				
B.				
C.				
D.				
E.				

These numbers represent the variability you would expect to see.

3. Now draw five systematic random 20% samples of 12 patients, and put your results in the table below:

Sample Number	Number complete records	Percent complete records	Number incomplete records	Percent incomplete records
A.				
B.				
C.				
D.				
E.				

How does the size of the sample affect the certainty of the data?

Discussion: The proportions are the sample statistics. For example, the proportion of complete records in your sample is the statistic that summarizes your sample.

1. How does this relate to the population parameter (the TRUE proportion of complete records in the register)?
2. Do you know the value of the Parameter when you start? Could you know the true value, i.e., the real number of incomplete records?
3. Do you know the values of the statistics?
4. Does the value of the parameter change each time you take a sample?
5. Does the value of the statistic change each time you take a sample?
6. Did all of the participants have the same proportion of complete records?
7. How do the actual sample values compare to the ones you estimated earlier?

Session 12: Exploring the Quality of Your Electronic TB Data

Objectives:

1. To grasp various means to detect (possible) errors.
2. To discover outliers and spurious values visually and quantitatively.
3. To learn best practices for finding errors.
4. To compare double entered data for errors in Epi-Data 3.0 and Excel.
5. To use Epi-Info to identify aberrations in the data.

Pre-requisites(s): Session 2 on assuring data quality prospectively and Session 11 on logic checks.

Background Preparation to be completed before the session:

1. Understanding and Using Tuberculosis Data. WHO 2014 Chapter 2 Analysis of case-based TB notification data

2. How to use Epi-Info Compare Utility v. 3.4.3 *(**NOTE for OLD VERSION OF EPI_INFO new Epi-info 7 has no way to check double entered data yet.**)*
   http://www.youtube.com/watch?v=24gBswHdZ8U

3. 2:36 min How to import Excel files into Epi_Info 7:
   http://www.youtube.com/watch?v=CgRCBord-YA

Time	Content	Methodology	Materials
30 min	Exploring quality of electronic data	Guided discussion	Power point
15 min	• Aberration detection for aggregated data • Aberration detection for individual data	Audiovisual stimulus	EPI-Info 7 videos
20 min	Demonstration of Epi-Data compare utility	demonstration	Two nearly identical data sets
30 min	Practice with comparing double entered data for errors	Hands on practice	Two nearly identical datasets

Homework assignment : Watch 2 videos to review key points:

1. 8:17 Epi Info 7 Aberration Detection with visit level data
   http://www.youtube.com/watch?v=kmCtyX5OlcU
2. 3:10 Epi Info 7 Aberration detection analysis with aggregated data
   http://www.youtube.com/watch?v=kmCtyX5OlcU

Session Notes

Data checking and cleaning

Check data
Detect (possible) errors
Validate
Correct errors
The whole process takes a lot of time!

Detection of errors

Every dataset contains errors
The goal is to find them and correct them
This improves the validity of your data and therefore of your results

Step 1: compare data entry files

Check for and correct typing errors

Compare data entry files

Comparison of values for all variables
If difference between the two files, find out the true value
Go back to the questionnaire
Give one of the original data entry files a new name and make all the corrections in this file
Document all the changes you make!
Which value is the true value and why

Missing values

Try to avoid missing values
10% missing values for height and 10% missing values for length gives 10-20% missing values for body mass index
Problem for multiple analysis with many variables: 10 variables with all 3% missing values gives up to 30% missing values for one of the variables and will not be used in the analysis
However, do not guess the value if you are not sure!

Example double data check in Epidata

Use
Document -- Validate duplicate files
or
Tools -- Prepare double entry verification

Example double data check in EpiData

Example double data check in Epi Info 3.02 (old version only)

Use
Utilities -- Data compare
File -- New script
Use the wizard to compare files

Example double data check in Epi Info

Example double data check in Excel

=IF (logical test, 'value if true', 'value if false')
=IF (Sheet1!A2-Sheet2!A2=0, 0, 1)

Example double data check in Excel

Step 2: descriptive statistics

Check for and correct errors in raw data (e.g. questionnaires)
Get a 'feel' for the data

Descriptive statistics

Single variables
Frequency tables
Histograms
Two variables
Cross-tabulation for discrete variables
Scatter plots for longitudinal variables
New variables
Compare values of old and new variable to make sure your new variable is correct
Frequency tables / histograms

Frequency tables

List of all unique values, including missing values
Check:
Lowest and highest values
Unlikely values
Missing values
Distribution of values: Likely? Peaks?
Duplicate entry of study subjects
…..

Frequency tables

Lowest and highest values
Out of normal range?
Within inclusion criteria?

Unlikely values
Negative values
Character instead of numeric values
‘O’ instead of ‘0’
Comma instead of point in figure or the other way around
Incorrect dates
Small vs. capital letters
Missing values
All coded in the same way?
Blank, 9, 99 or 999
Duplicate entry of study subjects?
Same ID number, or same date of birth, sex, and city

Cross-tabulation checks

Some errors will appear only when looking at two variables at the same time
For example:
Man who is pregnant
Woman of 60 years old who is pregnant
Length of 1.90 metres and weight of 40 kilo’s

Frequency table

Cross-tabulation

Scatter plot

New variables

Always check
Number of observations
Number of missing values
Consistency old and new value
A new variable usually has at least the number of missing values of (one of the) old variable(s)
Example: Age from ‘date of birth’ and ‘date of diagnosis’
List the three variables next to each other
No negative values?
How have the missing values been converted?
Age in age groups
List age in years and age group

Other possible errors

Case not according to inclusion criteria

E.g. too young, specific co-morbidity, …
Duplicate cases
Is OK in some studies
When putting together several files, the matching criteria for the same person incorrectly do not match
Variable coded as character instead of numeric

Correction of errors

Check the raw data (e.g. besides the questionnaire there may be the notification record)
Again, make a new file based on your final data-entry file (and save all the old files)
Again, if you are not sure about the true value, make it 'missing'
Again, record all the changes you make

Discussion points for debriefing:

1. Which techniques are good for the preliminary exploration of your data? Why?
2. Under what conditions would fast aberration detection be important?
3. What are the relative strength of Epi-Info 3 versus Epi-Data 3 for comparing double entered data?

Multiple choice questions to assess comprehension:

1. What are exploratory frequencies good for?
   a. To calculate the error rate
   b. To look for missing data
   c. To do multiple imputation
   d. To do double data entry

2. What are exploratory cross tabulations (2x2 tables) good for?
   a. To derive standard deviations
   b. To do logic checks
   c. To calculate the means
   d. To apply weights

Session 13: Introduction to Routine Data Quality Assessment

Objectives:

1. Describe the RDQA process.
2. Describe how to prepare for and conduct an RDQA.
3. Introduce an RDQA Tool.

Pre-requisites: Session on theory, structure, and process of data quality audits.

Background Preparation to be completed before the session:

1. *Manual on use of routine data quality assessment (RDQA) tool for TB monitoring*, WHO.
2. MEASURE Evaluation RDQA tool.
3. *Checklist of Standards and Benchmarks for TB Surveillance and Vital Registration Systems Version 2.4*. September 2013.

Time	Content	Methodology	Materials
10 min	Overview of the RDQA process	Presentation	Powerpoint
12 min	Preparing for and conducting an RDQA		
18 min	Overview of the MEASURE Evaluation RDQA tool		

Homework Assignment: None

Session Notes

Topics to be covered

RDQA process
Preparing for an RDQA
Conducting an RDQA
Overview of the MEASURE Evaluation RDQA tool

Purposes of RDQA

RDQA Implementation Steps

Site Selection

Not necessary to visit all reporting sites to determine the quality of data
If RDQA is part of on-going monitoring
Select sites in parallel with existing supervision visit schedule
If targeting issues (delays in reporting, incomplete reports, questionable data, etc.)

Select sample of problematic sites
Select sample of high functioning sites
If preparing for an audit
Select a representative group using random sampling techniques

RDQA Team

RDQA Team Leader's Responsibilities

RDQA Team Member's Responsibilities

At the Beginning of Each Visit

Conducting the RDQA

Conducting Site Visits

Start verifications at the highest level being evaluated
Complete relevant sections of RDQA tool through interviews & document review

Complete Tool During Visits

Provide a brief overview of the tool
Facilitate a discussion based on the questions in the RDQA tool
Where staff say they have documentation available, ask to see a copy of the documentation at the end of the discussion

Debrief at Assessment Site

Debriefs are provided to each reporting level so that
Staff can see and understand the results of the assessment at their office, i.e. the strengths and weaknesses of their M&E system
Staff have an opportunity to ask questions, correct any errors/misunderstandings, and provide additional clarification on the findings
Team can update the answers in the tool with any corrections or qualifying information
Help the staff generate an action plan appropriate to their site

Debrief Outline

Present findings to the site staff
Highlight and praise all areas of strength, i.e. don't just focus on weaknesses
Discuss each weakness and ask staff to comment on the findings
Develop action items with staff input
End on a positive note

RDQA Feedback by Level

Objectives of the MEASURE Evaluation RDQA Tool

Overview of the RDQA Tool

3 versions
Single indicator
Multiple indicators – up to 4
Longitudinal - single indicator over 4 reporting periods
May be implemented at up to 4 levels
Service Delivery Sites
Health Districts
Intermediate/Aggregate levels
National M&E

Attributes of the RDQA Tool

Components of the RDQA Tool

Part 1 – Data Verifications

Purpose
Assess if sites are collecting and reporting data to measure the selected indicator(s) accurately and on time
Cross-check the reported results with other data sources (service delivery level only)

Data Verification–Health Centre

Data Verification–Higher Levels

Part II—System Assessment

Purpose
Identify potential threats to data quality from the data management and reporting system due to
how it is designed
how it is implemented

Conducting a System Assessment

Apply the system assessment questionnaire in a participatory manner with all relevant M&E staff present
Discuss answers thoroughly
Take detailed notes to ensure a comprehensive understanding of the responses

RDQA Timeline for Monitoring

Discussion points for debriefing:

1. Under what circumstances is an RDQA appropriate to use?
2. Discuss considerations when selecting sites for an RDQA.

Multiple choice questions to assess comprehension.

1. In general, how long should you allow per site for an RDQA?
   a. 2-4 hours
   b. ½ -1 day
   c. 1-1½ days
   d. 1-2 days

2. Which is not part of the RDQA implementation process?
   a. Interpret results
   b. Indicator selection
   c. M&E framework development
   d. Action plan development
   e. Site selection

Session 14: Practice Session with Audit Case Study

1. **Objectives:** Practice data verifications using the RDQA tool.

Pre-requisites: Introduction to RDQA

Time	Content	Methodology	Materials
45 min	Skills practice in RDQA using TB patient dataset from fictional Island country of Maravilha	Hands on use of Part I- Data Verifications	Paperdatset-district.docx [SummaryRept] RDQA_Tool_TB.xls

Homework assignment: none

Discussion points for debriefing:

1. Under what circumstances is an RDQA appropriate to use?
2. Discuss considerations when selecting sites for an RDQA.

Session 15: Use of WHO Checklist of Standards and Benchmarks for TB Surveillance and Vital Registration Systems

Objectives:

1. To gain a broad understanding of the WHO standards and benchmarks.
2. To explore in groups (two countries/groups) the feasibility and challenges of implementing the standards and benchmarks in the learners' countries.

Background Preparation to be completed before the session:

1. *Standards and benchmarks for tuberculosis surveillance and vital registration systems: The checklist and user guide*, Version 4, 2013.

Time	Content	Methodology	Materials
40 mins	Overview of standards and benchmarks	didactic lecture	Power point
40 min	• Each country-group will discuss the methodology and how it would be applied. • Participants review the checklist to determine how to assess the system.	Group exercise aimed to apply the material to their actual context	Checklist and discussion questions

Homework assignment: None

Session Notes:

Task Force strategic areas of work

Surveys of the prevalence of TB disease
Methods to estimate disease burden
Strengthening routine surveillance
The goal: direct measurement of TB cases & death from notification & vital registration data
TB notifications in surveillance system ~ TB incidence
TB deaths in vital registration system ~ TB mortality

Why strengthen surveillance?

Estimates of disease burden are currently highly reliant on expert opinion
Two main reasons why this is the case
TB cases are diagnosed but not reported
TB cases are not diagnosed
Vital registration data not frequently utilized by National TB Programs
No systematic method for assessing data quality and coverage prior to 2010

Standards and benchmarks (S&Bs): Definitions

Standards: general statements about the characteristics that define a high-performance TB surveillance system
Benchmarks: define in quantitative terms wherever possible the level of performance that is considered good enough to meet the standard

Standards and benchmarks for TB surveillance: Purpose

Assess a surveillance system's ability to accurately measure TB cases and deaths in all settings in a standardized way
Use surveillance data for direct measurement
Identify and better quantify shortcomings in surveillance systems that need to be addressed

Standards and benchmarks for TB surveillance: Purpose

Inform TB program staff, policy-makers & partners about aspects of surveillance systems that need to be strengthened to improve TB control
Develop a M & E investment plan to address identified gaps in surveillance

Development of the standards and benchmarks for TB surveillance

Underlying principles

Built on experience of regional workshops (2010 – 2011)
TB epidemiology
Evidence-based (WHO data and literature)
High performing systems used as models
Aimed for a minimum set of standards
Applicable across different geographic areas (high & low burden settings) & systems (electronic & paper-based)
Involved partners from national programs & technical agencies

Lessons learned- Pilot Testing

Perceived to be useful and feasible
Some parts needed to be removed or changed
Can be done in about ~ 1 week, except for cross checking of source documents (paper-based)
Users required some epidemiology background to conduct assessment

Lessons learned- Pilot Testing

Some challenges to identifying S & B that are appropriate for all systems and settings
1 standard different for electronic and paper-based
Evidence from previous studies may be used for some standards, e.g.
B1.4 - requiring cross checking of source documents
B1.8 - assessing under-reporting
User guide needed

Implementation of standards & benchmarks

Methods

Standards & benchmarks for TB surveillance: Intended use

Designed to allow a national assessment for most recent complete calendar year
Lag time may range from no delay to one year
An assessment of a TB surveillance system using this checklist would take place at least every 3-5 years (or more often, if feasible)

Standards & benchmarks for TB surveillance: Intended use

Checklist can be used by in-country staff for self-assessment or by external reviewers, e.g.
Global Fund
National Program Reviews

Standards & benchmarks for TB surveillance: Method used

Desk review of documents, datasets, and electronic surveillance systems
Data quality audits

Standards & benchmarks for TB surveillance: Requirements

Description of the TB surveillance system
Data sources
Surveillance data for analyses
Program documents, manuals, SOPs
Facility & district level source documents
Previous studies (e.g. TB, HIV, DRTB surveys, inventory & mortality studies)
Data external to the program

Standards & benchmarks for TB surveillance: Requirements

Personnel
National, district and facility levels
M&E officers, data managers, lab staff, epidemiologists, statistician, TB program officers
Vital registration & HIV staff

Standards & benchmarks for TB surveillance: Interpretation

For a country's TB surveillance system to be certified as providing a direct measurement of TB cases:
10 standards need to be met
1 is specific to paper-based systems
1 is specific to electronic case-based systems
2 assess system coverage

Standards & benchmarks for TB surveillance: Interpretation

For a country's TB surveillance system to be certified as providing a direct measure of the # of DR-TB, TB/HIV, and TB cases in children specifically, 3 additional standards must be met
For surveillance system to provide a direct measure of TB deaths there is 1 standard that must be met

Overview of the standards and benchmarks for TB surveillance: A checklist

Standards and benchmarks for TB surveillance: Overview

Checklist includes standards and benchmarks related to data quality, system coverage, TB mortality data, drug resistant TB (DRTB) surveillance, TB/HIV & TB cases in children
Checklist consists of a set of 13 standards & associated benchmarks
9 standards: related to TB cases measurement
1 standard: related to TB deaths measurement

Standards and benchmarks for TB surveillance: Data Quality

Standards and benchmarks for TB surveillance: Data Quality

Standards and benchmarks for TB surveillance: Data Quality

Standards and benchmarks for TB surveillance: Data Quality

Standards and benchmarks for TB surveillance: Coverage

Standards and benchmarks for TB surveillance: Vital Registration

Standards and benchmarks for TB surveillance: DR TB, TB/HIV & children

Standards and benchmarks for TB surveillance: DR TB, TB/HIV & children

What have we learned so far in rolling out the TB surveillance checklist?

Common findings from roll-out of TB surveillance checklist

Sub-optimal or unknown data quality at facility and district levels, based on available information, but difficult to assess
Need to conduct national level data quality audits
Electronic recording and reporting systems needed
Limited use and analysis of TB surveillance data
Guidance (TB surveillance analysis handbook) is being developed

Common findings from roll-out of TB surveillance checklist

Limited understanding of level of underreporting of TB
Inventory studies can be used to measure unreported cases
Poor measurement of TB mortality
Need to strengthen vital registration systems and coding of causes of death

Importance of linkages with other initiatives and closely related efforts

Supports Global Fund approach to strengthening impact measurement
Uses the Service Availability and Readiness Assessment (SARA) tool to systematically assess data quality nationally
Tracks progress in health systems strengthening

Importance of linkages with other initiatives and closely related efforts

Feeds into workshops by the Commission on Information and Accountability for Women's and Children's Health (COIA)
Developing country roadmaps for health systems strengthening

Acknowledgements

Emily Bloss, PhD; Division of Tuberculosis Elimination
Centers for Disease Control and Prevention
Contributors to the development of the standards and benchmarks checklist and/or user guide: Members of the Task Force
Countries contributing to the work around standards and benchmarks: Brazil, China, Côte d'Ivoire, Egypt, Estonia, Ghana, Indonesia, Japan, Kenya, Netherlands, Nigeria, Thailand, Uganda, UK, USA, Viet Nam

Discussion points for debriefing:

1. Which of the benchmarks do you think are the least likely to be measurable and why?
2. What other methods might you try in order to capture the mortality information?

Multiple choice questions to assess comprehension:

1. Standards and benchmarks focus on strengthening various aspects of a country's TB routine surveillance system. What other aspects does it focus on?
   a. S & B checklist focuses on TB mortality data, drug resistance, data quality, system coverage, and TB/HIV & TB cases in children.
   b. Standards define wherever possible in qualitative terms performance levels that is good enough to meet the standard.
   c. All the above.
   d. None of the above.

True/False: The checklist consists of a set of 13 standards and associated benchmarks, nine standards related to TB cases measurement and one standard related to TB deaths measurement.

2. Lab-confirmed cases vs. clinically diagnosed cases is an example of:
   a. Standard
   b. Benchmark
   c. Both

d. None of the above

Session 16: M&E of TB Mortality

Objectives:

1. To discuss the M&E implications of the new Post-2015 Stop TB targets.
2. To discuss the limitations and challenges of mortality measurement including inadequacy of verbal autopsy and civil registries.
3. To describe a participatory approach for assessing the root causes of high mortality in specific facilities or communities.

Background Preparation to be completed before the session:

1. *STOP TB partnership Post-2015 Targets.*
2. http://www.who.int/tb/post2015_strategy/en/
3. The Lessons from Loss Tool: www.tbcare1.org/access
3. Understanding and Using Tuberculosis Data. WHO 2014 Chapter 5 Estimating tuberculosis mortality using vital registration and mortality survey data

Time	Content	Methodology	Materials
30 min	Discussion of mortality projections and new Post-2015 Stop TB targets	Presentation	Power point
30 min	Review of the lessons from Loss tool	Presentation	The Lessons from Loss Tool

Homework Assignment: None

Session Notes

Objectives of this session
To explore the new STOP TB targets
To discuss potential approaches to measurement
To describe a new tool for mortality audits– the Lessons from Loss tool

Post-2015 tuberculosis strategy

The vision for the post-2015 tuberculosis strategy is "A world free of tuberculosis", also expressed as:

"Zero deaths, disease or suffering due to tuberculosis"

The goal is to end the global tuberculosis epidemic

The global TB targets for 2035 is:

a **95% decline in the deaths** due to tuberculosis, compared with 2015,

2025 Mortality Milestone

A key milestone is a **75% reduction in tuberculosis deaths by 2025**, compared with 2015.
This requires two things.
First, the annual rate of decline in global tuberculosis incidence rates must accelerate from an average of 2% per year in 2015 to 10% per year by 2025.
Second, the proportion of incident cases dying from tuberculosis (the case fatality ratio) needs to decline from a projected **15%** in 2015 to **6.5%** by 2025.

Which means: countries need a TB mortality baseline value by 2015

This is not just deaths in the treatment cohort, *but also deaths of people with TB who were never diagnosed*

Mortality: Notification Ratio

M:N is a proposed indicator
Once we find out # of TB deaths, we see how many of them were diagnosed and entered in the surveillance system.

ICD-10 Mortality Measurement

According to the latest revision of the inter-national classification of diseases (ICD-10), TB mortality is the number of deaths caused by TB in HIV-negative people.
TB deaths among HIV-positive people are classified as **HIV deaths** in ICD-10.
For this reason, estimates of deaths caused by TB in HIV-positive people are presented separately from those in HIV-negative people

How can these baseline rates be obtained?

The two approaches promoted globally to measure the TB mortality will be:

- Health and demographic surveillance system (HDSS) conducted on sentinel populations;
- Sample vital registration with verbal autopsy (SAVVY) conducted on statistically
- sampled population clusters representative of the whole population.

Both of these approaches use verbal autopsy (VA) to determine cause of death

What is Verbal Autopsy?

Does it work for TB? No.

Verbal Autopsy for TB

Definition: A verbal autopsy is an interview of relatives or caregivers regarding:

- the signs,
- symptoms,
- behaviors and
- other circumstances experienced by the deceased before their death

What might be some problems with human coders using ICD-10?
What might be some problems with a computer algorithm?

Verbal autopsy (VA) may not work and Necropsy is not popular or feasible

Without a pathognomonic sign for TB it is hard to use an algorithm-based VA
In many settings family members refuse necropsy at rates > 70% and there is no capacity to do this worldwide.

Minimally invasive autopsy techniques are being tested and may be a good option for pulmonary TB in the future.

Sample Vital Registration will be promoted continent-wide

Draft of the STOP TB Strategy post-2015
"An interim solution being adopted by an increasing number of countries is the introduction of a sample Vital Registration system. In the coming decade, the biggest challenge will be the expansion of Vital Registration systems in African countries. "

WHO Civil Registry 2013 Resource Kit already promoting InterVA as better than coders

" *One approach – **InterVA** – is now in widespread use in HDSS sites... ...new automated techniques that perform even better than physicians.."*

Validity Problems with Verbal Autopsy

We know that verbal autopsy does not work well for children with TB and we suspect it does not distinguish adult TB deaths well either.

Discussion

What are the possible alternatives ?
Piloting TB Patient Mortality Audits using a Patient-Centered Approach
BY MOH Ethiopia & TB CARE I Ethiopia
2013

Purposes of TB Mortality Audit…

Both health care and community workers and family members should be assured that the sole purpose of the audit is to learn valuable lessons from the tragic death of the patients and to save lives in the future.
These reviews seek only to identify barriers to accessing and receiving quality care in the health care system. They must never be used to provide the basis for litigation, management sanctions or personnel decisions.

Pinpoint the Missteps on the TB journey

Methods of TB Mortality Audit

Definitions, advantages and disadvantages of audit methods to facilitate decision making.

Part A: Community-based Death Review (CBDR)

Definition: A method of ascertaining the personal, familial, community, and quality of care factors that may have contributed to the deaths.

Part B: Facility-Based Mortality Audit (FBMA)

Definition: An in-depth investigation of
-care provided
Thanks
Acknowledgements:
Osman Abdulahi, Eveline Klinkenberg, Eliud Wandwalo, Jacques van den Broek, Charlotte Colvin, Hillary Kipruto,

Discussion points for debriefing:

1. What are the limitations of verbal autopsy for TB?
2. What are some possible alternatives to verbal autopsy? Why would these work?

Multiple choice questions for assessing comprehension:

2. How does the current system of death classification bias the measurement of TB mortality?
   a. People who die of HIV are always classified as TB.
   b. People who have TB and HIV are always classified as having died from HIV.
   c. People who have any cough are always classified as pneumonia deaths.

3. A key milestone in the 2015 strategy is a 75% reduction in tuberculosis deaths by 2025. What would this require?
    a. The annual rate of decline in global tuberculosis incidence rates must accelerate from an average of 2% per year in 2015 to 10% per year.
    b. The proportion of incident cases dying from tuberculosis (the case fatality ratio) needs to decline from a projected 15% in 2015 to 6.5% by 2025.
    c. A lot of money is needed.
    d. <u>All of the above</u>.

Theme 3: New Challenges in TB M&E

Session 17: Revised WHO Case Definitions and Reporting Forms for 2013

Objectives:

1. Define the key changes in the TB definitions released by WHO in 2013.
2. Explain WHO-recommended changes to reporting forms.
3. Analyze the implications of adopting the revised definitions and forms at country level for the NTP.

Background Preparation to be completed before the session:

1. *Definitions and reporting framework for tuberculosis – 2013 revision* (WHO): www.who.int/iris/bitstream/10665/79199/1/9789241505345_eng.pdf

2. *Companion Handbook to the WHO Guidelines For The Programmatic Management Of Drug-Resistant Tuberculosis. 2014. Chapter 2. Monitoring the detection, enrolment and treatment outcomes of drug-resistant TB patients www.who.int*

Time	Content	Methodology	Materials
20 min	Summarize key changes to TB definitions (focus on the changes as opposed to reviewing every definition).	Presentation/ discussion	PowerPoint presentation
10 min	Introduce WHO sample forms that incorporate the new definitions (TB/HIV and PMDT-related forms will not be discussed in this session).		
30 min	What do these changes mean for my country? (Countries share experiences with updating forms, discuss steps for adopting new definitions/forms, and examine how their results may be affected by the changes).	Group discussion	

Homework assignment:

Review this website: Questions and answers: the 2013 revision of the WHO definitions and reporting framework for tuberculosis

http://www.who.int/tb/publications/definitions_faq/en/

Please read *Definitions and reporting framework for tuberculosis – 2013 revision* (WHO):
(www.who.int/iris/bitstream/10665/79199/1/9789241505345_eng.pdf)

Session Notes

Objectives
Outline the main changes in the 2013 document:
1) Definitions: Basic TB & rifampicin-resistant TB (RR-TB)
2) Reporting framework: Basic TB & RR-TB
What do the changes mean for our NTPs?

Revision process

Collaborative work of World Health Organization (WHO) staff at different levels, technical partners and national staff.
May 2011: expert consultation in Geneva, Switzerland.
June 2011: WHO's Strategic and Technical Advisory Group on TB (STAG-TB)
July 2011: presentations and discussions with WHO regional and country staff, Geneva, and subsequent further consultation with WHO staff.
October 2011: meeting of the DOTS Expansion Working Group, Lille, France.
E-mail consultation with a wide range of countries and technical partners between November 2011 and March 2013.
Seven countries (Belarus, Brazil, Cambodia, Djibouti, Estonia, Pakistan, Philippines) pilot the definitions and forms in 2012 and provide feedback

Definitions

Bacteriological confirmation needs to consider results from new WHO-approved rapid diagnostics (WRD), including Xpert MTB/RIF;
Differentiate b/w rifampicin-resistant TB (RR-TB) and confirmed MDR-TB cases;
Simplification of definitions of 'Cured' and 'Treatment Failed' in RR-TB cohorts to allow for their application while patient is still on treatment;
Less judgmental language: **'Defaulter'** replaced by **'Lost to follow-up'** and **'TB suspect'** by **'Presumptive TB'**

- Classification based on anatomical site of disease
- Classification based on history of previous TB treatment (patient registration group)
- Classification based on HIV status
- Classification based on drug resistance
- Treatment outcome definitions

- Treatment outcomes for TB patients (excluding patients treated for RR-TB or MDR-TB)
- Outcomes for RR-TB/MDR-TB/XDR-TB patients treated using second-line treatment

Presumptive TB:

A patient who presents with symptoms or signs suggestive of TB **(previously TB suspect)**

TB case:

A bacteriologically confirmed TB case: a biological specimen is positive by smear microscopy, culture or WRD. All such cases should be notified, regardless of whether TB treatment has started (**previously *Definite TB case*; now includes explicit mention of WRD**)
A clinically diagnosed TB case : not bacteriologically confirmed but diagnosed with active TB by a clinician or other medical practitioner who has decided to give the patient a full course of TB treatment (**previously a case of TB, not considered *Definite***)
Pulmonary tuberculosis (PTB): **specific mention of tracheobronchial tree**

Extrapulmonary tuberculosis (EPTB)

Focus is now on previous treatment history*, independent of bacteriological confirmation or site of disease* (NB: for RR-TB these groups are different).
New: never been treated for TB or have taken anti-TB drugs for less than 1 month
Previously treated: have received 1 month or more of anti-TB drugs in the past
(changes for sub-category definitions)

- Relapse
- Treatment after failure
- Treatment after loss to follow-up
- Other previously treated
- Patients with unknown previous treatment history **(new group)**

Relapse: previously treated for TB, were declared cured or treatment completed at the end of their most recent course of treatment, and are now diagnosed with a recurrent episode of TB (either a true relapse or a new episode of TB caused by reinfection); **(reworded & removed mention of bacteriological positive TB)**

Treatment after failure: **(reworded, but similar meaning as before)**

Treatment after loss to follow-up: **(previously known as 'treatment after default')**

Other previously treated: **(cases with unknown previous TB treatment history classified separately)**

HIV-positive TB patient: any TB case who has a positive result from HIV testing conducted at the time of TB diagnosis or **other documented evidence of enrolment in HIV care**[1], such as enrolment in the pre-ART register or in the ART register once ART has been started.
HIV-negative TB patient: any TB case who has a negative result from HIV testing conducted at the time of TB diagnosis. **(not previously defined)**

HIV status unknown TB patient: any TB case who has no result of HIV testing and **no other documented evidence of enrolment in HIV care**[1]**.**
1) A guide to monitoring and evaluation for collaborative TB/HIV activities (whqlibdoc.who.int/publications/2009/9789241598194_eng.pdf)
Main change: **inclusion of rifampicin-resistant TB (RR-TB).** RR-TB includes any resistance to rifampicin, whether mono-resistance, multidrug resistance, poly-drug resistance or extensive drug resistance.
Category is not mutually exclusive with the others.

Rifampicin resistance: resistance to rifampicin detected using phenotypic or genotypic methods, with or without resistance to other anti-TB drugs. **(new definition)**

- Mono-resistance
- Poly-drug resistance
- Multi-drug resistance
- Extensive drug resistance
-

Classification based on drug resistance (3)
NOTE: Mono-resistance and poly-drug resistance are usually applied to first-line drugs only (R, H, E and S). Future drug regimens may make it important to classify patients by their strain resistance patterns to fluoroquinolones, second-line injectable agents and any other drug for which reliable DST becomes available.
Two sets of definitions for **two, mutually-exclusive treatment outcome cohorts**:
Outcomes for TB patients, excluding patients treated for RR-TB ("Basic TB")
Outcomes for RR-TB/MDR-TB/XDR-TB patients treated using second-line treatment **("RR-TB")**
The first group **("Basic TB")** may include cases who have **drug-susceptible TB**, or other forms of mono-resistance (e.g., INH-resistance) not requiring a full second-line

regimen for MDR-TB. Outcomes are assigned to all bacteriologically confirmed and clinically diagnosed TB cases including those who die or who are lost to follow-up before starting treatment.
The second group ("**RR-TB")** includes **all RR-TB, MDR-TB and XDR-TB cases, confirmed or presumptive**, started on combination second-line regimen for MDR-TB as per the local policy. Outcomes are assigned to all.

For treatment outcome monitoring, only laboratory confirmed RR-TB (+ MDR-TB/XDR-TB) are enumerated.

1. Cured **(only pulmonary; initial bacteriological confirmation may be based on WRD)**
2.
3. Treatment failed **(no longer includes systematically any case with confirmed MDR-TB)**
4. Lost to follow-up **(previously 'Default')**
5. Not evaluated **(now includes previous 'Transfer out' category)**
6. Treatment completed
7. Died

Treatment success

Treatment failed: a case confirmed to be MDR-TB is no longer automatically assigned this outcome. If the patient is started on a combination second-line regimen for MDR-TB the case is excluded from the "Basic TB" cohort when calculating treatment outcomes and transferred to the "RR-TB cohort". If treatment with a combination second-line regimen for MDR-TB is not possible, the patient is kept in the "Basic TB" cohort and assigned an outcome from among those on the previous page.

1. Cured
2. Treatment completed
3. Treatment failed
4. Died
5. Lost to follow-up
6. Not evaluated

- Treatment success

Old definition of Cured
(Cat IV)
Cured: **(negative cultures counted after the intensive phase no longer limited to last 12 months of treatment)**

Treatment completed: **(changes only insofar as applied to 'Cured')**

Old definition of Failed
(Cat IV)
Definition now determined primarily by changes required to the regimen as a result of non response as determined by lack of conversion, reversion, amplification or ADRs.

Treatment failed: Treatment terminated or need for permanent regimen change of at least two anti-TB drugs because of:
lack of conversion by the end of the intensive phase, or
bacteriological reversion in the continuation phase after conversion to negative, or
evidence of additional acquired resistance to fluoroquinolones or second-line injectable drugs, or
adverse drug reactions (ADRs)

Lack of conversion by the end of the maximum intensive phase used by the program.
If no maximum duration is defined, an 8-month cut-off is proposed.
If regimens do not have a clear distinction between intensive and continuation phases, a cut-off 8 months after the start of treatment is suggested to determine when the criteria for Cured, Treatment completed and Treatment failed start to apply.
Reversion (to positive): after an initial conversion, 2 consecutive cultures, at least 30 days apart, are positive. For Treatment failed, reversion considered only when it occurs in the continuation phase.

Revised Recording & Reporting Forms

Inclusion of TB cases detected using WRD as well as RR-TB cases
Combining outcome reporting for drug-sensitive and RR-TB for countries where PMDT is incorporated ("mainstreamed") in the NTP
Childhood TB reporting was incomplete because age disaggregation was previously limited to sputum smear-positive TB, which is uncommon in children
There was a delay of 2 years in the reporting of cotrimoxazole preventive therapy (CPT) and antiretroviral therapy (ART) in TB/HIV because these data were collected only in the treatment outcome reports and not in the case registration reports

Revised forms in document

1. Request for examination of biological specimen for TB
2. Basic management unit TB register
3. Second-line TB treatment register
4. Laboratory register for smear microscopy and Xpert MTB/RIF
5. Laboratory register for culture, Xpert MTB/RIF and drug susceptibility testing (DST)
6. Quarterly report on TB case registration in the basic management unit

7. Quarterly report on TB treatment outcomes in the basic management unit
8. Combined annual outcomes report for basic TB and for RR-/MDR-TB

Tools for patient management not included;

1. Forms for human resource or management of consumables not covered in this document;
2. Forms for community-based management of TB
3. Registers for persons with presumptive TB

Revised forms and reports for RR-TB will be discussed in greater detail in the forthcoming "Companion handbook to the 2011 WHO guidelines for the programmatic management of drug-resistant tuberculosis".
The register is intended to **record all patients diagnosed** with TB and eligible for TB treatment, including those diagnosed with RR-TB or MDR-TB, regardless of whether treatment was actually started. **This is different from the previous advice to include only cases starting treatment.**

Bacteriological examination before the start of treatment ("month 0") now **allows for results from Xpert MTB/RIF test**. Record if the case is RR-TB or MDR-TB, replacing X-ray result.
Dates for **HIV testing and start of ART/CPT removed** from register
Change in **treatment categories**

The terms 'Lost to follow-up' and 'presumptive TB' are less judgmental towards the patient than previous terminology.
A patient can be considered RR-TB and XDR-TB at the same time.
HIV-positive status can only be counted if the test is done at the time of TB diagnosis.
There are 2 mutually exclusive treatment cohorts – 'Basic TB' and 'RR-TB'.
MDR-TB patients must always be moved to the 'RR-TB' cohort.
The BMU register (i.e. District TB register) records all patients diagnosed with TB and eligible for TB treatment, regardless of whether treatment was actually started.
'Category IV' treatment has been replaced by 'second-line treatment regimen';
Forms should be used only as WHO has designed them (no changes possible).
What does this mean for our countries?

DISCUSION

- How do you think the changes in definitions and reporting framework will affect your data?
- Which indicators do you think you'll have better data for? Worse?
- Which indicators will be impacted the most?

POTENTIAL RESPONSES

- Increase in HIV results (HIV+ and known status)
- Data on ART/CPT available more quickly (quality still an issue?)
- Including all diagnosed patients (regardless of treatment initiation) in register may result in greater case notification, but outcomes may decline (more 'loss to follow-up', 'not evaluated'.
- Removing RR-TB patients from Basic TB cohort will reduce basic treatment failure.

What else?

DISCUSSION OF Country experiences

1. What are your experiences with transitioning to new definitions or reporting forms?
2. What was (or will be) the most challenging part?
3. What would you do differently, or emphasize the importance of, this time around?
4. Will you need technical or financial support to implement the changes?
5. What questions do you have?

REMEMBER:

- Forms are illustrative; shows minimum dataset.
- Countries will need to adapt the forms to fit their needs.
- Potential modifications?
- Translate
- Add new data items (ID number, dates, etc.)
- Remove HIV data (due to confidentiality laws)
- Add logos
- Change format
- Adapt terminology to local situation

What else?

Pilot test (and retest!) forms at all levels where they will be used;
Develop roll-out plan for new definitions and revised forms;
The revised definitions should be applied by the NTP at a set changeover date (e.g. 1 January);
All cases on treatment on that date will be assigned outcomes according to the revised definitions. This means that patients started on treatment in the previous year may be assigned outcomes according to two different definitions of *cured* or *treatment failed*, depending on whether they completed treatment before or after the changeover date. (More practical than retrospective outcome reassignment).

Acknowledgements
This presentation has been adapted from a WHO-developed presentation on the document. Special thanks to Dennis Falzon for sharing it.

Thank you!
Objective: minimum indicators for national or project level monitoring and suitable for different partners (WHO, TGF) easily extracted manually or electronically conform to what was used in past and DOTS system

Focus on indicators rather than forms

Indicators for RR-TB / MDR-TB / XDR-TB - Detection

Discussion points for debriefing:

1. What are some of the reasons for changing the case definitions at this time?
2. How have the treatment categories changed (i.e., Category I-IV)?

Multiple choice questions for assessing comprehension:

1. TB terminology was changed to be less judgmental. Select the correct change(s):
   a. *MDR TB* is now known as *RR TB*
   b. *Defaulter* is now known as *Lost to follow-up*
   c. *TB suspect* is now known as *presumptive TB*
   d. b and c
   e. all of the above

ANSWER: D. MDR-TB is included in the general category of rifampicin-resistant TB (as is mono-resistance, multidrug resistant TB (XDR-TB), etc.), but is still defined as resistance to at least INH and RIF.

2. True/False: One of the main reasons WHO has revised reporting forms is to permit the inclusion of TB cases detected using WHO-approved rapid diagnostics.

ANSWER: True. The introduction of diagnostics like GeneXpert requires countries to adjust how they define, collect, and analyze their TB data.

Session 18: M&E of Contact Investigations & Screening Programs

Objectives:

1. To introduce the contact investigation (CI) guidelines and the Active Case Finding Guidelines to M&E Officers.
2. To learn both methods and indicators to measure the effectiveness of screening.
3. To compare approaches to gathering these variables.
4. To review data collection tools and forms.

Back ground Reading:

1. *Systematic screening for active tuberculosis: principles and recommendations*: http://www.who.int/tb/tbscreening/en/
2. *Recommendations for investigating contacts of persons with infectious tuberculosis in low- and middle-income countries*. WHO 2012: http://www.tbcare1.org/publications/toolbox/tools/hss/Guidelines_on_TB_Contact_Investigation.pdf

Time	Content	Methodology	Materials
45min	Overview of the WHO Screening & Contact Investigation Guidelines	Guided discussion/interactive	27ppt slides

Homework assignment:

Objectives

To introduce the new WHO Contact investigation and Screening Guidelines
To familiarize ourselves with the key concepts and variables to track the effectiveness of screening
To compare approaches to gathering this information
To practice calculating the main indicators with sample data

Why do we have to go out and look for people when our TB clinics are full?

- Why contact tracing?
- Why active case finding?
- Why now?

What do we *mean* by "TB screening" ?

Screening tests vs. diagnostic tests?

•Screening tests sort out apparently well persons who probably have a disease from those who probably do not, and are not intended to be diagnostic.
Persons with positive or suspicious findings should be tested with a confirmative diagnostic test.

Definition

"Systematic screening for active TB"

Systematic identification, in a predetermined target group, of people with suspected active TB, by the use of tests, examinations, or other procedures which can be applied rapidly

Among those screened positive, the diagnosis should be established through diagnostic tests and clinical assessments with combined high specificity.

Can target:

people who do not seek care (access barriers, not recognizing symptoms as serious, etc), *or*
people who seek care (with or without symptoms/signs compatible with TB), e.g. specific clinical risk groups

Define "an index TB case"

Define "a contact"

Guidelines Overview

contact investigation should be conducted for household and close contacts when the index case …

• has sputum smear-positive pulmonary TB,
• has multi-drug-resistant TB (MDR-TB or extremely-resistant TB (XDR-TB) (proven or suspected),
• is a PLHIV or
• is a child < 5 years of age.

Whom to screen?

Priority for contact investigation should be given to:
People of all ages with symptoms suggestive of TB,
Children <5 years of age,
People with known or suspected immune-compromising conditions (especially PLHIV)
Contacts of index cases with MDR-TB or XDR-TB (proven or suspected).

HIV testing as part of the C.I.

In settings of high HIV prevalence it is recommended that all household and close contacts be counseled and tested for HIV.

What if the contacts do not have TB yet?

PLHIV and children < 5 years of age who are household or close contacts should be treated for presumed Latent TB Infection(LTBI)

Various regimens

Convert these into M&E indicators…

2013 WHO Screening Guidelines

WHO Guideline Review Committee approval March 2013

What are some examples of types of TB screening programs from your countries?

What do we want to know?

1. Yield of our efforts?
2. Is the benefit worth the higher cost?
3. Are we helping individuals?
4. Are we having an impact on the epidemic?
5. Are we screening the right people?
6. Are we successfully treating the people we diagnose?

What might be some *risks* or *problems* with doing a lot of "TB screening" ?

1. Screening the wrong people
2. Screening people with the wrong test tests
3. Under-diagnose TB
4. Over diagnose TB – expose people to medicines they don't need
5. Screening people too often or in ways that harm them

How will you decide whom to screen?

Screen: Cough >2weeks (35% sens. ; 95% spec.)
Diagnosis: Sputum Smear Microscopy (61% sens. ; 98% spec.)

What information do we want to capture in our M&E system?

Brainstorming on existing M and E systems for Contact investigation

Brainstorm Basic Outcome indicators?

% of those screened that are identified as having active tuberculosis
Number Needed to Screen to detect 1 case
Numbers screened / yield = number needed to screen

Some Process indicators?

- % of index TB patients with any contacts identified by name
- % of index TB patients with any contacts screened for TB according to the prioritization plan
- % of contacts that are screened for active TB, among all contacts identified
- % of contacts that have symptoms suggestive of TB, among all contacts screened

Double Danger: Tracking Risk Groups among TB contacts

- *% of contacts that are under the age of 5 years*
- *% of contacts that are tested or known HIV positive*
- *% of contacts classified as having diabetes or other immunosuppressive condition*

What data collection instruments would be needed to get all these data?

Show samples from Namibia

- Contact investigating slip
- TB treatment card: list of contacts for each diagnosed case
- DR-TB register
- Quarterly reporting form
- IPT register/IPT quarterly reporting form

Routine data collection tools for contact investigation

- Contact investigating slip
- TB treatment card: list of contacts for each index case
- DR-TB register (# of contacts and status of TB treatment)
- Quarterly reporting form (summary of contacts and their TB status)
- IPT register/IPT quarterly reporting form

Contact investigation slip

- Information leaflet for contacts
- Introducing TB patient to contact

Explaining risk of transmission and benefits of screening
Sharing information on common symptoms
Administering the questionnaire and possible actions based on the result
If yes to any of the listed questions below; visit to nearest health facility for further assessment

- Extract from TB Treatment Register
- Extract from DR-TB Register
- Extract from Quarterly reporting format for TB

IPT Register

OR: Contact investigation in Namibia

One of **5** research projects planned for Namibia before end of year
Using a stepped wedge design
RCT interventions during routine implementation
Cross-over design
Roll-out of interventions over a # of time periods
To be done in 2 of 13 regions

- Data collection starts July 2013
- Data analysis planned for Sept 2013
- Publishable by end of 2014

THANKS
Acknowledgements:
Knut Lönnroth (WHO)
Nanurai Ruswa (KNCV)

Discussion points for debriefing:

1. What are the main differences between a screening test and a diagnostic test?
2. Why is the choice of screening algorithm so important for M&E?

Multiple choice questions for assessing comprehension:

1. What are key impact indicators for M&E of contact investigation?
   a. Percentage of eligible child contacts under 5 year placed on IPT
   b. Percentage of people with HIV put on IPT
   c. Percentage of smear positive TB patients who report names of contacts
   d. Percentage of TB patients who complete treatment
   e. Percentage of screened contacts who are diagnosed with TB
   f. All of the above

2. What are the main differences between a screening test and a diagnostic test?
    a. A screening test is for ruling-out disease, but a diagnostic test is for ruling a disease in.
    b. A screening test should always be low tech, and a diagnostic test is high tech.
    c. A screening test needs to be inexpensive, but a diagnostic test can be expensive.
    d. A screening test should be sensitive, and a diagnostic test should be both sensitive and highly specific.

Session 19: M&E of TB in Health Care Workers and other Occupational Groups

Objectives:

1. To discuss whether and why we screen HCW for TB
2. To learn how we should we monitor and evaluate HCW screening
3. To discuss various ways of assessing and reporting on HCW screening efforts

Pre-requisites: Epi-Info 7 intro, session on contact investigation and screening guidelines.

Background Preparation to be completed before the session: •

1. *WHO Policy on TB Infection Control in Health-Care Facilities, Congregate Settings and Households*. WHO/HTM/TB/2009.419. Available from: http://whqlibdoc.who.int/publications/2009/9789241598323_eng.pdf
2. *Joint WHO-ILO policy guidelines for improving health worker access to HIV and TB prevention, treatment, care and support*. Geneva: World Health Organization and International Labour Organization, 2010. Available from: http://whqlibdoc.who.int/publications/2010/9789241500692_eng.pdf
3. *Guide on the Monitoring of TB Disease Incidence Among Health Care Workers*: http://www.tbcare1.org/publications/toolbox/tools/hss/HCW_TB_Incidence_Measuring_Guide.pdf
4. *Guide to Measure the Prevalence of Active TB Disease Among Health Care Workers*: http://www.tbcare1.org/publications/toolbox/tools/hss/HCW_TB_Prevalence_Measuring_Guidelines.pdf

Time	Content	Methodology	Materials
30 min	Overview of the M&E implications of the HCW screening guidelines	Presentation	Slides
15 min	Review of the Ndola Screening M&E forms	Discussion	Excel file
15 min	Good practices in data security and confidentiality and prevention of deductive disclosure.	Practice filtering out names in Epi-Info.	

Homework assignment: None

Session Notes

Objectives of this Session

- Why Screen Health Care Workers?
- How do we screen them?
- How do we monitor & evaluate HCW screening?
- How do we report on it?

Why? Justification

In many settings the burden of TB is higher among HCW than among the general population [Menzies 2007, Joshi 2006, Baussano 2011]
Workers' rights: Priority access for HCW to services for the prevention, treatment, and care of tuberculosis,
HIV, through Provider Initiated Testing and Counseling (PITC) [WHO/ILO 2010]

The bigger picture

Monitoring of individuals employed in the healthcare sector for active TB is an essential component of infection control
Algorithm options
Finding the needle in the haystack
Pilot project on HCW screening
Ndola district Zambia pilot implementation
Part of larger TB infection control demonstration project
Objective of intervention
To assess feasibility and acceptance of performing the HCW screening

Will they come?

Participation rates in screening among Healthcare workers (HCWs) have been considered in a systematic literature review in 2007. A total of 25 publications were identified and the participation rates of HCWs in TST surveys ranged from 80 to 100%.[77]
Across 14 hospitals in Melbourne, participation in LTBI screening varied widely from, 13%-66% suggesting a role of hospital administration and context in determining acceptability.

Acceptability varies by gender & cadre

Even when screening was compulsory among health workers, participation rates could be as low as 75% and not conform to the prescribed periodicity.[82-83]
Participants: all HCW
All staff in 15 facilities of the Ndola
Including:

- support and administrative staff, laboratory workers
- cleaners
- TB treatment supporters

Screening process

Annual screening at own facility

HCW may choose to be screened elsewhere
We want to monitor how often this happens
HCWs with symptoms of TB (either throughout the year or during the screening process) will be referred to staff clinic of Ndola Central Hospital for
Sputum for culture at TDRC [priority]
Chest X-ray at Ndola Central Hospital [priority]

HIV screening

Provider Initiated Counseling and Testing
HCW are encouraged to know their status
Annual screening of all
Those who already know they are HIV infected do not need to be screened again
There is no need to disclose HIV test results in order to ensure confidentiality

Relevant staff & tasks

Should be an assigned trusted clinician: screening person
Encouraged HCW to report when having any symptom that may be related to TB
To invite all HCWs for annual screening & refer for CXR and sputum sample taking (use referral form and indicate this is project participant)
To keep records in a lockable cabinet
To ensure all results come in and those with TB start treatment
1 or 2 Cough monitor(s) assigned per facility
To observe all staff daily; and in case of cough encourage to report to above screening person
May be same person as above

HCW indicators

Collect number (#) and percentage (%) for the following indicators:
HCW who were TB screened
Cases of active TB
HCW with active TB disease placed on TB treatment
Cases of drug-resistant TB
TB related mortality
Four provisional core indicators
1. The percentage of HCWs who had a documented TB screening according to national and/or institutional screening algorithms/guidelines in the past 12 months
Indicator #2
2. The percentage of HCW TB cases placed on TB treatment consistent with national guidelines out of all registered TB patients in the past 12 months.
Indicator #3
3. The number of TB cases (all forms) among HCWs during the past year divided by the total
Number of registered HCWs (mid- or end-year population)

Indicator #4

4. The total number of TB deaths per year among HCWs divided by the total number of HCWs (mid- or end-year population in the past year)

(We will discuss the feasibility of this in the M&E of mortality session)

UN data Security protocols

Ensure confidentiality -2

Laboratory results will be anonymized and recorded in a provisional HCW TB register by the facilities

All forms containing personal identifiers, such as names, addresses, and telephone numbers will be kept confidential

Each HCW is assigned a unique and personal ID number; that consists of facility number and individual number

Ensure confidentiality -3

One form will link the name and address of the HCW with the unique ID number to allow for identification of the participant if follow-up activities are required.

Unique Identifiers

> Programs should consider the use of a unified unique identifier that combines patient and facility characteristics such as a 17 digit alphanumeric code that makes duplication unlikely and deductive disclosure very difficult.
>
> *Facility code(3 digits)+ BMU code(3 digits)+date of birth(8 digits)+years of education (2 digits)*

A randomly generated Unique identifier is preferable in cases where stigma is a major issue and people are well known to each other. However, an advantage of a non-computer generated ID is that patients who access multiple health facilities will be assigned the same code if the Unique Identifier is based upon immutable patient characteristics.

This one form/database will be kept in a lockable cabinet by a person trusted by HCW

Summary reports (without names) will be developed by facilities, with assistance from DMO office and FHI360 for the duration of the project

Data tools

The proposed screening forms consist of

2 questionnaires,

- one for all HCW;
- one only for those with symptoms: TB suspects

3 registers:

- Register to keep track of those who show up for screening (has link name and ID number)

- Register of TB suspects
- Register of HIV data

Summary form with indicators

Forms for the intervention- 1

Form 1: questionnaire for screening of HCW for TB
This form should be filled for every HCW that shows up for screening; it is needed to assess risk of TB and give proper care to HCW
It uses ID number from form 3

Forms for the intervention - 2

Form 2: HCW TB suspect form
This form should be filled for HCW who are TB suspects; in order to make sure all results are collected
Ensure that if TB is detected, treatment is started and completed
It uses ID number from form 3

Registers

Form 3: Register of TB and HIV screening of health facility staff
List of all staff that should be provided by head of facility, then be kept by screening person
Objective: to keep track of who showed up for screening and who did not; and encourage those that did not come; and keep list of reasons for refusal

Register of TB suspects among Health Facility staff (form 4):

This anonymous register is to summarize data about TB suspects and patients among HCW
It uses ID number from form 2
This is to have an overview of data to be collected and analyzed where gaps in the screening system may occur

Register to summarize HIV data among HCW (form 5)

This is to summarize HIV data from form 1 for the 3 demonstration project indicators
This is kept separate from TB registers since
It cannot have names
It is needed for all HCW (not only suspects as form 4)

Register to summarize HIV data among HCW (form 5)

This is to summarize HIV data from form 1 for the 3 demonstration project indicators
This is kept separate from TB registers since
It cannot have names
It is needed for all HCW (not only suspects as form 4)

Summary register of TB and HIV among (form 6)

This is an anonymous register to summarize the screening results from forms 1, 2 3 and 4; including the indicators
Any TB patient should also be written in TB treatment register at own facility or other facility
Data collection: location
Screening register, TB suspect forms and diagnostic test results to be kept by own facility
Summary forms to be collected by DMO staff from facilities and NCH staff clinic

Flow of forms

Thank you

Acknowledgements:
Max Meis, Suzanne Verver, Daniel Chertob

Discussion points for debriefing:

1. Measurement of TB infection is considered the most robust measure of transmission in a health care setting. Describe two pros and cons to measuring latent TB infection in health care workers.
2. What three risks to data quality might we expect when screening health care workers? How might these be minimized in the collection and analysis processes?
3. Why does the incidence of TB in health care workers need to be adjusted for age and gender? What is meant by adjustment? What could happen if these data were not adjusted?
4. Any questions on forms?
5. What kind of data quality challenges do you foresee?

Multiple choice questions for assessing comprehension:

1. Why do we consider HCW a priority population for TB screening?

a. They are a vital resource for every country.
b. They work in a congregate setting.
c. They work with vulnerable populations.
d. They have a right to work in a safe environment.
e. All of the above

2. If there are 40,000 health workers in Maravilha, 10,000 are screened each year, and 200 cases of TB are found, what is the number needed to screen (NNS) to find 1 case of active TB among the health care workers?
   a. 100
   b. 50
   c. 200
   d. 1000

Session 20: Screening M&E Skills Practice with Electronic Data

Objectives:

1. To integrate a lot of the skills and concepts learned in the preceding sessions.
    a. Cleaning data
    b. Preparing datasets for merging
    c. Merging data sets with Unique IDs
    d. Deriving yield
2. To practice working with screening data to derive numerators and denominators

Pre-Requisites:

1. Active participation in session 9 on merging

Background Reading:

1. Understanding and Using Tuberculosis Data WHO 2014. Ch. 1 Analysis of aggregated TB notification data. www.who.int

Time	Content	Methodology	Materials
60 min	Integrating cleaning, merging and analytic steps	Hands on skills practice	2. TB patient register data 3. Presumptive TB register data

Home work assignment: None

Exercises with Sample data

Using the dataset from the Maravilha
Compare the screening register with the TB patient register to calculate the overall yield of TB screening.
Calculate the NNS for the 2 risk groups--

- Looking at the number of contacts identified per case, do you notice any patterns?
- What can you conclude about the quality of these CI data?

Discussion points for the debriefing:

1. Effective contact investigation relies on TB patients disclosing the names and addresses of people they live and spend time with.

a. Looking at the number of contacts identified per case by risk group. Do you notice any patterns?
b. What might you infer about the data quality of these CI data?
c. Which risks to data quality might we find in contact investigation data?

Handout Exercise 1:

Step 1: Open the TB screening register data and the TB patient register data.

Step 2: Clean the data in the following ways:

Clean/prepare presumptiveTBRegister.xlsx

6. Open worksheet in Excel

7. Sort worksheet by Date of TB Screening (oldest to newest)
    a. Delete first record due to invalid and missing data

8. Sort worksheet by Region (largest to smallest)
    a. Re-enter Region in rows two through four

9. Sort worksheet by Suspect Number (largest to smallest)
    a. Delete row with Suspect Number equal to all 9s, i.e., 9999999999

10. Save cleaned file, i.e., SuspectRegister-clean.txt (n=17,668)

Step 3: Merge the presumptive TB and TB patient data registers using Excel vlook up or EPI-Info 3, matching either with a UNIQUE ID or with another set of variables, whichever is easier.

Exercise 2: The goal of this exercise is to generate the yield of TB screening by risk group. There are seven risk groups. In the data set, each risk group is denoted by a number.

1. Compare the screening register to calculate the overall yield of all types of TB screening.

2. Calculate the NNS for the following risk groups:
    1. HIV positive persons = 1
    2. Miners = 2
    3. Contacts of a smear positive case = 3
    4. Health care workers = 4
    5. Children under 5 = 5
    6. Sex workers = 6
    7. Health care workers = 7

Categories 8 and 9 are non-risk groups.

3. Looking at the data, do you notice any patterns? What can you conclude about the quality of these screening data?

Discuss in small groups how you would monitor and evaluate a contact investigation.

- What would be the main ways for measuring process, outcome, and impact?
- What might be some indicators that you would include?
- Compare the results with the indicators below. How similar or different are they?

M&E of TB Screening and Contact Investigation

Some potential indicators to consider in a screening program include:

Input Indicators

1. Total number/proportion of nurses in the district.
2. Number/proportion of nurses providing TB services.
3. Number/proportion of registered nurses.
4. Is there any nurse at this facility who received focused training on TB contact investigation?
5. Number/proportion of TB field promoters serving this facility's catchment population.
6. Has this facility received any telephone call from an external supervisor regarding a TB contact investigation?
7. How many visits have been conducted for the TB contact investigation, excluding visits by data collectors or interviewers for this study?
8. Is a copy of the National Guidelines on the Management of TB 3rd edition (2012) available and visible?
9. Are blank copies of the revised TB contact tracing slip available?
10. Are blank copies of the revised TB treatment card available?
11. Is the revised Facility TB register with contact tracing columns available?
12. Total number/proportion of patients registered in the facility TB register in the last two calendar months.
13. From (date).
14. To (date).
15. Number/proportion of patients with smear positive TB.

Process Indicators

1. Number/proportion of patients registered in the period, with documented contacts identified in that register.
2. Number/proportion of index patients with contacts documented or identified in the treatment card.
3. Number/proportion of index patients with documented contacts traced or investigated in the treatment card.

4. Total number/proportion of contacts investigated using the contact investigation slip (count completed slips).
5. Total number/proportion of contacts identified as TB suspects (yes to any screening questions).
6. Total number/proportion of contacts identified as being less than 5 years of age.
7. Total number/proportion of contacts identified as HIV positive.
8. Total number/proportion of contacts identified as having an immunosuppressive condition other than HIV, including diabetes mellitus.
9. Number/proportion of index patients who are:
    - Under 5 years of age
    - Between 5-15 years of age
    - Over 15 years of age.

Outcome Indicators

1. Percentage of index TB patients with any contacts screened for TB, according to the prioritization plan.
2. Percentage of contacts that have symptoms suggestive of TB, among all contacts screened.
3. Percentage of contacts that undergo proper diagnosis for active TB, among all contacts identified.
4. Percentage of those screened that are identified as having active tuberculosis.
5. Number needed to screen.

Impact Indicators

1. Trends in ratio of notification rates in children and adults, i.e., trends in the number/proportion of cases in children under 15 years of age (recent trans).
2. Trends in cluster size, smaller clusters, and DNA fingerprinting.

Session 21: M&E of Programmatic Management of MDR (PMDT)

Objectives:

1. Review the new WHO suggested forms/templates and identify any associated challenges.
2. Understand the complexities of monitoring and reporting on PMDT cohorts.
3. Review the TB CARE I interim outcome tracking tool.

Background reading(s):

1. *The companion handbook to the 2011 WHO guidelines for the programmatic management of drug-resistant tuberculosis*.

Time	Content	Methodology	Materials
10 min	Identify and discuss the participants' current PMDT recording and reporting challenges	Guided discussion/interactive	None
20 min	Discuss the new recommended reporting framework for PMDT (including WHO forms)	PowerPoint presentation/ interactive discussion	PowerPoint/paper copy of draft WHO forms
20 min	• Draw a diagram illustrating the PMDT reporting framework for one three-month cohort • discuss participants' diagrams and the 'gold standard'	Skills-building exercise	Paper, markers, pens
10 min	Review the TB CARE I-developed tool to assist with PMDT interim outcome R&R	Guided discussion/interactive	TB CARE I-developed PMDT R&R tool

Homework Assignment: None

Session Notes

Overview

Brainstorm: What are your current challenges with PMDT M&E?

Elements of PMDT M&E
Complexities of PMDT cohorts
Review the new suggested forms/templates
TB CARE I tool to help with tracking of interim/final outcomes
PMDT Challenges
Low treatment success
Prevent emergence of XDR-TB!!
Further scale-up of diagnosis of MDR-TB

- Sputum transportation
- Laboratory feedback to clinician
- GeneXpert MTB/RIF expansion and C/DST
- Minimize cascade between Dx and Rx
- Expand treatment capacity
- Ambulatory Rx: admission short and only on indication
- Patient-tailored support during full treatment period

TB-IC: ➔ FAST

M&E
Close supervision during full treatment period
Closer monitoring of interim results!?

Resources
DISCLAIMER!!!

The "*Companion handbook to the 2011 WHO guidelines for the programmatic management of drug-resistant tuberculosis*" is due out in 2014.

DISCUSSION
What are your current Challenges with PMDT recording & Reporting?
What are we measuring in PMDT?

- Detection
- Enrollment
- Interim results
- Final outcomes

Objective: <u>minimum</u> indicators for national or project level monitoring and suitable for different partners (WHO, GF)
easily extracted manually or electronically
conform to what was used in past and DOTS system
Focus on indicators rather than forms

Detection – Form 05

Let's create an MDR cohort

Draw a timeline to explain the reporting timeframes for:

1. Detection
2. Enrolment
3. Interim results
4. Final outcomes

TB CARE I interim results tool

Discussion points for debriefing

1. Why is the tracking of interim results so important for RR-TB patients?

ANSWER: The treatment period is so long, you need to know how treatment is going before it is too late. It gives you an opportunity to adjust your PMDT approach or an individual's treatment/care mid-course, instead of waiting until the end. This can improve final outcomes and help to prevent XDR-TB from developing.

2. Why are interim results reported nine months after the end of the quarter?

ANSWER: A three-month cohort (Jan-Mar 2011) should be monitored after six months of treatment (after Oct 2011). Allowing for three months to collect the data, this report would be due nine months after the end of the quarter (Jan 1, 2012).

Two **Multiple choice questions for assessing comprehension:**

1. Which one of the following elements of PMDT R&R is not a part of basic TB R&R?
    a. Detection
    b. Enrollment
    c. Interim results
    d. Final outcomes

2. **True** or False: The new *Companion handbook to the 2011 WHO guidelines for the programmatic management of drug-resistant tuberculosis* will include guidance on RR-TB reporting to align with the new *Definitions and reporting framework for tuberculosis – 2013 revision*.

Theme 4: Fix Problems

Session 22: How to Cope with Poor Quality Data

Objectives:

1. To derive some approaches to triaging inaccurate, inaccessible, insecure, imprecise, and missing data.
2. To empower learners to make good decisions about how to handle missing data by exploring the guiding principles, the types of missing data, and how the underlying causes should drive the triage plan.
3. To teach one imperfect method of inputting missing data, the hot deck.

Pre-requisites: Elements of Data Quality and Intro to Epi-Info.

Background Preparation to be completed before the session:

1. *EPI-Info 7 User Manual Data Quality Appendix 2012.*

2. **Watch: 3:55 Epi Info 7 2x2 Table Analysis**
http://www.youtube.com/watch?v=8BZZvU4zy_c

Time	Content	Methodology	Materials
12 min	Overview of inaccurate data	Interactive/didactic	Power point slides
15 min	Overview of missing data issues	Interactive/didactic	Power point slides
20 min	Missing data Scenario 1	Discussion in groups of 3	Scenario 1 slide Big newsprint
15 min	Attempt to fix missing data with a HOT DECK	Hands on computer exercise	Maravilha dataset Epi_Info 7 or Excel
10 min	Other types of bad data	Interactive/didactic	Power point slides

Session Notes:

Objectives:

1. To derive some approaches to triaging inaccurate, inaccessible, insecure, imprecise, and missing data
2. To empower learners to make good decisions
3. To teach one imperfect method of imputing missing data – hot deck

What can be done?

Documenting changes to a data set

Do not work on the original data set
Always create a unique version that you will clean, in case you make a mistake.

Inaccurate data

Examples of inaccurate data that you are coping with?

Inaccurate data

- What can be done?
- Why is it inaccurate?

Inaccurate data

Examples of inaccurate data that you are coping with?

Missing/Incomplete Data

- When do we encounter data missing?
- What do you currently do to cope with missing data?

Key elements of "missingness"

Handling missing data requires knowledge

2 Most important issues:

1. **WHY ARE THE DATA MISSING?**

2. **HOW DO THE MISSING DATA RELATE TO OTHER VARIABLES?**

3 categories of missing data

- Missing completely at Random (MC@R)-
- Missing 'at random' (M@R)-
- Missing not at random (MN@R)-

There is software that will run a "Little's Test" to tell you which you have.

Reference: R. Barbour (2013) Making Statistics Accessible: Approaches to Missing Data
http://www.youtube.com/watch?v=82hDeiG2D-c
What would be some TB examples of each type of missing data?

Missing completely at Random (MC@R)

=You can be 100% confident that the absence of data has no rhyme or reason to it
The best kind
As long as you don't have a lot of missing data (<5%), you can feel not very guilty about just dropping these cases

Missing at random (M@R)

= whether or not you have the data depends on some information that you do possess (e.g. age, gender,) i.e. the likelihood of missing data can be predicted.
Most (>80%) missing data are this variety
You need to consider finding stats help to see if you try to get the data, use non-response weighting, or use single or multiple imputation,

Missing not at random (MN@R)

=not having the data is related to information you also don't have
The worst kind
This kind can really bias your results
There is little that can be done to fix this situation except going out to get the missing data

Other inputs for decision making about Missing data

How critical are the data?
How much are missing?
Can the missing data be obtained?

Handling missing variables in a case

These approaches are almost always bad:

Ignore missing data by only focusing on complete cases

Mean imputation (e.g. just stick in the mean value from the whole sample)
Regression prediction-single imputation
"last observation carried forward" i.e. take the value from the last visit and paste it into the missing visit.

The Gold Standards:

Find a Statistician who can help you to perform:

Multiple Imputation Methods

1. Using all the data to help guess the missing values

   E.g." Amelia" Package in R

2. Non-response Weighting

   Figuring out how reduce the impact
   E.g. Sudaan in SAS, pweight in STATA

" Silver" Standard Options

If missing data are < 10% and you are sure you know and have all other variables that predict the missing value, you might try to impute using a " Hot Deck" approach–
Find all the cases with the same characteristics as the case with missing variables.
Randomly pick one of the cases, and paste in that value to replace the missing value

Example of a simple hot deck

Scenario 1

In Maravilha, there is a rural TB sanatorium for children under 5. When the pediatrician goes on holiday, no one collects samples and so all the smear or culture results for new patients starting treatment are missing (±4 patients/wk for 3 wks).

DISCUSS IN 3s and draft a short response to present to the group:

1. What (if anything) could be done?
2. How essential are these data?
3. How much data are missing?
4. Can it be imputed? inferred? triangulated?
5. What is your team's recommendation for this situation?

Find Missing Data

Using the SELECT command, this technique allows you to find missing records for specific fields or variables.

How to:
READ {C:\Epi Info 7\Projects\maravilha
SELECT treatmentoutcome = (.)
LIST * GRIDTABLE

Classical Analysis in Epi-Info 7

Include Missing Values feature allows you to determine if you want missing values to be included in statistical calculations or not
Inconsistent Data

What are some examples of data that you consider to be "inconsistent" ?
Examples of inconsistent data

HIV negative TB patients who are reported to be taking ARVs
Solutions for inconsistent data?

Triangulate to try to find the "truth"
Recode all inconsistent as *missing*
Imprecise (or overly precise) Data

Examples?
Old forms collapsed 0-14 yrs
Date-time stamps
Some solutions for imprecise or irrelevant data

converting open-ended questions into closed-ended questions
Disaggregation of a compound variable, such as a combined date and time variable (time stamp).
Aggregation of text via Wild card searching
Inappropriately presented

Provided as 'string' (text) when you need it to be numeric in order to do calculations
Epi-Info 7 uses 'fuzzy' searches

Searches are not case sensitive and are designed to be more inclusive than exclusive.
Spelling variations are automatically accommodated
Wild card searching

Enhance Searches with Find
When searching a numeric or date field, the accepted search format for the field is displayed to the right of the entry field.

Embedded text items in multiline and text fields can be found by searching for *word* where "word" is the text string being sought. This type of search is called a Wild Card search.
In a Wild Card search, the asterisk represents any letter or string of letters (i.e., a search for DIAGNOSES *pulm* would identify all records for a large text field called DIAGNOSES which contained the letters “pulm)."

Biased data & the need to show positive results

Unbelievable data

What do you do when the data are too good to be true?

- Too clean.
- Too complete.
- Too high performing.

Discussion

What are some of the ways to address pressure to meet targets without sacrificing data quality?
What can we do with data that are simply unbelievable?

thanks

Discussion points for debriefing:

1. What are some potential problems we can have if I de-duplify in Excel? How can they be prevented by using other software?
2. Why is it problematic to simply exclude those cases with missing treatment outcome data from a cohort analysis?

Multiple choice questions for assessing comprehension:

1. Missing TB data on risk behavior such as injection drug use due to participant non-response is very unlikely to be which type of missing data:
   a. Missing completely at random
   b. Missing at random
   c. Missing not at random

2. To determine what to do about missing data, the following information is needed:
   a. When is the report due?
   b. Who will be reading the report?
   c. How much data are missing?
   d. Whether the data are paper or electronic

Session 23: How to Link Datasets When There are No Unique IDs

Objectives:

1. Introduce concept of record linkage using non-unique data identifiers.
2. Compare deterministic and probabilistic linking methods.
3. Describe Link Plus software.
4. Demonstrate the use of Link Plus software for de-duplicating and linking datasets.

Back ground Reading:

1. LINK PLUS instructions (Link Plus 2.0 requires administrator privileges for installation, so it is essential that you work with your systems administrator to install the program prior to leaving your workplace): http://www.cdc.gov/cancer/npcr/tools/registryplus/lp_tech_info.htm

Time	Content	Methodology	Materials
50 min	Overview of record linkage methods	Guided discussion/interactive	Power point
10 min	Overview of Link Plus software		
30 min	Demonstration of Link Plus software		Link Plus software

Homework assignment:
Several methods exist to increase the ability of record linkage to find related records. List them and distinguish their functions.

Session Notes:

Record Linkage with LINKPLUS software

Objectives

In a previous session, you learned how to match, append, and link datasets when there is a unique identifier, in this session you will
Grasp the 2 main approaches to data linking
Learn one software of linking data (LINKPLUS)

Introduction

3 main processes exist
Manual/Clerical matching of data

Labor intensive, not always feasible, error prone, slow
Deterministic
Probabilistic

Record linkage methods

Probabilistic record linkage methods can be "trained" to perform well with much less human intervention

Deterministic record-linkage

Rules-based record linkage
Binary result, i.e. match or no match
Used to examine vital status of individual patients
Used if identifier(s) in different records identical
Hierarchical rules give better control over specificity of matches

Deterministic record-linkage

Based on personal identifier:
Unique
Universal
Permanent
Accurate
Reasonable
Simple
Known

Deterministic record linkage

A small decrease in data quality or small increase in the complexity of the data can result in a very large increase in the number of rules necessary to link records properly. Eventually, these linkage rules will become too numerous and interrelated to build without the aid of specialized software tools.

Deterministic record-linkage

Linkage rules are often specific to the nature of the data sets they are designed to link together
New data exhibiting different characteristics than was initially expected may require a complete rebuilding of the record linkage rule set, which could be very time-consuming and expensive.

Probabilistic record-linkage

Fuzzy merging/matching
Determines probability of a match
Used for the purpose of studying popn-based characteristics
Unique identifiers unavailable
Uncertain in nature therefore should only be used in this case

Probabilistic record-linkage

Based on a score reflecting probability (P) records relate to same entity
Based on comparison of a wider range of potential identifiers/variables & calculating a maximum likelihood estimator to give a score for similarity btw records

Probabilistic record-linkage

Compute weights for each identifier based on its estimated ability to correctly identify a match or a non-match,
Use weights to calculate probability (P) that 2 given records refer to the same entity.
MATCH: records pairs with probabilities above a certain threshold
NON-MATCHES: pairs with P below another threshold
POSSIBLE MATCHES: P that exists btw these 2 thresholds

Probabilistic record linkage

Record linkage methods

Choice between 2 methods depends on characteristics of datasets:

Several methods exist to increase the ability of record-linkage to find related records. List them & distinguish their functions.

Selecting matching variables

Selected variables must be suitable
Allow for records to be matched but also have the ability to discriminate btw diff records
e.g. a comparison of 2 different records containing same last name has greater discriminating power if the name is rare

Selecting matching variables

Use of probability matching depends on variable: “name”, “DOB”, “name of mother” and “address”
(word of caution if one cannot re-arrange address in pre-processing phase, it should not be used as matching variable)

Grouping/Blocking

Theoretically 1 record in 1st file to be compared with every record in 2nd file
So, combination of records & time need to be searched increases quadratically

Grouping/Blocking

Blocking/indexing techniques are used to reduce no of compared with & only within these blocks are records compared between files
E.g. use district codes (remember people move); DOB, last names; sex: good code but get 2 groups, too numerous

Searching and Scoring
Used in probabilistic process
Computer based
Very important step: core of linkage process
Computer searches for probable pairs of records, calculates P of being same person
Basically, objective is find matches

Searching and Scoring

Realistically, not possible to know which comparisons are matches and non-matches
So, combination (matches & non-matches) at any given total weight score is given
2 cut-off weights are set
> cut – off weight= LINK
< cut- off weight+ NON- LINK
Any record between two cut-offs is manually reviewed

Review Process

Intuition & intrinsic knowledge of data needed
Access to additional variables not used in search necessary

Do not delete duplicate record, should be marked and kept on file for further re-assessment
Whole process of linkage exercise should be accounted by full documentation

Software

- SAS
- Link King/Link Plus/ Link Pro
- Access
- CODES 2000/LinkSolv
- Other
- FEBRL/Integrity/home

Pricing
Free to USD

Link Plus Introduction
Software developed at CDC by a statistician
Statistical specifications based on research in the published literature
Tested by researchers experienced in record-linkage
Originally designed for use by cancer registries, but can be used with any type of data in *fixed width* or *delimited* format
Windows interface that includes help and samples

Link Plus Features

Handles missing values of matching variables by treating null or empty values as missing data automatically
and allows the user to indicate additional values to treat as missing data
Facilitates a simple and efficient blocking mechanism by
indexing the variables for blocking
comparing pairs with identical values on at least one variable

Link Plus Matching Methods

Value-specific (frequency-based): Sets weights for matching values based on the frequencies of values in the files being compared. A match on a frequent value is associated with a low weight, but a match on a rare value is associated with a high weight.
Last name and first name: Incorporates both partial matching and value-specific matching and NYSIIS phonetic code to account for minor typographical errors,

misspellings, and hyphenated names. For first names, nicknames are matched with formal names.

Link Plus Matching Methods

Date: Incorporates partial matching on separate date components, and accounts for transposition of date components, as well as missing month or day values.
Generic string: Uses an edit distance function and incorporates partial matching to account for typographical errors.

Link Plus Modes

Two modes

- Detect duplicates in a single dataset
- Link records from two datasets

Deduplication

Records in the same file are blocked, compared, and scored against each other
The result is a ranked list of record pairs
High-scoring pairs may be duplicate

Linking Records

Find the records in File A that seem to match records in File B
Calculate a score that indicates, for any pair of records, how likely it is that they both refer to the same person
Discard unlikely matched pairs (low scores)
Sort the likely and possible matched pairs in order of their scores
Visually review a range of uncertain matches

Link Plus Demo

Discussion points for debriefing:

1. Describe the main differences between deterministic and probabilistic linking methods.
2. Describe a scenario in which you would use probabilistic linking.

Multiple choice questions for assessing comprehension:

1. The key phases in the record linkage process are:
    a. Data cleaning, creating unique identifiers, and selecting matching variables.
    b. Pre-processing, selecting matching variables, and manual scoring.
    c. Pre-processing, decision-making, grouping, searching/scoring, and reviewing results manually.

d. Selecting matching variables, grouping, and manual scoring.

2. Please select the correct statement below:
    a. Link Plus is a deterministic record linkage program developed by the CDC.
    b. Link Plus can run in two modes: de-duplication and record linkage.
    c. Comma delimited files are the required format for Link Plus.
    d. Link Plus saves unique versions of the Linkage and Non-matches report files automatically after each run.

Session 24: Skills Practice with Link Plus Software

Objectives:

1. Practice de-duplication using Link Plus with sample TB data.
2. Practice linking records using Link Plus with sample TB data.

Time	Content	Methodology	Materials
20 min	Skills practice with Maravilha dataset.	Hands-on use of Link Plus De-duplication function	• Link Plus software • Screening data set
40 min	Skills practice with Maravilha dataset.	Hands-on use of Link Plus Record Linkage function	• Link Plus software • Screening data set • TB register data set

Homework assignment:

Go into Link Plus and reopen the Link Plus configuration file created during the de-duplication skills practice. Add/change blocking and matching variables and note the effect on the results.

Discussion points for debriefing:

1. How does this type of linking differ from the linking we did in session 10 (unique IDs)?
2. What are the main steps in linking using Link Plus?

Multiple choice questions for assessing comprehension:

1. What is a blocking variable in Link Plus?

    a. A means of pre-selecting key variables to reduce the number of comparisons.
    b. A confounder.
    c. An obstacle to assessing data quality.
    d. A sampling strategy.

Theme 5: M&E as Collaboration

Session 25: Data Are Human – The politics and Practice of TB Data Exchange

Objectives:

1. To explore and improve upon the politics and practices of TB data exchange among partners.
2. To discuss best practices for the ethical and safe sharing of TB data.
3. To improve data exchange negotiation skills.

Background Preparation to be completed before the session:

1. FHI research ethics curriculum research ethics online course:
   http://www.fhi360.org/training/en/RETC2/index.html
   http://www.fhi.org/en/rh/training/trainmat/ethicscurr/index.htm

Time	Content	Methodology	Materials
4min	• Safe and ethical sharing of patient information among institutions. • Packaging data for sharing.	Video	Epi-Info 7 packaging data for sharing videos
1 hr	Working in teams, participants role play the negotiation of TB notification data.	Role playing	Nine scenarios list

Homework assignment:

Practice negotiating a data exchange with your colleagues.

Discussion points for debriefing:

1. You have calculated a notified incidence of 267/100,000 for 2012. A senior manager at the NTP tells you that the notified incidence has to be above 300/100,000 or the program may lose donor funding. What options might you consider in this situation?

2. What does the term deductive disclosure mean?
3. What is a data transfer agreement, and what does it entail?

Multiple choice questions for assessing comprehension:

1. When sharing TB data, what ethical principles should be kept in mind?
   a. Justice
   b. Non-Malfeasance
   c. Respect for persons
   d. Generalizability

2. How can deductive disclosure be prevented?
    a. Remove all identifying information from a data set.
    b. Keep linking tables under lock and key.
    c. Use password protected data bases and computers.
    d. All of the above

Handout

Data Sharing Role Plays

Instructions:

Working in teams, develop a six minute role play based on one of the scenarios described below. In the role play, articulate the reluctance to share data. The NTP Manager must try to reassure the partner about the data management and data security/sharing practices of the TB program. In a four minute dialog, present the concerns and potential solutions, articulating how the use of high data quality standards and best practices can facilitate data sharing. Try to convince the resistant party to share the TB data.

1. The Diabetes Program: The manager sees TB reporting as an additional burden for her staff. She has suggested that she may be willing to report on the patients diagnosed through the new TB screening program if the TB clinics are also willing to institute a new diabetes screening program among TB patients. The diabetes program is new to TB diagnosis and is worried that they may be over- or under-diagnosing TB, and they prefer to keep their records internal until they have more experience.

2. The Mining Company: This company does not want to report their TB treatment outcome data because they worry that they will be criticized over their incomplete treatment outcome data. The RR TB patients diagnosed through the mining screening program are later transferred to the public referral hospital, so the Health Director insists there will be double counting if he reports. Discuss with the Mining Health Director to convince him of the need for data exchange.

3. The National Pediatric Referral Hospital: The hospital is well regarded, and their diagnostic and treatment facilities are first rate. However, they do not regard the national program seriously because they find the pediatric contact investigations to be sub-standard. They do not see how their data would be compatible with the national surveillance data because they use more sophisticated diagnostics, including induced sputa, gastric aspirates, and liquid culture on all children. They collect many more variable and they find the recording and reporting system of the national TB program too simplistic. Discuss with the Chief of Pediatrics to convince him of the need for data exchange.

4. The Private Sector: This sector does not want to share data because they are concerned that there will be undue scrutiny of their diagnostic practices. Some practitioners diagnose using multiple and unnecessary examinations.

5. The HIV Program: This program is concerned with sharing the names of HIV positive patients. The program manager believes that HIV positive clients face special threats of discrimination and ostracism, and therefore should be exempt from routine reporting. Moreover, the country's HIV law has harsh penalties for deductive disclosure, and she does not want to be sued if the TB program's

security is lax. Discuss with the HIV Program Manager how you might address her concerns.

6. The Indigenous Health Service: This service was formed because ethnic minority groups were not receiving good treatment in the problem sector. The leadership is wary of data collection because of past ethnic conflict in which lists were used for committing human rights violations. Role play a discussion between the Chief of the Indigenous Health Service and the TB Program Manager to resolve the issue.

7. The NGO serving injection drug users: This NGO has a TB screening program, and they contribute TB data on those who are diagnosed and the number of people screened. However, they do not differentiate between persons screened repeatedly and unique individuals. They do not collect names. Role play a discussion with the head of the NGO and the TB Manager to resolve the issue of data exchange.

8. The Ministry of Finance: The Ministry of Finance receives money from the Global Fund and keeps data on the costs of TB services offered through prisons and the HIV program, but they do not like to report the financial figures to the TB program for reasons that are unclear to the TB Manager. It may be technically difficult to determine the costs of a health worker who treats many diseases, but the TB Manager suspects that there may be other issues that explain why financial data are not shared. The TB program needs to know how much all TB services cost in order to prioritize strategically. Role play a discussion between the Chief Financial Director for Health and the TB Manager to resolve the issue.

Session 26: M&E for TB/HIV Data Integration

Objectives:

1. To show concrete examples of TB/HIV data verification and how it changes policy.
2. To understand visual benchmarking as a means of identifying data quality problems.
3. To conceptualize the problem of double-counting generally, and in TB and HIV service delivery specifically.

Background Preparation to be completed before the session:

1. Brouwer et al. 2013. *Benchmarking to Assess Potential Under-Diagnosis of Smear-Negative and Extrapulmonary Tuberculosis. A Case Study from Mozambique:* http://benthamscience.com/open/toidj/articles/V007/1TOIDJ.pdf

2. Brouwer et. al. . 2013. *Are routine tuberculosis programme data suitable to report on antiretroviral therapy use of HIV-infected tuberculosis patients*? BMC Research Notes 2013, 6:23: www.biomedcentral.com/1756-0500/6/23

Time	Content	Methodology	Materials
45min	Introduction to benchmarking and examples of ecological approaches to TB/HIV	Presentation	Power point

Homework assignment: None

Session Notes

Objectives

- To show some examples of TB/HIV verification
- Describe visual benchmarking as a means of identifying data quality problems
- To discuss double-counting
- To de-duplify TB/HIV data in Epi Info 7 to address double counting

Data Quality implies looking not only at the completeness of individual columns and rows, but also looking at data ***in relation to each other***.
Benchmarking..
"Ecological" comparison of 2 values
Similar to histograms for "outliers" but with 2 or more variables
What do you see in this graph?
Benchmarking
Operational issues—discuss in 3s
How important are those dates of ART/CPT start for M&E?

What are the possible M&E implications of reporting ART/CPT status by quarter instead of at the end?

Discussion points for debriefing:

1. Name three types of double counting and how they differ.
2. Is it important to have start dates of ART (antiretroviral treatment)/cotrimoxizole (CPT) in the TB register? Why or why not?
3. What are the possible M&E implications of reporting ART/CPT status by quarter instead of at the end?

Multiple choice questions to assess comprehension:

1. What is benchmarking? *(*Choose one*)*
   a. A means of identifying outliers through ecological comparison.
   b. A method for triangulation of qualitative data.
   c. A method of verifying the accuracy of data.

2. What are some potential pitfalls of de-duplifying TB data using Excel?
   a. No syntax to re-run if needed.
   b. No audit trail.
   c. Duplicates are removed in order.
   d. All of the above.

Session 27: Assessing Under-Reporting through Inventory Studies

Objectives:

1. To familiarize learners with the strengths and weaknesses of four techniques for measuring undercounting.

Pre-requisites: Sessions 9, 10, 25, and 26 on linking datasets

Background Preparation to be completed before the session:

1. *Assessing Under-Reporting through Inventory Studies*: http://www.tbcare1.org/publications/toolbox/tools/hss/Assessing_TB_under-reporting_through_inventory_studies.pdf

Time	Content	Methodology	Materials
10 min	Overview of Under-counting: justification and studies to date	Presentation	8 CDC/WHO slides
5 min	Method 1: ***Prospective Random Sentinel Verification***		2 slides
5 min	Method 2: ***Prospective Mass Verification***		2 slides
5 min	Method 3: ***Retrospective 3D Verification***		4 slides
45 min	3D verification simulation game using Maravilha data	Game	21 cups, labeled candies, and a venn diagram

Homework assignment: None

Session Notes

(i.e. low level of completeness)

What is an inventory study?

A study of the level of under-reporting of TB cases

What is under-reporting?

Dimensions of Data Quality

Why are inventory studies important?

TB incidence is best measured from state-of-the-art TB surveillance systems linked to well performing health systems

High coverage of health/social protection and good diagnostic services -> limited under-diagnosis
High coverage of reporting (including from the private sector) -> limited under-reporting*

Under-reporting in India

3 possible aims

Quantify the level of under-reporting of diagnosed cases of TB to national surveillance systems
Demonstrate under-reporting is minimal
Estimate TB incidence using capture-recapture methods, if applicable

Where should inventory studies be done?

Inventory studies are especially helpful in countries with

High TB burden
Robust TB surveillance system not yet in place
High utilization of private providers and/or private sector / hospitals / facilities suspected to not report TB cases
How can inventory studies help strengthen surveillance?

Inventory studies can help engage private providers

Measuring levels of under-reporting can

identify gaps to target available resources
be used to improve estimates of TB incidence
identify countries where TB incidence can be measured directly from surveillance data

Recent inventory studies

Capture-recapture

Netherlands
UK
Egypt
Syria
Yemen
Iraq

How do inventory studies work?

TB cases detected by health providers are recorded

NTP providers (e.g. TB dispensaries)
General hospitals
Private doctors
Health insurance
…

Match cases in non-NTP list with cases in NTP list
Three main study designs
Prospective
Prospective Random Sentinel Verification
2 data sources
Prospective Mass Verification 3 data sources
Retrospective
Retrospective 3D Verification using existing computerized records
3 data sources

Method 1: *Prospective Random Sentinel Verification*

Quantify under-reporting, *no incidence estimation*
2 data sources

Method 2- *Prospective Mass Verification*

Quantify under-reporting and estimate incidence
At least 3 data sources
Whole country
Very intensive
Simple Random Sampling of Providers (50%)

Method 3- *Retrospective 3D Verification*

Quantify under-reporting *and estimate incidence* using existing computerized records
3 data sources

Sampling

Limitations of capture-recapture methods: Four conditions

1. No change in study population during study
2. Cases can be matched across data sources (i.e. no misclassification)
3. Probability of being included in a data source is the same for all members of the population
4. Data sources are independent

Essential ingredients

Prospective:

Providers outside NTP network can be mapped and convinced to participate

Essential ingredients

Case-based data with reliable personal identifiers
Standard case definitions across all care providers
Capacity in sampling design, data management and data analysis
Adequate staffing and funding
At least three fairly independent data sources, if capture-recapture methods used
Proportion of patients with active TB reported only to the HIV surveillance database

Capture-recapture study in Iraq

Capture Recapture Game

How complete is Maravilha's surveillance data?

Capture-recapture game

3 cups filled with candies with TB patient ID labels,
Working in teams
complete the venn diagram

What proportion of TB cases were under-reported?
Thank you

Acknowledgements:
Emily Bloss (CDC), Phillipe Glaziou (WHO), Rob van Hest (KNCV),
Homework

1. What do you think is the biggest challenge with carrying out an inventory study?
2. Do you think retrospective or prospective methods are most appropriate for your setting, and why?

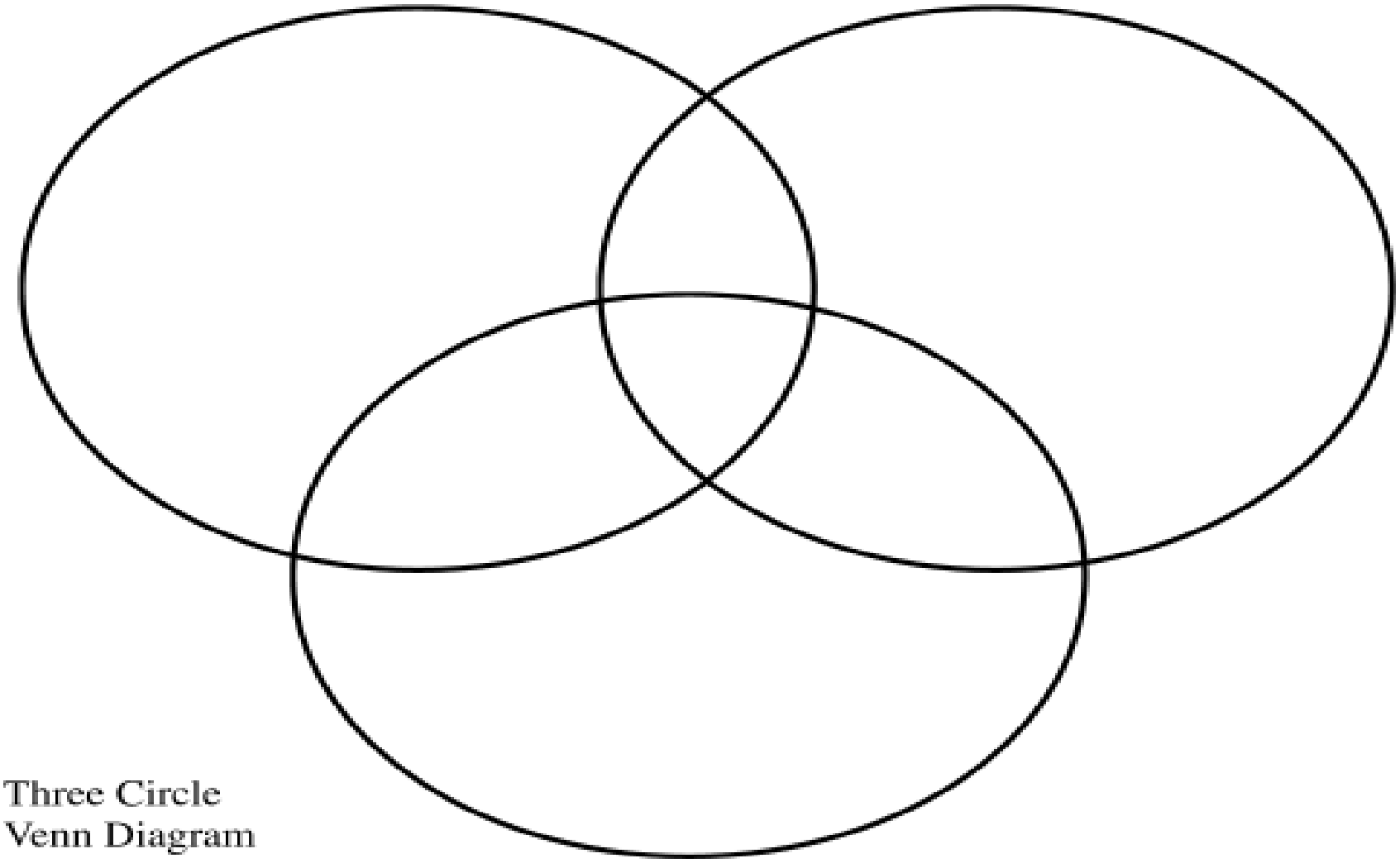
Discussion points for debriefing:

1. What do you think is the biggest challenge in carrying out a prospective inventory study?
2. Do you think a retrospective or prospective method is more appropriate for your setting, and why?

Multiple choice questions for assessing comprehension:

1. What are the pre-requisites for a capture–recapture study? (check all that apply)
    a. Ethical permission to interview TB patients
    b. Unique identifiers
    c. Three independent sources of TB notification data

2. What are the potential benefits of measuring underreporting ?(select one)
    a. Can identify gaps to target available resources
    b. Can help improve estimates of TB incidence
    c. Can identify countries where TB data are so good that incidence can be measured directly from surveillance data
    d. All of the above



Three Circle
Venn Diagram

Answer Sheet for Inventory Studies Simulation Game — 21 Candies

Only source 1: National TB surveillance system		Only source 2: TB Patient support group membership list		Only source 3: Patients' Compensation claims from the Maravilha miners union	
58315	sonomidoso	54204	sofarayufa	81190,0	nododoKeu
67688	latilanono	58045	sonoufaso	81197,0	nododoKeti
67713	latitidomi			57794,0	sotitiKefa

1+2		2+3		1+3		all	
11357	dodomisoti	11044	dodoufafa	10744	doutifafa	10645	doulafaso
38668	minolalano	11075	dodoutiso	10823	dounoraymi	10658	doulasono
38670	minolatiu	11258	dodoraysono	11005	dodouuso	10672	doulatiray
						67723	latitiraymi

Labels for the Candy or items to represent TB patients.
Each color represents and independent data source.

58315 Sonomidoso	67688 Latilanono	67713 Latitidomi	11357 Dodomisoti	38668 Minolalano
38670 Minolatiu	10744 Doutifafa	10823 Dounoraymi	11005 Dodouuso	10645 Doulafaso
10658 Doulasono	10672 Doulatiray	67723 Latitiraymi	54204 Sofarayufa	58045 Sonoufaso
11357 Dodomisoti	38668 Minolalano	38670 Minolatiu	11044 Dodoufafa	11075 Dodoutiso
11258 Dodoraysono	10645 Doulafaso	10658 Doulasono	10672 Doulatiray	67723 Latitiraymi
81190 Nododokeu	81197 Nododoketi	57794 Sotitikefa	11044 Dodoufafa	11075 Dodoutiso
11258 Dodoraysono	10744 Doutifafa	10823 Dounoraymi	11005 Dodouuso	10645 Doulafaso
10658 Doulasono	10672 Doulatiray	67723 Latitiraymi		

Reference Materials

Service Availability and Readiness Assessment (SARA-WHO):
http://www.who.int/healthinfo/systems/sara_introduction/en/index.html

Assessing tuberculosis under-reporting through inventory studies. WHO 2012:
http://www.tbcare1.org/publications/toolbox/tools/hss/Assessing_TB_under-reporting_through_inventory_studies.pdf

Checklist of Standards and Benchmarks for TB Surveillance and Vital Registration Systems.
Compendium of indicators for monitoring and evaluating national tuberculosis programs. WHO, 2004, page 10:
http://www.who.int/tb/publications/tb_compendium_of_indicators/en/

Guide to monitoring and evaluation for collaborative TB/HIV activities. Stop TB Department and Department of HIV/AIDS, WHO 2009.

Data Quality Audit Tool. Guidelines for implementation. MEASURE Evaluation
Data Quality: Standard Operating Procedure. 2012 MEASURE Evaluation.

Definitions and reporting framework for tuberculosis – 2013 revision. WHO 2013:
http://apps.who.int/iris/bitstream/10665/79199/1/9789241505345_eng.pdf

WHO External
M&E Guide to ACSM (PATH/WHO) 2013.

Manual on use of routine data quality assessment (RDQA) tool for TB monitoring. WHO Monitoring and Evaluation for National Tuberculosis Programs- curriculum MEASURE 2005.

Monitoring and evaluation tool for tuberculosis programs in refugee and post-conflict settings. CDC 2010.

Monitoring and Evaluation Toolkit - World Health Organization:
www.who.int/hiv/pub/epidemiology/en/me_toolkit_en.pdf

Routine Data Quality Assessment: Standard Operating Procedure. Botswana MEASURE Evaluation.

Standards and benchmarks for tuberculosis surveillance and vital registration systems:

TB Infection Control Monitoring and Evaluation | I-TECH TB.

TB Program Evaluation Handbook - Centers for Disease Control. The Checklist and User Guide.

Tuberculosis and Malaria Monitoring and Evaluation Toolkit - 4th edition. 2011. The Global Fund to Fight AIDS.

Tuberculosis prevalence surveys: A handbook. WHO 2011. (aka The Lime Book)

ICEBREAKER: BUDDY BINGO

Has 2 or more brothers	Has a bicycle	Has been to Mexico	Plays football	Has a rabbit
Doesn't drink alcohol	Is afraid of spiders	Is left handed	Is younger than you	Is wearing earrings
Is a good dancer	Plays an instrument	**Free space**	Likes to sing	Doesn't like chocolate
Plants a garden	Wishes they were sleeping	Speaks more than 3 languages	Likes onions	Has never been to Kenya before
Is wearing black socks	Has a child	Likes the color pink	Is a twin	Is afraid of heights

Has a cat	Has 2 or more brothers	Plays an instrument	Likes spicy food	Is afraid of spiders
Is left handed	Likes the color orange	Doesn't like cheese	Has run a marathon	Can write in more than 2 languages
Plants a garden	Is wearing a watch	**Free space**	Is afraid of heights	Plays football
Is younger than you	Is wearing black socks	Is a good cook	Wishes they were sleeping	Likes to sing
Has a child	Has been to Mexico	Has never been to Kenya before	Doesn't drink coffee	Is a twin

Plays an instrument	Is wearing earrings	Is afraid of heights	Plants a garden	Is a good cook
Likes to sing	Is a twin	Is afraid of spiders	Has a cat	Has been to Mexico
Speaks more than 3 languages	Plays football	**Free space**	Is left handed	Is wearing black socks
Has 2 phones	Wishes they were taller	Has 2 or more brothers	Has a child	Is younger than you
Likes the color pink	Has run a marathon	Likes spicy food	Doesn't like bananas	Has never been to Kenya before

Doesn't drink coffee	Wishes they were sleeping	Likes to sing	Plays football	Has run a marathon
Is wearing earrings	Is wearing blue socks	Is a good cook	Likes spicy food	Likes the color pink
Has never been to Vietnam before	Is afraid of monkeys	**Free space**	Speaks more than 3 languages	Is left handed
Has been to Mexico	Is afraid of heights	Is younger than you	Plays an instrument	Has a cat
Is a twin	Doesn't like chocolate	Has a child	Has 2 or more brothers	Plants a garden

Speaks more than 3 languages	Is younger than you	Plants a garden	Is a good cook	Has never been to Kenya before
Plays an instrument	Doesn't like chocolate	Has been to Mexico	Is afraid of spiders	Is first born I their family
Has a cat	Is afraid of heights	**Free space**	Wishes they were sleeping	Has run a marathon
Likes spicy food	Has 2 or more brothers	Is wearing black socks	Is left handed	Is wearing earrings
Has a child	Plays football	Likes to sing	Likes the color pink	Doesn't drink tea

Is left handed	Has never been to Kenya before	Has run a marathon	Likes the color pink	Is wearing earrings
Is younger than you are	Plants a garden	Plays an instrument	Doesn't like chocolate	Is a good cook
Likes to sing	Wishes they were sleeping	**Free space**	Plays football	Likes spicy food
Speaks more than 3 languages	Has a cat	Is a twin	Doesn't drink coffee	Is wearing black socks
Is afraid of heights	Is afraid of spiders	Has been to Mexico	Has a child	Has 2 or more brothers

TB M&E Data Quality Glossary

Absorptive capacity	A program's ability to take in, assimilate, and apply new information.
Accessibility	The degree to which data are available to the people who need them.
Appreciative inquiry	An organizational development method which focuses on increasing what an organization does well rather than on eliminating what it does poorly.
Attribution	An ascribed quality, character, or right.
Audit:	A formal examination of an organization's or individual's accounts or financial situation.
Behavior change	A broad range of activities and approaches which focus on the individual, community, and environmental influences on behavior.
Believability	Capable of being believed, especially as within the range of known possibility or probability
Bias	An inclination of temperament to present or hold a partial perspective at the expense of (possibly equally valid) alternatives in reference to objects, people, or groups.
Branding	The promoting of a product or service by identifying it with a particular name, term, design, symbol, or any other feature that identifies the good or service as distinct from those of other sellers.
Capture-recapture	Methods having their antecedents in animal ecology and related to attempts to estimate or adjust for the extent of incomplete ascertainment, using information from the overlapping lists of cases from distinct sources from the affected population.
Case study	An intensive analysis of an individual unit (as a person or community) stressing developmental factors in relation to the environment.
Catchment area	The geographical area served by an institution.
Causal inference	The process of drawing a conclusion about a relationship or causal connection based on the conditions of the occurrence of an effect.
Certification	The confirmation of certain characteristics of an object, person, or organization.
Change Agent	Any person within an institution that has enough social capital, respect, and leadership to catalyze new behaviors among the staff through example, mentoring, advocacy or other means.
Change process	An effective change process is a recipe for selecting, adapting, implementing, and scaling up effective practices in a way that will achieve health results and sustain those results over the years.

Civil society Also known as non-governmental organizations (NGOs), these are critical for the advancement of universal values around human rights, the environment, labor standards, and anti-corruption.

Codebook A type of document used for gathering and storing codes.

Cohort analysis The analysis of data about a particular group. Cohort analysis may involve comparing successive groups passing through a cycle of activity.

Comma separated values A file that stores tabular data (numbers and text) in plain-text form separated by commas.

Commitment to change

Commitment to change is the determination to carry the process to the end. The change is complete when all program levels, working together, continually produce desired results as they implement, or support, the changed practices. When stakeholders are committed to change, they don't give up when they encounter barriers, nor do they stop when donors turn their resources toward other needs.

Communication channels A medium through which a message is transmitted to its intended audience, such as print media or broadcast.

Community empowerment The process of enabling communities to increase control over their lives.

Community systems Community-led structures and mechanisms used by communities through which community members and community-based organizations and groups interact, coordinate, and deliver their responses to the challenges and needs affecting their communities.

Community systems strengthening (CSS) An approach that promotes the development and sustainability of communities and community organizations, enabling them to contribute to the long-term sustainability of health and other interventions at the community level.

Completeness Having all necessary parts, elements, or steps.

Conceptual framework/program theory model A logic model that demonstrates how an intervention (a project, a program, a policy, or a strategy) is understood to contribute to possible or actual impacts.

Concise representation A set of coherent ideas or concepts organized in a manner that makes them easy to communicate to others.

Confirmability One of the standards in qualitative inquiry which refers to the quality of the results produced by an inquiry in terms of how well they are supported by informants/members who are involved in the study and by events that are independent of the inquirer.

Conformity The act of matching attitudes, beliefs, and behaviors to group norms.

Construct validity This refers to whether a scale measures or correlates with the theorized psychological scientific construct that it purports to measure.

Control group A set of items or people that serves as a standard or reference for comparison with an experimental group.

Coping skills Psychological mechanisms which allow for constantly changing cognitive and behavioral efforts to manage specific external and/or internal demands that are appraised as taxing or exceeding the resources of the person.

Cost-benefit analysis A systematic process for calculating and comparing benefits and costs of a project, decision, or governmental policy**.**

Counter factual This indicates what would be the case if its logic were true (although it is not true) or contrary to fact.

Credibility One of the standards and quality in qualitative inquiry that involves establishing that the results of qualitative research are credible or believable from the perspective of the participant in the research.

Cross-sectional data Refers to observations of many individuals (subjects, objects) at a given time. A type of one-dimensional data set.

Cut-off value A threshold value for a quantity.

Data augmentation Adds value to base data by adding information derived from internal and external sources within an enterprise, which can help provide more in-depth insight.

Data dictionary A centralized repository of information about data, such as meaning, relationships to other data, origin, usage, and format.

Data integration This has the goal of building and presenting a unified view of data owned by heterogeneous data sources in distributed, cooperative, and peer-to-peer information systems.

Data mining This is a field at the intersection of computer science and statistics that attempts to discover patterns in large data sets.

Data naming conventions Refers to a convention established to resolve problems with traditional data names.

De-duplication Refers generally to eliminating duplicate or redundant information to try and reduce the storage needed for backups by chunking the backup stream and storing unique segments once.

Dependability One of the standards and quality in qualitative inquiry for judging qualitative studies, this refers to the stability or consistency of the inquiry processes used over time.

Descriptive statistical analysis The analysis of data that help describe, show, or summarize data in a meaningful way such that patterns might emerge from the data. However, it does not allow one to make conclusions beyond the data analyzed or reach conclusions regarding any hypothesis made.

Deterministic A traditional branch of philosophy stating that for everything that happens there are conditions in which nothing else could happen.

Diagnostic Delay The time interval between the first TB diagnostic test and the patient receiving the tuberculosis diagnosis if it exceeds two days.

Diffusion The spread of cultural elements from one area or group of people to others by contact.

Disaggregated analysis Analysis by lowest level or by specific characteristics, such as sex or age.

Dissemination plan A part of the overall project plan that explains how the project will share outcomes with stakeholders, relevant institutions, and organizations, and how it will contribute to the overall dissemination strategy for the program.

Dosage: Administration of a therapeutic agent or exposure in prescribed amounts.

Double data entry A data entry quality control method where the same data is punched and verified by two different operators.

Dummy values: A numerical variable used in regression analysis to represent subgroups of the sample in your study.

Early adopters In the diffusion of innovation theory, the minority group (comprising about 14%) of the population which, after innovators, is first to try new ideas, processes, goods, and services.

Ease of manipulation Ease of use.

Effect size (Minimum acceptable effect size) A measure of the strength of a phenomenon (for example, the relationship between two variables in a statistical population) or a sample-based estimate of that quantity.

Emic Of, relating to, or involving analysis of cultural phenomena from the perspective of one who participates in the culture being studied.

Empirical Research A way of gaining knowledge by means of direct and indirect observation or experience.

Empowerment Refers to increasing the spiritual, political, social, educational, gender, or economic strength of individuals and communities.

Enabling environment- The expression that encompasses policies that focus on creating and maintaining an overall climate that facilitates high quality TB control implementation.

Entertainment-education- The process of purposely designing and implementing a media message to both entertain and educate, in order to increase audience knowledge about an educational issue, create favorable attitudes, and change overt behavior.

Error log A log of errors encountered by a system, which can be extremely useful tools for people who need to diagnose and manage systems such as web servers and office networks.

Etic Of, relating to, or involving analysis of cultural phenomena from the perspective of one who does not participate in the culture being studied.

Expectation management - A formal process to continuously capture, document, and maintain the content, dependencies, and sureness of the expectations for persons participating in an interaction, and to apply the information to make the interaction successful.

Experimental designs The design of any information-gathering exercises where variation is present, whether under the full control of the experimenter or not.

External validity The generalizability of study results to other groups, settings, treatments, and outcomes.

Feasibility That which is achievable.

Fidelity The degree to which an electronic device (record player, radio, or television) accurately reproduces its effect (as sound or picture).

Focus groups A form of qualitative research in which a group of people are asked about their perceptions, opinions, beliefs, and attitudes towards a product, service, concept, advertisement, idea, or packaging.

Formative evaluation This seeks to strengthen or improve a program or intervention by examining, amongst other things, the delivery of the program, the quality of its implementation, and the organizational context, personnel, structures, and procedures.

Gap analysis A technique that businesses use to determine what steps need to be taken in order to move from its current state to its desired future state.

Hard-to-Reach Groups Refers to entirely disparate populations and communities who pose difficulties to the conventional ways of doing things.

Harm reduction Refers to a range of public health policies designed to reduce the harmful consequences associated with human behaviors, even if those behaviors are risky or illegal.

Health information system (HIS) A system for the collection/processing of data from various sources, using the information for policy making and management of health services.

Health seeking delay A long time interval between initial symptoms and arrival to the first health care provider, e.g., more than 30 days.

Histogram This is a graphical representation showing a visual impression of the distribution of data.

Inference The act or process of deriving logical conclusions from premises known or assumed to be true.

Implausible value A data entry error with 'implausible' or 'impossible' values, for they make no sense when considering the expected range of the data.

Incentives Something that motivates an individual to perform an action.

Indicator Statistics used to measure current conditions as well as to forecast financial or economic trends.

Indirect effects Describes a situation where national courts are required to interpret national law in line with an unimplemented or badly implemented directive, as opposed to ignoring national law in preference to the directive, as occurs when direct effect is invoked.

Information and communications technology (ICT) A more specific term that stresses the role of unified communications and the integration of telecommunications (telephone lines and wireless signals), computers, and necessary enterprise software, middleware, storage, and audio-visual systems, which enable users to access, store, transmit, and manipulate information.

Inputs Resources such as people, raw materials, energy, information, or finance that are put into a system (such as an economy or computer system) to obtain a desired output.

Interaction effect The simultaneous effect of two or more independent variables on at least one dependent variable in which their joint effect is significantly greater (or significantly less) than the sum of the parts.

Internal consistency A measure based on the correlations between different items on the same test (or the same subscale on a larger test). It measures whether several items that propose to measure the same general construct produce similar scores.

Internal validity A property of scientific studies which reflects the extent to which a causal conclusion based on a study is warranted.

Interpretability Mathematical logic that explains the relationship between formal theories that expresses the possibility of interpreting or translating one into the other.

Intervening variables A hypothetical internal state that is used to explain relationships between observed variables, such as independent and dependent variables, in empirical research.

Lot quality assessment sampling (LQAS) A random sampling methodology, originally developed in the 1920s as a method of quality control in industrial production, that requires substantially smaller sample sizes.

Management information system (MIS) Computer systems used for managing three primary components: technology, people (individuals, groups, or organizations), and data (information for decision making) and used to analyze and facilitate strategic and operational activities.

Mapping Determination of the scale/level of detail and content of geographic or cartographic databases, entry criteria, and symbol specification for geospatial objects, generalization, and layout design.

Maturation The emergence of personal and behavioral characteristics through growth processes.

Measurement error The difference between a measured value of quantity and its true value obtained by a measurement.

Measurement validity
The degree to which a computation measures what it purports to.

Misclassification A flaw in measuring exposure, covariate, or outcome variables that results in different quality (accuracy) information between comparison groups.

Mixed-method designs To collect, analyze, and report both quantitative and qualitative data (multiple ways) to explore a research problem.

Monitoring Supervising activities in progress to ensure they are on-course and on-schedule in meeting the objectives and performance targets.

Non-probability sampling/purposive sampling
The researcher chooses the sample based on the knowledge of the population and the purpose of the study. This is used primarily when there is a limited number of people that have expertise in the area being researched.

Non-response This is when the required information is not obtained from the persons selected in the sample.

Non-response bias The bias that occurs in statistical surveys if the answers of respondents differ in a meaningful ways from the potential answers of those who did not answer.

Null hypothesis A type of hypothesis used in statistics that proposes that no variation exists between variables, or that a single variable is no different than zero.

Objectivity A lack of favoritism toward one side or another.

One-tailed test A statistical test in which the critical area of a distribution is one-sided so that it is either greater than or less than a certain value, but not both; such that the sample that is being tested falls into the one-sided critical area, the alternative hypothesis will be accepted instead of the null hypothesis.

Opinion leaders Leadership by an active media user who interprets the meaning of media messages or content for lower-end media users.

Optical scanning The process of interpreting data in printed, handwritten, bar-code, or other visual form by a device (optical scanner or reader) that scans and identifies the data.

Order bias Survey bias by which responses can be affected by the order of answer choices.

Organization development A deliberately planned effort to increase an organization's relevance and viability.

Outcome Indicator a benchmark that tracks the direct consequence of an activity or strategy upon the beneficiary or program.

Outlier An observation that is numerically distant from the rest of the data.

Outputs The term often referring to a tangible product produced following an activity, such as a manual or a poster.

Over-reporting To report (an event or instance of something) an occurrence (such as TB diagnosis or cure) more often that is actually occurs.

Panel design Refers to multi-dimensional data frequently involving measurements over time, i.e., participants are followed over multiple survey (observations on multiple phenomena observed) rounds for a specified period of time.

Parsing The process of analyzing a string of symbols, either in natural language or in computer languages.

Participant observation One type of data collection method typically done in the qualitative research paradigm, to gain a close and intimate familiarity with a given group of individuals (such as a religious, occupational, sub cultural group, or a particular community) and their practices through an intensive involvement with people in their cultural environment, usually over an extended period of time.

Participatory communication A term that denotes the theory and practices of communication used to involve people in the decision-making of the development process.

Patient identifier (unique identifier) This is a unique value assigned to an individual to facilitate positive identification of that individual for healthcare purposes.

Personal digital assistants (PDA) A mobile device that functions as a personal information manager.

Political will The exercise of an abstract feature of political authority to enforce certain acts for the benefit of its intention, usually for the public welfare.

Populate To fill or be present in a place, environment, or domain.

Positioning The process by which marketers try to create an image or identity in the minds of their target market for its product, brand, or organization.

Positive deviance An approach to behavioral and social change based on the observation that in any community, there are people who's uncommon but successful behaviors or strategies enable them to find better solutions to a problem than their peers, despite facing similar challenges and having no extra resources or knowledge than their peers.

Precision The measurement deviation from true value and its scatter.

Pretest A preliminary test administered to determine the baseline knowledge or preparedness of something, such as a questionnaire, product, or idea.

Propensity score matching This is a statistical matching technique that is used in the statistical analysis of observational data that attempts to estimate the effect of a treatment, policy, or other intervention by accounting for the covariates that predict receiving the treatment.

Prospective Likely to happen at a future date; concerned with or applying to the future.

Punctuality The characteristic of being able to complete a required task or fulfill an obligation before or at a previously designated time.

Quasi experimental design (QED) An empirical study used to estimate the causal impact of an intervention on its target population, but they specifically lack the element of random assignment to treatment or control.

Qualitative data Data that approximates or characterizes but does not measure the attributes, characteristics, or properties, of a thing or phenomenon. Qualitative data describes whereas quantitative data defines.

Quantitative data Refers to the systematic empirical investigation of social phenomena via statistical, mathematical, or computational techniques.

Queries A form of questioning in a line of inquiry.

Random sampling A subset of individuals (a sample) chosen from a larger set (a population) through a process of chance.

Referral delay An excessive time interval between arrival at first point of care and first TB diagnostic test, e.g., length over one day.

Reflexivity Sensitivity to the ways in which the researcher and the research process have affected the data

Register A formal or official recording of items, names, or actions, a book for such entries, or an entry in such a record.

Reliability coefficients The ratio of true score variance to the total variance of test scores.

Reliability The ability of a person or system to perform and maintain its functions in routine circumstances, as well as hostile or unexpected circumstances.

Representativeness The level of how well or how accurately something reflects upon a sample.

Reputation An overall quality or character of a social entity (a person, a group of people, or an organization) that forms the basis of an opinion about that entity, typically a result of social evaluation on a set of criteria.

Response rate Refers to the number of people who answered a survey divided by the number of people in the sample, usually expressed in the form of a percentage.

Retrospective To take a look back at events that have already taken place.

Return on investment A performance measure used to evaluate the efficiency of an investment or to compare the efficiency of a number of different investments.

Selection bias A statistical bias in which there is an error in choosing the individuals or groups to take part in a scientific study.

Self-efficacy The measure of one's own ability to complete tasks and reach goals.

Social Capital The expected collective or economic benefits derived from the preferential treatment and cooperation between individuals and groups.

Social Change Any significant alteration over time in behavior patterns and cultural values and norms.

Social marketing The systematic application of marketing, along with other concepts and techniques, to achieve specific behavioral goals for a social good

Social networks A social structure made up of a set of actors (such as individuals or organizations) and a complex set of the dyadic (interaction between a pair of individuals) ties between these actors.

Standard operating procedure (SOP) Detailed written instructions intended to achieve uniformity of the performance of a specific function.

Stakeholder A person, group, organization, member, or system that affects or can be affected by an organization's actions or by the results of that which they are said to have a stake in.

Standard deviation How much variation or dispersion exists from the average (mean) or expected value.

Stigma The extreme disapproval of (or discontent with) a person on socially characteristic grounds that are perceived, and serve to distinguish them, from other members of a society. A distinguishing mark of social disgrace.

Stratified sampling A probability sampling technique in which the researcher divides the entire target population into different subgroups, or strata, and then randomly selects the final subjects proportionally from the different strata. This is used to highlight specific subgroups within the population.

Summative evaluation- Refers to the assessment of the learning and summarizes the development of learners at a particular time. Looks at the impact of an intervention on the target group.

Survey A method for collecting quantitative information about items in a population.

Sustainability Sustainability creates and maintains the conditions under which project and programs can exist in harmony and that permit fulfilling the social, economic, and other requirements of present and future generations.

Target population The entire group of individuals or objects to which researchers are interested in generalizing the conclusions.

Terms of reference The purpose and structure of a project, committee, meeting, negotiation, or any similar collection of people who have agreed to work together to accomplish a shared goal.

Theoretical framework This collection of interrelated concepts that guides research, determining what will be measured and what statistical relationships will be looked for.

Time Series Design A quasi-experimental design that is a standard method of causal analysis in evaluating the impacts of interventions, health programs, and state/national policies, i.e., it focuses on processes and behavior during treatment.

Timeliness Occurring at a suitable or opportune time; with a desired frequency.

Transferability One of the standards and quality in qualitative inquiry that refers to the degree to which the results of qualitative research can be generalized or transferred to other contexts or settings.

Transparency Implies openness, communication, and accountability.

Treatment Delay A long time interval between the date that a TB diagnosis was given to a patient and when TB medicines were dispensed to the patient (treatment start), e.g., in excess of one day.

Treatment literacy Subcomponents of broader treatment education as defined by UNESCO and WHO that means people, both individually and in communities, understand what HIV drugs are, why they are needed, and what they can and cannot do. Treatment literacy translates medical information about ART into languages and formats that are accessible for everyone.

Trialibility The ability to test the intervention on a small scale in an organization, and to be able to reverse course (undo implementation) if warranted.

Triangulate Refers to a process of contrasting diverse sources of information and different data to identify divergent perspectives, validate key information, explore disparities, and yield a richer, more nuanced analysis of a situation.

Truncation The term for limiting the number of digits to the right of the decimal point by discarding the least significant ones.

Two-tailed test A statistical test used in inference in which a given statistical hypothesis, H_0 (the null hypothesis), will be rejected when the value of the test statistic is either sufficiently small or sufficiently large.

Type 1 error This occurs when the H_0 is true, but is rejected. It is asserting something that is absent, a false hit.

Type 2 error This occurs when the null hypothesis is false, but erroneously fails to be rejected. It is failing to assert what is present, a miss.

Under-reporting A type of reporting bias that reports unexpected or undesirable results as being less than is actually the case.

Validation rule A criterion used in the process of data validation that is executed after the data has been encoded onto an input medium. Inv

Value-added The amount by which the value of goods or services are increased by each stage in its production.

Vital Registration system A registration of births and deaths and the cause of death.