

Patient Pathway Analysis: How-to Guide

Assessing the Alignment of TB Patient
Care Seeking & TB Service Delivery

Table of Contents

Acknowledgments.....	5
Acronyms	6

INTRODUCTION 7

0.1 Background.....	7
0.2 Rationale.....	7
0.3 Objectives.....	9
0.4 Audience	9
0.5 Defining Health Facility Levels/Sectors.....	9
0.6 Patient Pathway Steps/Metrics.....	10
0.7 Assumptions and Limitations.....	15



STEP 1: PREPARE TO CONDUCT A PATIENT PATHWAY ANALYSIS 17

1.1 Assemble PPA Team.....	17
1.2 Access Companion Resources	18
1.3 Define Scope	18
1.4 Establish Timeline	18
1.5 Step 1 – Checklist	19



STEP 2: COLLECT DATA SOURCES 21

2.1 Data Sources for Core Metrics	22
2.2 Review Data Sources for Supplemental Metrics	32
2.3 Access and Save Data Resources.....	33
2.4 Step 2 - Checklist	33

Table of Contents (cont.)

	STEP 3: ORGANIZE DATA AND CALCULATE METRICS	35
	3.1 Introduction to the PPA Excel Workbook Template	35
	3.2 Raw Data Sheets.....	36
	3.3 Summary Data Sheets.....	37
	3.4 Mapped Data Sheets.....	39
	3.5 Create the Master PPA Data Sheet	44
	3.6 Step 3 – Checklist	45
	STEP 4: VISUALIZE RESULTS	47
	4.1 PPA Visual Design	47
	4.2 Tools for Visualization	49
	4.3 Step 4 – Checklist	49
	STEP 5: INTERPRET RESULTS AND APPLY THEM TO PLANNING AND STRATEGY	51
	5.1 Summarize and Evaluate the PPA Process.....	51
	5.2 Interpreting a PPA	51
	5.3 Applying Results to Planning and Strategy.....	54
	5.4 Step 5 – Checklist	55
	APPENDIX	57
	6.1 Collecting and Summarizing DHS Data	57
	References	62

Acknowledgments

The following people contributed to the writing, design, and review of this document:

- Dena Seabrook, George Durham, Jessie Brown, Metta Chaphiv, Mike Osberg (Linksbridge)
- Dr. Christy Hanson (Bill & Melinda Gates Foundation, Macalaster College)
- Cicilia Gita Parwati (KNCV Tuberculosis Foundation – Indonesia)
- Faith N Ngari (Kenya Ministry of Health, National Tuberculosis, Leprosy and Lung Disease Program)

Comments or questions can be addressed to:

Mike Osberg – mike.osberg@linksbridge.com

Christy Hanson – christy.hanson@gatesfoundation.org

Financial support for this work was provided by the Bill & Melinda Gates Foundation

Acronyms

CSV – Comma-separated Values File
DHS – Demographic and Health Survey
DST – Drug Susceptibility Testing
EPI – Expanded Programme on Immunization
FTE – Full-time Equivalent
HEUS – Health Expenditure Utilization Survey
HFML – Health Facility Master List
JEMM – Joint External Monitoring Mission
LSMS – Living Standards Measurement Survey
MDR – Multi-Drug Resistant
MoH – Ministry of Health
NGO – Non-Governmental Organization
NHA – National Health Accounts
NTP – National TB Control Program
PDF – Portable Document Format
PPA – Patient Pathway Analysis
RR – Rifampicin Resistant
SARA – Service Availability and Readiness Assessment
SPA – Service Provision Assessment
WHO – World Health Organization

Introduction

0.1 BACKGROUND

According to the World Health Organization (WHO), more than 40 percent of incident tuberculosis (TB) cases in 2015—over 4 million patients—were not notified to National TB Control Programs (NTP). Whether these patients received appropriate diagnosis and treatment or not is unknown.¹ These so-called ‘missing’ cases fall into three groups of patients:

- Patients who never accessed care because of financial, geographic, or other barriers to accessing care.
- Patients who sought care in the private (or non-state) sector and were diagnosed and treated there, but not notified to the NTP.
- Patients who were diagnosed and treated in the public sector but not notified to the NTP.

Finding these ‘missing’ patients is essential, if the goal of ending TB is to be achieved. Patient-centered care is a core principle of the WHO End TB Strategy. Where TB prevention and care services are planned from the top down, according to the available capacities of the health system, there is room to strengthen the patient-centered delivery of services by considering where services should be positioned in order to meet patients based on where they are. Essentially, instead of thinking about ‘where *can* we provide services?’ patient-centered care asks, ‘where *should* we position services?’ to most efficiently serve TB patients.

0.2 RATIONALE

The patient pathway analysis (PPA) methodology was developed to better understand the alignment between patient care seeking and TB service availability. The results can inform programmatic priority setting and planning for more patient-centered availability of services.

The PPA aims to describe the steps TB patients take from the initial point of seeking care to the point of being cured. At the same time, the analysis reviews the availability of TB screening, diagnosis, and treatment at various levels of the health system.

Table 1 Aligning Care Seeking and Service Availability

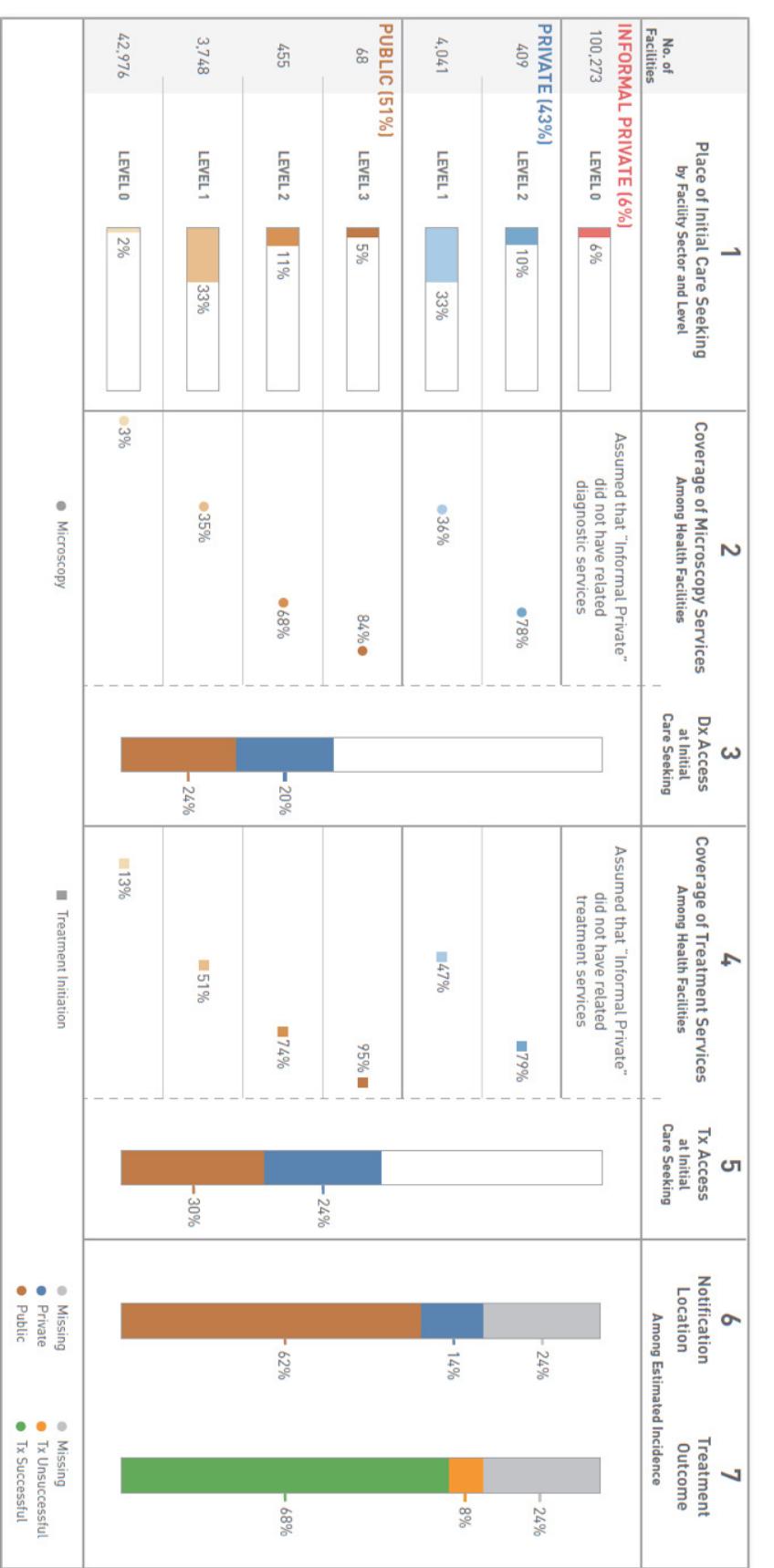
Patient Care Seeking	TB Specific Care Seeking		General Care Seeking
	TB Service Availability	How does the supply of services align with the demand for services?	
Alignment	Public Sector	Private Sector	Informal Sector

By examining the alignment of care seeking with service availability, the PPA may reveal where TB patients experience delay during care seeking or treatment initiation, access inappropriate care, or are lost to follow-up during their journey toward cure. The intent of the PPA is to help national tuberculosis programs more accurately identify some of the health systems alignment gaps that can be addressed through targeted program interventions. The results from a PPA complement other methods of understanding programmatic gaps, including TB prevalence surveys, routine surveillance, qualitative program reviews, care cascade analysis^{1,2} and special studies. Equipped with data from a PPA and other evidence, TB programs can plan TB prevention and care services in a manner that responds to patient care seeking preferences and accelerates the patient pathway to diagnosis and care. Figure 1 illustrates a hypothetical patient pathway analysis. This guide addresses each component of the PPA in detail throughout the document.

Introduction

Introduction

Patient Pathway Description: The PPA visual shows several steps along a patient's journey for TB care. Read from left to right, the PPA starts by showing the number of health facilities in each sector and level of the health system, included for additional context. Column 1 shows the proportion of TB patients that initiate care at each sector and level. Column 2 shows the proportion of facilities at each sector and level that are equipped with a diagnostic tool (in this example, microscopy). Column 3 shows the estimated proportion of all TB patients that access a facility with a diagnostic tool at initial care seeking. For each health sector and level, the proportion of patients that seek care at that sector and level (Column 1) is multiplied by the proportion of facilities at that sector and level that are equipped with a TB diagnostic tool (Column 2) to obtain the estimated proportion of patients that access TB diagnosis at that sector and level. Column 3 shows the total estimated proportion of TB patients that access diagnosis at initial care seeking in each sector. Column 4 shows the proportion of facilities in each sector and level that offer TB treatment. Column 5 shows the estimated proportion of patients that access treatment at initial care seeking. Column 5 is the sum of the products of Columns 1 and 2. The PPA finishes with Columns 6 and 7, showing notification by location and treatment outcomes among patients notified to the NTP (both as a proportion of the estimated incident TB cases at the national level).



0.3 OBJECTIVES

The objective of this document is to explain the process of completing a PPA and interpreting its results for use in programmatic priority setting and planning. Ideally, the process of completing a PPA will be done in conjunction with other data analysis activities (e.g., EPI reviews, inventory studies) and will be one of several inputs into the development of strategy and planning documents.

This document provides guidance for anyone involved in TB control who wishes to complete a PPA at a national or sub-national level. Using this document as a guide, readers should be able to:

1. Describe how a PPA can support strategy and planning work for TB control
2. Identify the best data sources to develop a PPA
3. Learn how to analyze the necessary data for use in a PPA
4. Complete the steps needed to create useful PPA visualizations
5. Facilitate the process of interpreting and applying the PPA results within the context of programmatic priority setting and planning

0.4 AUDIENCE

This document is intended for national TB program (NTP) or technical partner staff who support evidence-based decision-making in high-TB-burden settings.

0.5 DEFINING HEALTH FACILITY LEVELS/SECTORS

The PPA is built from commonly available population-based surveys and data sets about patient care seeking and service availability. When reviewing data on care seeking and service availability, data sets will often use unique naming conventions that are not standardized across sources. To allow for comparison across sources, the PPA uses a standardized naming convention to organize the data. For each data source, health facility types will be grouped into common categories that will indicate the sector and level of the health system to which a facility can be assigned.

The two primary sectors used in the PPA are public and private (or non-state). In many countries, it is useful to split private sector facilities into formal private facilities and informal private facilities, where the informal private facilities include practitioners with little or no formal health training, e.g., drug shops, traditional healers. The designation of non-governmental organizations (NGOs) into public or private sector should align with the country's norms for categorization.



TIP

Throughout the document, several terms related to health facility categorization are used. There is a hierarchy to this categorization that should be kept in mind. For instance:

Health Facility Name – refers to the proper name of a specific health facility. For example, 'Santa Ana Hospital' or 'Addis Ketema Health Center.'

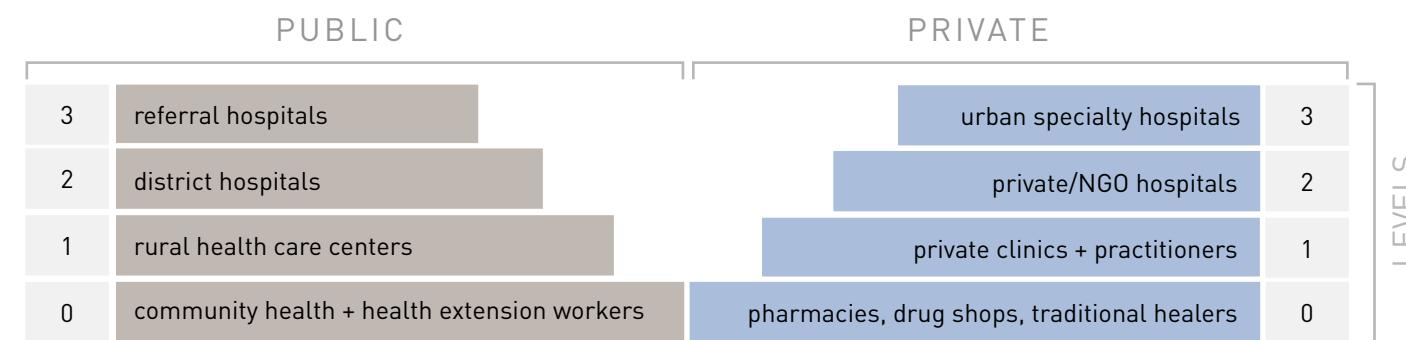
Health Facility Type – refers to the category of facility. For example, hospital, health clinic, pharmacy, or community health post.

Health Facility Sector – refers to the ownership of the health facility. For example, public or private.

Health Facility Level – refers to the hierarchical layers of the health system and designates each facility as a level. For example, level 0 or level 1.

Introduction

Fig. 2 Health System Levels and Sectors for PPA



Categorization may differ by country, reflecting the diversity of health systems.

Within each sector category, the PPA groups health facilities into standard levels according to the level of care provided (Figure 2). The standard levels used by the PPA are as follows:

- **Level 0 (L0)** refers to basic, community-based care. Level 0 services include basic triage, health information, and essential prevention and care. Services are commonly provided as an extension of facility-based care, and are provided by volunteers or paramedical staff with limited formal training. Sputum collection may occur at this level, and L0 staff may serve as treatment supporters for TB patients. Examples: community health worker (public); pharmacies, drug shops, and traditional healers (private).
- **Level 1 (L1)** refers to a facility that provides primary health care. L1 services are commonly provided on an outpatient basis by nurses, midwives, paramedical staff, or private doctors. Some basic diagnostic services, including basic microscopy and essential medicines, may be available. Rural health centers and private clinics are examples within public and private sector, respectively.
- **Level 2 (L2)** refers to facilities that provide primary health care as well as more advanced care. L2 facilities commonly have more extensive diagnostic and treatment options and can provide both outpatient and inpatient care. Examples: district hospital (public), rural or NGO private hospital (private).
- **Level 3 (L3)** refers to facilities that provide specialized care with large inpatient capacity. L3 facilities provide access to specialized doctors and have more sophisticated diagnostic and treatment capabilities. Examples: referral or teaching hospital (public), urban private hospital

0.6 PATIENT PATHWAY STEPS/METRICS

The PPA is intended to present some critical steps patients take on their journey to TB care. To do this, several core metrics are used by the PPA to describe different steps on a patient's pathway to care. All the core metrics can be segmented by health sector and facility level as described in section 0.5 and as shown in Figure 2.

Introduction

CORE METRICS	
1. Place of Initial Care Seeking	
Proportion of patients who initiate care, by facility level and sector.	$\frac{N}{D}$
Number of patients initiating care for illness or symptoms (preferably TB symptoms) in a specific facility level and sector	Total number of patients initiating care for illness or symptoms (preferably TB symptoms) in the survey population
2. Diagnostic Coverage	
Proportion of health facilities with a given TB diagnostic service, by facility level and sector.	$\frac{N}{D}$
Number of health facilities with a given TB diagnostic service, by facility level and sector. This can be calculated for each type of TB diagnostic services, e.g., smear microscopy, Xpert, X-ray	Total number of health facilities, by facility level and sector
3. Access to Diagnosis at Initial Care Seeking	
Estimated proportion of care seekers that access diagnostic services on their first visit.	$\sum \text{CM}_1 \times \text{CM}_2$
This metric is the product of multiplying initial care seeking (core metric 1) and diagnostic coverage (core metric 2) at each level and sector. It can be interpreted as the estimated proportion of care seekers that access TB diagnostic services on the first visit. Alternatively, it can be interpreted as the likelihood that a patient has access to TB diagnostic services on his or her first visit to the health care system. One can sum up the diagnostic access for all facility levels and sectors to derive an indicator of overall diagnostic access . Overall diagnostic access can be calculated for each type of TB diagnostic service or for a combination of diagnostic tools. The latter could be considered as access to any of the diagnostic tools available. Importantly, this metric assumes that care seekers visit all health facilities in a sector uniformly, which may or may not be true. These assumptions and related limitations are described in more detail in section 0.7.	Percent of patients who initially sought care at each level and sector \times Diagnostic coverage at each level and sector
4. Treatment Coverage	
Proportion of health facilities that have anti-TB medicines in stock or can supervise patients during treatment, by facility level and sector.	$\frac{N}{D}$
Number of facilities that have anti-TB drugs in stock or can supervise patients during treatment, by facility level and sector	Total number of health facilities, by facility level and sector

Introduction

CORE METRICS	
5. Access to Treatment at Initial Care Seeking	
Estimated proportion of care seekers that access diagnostic services on their first visit.	$\sum \text{CM} 1 \times \text{CM} 4$ <p>Percent of patients who initially sought care at each level and sector X Treatment coverage at each level and sector</p>
This indicator is the product of multiplying initial care seeking (core metric 1) and treatment coverage (core metric 4) at each level and sector. It can be interpreted as the estimated proportion of care seekers that access TB treatment services on the first visit. Alternatively, it can be interpreted as the likelihood that a patient has access to TB treatment services on their first visit to the health care system. One can sum up the treatment access for all facility levels and sectors to derive an indicator of overall treatment access. Overall treatment access can be calculated for anti-TB drugs, treatment supervision, or a combination of the two. Importantly, this metric assumes that care seekers visit all health facilities in a sector uniformly, which may or may not be true. These assumptions and related limitations are described in more detail in section 0.7.	
6. Notification Location	
Location of case notification for patients notified to NTP/WHO. This indicator reflects the location of notification for the patients notified to the NTP/WHO. Notifications are segmented by public and private sector. The number of patients notified in each sector are divided by a denominator of the total estimated incident TB cases in the country. The resulting metric is the percent of the estimated incident cases made up of patients notified by the public sector and the private sector. The remaining estimated patients that are not notified by either the public or private sector health facilities are reported as 'missing' patients.	$\frac{N}{D}$ <p>Number of patients notified by public sector, number of patients notified by private sector, number of estimated patients not notified.</p> <p>Estimated number of incident TB patients.</p>
7. Treatment Outcomes	
Treatment outcomes among patients notified to NTP/WHO. Treatment success rates are applied to the notified patients for a given year in the country. The notified patients are divided into those that were successfully treated and those that were not successfully treated. Each of these numbers are then calculated as a share of the total estimated incident cases in the country. The share of estimated patients that do not have a known treatment outcome of either successfully treated or not successfully treated are reported as 'missing'. The percent of estimated incident cases that is 'missing' in core metric 6 should be equal to the percent of estimated incident cases that is 'missing' in this metric.	$\frac{N}{D}$ <p>Number of notified patients that are successfully treated, number of notified patients that are not successfully treated.</p> <p>Estimated number of incident TB patients.</p>

Introduction

SUPPLEMENTAL METRICS	
In addition to the core metrics, there are several supplementary metrics that can provide useful insights, but for which data are often limited or not available. Several of these metrics would enable a more robust PPA and should be included along with the core metrics, if data are available. The others will not be included in the PPA visual or analysis process, but could help with the interpretation of the PPA described in Step 5, later in this document.	
METRICS TO INCLUDE IN PPA	
1. Location of Diagnosis by Sector and Level	$\frac{N}{D}$ <p>Number of participants in a nationally representative survey who received a diagnosis for TB, separated by location of diagnosis (sector and level)</p>
This metric is the proportion of TB patients diagnosed at each sector and level of the health system, among a cohort of patients that received a diagnosis. Different from the core PPA metrics, which suggest a likelihood of diagnosis based on the health system structure, this supplemental metric provides data on the location where patients were actually provided a TB diagnosis. It is less likely that data are available for this metric, but when they are, they can be useful to add to the PPA.	$\frac{N}{D}$ <p>Total number of participants in a nationally representative survey who received a diagnosis for TB</p>
2. Location of Treatment by Sector and Level	
Similar to supplemental metric 1, this metric is the proportion of TB patients treated at each sector and level of the health system, among a cohort of treated patients. When the data for this metric come from a national TB prevalence survey, this metric can be used as a representation of the treatment location of patients, whether or not they were notified to the NTP. Thus, it can serve as an important point of comparison to the core metric 6 that estimates the location of notification for patients notified to the NTP. Using core metric 6 (location of notification) and this supplemental metric, the PPA can provide helpful suggestions on where patients might be receiving treatment but not being notified.	$\frac{N}{D}$ <p>Number of participants in a nationally representative survey who were treated for TB, separated by location of treatment (sector and level)</p>
3. Referral Patterns Between Sectors and Levels	
The core metrics of the PPA focus on patient care seeking patterns during their initial visit to the health system. However, many patients often move through several referrals and facilities before reaching one that meets their needs. This metric is intended to capture any available information on where patients are most likely to be referred to after their initial care seeking visit.	$\frac{N}{D}$ <p>The number of patients referred from provider, segmented by sector and level they are referred to</p>
This metric could be presented in several different ways, depending on the availability of data. Ideally, the data would provide information on the types of health facilities to which providers typically refer their patients if a referral is necessary.	$\frac{N}{D}$ <p>Total number of patients who are referred to other facilities</p>
An example of how this metric might be described: 'Of patients who initially visit public level 1 facilities and are referred elsewhere for care, 65 percent are referred to public level 2 facilities, 25 percent are referred to other public level 1 facilities, and 10 percent are referred to private level 2 facilities'.	

Introduction

METRICS TO SUPPORT PPA INTERPRETATION	
4. Percent of Patients Who Do Not Initially Seek Care	
N	Number of patients not initiating care for illness or symptoms (preferably TB symptoms)
D	Total number of patients with illness or symptoms (preferably TB symptoms) in the survey population
5. Reasons for Not Initially Seeking Care	
N	Number of participants in a nationally representative survey who reported not initially seeking care for a particular reason
D	Total number of participants in a nationally representative survey who reported not initially seeking care
6. Reasons for Initial Care Seeking Patterns	
N	Number of participants in a nationally representative survey who reported seeking care for a particular reason by facility level and sector
D	Total number of participants in a nationally representative survey who reported seeking care by facility level and sector
7. Delays Between Pathway Steps	
A	Average time reported between pathway steps by facility level and sector

KEY	
N	numerator
D	denominator
A	average

CM 1 core metric 1
CM 2 core metric 2
CM 4 core metric 4

Introduction

0.7 ASSUMPTIONS AND LIMITATIONS

When building and interpreting a PPA, it is important that the analysis team keeps in mind several key assumptions and limitations of the PPA. Some teams may have access to data where the appropriateness of these assumptions can be tested for their specific country context, in which case it is highly recommended that the team do so.

Assumption #1 – Care seeking patterns for general illness mirror those for TB-specific symptoms

When countries do not have data on patient care seeking for TB-related symptoms, the PPA will often use care seeking data for general symptoms or other illnesses. As tested in prior PPA case studies, care seeking for a general set of illnesses can be used as a reasonable proxy for TB care seeking. However, in countries where there are unique health systems for a specific illness that are not well integrated in the broader health system, care seeking data for these illnesses should be treated with caution and may not be appropriate to use as a proxy for TB.

Assumption #2 – Health facilities in each sector and level are relatively homogeneous

The PPA metrics described in section 0.6 assume homogeneity across health facilities in each level and sector. Specifically, the underlying assumption suggests that factors like care seeking, lab throughput, or service coverage area are the same across health facilities at each level and sector. In some cases, this assumption may lead to an overestimation or an underestimation of diagnostic and treatment access at initial care seeking. If PPA teams have access to patient-level records on care seeking or treatment location, they could conduct more detailed analysis to better understand the differences between health facilities.

Assumption #3 – Data sources can be compared across time to help elucidate a patient's care seeking journey

The PPA uses a wide variety of data sources, which are typically collected at different time periods and from different cohorts. By bringing together these data sources, the PPA can help highlight how patients might move through the health system.

The data sources used for a PPA should be as recent as possible and a time limit should be put on how old the data sources can be.

The analysis team should determine a programmatically relevant year in the history of their program and not allow data prior to that point in time to be used. The PPA will then assume that it is reasonable to use data collected after that time and to compare the results of those different data sources.

Additional Limitations

There are several additional limitations related to the PPA framework that should be noted:

- There is generally a paucity of sub-national data available for care seeking and service availability. When these data are available, if they are further broken down by health facility level and sector, the sample sizes can often become quite small and unsuitable for analysis, making sub-national PPAs more challenging to interpret.
- Unless delay studies are available in the country, the PPA does not provide information on the delay between each step of the pathway. As shown elsewhere, delays in patient care seeking can be an important contributor to the transmission of disease in a community.^{3,4}
- The PPA cannot be used to estimate the drop-offs of a cohort of patients between the steps of a patient's pathway to care. Other cascade of care studies have been conducted which do provide these estimates and describe the methodology in more detail.^{1,2}
- Finally, the PPA focuses on the availability of services where patients seek care, but it does not try to estimate other factors that may influence whether or not a patient will receive proper care. Other factors, such as the quality of diagnostics and treatment, the capacity of health care workers to provide the service, or the cost of the services may impact whether or not a patient actually is able to receive a necessary TB service at the place they initiate their care seeking journey.

Step 1: Prepare to Conduct a Patient Pathway Analysis

1

Prior to starting the PPA process, confirm the following:

- A. The timing is right to conduct such an analysis, i.e., results can be used to inform an upcoming planning, resource allocation, or decision-making process
- B. Decision-makers and partners will allow and facilitate access to national and sub-national data for the PPA
- C. Human resources are sufficient to dedicate to the analytical process
- D. Resources are available to involve stakeholders in the interpretation and dissemination of results, and for their translation into programmatic planning
- E. PPA resources have been downloaded, including country PPA case studies, an Excel workbook, R scripts, and demo PPA built in Tableau, available at www.linksbridge.com/work/tb-ppa

1.1 ASSEMBLE PPA TEAM

The process for completing a PPA is likely to require between .5 and .75 FTE staff between three to five weeks, depending on data availability. An individual with skills in data management, data cleaning and analysis, and interpretation should be identified to lead the exercise. This person will be responsible for ensuring that the data are collected, analyzed, and visualized as needed by the broader NTP and technical partners team. There should also be someone on the team who has experience with Excel and is willing to learn new analytical tools as needed (e.g., Tableau and R). Ideally, one person on the PPA team will have relationships with government counterparts to fulfill requests for non-publicly available data and for interpretation of the findings.

Finally, it is beneficial but not essential for some of the PPA team members to also have familiarity with broader health systems data. These might include Demographic and Health Surveys (DHS), Service Provision Assessments (SPA), Health Expenditure Utilization Surveys (HEUS) or National Health Account (NHA) data sources (described in more detail in Step 2). For those who aren't familiar with other non-TB data sources, this guide will explain how they can be used in support of the PPA.

ASSEMBLE YOUR PPA TEAM



PPA Lead

Skill Set

- » data management
- » data cleaning
- » data analysis + interpretation
- » familiarity with broader health systems data (DHS, SPA, HEUS or NHA data sources)



Analyst

Skill Set

- » proficient in: Excel, Tableau, R
- » familiarity with broader health systems data (DHS, SPA, HEUS or NHA data sources)



Researcher

Skill Set

- » relationships with government counterparts (for obtaining and interpretation of data)
- » familiarity with broader health systems data (DHS, SPA, HEUS or NHA data sources)

Step 1: Prepare to Conduct a Patient Pathway Analysis

Table 2 PPA Scope Selection

	Drug Sensitive	Drug Resistant
National	Need facility-level data on the availability of microscopy (or other diagnostics) and first-line drugs	Need facility-level data on availability of Xpert, culture, drug susceptibility testing (DST), and second-line drugs
Sub-national » Urban / Rural » Administrative area; e.g. province, region, county	All care seeking and service availability data need to be disaggregated by facility type for each sub-national unit For survey data, sample sizes at the sub-national level need to be sufficient	All care seeking and service availability data need to be disaggregated by facility type for each sub-national unit For survey data, sample sizes at the sub-national level need to be sufficient

1.2 ACCESS COMPANION RESOURCES

There are several companion resources to this guide that will be referenced throughout the document. The lead analyst should ensure that the team has access to these resources:

1. PPA Excel Workbook Template
2. Published Tableau PPA workbooks
3. GitHub library with R scripts for DHS data collection (if DHS data are used)
4. Examples of PPA case studies from The Journal of Infectious Diseases supplement

These resources can be accessed at:

www.linksbridge.com/work/tb-ppa

1.3 DEFINE SCOPE

Depending on the availability of data and the programmatic objectives for completing a PPA, the scope of the analysis will differ. In particular, consideration should be given to whether the analysis will be done at the national or sub-national level, and if it will cover all forms of TB or a subset of drug-sensitive or drug-resistant patients. Defining this scope early in the process will influence the types of data points that the PPA team will need to collect later in the process (Table 2).

It is important to consider the population that is most important for the TB program context. Many TB programs have become increasingly decentralized, in which case a sub-national level PPA could be most useful to guide differentiated planning.



TIP

Before moving to *Step 2: Collect Data Sources* of this document, open and review the PPA Excel Workbook Template. This will provide an idea of the output of the data collection, organization, and analysis steps of the PPA process. This workbook will also be used to create the PPA visuals and any accompanying visuals the PPA team will need for interpretation and recommendations.

In other situations, there are important urban and rural differences in health care access, thus this level of detail in the PPA could be more useful. Finding the necessary data for the PPA is often more difficult for sub-national analyses.

1.4 ESTABLISH TIMELINE

On the next page is an outline of the estimated timeline for completing a PPA. Note that this timeline incorporates the time that may be involved waiting for data requests to be processed. If all data are immediately accessible, the timeline could be shorter.

Step 1: Prepare to Conduct a Patient Pathway Analysis

Table 3 PPA Timeline

Timeline	Action Items
Week 1 Pre-PPA Work	» Hold kick-off meeting with team involved in PPA to discuss scope (national or sub-national) and metrics of importance » Identify key data sources » Contact counterparts who might control access to necessary data » Schedule a time for interpretation with TB team leadership (two to five weeks later)
Weeks 2-3 Collection, Cleaning and Analysis of Data	» Attribute common health sector and level categories to data from each data source » Clean and analyze data » Complete full master spreadsheet
Week 4 Creation of PPA Visuals and Summary of Key Results	» Generate PPA visualizations and supplementary figures » Write summary of key results
Week 5 Interpretation and Finalize Recommendations	» Workshop with TB team members to interpret the PPA – Facilitated by question prompts – Use additional programmatic data or reports as needed (e.g., JEMM recommendations, country specific studies) » Develop recommendations for strategy and planning

1.5 STEP 1 – CHECKLIST

During Step 1, each of the following activities should be completed:

STEP 1 CHECKLIST

- Identify a PPA team lead
- Download all companion resources
- Define the scope of the PPA (e.g. national, sub-national, urban/rural)
- Outline a timeline and workplan
- Review the PPA Excel Workbook Template

Step 2: Collect Data Sources

The second step of the PPA is to research and collect the data sources needed for each of the core metrics. Each metric will likely come from a different data source and some metrics will require several data sources. This step should also include research on data that could be used to calculate any of the supplemental metric.

In general, TB-specific data should be prioritized over general data. For example, where both TB-related care seeking data are available from a TB prevalence survey as well as general illness care seeking data from a DHS survey, the data from the TB prevalence survey should be prioritized.

Further, where sub-national data are available, they should be prioritized over national-level data. In some cases, care seeking data are available for TB only at the national level, but general care seeking data are available at the sub-national level. In these situations, the scope of the analysis decided on in section 1.3 should help decide which data to use. For example, if the scope of the PPA is the national level, TB-specific data at the national level should be prioritized. However, if the scope of the PPA is sub-national, care seeking for general illness can often be a useful proxy for TB care seeking. This relationship between TB and general care seeking can be tested by comparing the two data sources at the national level.

When reviewing data sources, the team should put a limit on the age of the data sources. It is often useful if this time limit is programmatically informed. For example, if major changes to the TB program were implemented six years prior to the time of conducting the PPA, it might be useful to limit the data sources from within the last six years.

As mentioned earlier, collecting data for the PPA will require between one to three weeks of effort. As soon as the team has identified available data sources, they should determine whether or not the data need to be requested from other partners. If so, those requests should be made immediately, as it can take some time for access to be granted.

The rest of this section will describe the most common data sources used for each of the core metrics.

The goal of this section is to develop a list of all the data sources that will be used in the PPA, the types of data files that are available (e.g., a CSV file, an online database, or a table in a report), and whether or not the team will need to request the data from a third party in order to access them.



TIP

Throughout this document, the terms **raw data**, **summary data**, **mapped data** and **master PPA data** are used.

Raw data refers to *direct results from a census or survey*. Typically, raw PPA data are stored as a CSV file or other tabular format in which each row represents either an individual person (patient or survey respondent) or a unique health facility. There may be many thousands of rows (or more). For example, in raw DHS data, each line represents a survey respondent; in raw SARA data, each line represents a health facility; in TB patient records, each line represents a patient.

Summary data refers to *data that are aggregated by facility type*. Sometimes, survey reports contain **summary data**. This is most common at the national level. For example, a DHS survey report may contain a table summarizing care seekers according to facility type visited. Other times, and often at the sub-national level, the PPA implementers will need to do the summarizing. For example, a PPA implementer may summarize DHS care seeking data by facility type at the sub-national level using R, SPSS, or other software, and output the results in a summary table.

Mapped data refers to *data that is aggregated by health sector and/or level as required by the PPA*. Facility types have been mapped to health sectors and levels as described in Section 0.5.

Master PPA data refers to *PPA data that is ready for visualization*. Calculations for Steps 3, 5, 6, and 7 of the pathway have been made. Data for all steps of the pathway are together in one worksheet.

Step 2: Collect Data

The PPA will be stronger when the team is able to access **raw data**. This list of data sources is not exhaustive, but represents the most common data sources available. In many cases, countries will have additional data sources not mentioned here that could be used for each of the PPA metrics.

2.1 DATA SOURCES FOR CORE METRICS

2.1.1 NUMBER OF HEALTH FACILITIES

Collecting data on the number of facilities will provide helpful context on the structure and size of the health system. In some cases, this metric will also need to be used to calculate the metrics that estimate coverage of TB services.

National Health Facility Master Lists (HFML)

The best source of data on number of health facilities is a national health facility master list or registry. These databases often provide a list of all registered health facilities in a country, along with important details about each facility, which might include facility name, ownership, location, or type of facility (e.g., hospital, clinic).

National Health Facility Master Lists (HFML)	
Where to Find Data	These databases are typically publicly accessible and are available to download from a Ministry of Health website. In some cases, they will not be accessible publicly and will need to be requested from a MoH counterpart
Typical File Type	Most websites have an option to download the entire list rather than just view on the website. If possible, it is best to access an Excel or CSV file that lists all of the facilities in the database
Examples	Kenya Health Facility List (http://kmhfl.health.go.ke/#/home) (Raw data shown in Figure 4) ⁵ , Tanzania Health Facility Registry (http://hfrportal.ehealth.go.tz/) ⁶ , Philippines National Health Facility Registry (http://nhfr.doh.gov.ph/rfacilities2list.php) ⁷

National Health Accounts

According to WHO, National Health Accounts 'depict the current use of resources in the health system.'^{8,9} While the bulk of NHAs focus primarily on the financial picture of the health system, they will, on occasion, also provide information about the health facilities that make up the health system. It is particularly useful to review NHAs when government census results or HFMLs are not available.

National Health Accounts
Where to Find Data
NHA reports are typically available from either a Ministry of Health or a Ministry of Statistics website
Typical File Type
NHA are most regularly available as summary data in a final report, as shown in Figure 3. It may be possible in some cases to download or request the raw data.
Examples
Pakistan National Health Accounts (http://www.pbs.gov.pk/content/national-health-accounts) ¹⁰

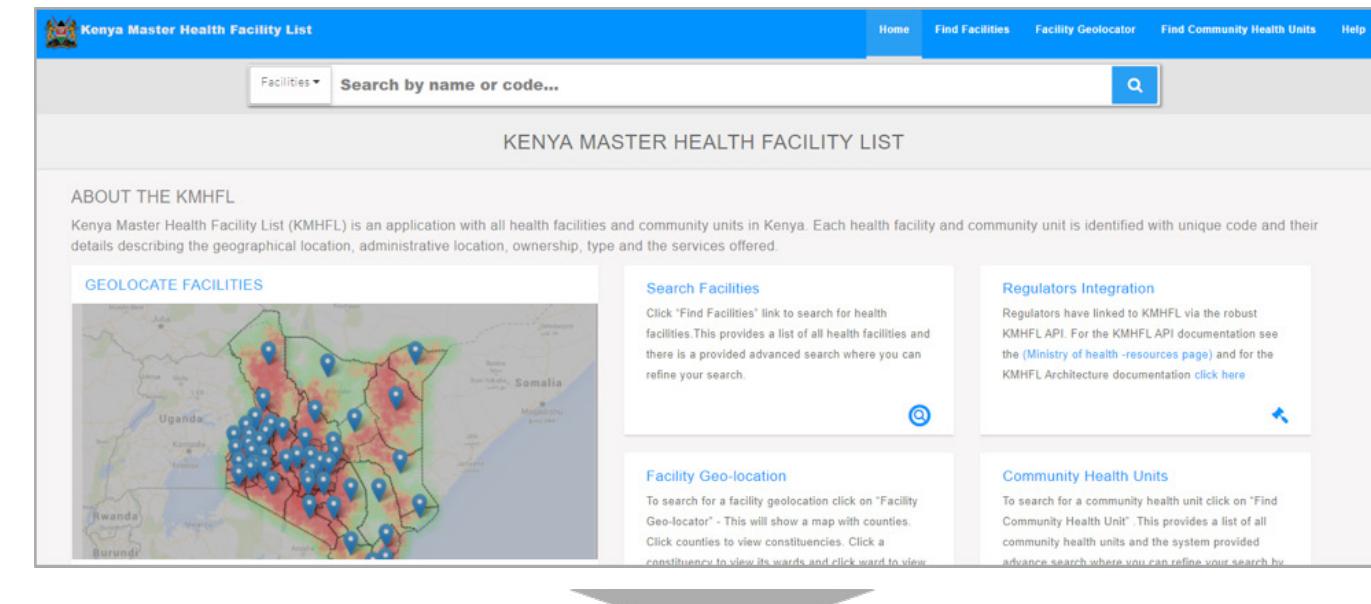
SUMMARY DATA: EXAMPLE
Fig. 3 Number of Health Facilities (NHA – Pakistan) ^{10,24}

Table 38: Public health facilities in Pakistan 2012	
Type	Number
Hospitals	980
Dispensaries	5039
Basic Health Units	5,449
Rural Health Centres	579
MCH Centres	851
TB Clinics	345
Beds in hospitals & dispensaries etc.	90,712
Population per bed	1,665

Step 2: Collect Data

RAW DATA: EXAMPLE

Fig. 3 Kenya Health Facility List



Code	Name	Registration_number	Keph level	Facility type	Owner
23021	Potter's Family Hospital	077263	Level 3	Basic primary health care facility	Private Practice - Clinical Officer
22998	Kaka Medical Clinic	None	Level 2	Dispensaries and clinic-out patient only	Private Practice - Medical Specialist
22997	Makutano Family Health Care Services	None	Level 2	Dispensaries and clinic-out patient only	Private Enterprise (Institution)
22996	Jambo Medical Centre	None	Level 2	Dispensaries and clinic-out patient only	Private Practice - Clinical Officer
22985	KOPANGA DISPENSARY	None	Level 2	Basic primary health care facility	Ministry of Health
22982	Tamba Pwani - Kilifi	None	Level 2	Basic primary health care facility	Non-Governmental Organizations
22977	Fairview Medical Centre	None	Level 2	Secondary care hospitals	Private Practice - Unspecified
22976	RADIANT GROUP OF HOSPITALS-UMOJA	None	Level 3	Secondary care hospitals	Private Practice - Unspecified
22971	Songet Medical Clinic	None	Level 2	Basic primary health care facility	Private Practice - Nurse / Midwifery
22969	KANCO Drop in Centre	None	Level 2	Basic primary health care facility	Non-Governmental Organizations
22967	Vital Solutions Health Centre	None	Level 3	Basic primary health care facility	Non-Governmental Organizations
22966	St.Ann's Catholic Dispensary Kapsorok	None	Level 2	Dispensaries and clinic-out patient only	Kenya Episcopal Conference-Catholic Secretariat
22964	Consalata Hospital Kyeni- Iriamurai Dispensary	007734	Level 2	Dispensaries and clinic-out patient only	Private Enterprise (Institution)
22963	Ankara Medical Centre	None	Level 2	Dispensaries and clinic-out patient only	Private Practice - Clinical Officer
22949	George Mudenyo Kadima Nasio Trust (Mumias West)	None	Level 2	Basic primary health care facility	Private Practice - Clinical Officer
22948	Nazareth Medical Services-Githurai	004231	Level 2	Dispensaries and clinic-out patient only	Private Practice - Clinical Officer
22946	Penda Medical Care-Kahawa West	None	Level 2	Basic primary health care facility	Private Practice - General Practitioner
22938	The Mater Hospital-Satelite Clinic(TRM)	None	Level 2	Basic primary health care facility	Kenya Episcopal Conference-Catholic Secretariat
22936	NAMUNYAK MEDICAL CENTRE	None	Level 2	Dispensaries and clinic-out patient only	Private Practice - Nurse / Midwifery
22934	THE MAZIRA MEMORIAL HOSPITAL	None	Level 2	Dispensaries and clinic-out patient only	Private Company

Service Availability and Readiness Assessment (SARA)

SARAs often provide data on the number of facilities and should be reviewed as a potential data source for this metric. SARAs are described in more detail in section 2.1.3.

2.1.2 PLACE OF INITIAL CARE SEEKING

As defined in section 0.6, the first core PPA metric is the proportion of patients that initiate their care seeking journey at each sector and level of the health system (e.g., public level 1 or private level 0).

This section describes some of the most common data sources that provide information on care seeking patterns, listed in order of priority. As described in the introduction for Step 2: Collect Data Sources, if multiple data sources are available, it's best to prioritize TB-specific data wherever possible. After that, using data on general illness (rather than specific non-TB illness or child illness) should be prioritized.

Step 2: Collect Data

SUMMARY DATA: EXAMPLE

Fig. 5 Care Seeking Data from Prevalence Surveys (Indonesia)¹²

Table 15 Treatment-seeking pattern of participants who reported cough for 14 days or more or haemoptysis

Place of treatment	Male		Female		Total	
	n	%	n	%	n	%
Public hospital	141	2.8	112	3.2	253	3.0
Private hospital	53	1.1	52	1.5	105	1.2
Public primary health facility	440	8.7	485	13.9	925	10.8
Private practitioner/clinic	283	5.6	284	8.1	567	6.6
Nurse or midwife	168	3.3	213	6.1	381	4.5
Pharmacy, drug store, shops etc	1 575	31.1	1 061	30.4	2 636	30.8
Did not seek treatment	2 404	47.5	1 281	36.7	3 685	43.1
TOTAL	5 064	100	3 488	100	8 552	100

National TB Prevalence Surveys

WHO describes prevalence surveys as 'cross-sectional and population-based surveys of a random sample of the population in which the number of people with TB disease in the survey sample is measured'.¹¹ In addition to measuring the prevalence of TB in the population, many prevalence surveys ask questions about where participants sought care when they first recognized symptoms of illness related to TB. The responses to this question provide the most useful care seeking data for a PPA, because they are specific to TB and are often representative of the population at the national level. In some cases, prevalence surveys will also include responses at sub-national levels, although based on the power of the survey, sub-national data should be reviewed with caution.

Final reports for prevalence surveys are often made publicly available in PDF form, but in some cases, requests need to be made to the MoH or NTP to access the report. If possible, PPA teams should request access to the raw underlying data for the care seeking questions included in a prevalence survey. If the raw data are available, the team may need to refer to a data dictionary to explain some of the variables or responses included in the dataset.

Step 2: Collect Data

Health Expenditure and Utilization Surveys (HEUS)

HEUS provide data on patient use of and expenditure in the health care system of a country. Occasionally, these surveys include TB-specific expenditure and utilization data. Often, they include information about where patients seek care for general illness or outpatient care. The latter can be used as a proxy for TB care seeking. In some cases, these surveys are powered at the sub-national level, and if the scope of the PPA exercise is sub-national, the HEUS is a good data source to use for care seeking patterns.

SUMMARY DATA: EXAMPLE

Fig. 6 Care Seeking Data from HEUS Reports (Ethiopia)¹⁴

Table 19: Outpatient Visits by Types of Health Service Providers

Type of Health Service Provider	Urban	Rural	National
Government	59.28%	78.37%	76.52%
Hospital	25.07%	6.03%	7.86%
Health center	28.97%	35.82%	35.16%
Health post	0.74%	28.66%	25.97%
Pharmacy	0.31%	0.37%	0.36%
Kebele health worker	4.19%	7.49%	7.17%
Private	34.54%	18.23%	19.8%
Hospital	8.04%	1.01%	1.68%
Clinic	25.51%	16.32%	17.21%
Pharmacy	0.99%	0.90%	0.91%
NGO	3.13%	1.2%	1.38%
Hospital	0.95%	0.27%	0.33%
Health center	0.65%	0.33%	0.36%
Clinic	1.53%	0.60%	0.69%
Employer organization clinic	0.19%	0.00%	0.02%
Traditional and religious healers	2.74%	2.07%	2.13%
Mobile HIV testing	0.12%	0.00%	0.01%
Not stated	0.00%	0.15%	0.13%
Total	100.00%	100.00%	100.00%

National TB Prevalence Surveys
Where to Find Data
These reports are typically available from the website of a National TB Program, Ministry of Health, or Ministry of Statistics. If they are not available on the website, they could also be available upon request from one of these Ministries. The WHO TB team may also be able to provide reports or data from prevalence surveys, upon request.
Typical File Type
Prevalence surveys commonly include a final report containing the summary data for care seeking data. If the scope of the PPA is sub-national, the raw data files will need to be requested along with a data dictionary for analysis of the data.
Examples
Indonesian National Prevalence Survey (summary data shown in Figure 5), ¹² Nigerian National Prevalence Survey (www.who.int/tb/publications/NigeriaReport_WEB_NEW.pdf) ¹³

Health Expenditure and Utilization Surveys (HEUS)

Where to Find Data

HEUS containing **summary data** are typically available from the website of the Ministry of Health or Ministry of Statistics. If they are not available on the website, they could also be available upon request from one of these Ministries.

Typical File Type

HEUS commonly include a final report which will contain the **summary data** for care seeking. If the scope of the PPA is sub-national, the **raw data** files will need to be requested along with a data dictionary for analysis of the data.

Examples

Ethiopia Health Expenditure Utilization Survey (**summary data** shown in Figure 6)¹⁴, Kenya Household Health Expenditure and Utilization Survey¹⁵ (<https://www.healthpolicyproject.com/index.cfm?id=publications&get=pubID&pubID=745>)

Step 2: Collect Data

SUMMARY DATA: EXAMPLE **Fig. 7** DHS Care Seeking Data for Any Illness (Philippines)¹⁶

Table 13.4 Specific types of health facilities utilized															
Background characteristic	Regional hospital/ medical center		Public			Private			Other			Total	Number		
	Provincial hospital	District hospital	Municipal hospital	RHU/ Urban health center	Barangay health station	Mobile clinic/ other public	Private hospital/ clinic	Private clinic	Other private	Alternative medical	Non-medical	Other/ missing			
Sex															
Male	5.2	4.9	3.0	3.7	18.7	27.6	0.9	19.8	11.8	1.6	1.5	0.8	0.4	100.0	3,260
Female	4.6	4.0	3.3	2.7	17.4	32.0	0.9	19.2	11.5	2.1	1.0	0.8	0.5	100.0	4,221
Age															
0-4	3.8	3.0	2.0	2.8	21.8	38.2	0.4	14.4	9.8	1.4	1.4	0.5	0.4	100.0	2,380
5-29	4.7	4.5	3.6	3.7	19.3	31.2	1.2	15.6	11.3	2.0	1.2	1.0	0.6	100.0	2,511
30-59	6.6	5.4	3.3	3.0	14.0	24.6	0.9	24.8	13.0	2.0	1.2	0.8	0.3	100.0	1,726
60+	5.1	5.7	4.6	2.6	11.7	15.3	1.2	34.3	14.9	2.5	0.7	0.7	0.7	100.0	864
Residence															
Urban	6.3	4.4	2.6	2.6	14.6	24.0	0.8	26.8	13.7	2.4	0.6	0.7	0.5	100.0	3,108
Rural	3.9	4.3	3.6	3.5	20.4	34.4	1.0	14.3	10.2	1.5	1.6	0.8	0.5	100.0	4,374
Region															
National Capital Region	11.1	2.0	2.6	2.5	26.3	14.3	0.8	24.3	12.0	2.5	0.5	0.3	0.9	100.0	848
Cordillera Admin Region	7.4	12.3	4.0	4.2	14.4	15.8	0.0	22.3	15.8	2.1	0.0	1.6	0.2	100.0	163
I - Ilocos Region	3.4	5.1	3.8	1.9	28.5	28.7	0.4	13.7	11.4	1.5	1.1	0.2	0.2	100.0	457
II - Cagayan Valley	4.4	7.5	6.9	6.4	24.4	21.4	0.2	13.4	13.4	0.7	0.2	0.5	0.5	100.0	318
III - Central Luzon	3.8	7.2	3.9	1.5	15.3	22.4	0.9	29.4	14.5	0.9	0.0	0.3	0.0	100.0	836
IVA - CALABARZON	5.5	4.1	2.7	4.7	5.9	34.8	0.3	24.3	13.5	2.1	0.5	1.3	0.3	100.0	953
IVB - MIMAROPA	1.8	5.9	3.6	4.9	23.9	31.7	3.8	13.3	7.7	1.3	2.0	0.0	0.3	100.0	257
V - Bicol	2.7	3.2	1.1	6.0	22.6	32.4	0.6	7.3	11.9	2.9	4.8	3.9	0.5	100.0	625
VI - Western Visayas	2.7	4.1	4.8	1.9	20.3	27.7	1.0	16.6	15.9	3.6	0.3	0.5	0.5	100.0	628
VII - Central Visayas	3.8	3.7	3.7	0.9	20.8	36.5	0.7	23.0	4.0	1.4	0.5	0.0	0.9	100.0	460
VIII - Eastern Visayas	1.0	3.1	3.3	3.1	26.2	22.9	3.6	13.7	13.9	1.8	5.2	0.8	1.5	100.0	377
IX - Zamboanga Peninsula	3.0	3.8	1.1	2.4	10.7	54.9	0.5	17.0	5.6	0.0	0.0	0.5	0.5	100.0	292
X - Northern Mindanao	10.9	4.6	0.9	0.3	12.7	39.6	0.6	22.1	4.6	3.1	0.6	0.0	0.0	100.0	331
XI - Davao	10.1	1.9	4.7	2.7	10.1	30.3	1.2	25.3	12.1	0.8	0.8	0.0	0.0	100.0	260
XII - SOCCSKSARGEN	1.1	2.1	2.1	4.2	5.3	51.6	1.3	15.9	11.4	1.3	2.1	0.8	0.8	100.0	377
XIII - Caraga	2.7	3.7	2.3	2.0	23.3	35.3	0.0	17.7	8.7	2.3	2.0	0.0	0.0	100.0	164
ARMM	5.0	5.1	2.8	7.1	18.9	36.4	0.0	11.1	10.9	0.0	1.1	0.4	1.3	100.0	134

Demographic and Health Surveys (DHS)

The DHS are large, nationally representative surveys that collect a wide variety of information on health, population, and—in some cases—service utilization. They are conducted for many countries at roughly five-year intervals. In each iteration of the DHS, there is typically a survey given to a household, as well as to males and females separately. The data from the female survey are referred to as the individual recode and provide the widest variety of survey questions on care seeking patterns. There are four questions about care seeking that are most commonly included in an individual recode data set. A fifth question is less common, but can be useful for the PPA. These care seeking questions are:

1. Place of care seeking for a child with diarrhea
2. Place of care seeking for a child with fever
3. Place of care seeking for an HIV test
4. Place of care seeking for an antenatal HIV test
5. Place of care seeking for any illness or treatment in the last 30 days, among adults [less common]

If question 5 is available, this is the best DHS metric to use in a PPA as it focuses on adult care seeking patterns (rather than children). However, if it is not available, question 2 can be a reasonable proxy for TB-related care seeking.

Step 2: Collect Data

SUMMARY DATA: EXAMPLE **Fig. 8** Care Seeking Data – NHA (Pakistan)

Table 27: Type of health care provider assessed by the households 2011-12									
Province	Private Hospital	Private Doctor clinic	Homeopath/ ha-keem/ herbalist etc.	Pharmacy/ shops	Govt Hosp/THQ/D HQ/Tertiary/Teaching Hosp	Dispensary/Maternal and child health center/ BHU/RHC	Laboratory	Others*	Total
In % of population									
Punjab	10.71	65.45	7.18	4.76	8.86	2.22	0.16	0.28	100
Sindh	6.84	44.91	0.92	32.50	10.96	0.34	0.08	3.34	100
KP	5.33	37.81	1.94	32.91	15.31	5.56	0.06	0.02	100
Balochistan	7.45	44.23	1.51	25.92	10.41	7.85	0.00	2.46	100
Pakistan	8.36	53.09	4.03	19.56	10.80	2.36	0.11	1.26	100
In % of expenditures									
Punjab	26.54	54.49	2.77	1.24	13.70	0.74	0.03	0.05	100
Sindh	43.75	36.83	0.81	6.45	11.47	0.12	0.00	0.44	100
KP	21.30	42.83	0.84	7.07	18.47	8.40	0.01	0.10	100
Balochistan	18.67	66.20	0.55	4.95	7.38	1.72	0.00	0.30	100
Pakistan	29.36	48.50	1.94	3.55	13.99	2.03	0.02	0.15	100

National Health Accounts (NHA)

In some cases, NHAs will include a care seeking information. The data are the same type of data as outlined in the HEUS description. They should include data about general care seeking at the national level and will in some cases include sub-national data.

National Health Accounts (NHA)
Where to Find Data
NHA reports are typically available from either a Ministry of Health or a Ministry of Statistics website
Typical File Type
NHAs are most regularly available as final reports, containing summary data on care seeking. It may be possible in some cases to download or request the raw data.
Examples
Pakistan National Health Accounts (http://www.pbs.gov.pk/content/national-health-accounts) (summary data shown in Figure 8) ¹⁰

Demographic and Health Surveys (DHS)

<tbl_r cells="1" ix

Step 2: Collect Data

SUMMARY DATA: EXAMPLE **Fig. 9** Example Care Seeking – LSMS (Nigeria)

TABLE 2.12 • Type of Health Facility Visited, among those Reporting any Illness in the Last 4 Weeks (%)

Reason	Hospital		Dispensary		Pharmacy		Chemist		Clinic		Maternity		Traditional		No facility	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
North Central	11.2	49.1	11.2	5.1	1.0	2.5	11.9	10.9	18.8	18.4	1.4	3.2	6.5	4.3	0.0	1.4
North East	12.9	22.2	12.9	17.0	1.4	0.7	22.7	15.8	8.7	11.3	1.7	3.7	5.7	3.7	11.1	10.2
North West	12.5	37.3	12.5	15.4	4.0	1.6	30.8	24.4	2.8	6.8	1.9	1.7	3.1	3.8	0.0	0.9
South East	0.0	27.3	0.0	0.3	1.0	0.3	59.0	63.0	1.9	0.9	1.7	1.6	1.1	1.5	0.0	0.4
South South	0.3	30.7	0.3	1.3	1.1	2.0	49.9	39.6	5.3	5.2	0.0	1.1	5.8	2.6	4.1	4.9
South West	0.7	34.5	0.7	3.4	5.3	6.8	24.3	26.3	3.8	5.3	1.4	2.6	2.1	3.7	0.4	3.7
Urban	0.6	40.0	0.6	3.4	3.9	4.1	34.9	30.8	4.4	5.2	1.3	1.3	2.3	3.1	4.2	3.7
Rural	9.3	29.2	9.3	8.2	1.7	1.3	33.8	34.3	7.1	8.1	1.4	2.7	4.8	3.1	1.2	2.7
NGA	6.0	33.4	6.0	6.3	2.5	2.4	34.2	33.0	6.1	6.9	1.4	2.1	3.9	3.1	2.3	3.1

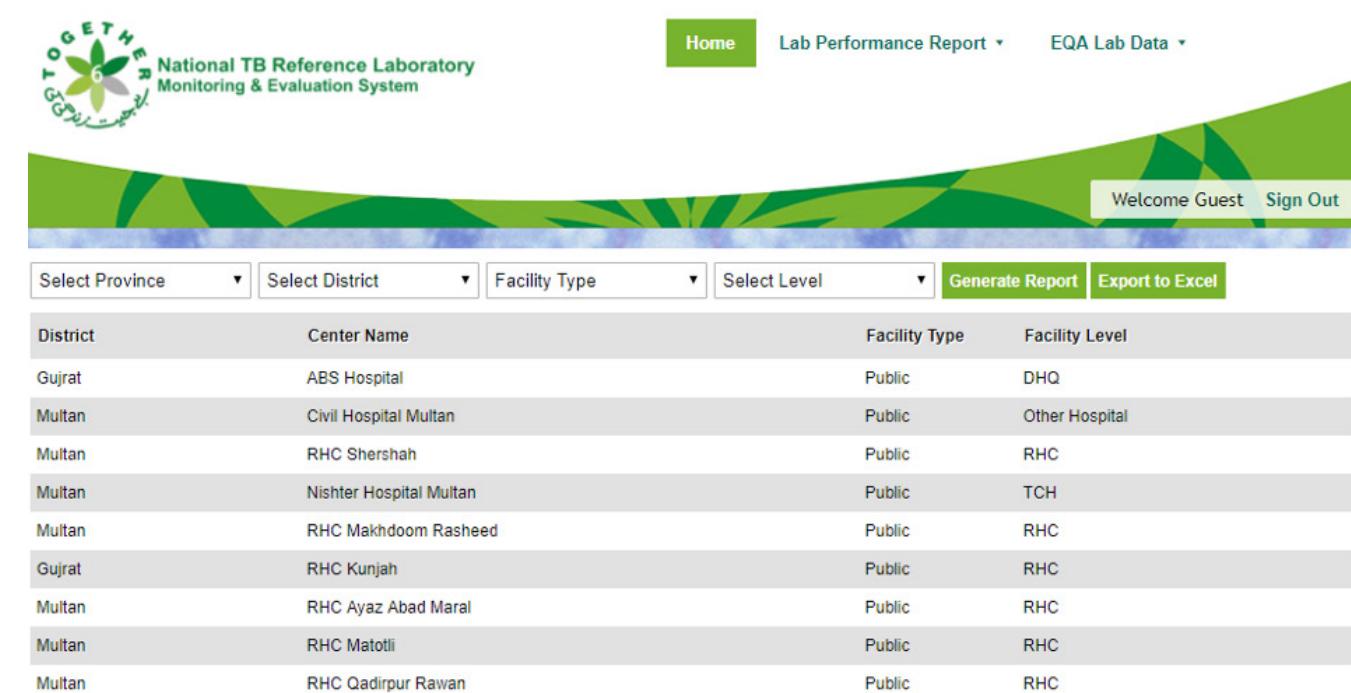
Living Standards Measurement Survey (LSMS)¹⁷

LSMS are similar surveys to the HEUS. They are focused on broader socioeconomic indicators; however, they often provide data on general care seeking patterns at the national level.

Living Standards Measurement Survey (LSMS)	
Where to Find Data	
Most LSMS are supported by the World Bank; reports and final data sets are available here	
Typical File Type	
LSMS are most regularly available as final reports, which typically include tables of summary data for care seeking. The World Bank also provides access to the raw Microdata files if more detailed analysis is needed.	
Examples	
Nigeria General Household Survey (http://microdata.worldbank.org/index.php/catalog/1952) (summary data shown in Figure 9) ¹⁸	

Step 2: Collect Data

RAW DATA: EXAMPLE **Fig. 10** TB Services Coverage – NTP Lab Database (Pakistan)



2.1.3 TB SERVICES COVERAGE

The next metric of the PPA is the percent of health facilities in each health sector and level that have TB services available. Typically, this metric represents the availability of the tool or service within a facility, but does not capture whether there is capacity to use the tool (e.g., diagnose) or the quality of the tool or service. Several common data sources can be used to calculate TB services coverage.

National TB Lab Database

A national TB lab database will provide a list of all the known TB labs in the country, with information about the location, sector, and, typically, the health facility name. These lists include the types of diagnostic tools available, such as microscopy, Xpert, or DST. This information can be used in conjunction with the National Health Facility Master List (HFML) to calculate TB services coverage. The lab lists will often include information about whether treatment initiation or TB drugs are also available in this facility.

National TB Lab Database
Where to Find Data
Most often, lab databases will need to be accessed by a staff member of the NTP, but in some cases, these databases are publicly available (see Figure 10)
Typical File Type
Most websites have an option to download the entire list rather than just view it on the website. If possible, it is best to access an Excel or CSV file that lists all of the facilities in the database.
Examples
Pakistan NTP Reference Lab Database (http://mdrtb.ntp.gov.pk/nrl/new_report.php) (raw data shown in Figure 10) ¹⁹

Step 2: Collect Data

Service Availability and Readiness Assessment (SARA)²⁰ or Service Provision Assessment (SPA)
 SARA and SPA are both assessment tools that capture the availability of health services among health facilities across a country. TB services, both diagnosis and treatment, are commonly included in both types of surveys. SARAs can be conducted as either a census or a representative survey, and SPAs are conducted as a representative survey of health facilities in a country. Both tools will often report coverage of services at a sub-national level, but may not provide sub-national breakdown by facility type in the final report. If the team is developing a sub-national PPA, it will need to request the raw data for the SARA or SPA to complete the sub-national analysis.

Service Availability and Readiness Assessment (SARA) or Service Provision Assessment (SPA)													
Where to Find Data													
SPA reports and raw data are available at the DHS program website (http://dhsprogram.com/What-We-Do/Survey-Types/SPA.cfm). SARA reports are available from WHO (http://www.who.int/healthinfo/systems/sara_reports/en/). Access to raw data for SARAs needs to be requested from WHO or may be available from the Ministry of Health or Ministry of Statistics in the relevant country.													
Typical File Type													
Summary reports are produced for both SPAs and SARAs. For more detailed analysis, the raw data should be requested from the managing organizations.													
Examples													
Ethiopia SARA, ²¹ Kenya SARAM ²² (summary data shown in Figure 11)													

SUMMARY DATA: EXAMPLE

Fig. 11 TB Services Coverage – SARA (Ethiopia and Kenya)

Table 3.5. 1 Percentage of facilities that offer tuberculosis services, by region, Facility type, Managing authority, and residency, Ethiopia SARA 2016 (N=547)

Category	Offers TB services	TB diagnosis	TB diagnostic testing	TB diagnosis by clinical symptoms	TB diagnosis by sputum smear microscopy	TB diagnosis by culture	TB diagnosis by rapid test (Gene Xpert MTB/RIF)	TB diagnosis by chest X-ray	Prescription of drugs of TB patients	Provision of drugs to TB patients	Management and treatment follow-up for TB patients	Total number of facilities
Regions												
Tigray	80	80	76	46	76	7	7	24	62	55	67	42
Afar	73	66	58	53	58	5	2	23	55	55	55	38
Amhara	62	57	48	39	48	1	10	19	50	57	57	61
Oromiya	69	66	59	66	59	4	5	12	60	61	61	99
Somali	73	68	67	68	67	0	1	9	60	64	64	43
Beni-Gumuz	55	55	55	55	55	0	6	5	47	47	51	30

Table 67: Percentage of facilities having the different TB tracer commodities, by County

County	Total number of facilities providing service	Mean availability of tracer items	Percent of facilities with respective tracer items for TB commodities						
			Streptomycin 1gm Injection (for re-treatment)	Isoniazid	Rifampicin	Isoniazid + Rifampicin (2FDC)	Isoniazid + Rifampicin + Pyrazinamide (RHZ) (3FDC)	Isoniazid + Rifampicin + Ethambutol (RHE) (3FDC)	Isoniazid + Rifampicin + Pyrazinamide + Ethambutol (4FDC)
Baringo	174	14%	13%	7%	6%	17%	19%	18%	21%
Bomet	110	23%	22%	17%	15%	24%	27%	23%	31%
Bungoma	121	45%	53%	36%	27%	49%	43%	45%	60%
Busia	74	46%	49%	35%	27%	51%	47%	54%	57%
Elgeyo-Marakwet	119	14%	15%	10%	10%	16%	15%	16%	18%
Embu	99	43%	52%	31%	20%	42%	55%	46%	56%
Garissa	79	24%	25%	16%	20%	25%	27%	29%	29%
Homa Bay	158	59%	56%	57%	42%	62%	66%	62%	67%
Isiolo	37	36%	16%	30%	30%	35%	51%	35%	54%
Kajiado	156	13%	17%	9%	7%	13%	15%	13%	15%

Step 2: Collect Data

2.1.4 LOCATION OF NOTIFICATION BY SECTOR

After calculating the coverage of TB services, the next core metric to calculate is the location of case notification for patients who are notified to the NTP. The denominator of this metric is the total estimated annual incidence of TB, which is also available from WHO. Unless sub-national incidence estimates are available, this metric cannot be calculated at the sub-national level.

National Electronic TB Surveillance System

National TB programs maintain a database of all cases notified to the program and start on treatment. In some cases, this database will provide records indicating the specific facility where a patient was treated. These data can then be used to calculate the numerator of treatment location for notified patients, segmented by sector and level. If these data are available, they should be the priority source used to calculate the numerator of this PPA metric.

National Electronic TB Surveillance System

This database is not typically publicly available and will need to be requested from the NTP. If individual patient records are provided, the team should ensure the data are properly secured.

Typical File Type

Raw data from the NTP case management system will likely need to be extracted from a central database and put into a CSV file.



TIP

When using the WHO Global TB Database, the team should also download the data dictionary which provides helpful context and information about the variables in the other data sets.

WHO Global TB Database

If the data are not available from the NTP, WHO reports national-level notification data segmented by sector. WHO will also provide national-level estimates of TB incidence.

WHO Global TB Database

Where to Find Data

The data on notifications and burden estimates are available for download from the WHO Global TB database (<http://who.int/tb/country/data/download/en>).²³ Incidence estimates are available in the *burden estimates* file, notifications by sector are available in the *TB policies and services* file, and overall notifications are available in the *case notifications* file.

Typical File Type

The WHO Global TB Database provides CSV files for download.

2.1.5 TREATMENT OUTCOME

The final metric of the PPA is a calculation of the number of patients who were successfully treated and those who did not successfully complete treatment; these are taken from the most recent annual cohort of notified patients. This metric is also calculated with the total annual incidence as the denominator, so this metric will include a portion of patients that do not have a treatment outcome or are ‘missing.’

National Electronic TB Surveillance System

Similar to the previous metric, a national electronic TB surveillance system should report treatment outcomes for those patients who were started on treatment in the last year. These data can then be segmented by sector and level as needed for the PPA.

National Electronic TB Surveillance System

Where to Find Data

This database is not typically publicly available and will need to be requested from the NTP. If individual patient records are provided, the team should ensure the data are properly secured.

Typical File Type

Raw data from the NTP case management system will likely need to be extracted from a central database and put into a CSV file.

Step 2: Collect Data

WHO Global TB Database	
Where to Find Data	
The data on notifications and burden estimates are available for download from the WHO Global TB Database (http://who.int/tb/country/data/download/en/). ²³ Incidence estimates are available in the burden estimates file and treatment outcomes are available in the treatment outcomes file.	
Typical File Type	
The WHO Global TB Database provides CSV files for download.	

WHO Global TB Database

If the data are not available from the NTP, WHO reports national-level treatment outcomes data. WHO will also provide national-level estimates of TB incidence.

2.2 REVIEW DATA SOURCES FOR SUPPLEMENTAL METRICS

As described in section 0.6, the supplemental metrics include information that can provide helpful context to the core metrics of the PPA, but for which data are less commonly available. Potential data sources for these supplemental metrics are listed in the table below.

Supplemental Metric	Potential Data Source
Metrics to Include in the PPA	
Location of Diagnosis by Sector/Level	» TB Prevalence Survey » General Health Survey (e.g., Riskesdas in Indonesia) » TB Laboratory Data for public sector
Location of Treatment by Sector/Level	» TB Prevalence Surveys » General Health Surveys (e.g., Riskesdas in Indonesia) » TB Facility Data / routine reporting for public sector
Metrics to Collect in Support of PPA Interpretation	
Percent of Patients Who Do Not Seek Care	» TB Prevalence Survey » Demographic and Health Survey (DHS) » Health Expenditure and Utilization Survey (HEUS)
Reasons for Not Seeking Care	» Demographic and Health Survey (DHS) » Health Expenditure and Utilization Survey (HEUS) » Published studies (may be sub-national)
Reasons for Initial Care Seeking Patterns	» Demographic and Health Survey (DHS) » Health Expenditure and Utilization Survey (HEUS) » Living Standards Measurement Survey (LSMS)
Delays Between Pathway Steps	Very few national surveys capture delays in care seeking between different steps of a patient's journey. Most often, these data will come from individual cohort studies done at a sub-national level. The PPA team should conduct a literature review of available studies to see what delay information is available. <i>Common search terms include:</i> patient care seeking, delays in care seeking, TB system delays, diagnostic delay, treatment delay, patient delays.
Referral Patterns Between Sector and Level	Sometimes patient records capture the referring facility. Very few national surveys capture referral patterns or practices between health facilities. Individual studies that interview patients or providers can provide useful information on the practices of referral and sputum transport between different health facilities. <i>Common search terms include:</i> provider referral, referral patterns, care seeking patterns, referral systems

Step 2: Collect Data

Table 4 Data Source Tracking Table (Example)

PPA Metric	Data Source	Available?	Data File Type	Priority?
Number of Health Facilities	National Health Facility Master List			
	National Health Accounts			
	Service Availability and Readiness Assessment			
	Other			
Place of Initial Care Seeking	TB Prevalence Survey			
	Health Expenditure and Utilization Survey			
	Demographic and Health Survey			
	Living Standards Measurement Survey			
	National Health Accounts			
	Other			
TB Services Coverage	National TB Lab Database			
	Service Availability and Readiness Assessment			
	Service Provision Assessment			
	Other			
Location of Notification	National TB Surveillance Database			
	WHO Global TB Database			
Treatment Outcomes	National TB Surveillance Database			
	WHO Global TB Database			

2.3 ACCESS AND SAVE DATA RESOURCES

The team will need to identify a data source for each core PPA metric. Some of the data sources may require making requests to partners or other organizations in order to access the raw data. As a result, accessing the data can take some time. Once the team has access to all of the data sources for the PPA, save them in a shared location for the team to use throughout the rest of the PPA. If multiple data sources were found that could be used for a given PPA metric, agree on the rationale for selecting one data source over another (see the introduction for Step 2 for considerations) and select the priority data source for that metric.

Table 4 will help track which data sources are available and the format in which they are available. Under each metric, one data source should be prioritized based on its relevance to TB or its use in a sub-national PPA. At a minimum, a PPA needs to have care seeking and service availability data. If this data is not available, it will not be possible to do even a high-level PPA. If data are missing for any of the other core or supplementary metrics, those metrics can be excluded from the PPA.

2.4 STEP 2 - CHECKLIST

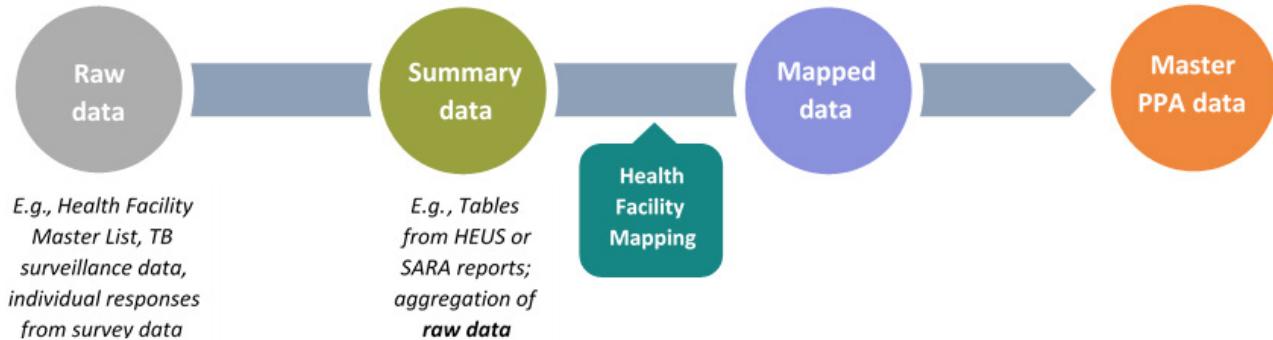
During Step 2, each of the following activities should be completed:

STEP 2 CHECKLIST

- Identify data sources for each of the core PPA metrics
- Save each data source (PDF report or raw data file) in a PPA folder (shared with all members of the team)
- Where multiple data sources are available for a given metric, make sure that one data source has been prioritized

Step 3: Organize Data and Calculate Metrics

Fig. 12 Process of Data Organization and Calculation



Step 3 will navigate the process of turning the **raw** and **summary data** sources collected in Step 2 into a final PPA Excel Workbook that can be used to create PPA visuals. There are four overarching parts to Step 3, captured in Figure 12:

1. Summarize **raw data** that is not already summarized. This can be done in Excel, R, or any other software with which the team is comfortable and which is suited to the type of analysis needed. This part may be skipped if the team is working only with **summary data** from the start.
2. Input **summary data** into an Excel workbook.
3. Map the health facility types from the **summary data** to the health sectors and levels that are described in section 0.5.
4. Calculate the core metrics needed for the PPA (as described in section 0.6). This will result in a 'Master Data Sheet' that will be used to create the PPA visuals described in Step 4.

3.1 INTRODUCTION TO THE PPA EXCEL WORKBOOK TEMPLATE

The PPA Excel Workbook Template provides an example of a useful way to structure the data, and helps organize the analysis for a PPA. For users who are more comfortable in other data management software (e.g., R or Python), each of these steps can be replicated outside of Excel. The PPA Excel Workbook Template described in this section includes tips and instructions directly in the workbook.

Open the PPA Excel Workbook Template and review the four main types of worksheets included:

Raw Data Sheets: The first type, in gray, are examples of sheets that can optionally store the **raw data** collected in Section 2. These sheets contain the original versions of the **raw data** that will be used in the PPA. After importing the data, the PPA team will not spend much time working in these sheets, aside from any data cleaning that the team opts to do in Excel. If the data set requires cleaning, make sure that the clean version is stored in the **raw data sheet** (i.e., misspellings are corrected, duplicate entries are removed, etc.).



TIP

The PPA Excel Workbook Template is available for download at:

<http://linksbridge.com/work/tb-ppa/>

The PPA Excel Workbook Template includes examples of raw data, summary data, mapped data, a lookup table and a master PPA sheet.

Step 3: Organize Data and Calculate Metrics

Summary data sheets will pull data from the clean raw data source sheets, leaving the integrity of the raw data preserved. Some examples of data that could go into **raw data** sheets include a Health Facility Master List (section 2.1.1) and TB lab records (section 2.1.3). Note that some **raw data** sets may be too large to work with comfortably in Excel (such as an entire DHS or HEUS data set). It may be possible to subset to only the **raw data** needed and use Excel, or use alternate software such as R.

Summary Data Sheets: The second type of data sheet, in **green**, are used to store the data that is summarized by health facility type (see the Tip box in section 0.5 for reference). At minimum, a **summary data** sheet will be required for core metrics 1, 2, and 4: care seeking, diagnostic coverage, and treatment coverage. A **summary data** sheet may also be used for number of health facilities or some supplemental metrics. There are three ways to arrive at a **summary data** sheet:

1. Summarize the **raw data** outside of Excel. In this case the PPA team would input the resulting **summary data** table directly into a green **summary data** sheet. For example, appendix 6 provides detailed instructions for summarizing DHS data in R. The output of such an exercise may be included in the PPA Excel Workbook as a **summary data** sheet for Care Seeking (core metric 1).
2. Summarizing **raw data** from a gray **raw data** source sheet contained within the PPA Excel Workbook, as referenced in the paragraph above.
3. Utilize data that is already summarized, as indicated in section 2.1. In this case the PPA team would input the data directly to a green **summary data** sheet. An example of such a scenario would be using place of initial care seeking data from national prevalence surveys, referenced in section 2.1.2.

Mapped Data Sheets: The third type, in **purple**, are sheets that will be used to translate the **summary data** aggregated by health facility type into data aggregated by health sector and/or level, as required for the PPA. Core metrics 1, 2, and 4 (and possibly number of health facilities, or other metrics) will have a **mapped data** sheet that references the corresponding **summary data** sheet. The **mapped data** sheet may reference a **lookup table** to assign health facility types from the **summary data** sheet to a health sector and level.

For core metrics 6 and 7, data from a report will likely already be summarized by health sector, and it is not a requirement that they are further segmented by level. In this case, the team may input data from a report directly into a **mapped data** sheet. If the team has **raw data** for metrics 6 or 7 (such as patient treatment records), it might be possible to skip the **summary data** sheet if the **raw data** contains a field for health sector. In this case, the team can map and summarize the data in one step within the **mapped data** sheet and there is no need to reference a **lookup table**.

Master PPA Data Sheet: The fourth type, in **orange**, is the **master PPA data** sheet. Included in this sheet are all the metrics from the **mapped data** sheets brought together into a single sheet. In addition, core metrics 3 and 5 are calculated based on core metrics 1, 2, and 4. The purpose of the **master PPA data** sheet is to arrange the final data in one place such that it is easy to work with in Tableau or other data visualization software.

Lookup Tables: Finally, other reference sheets or **lookup tables**, such as those in **teal**, might be used to organize or categorize the data. The facility mapping sheet is an example of a **lookup table**, which is used by formulas in this workbook map health facility types to the health sector and level to which they belong.

3.2 RAW DATA SHEETS

The first step in creating a PPA Excel Workbook is to put all the **raw data** sources that the team wishes to collect, if any, into the PPA Excel Workbook. Using Table 4 as a reference for which data sources are needed, create a worksheet/tab for each **raw data** source. As seen in the PPA Excel Workbook Template, it is helpful to color code the **raw data** source worksheets as gray tabs to keep them clearly separated from other types of worksheets. If the data source is in the form of a raw Excel or CSV file, it is best to copy and paste the whole Excel sheet from the **raw data** source into its own tab in the PPA Excel Workbook. If the team has access to the **raw data** from a large survey, the team will likely need to summarize only the data needed before bringing them into Excel as **summary data**, since the **raw data** files are typically larger than Excel can handle. An example of this process is described using DHS data in the Appendix (section 6.1).

Step 3: Organize Data and Calculate Metrics

Raw data sources containing health facilities, such as health facility master lists (HFML), list the proper names of all facilities (e.g., Jakarta Municipal Hospital or South Nairobi Primary Health Clinic). If the **raw data** does not contain a field (column) for health facility type or health sector, then one—or likely both—of these fields will need to be created. Each individual facility will need to be coded (depending on which metric the **raw data** applies to).

3.3 SUMMARY DATA SHEETS

Any metric that needs to be summarized by health facility type will need to be included in a **summary data** sheet within the PPA Excel Workbook. Usually, this would include core metrics 1, 2, and 4 of the PPA, and possibly others. Each **summary data** sheet will represent a data source (because each data source has a unique way of categorizing health facility types). Thus, multiple metrics may be represented on one sheet if the data come from the same source, and/or one metric may be split across two sheets if it has two components from different data sources. The latter would be the case if, for example, core metric 2 (diagnostic coverage) incorporated data on Xpert coverage from one source and microscopy coverage from another source.

If you have summarized data by health facility type outside of excel, you may paste the resulting table into your PPA Excel Workbook as a **summary data** sheet. Otherwise, you will need to follow one of the two processes that follow: Summarizing **raw data** within the PPA Excel Workbook, or pasting **summary data** from a report directly into the PPA Excel Workbook.

3.3.1 SUMMARIZE RAW DATA WITHIN THE PPA EXCEL WORKBOOK

Raw data sources collected and stored in the PPA Excel workbook may need to be summarized by health facility type in a **summary data** sheet. Some exceptions are discussed at the end of this section. Each **raw data** source for the PPA is either a survey or census containing records of unique health facilities or people. Sources typically include a variable for the type of health facility (e.g. the type of health facility did a patient first visit for illness). This is essential for generating the **summary data** sheet.



TIP

Refer to section 0.6 for definitions of each of the core metrics for the PPA. Each of the definitions included in that section provide the calculation for the core metrics, including the numerator and denominator that will need to be acquired from the raw data.

For a subnational PPA, the other essential variable is subnational geographic unit such as region, province, county, or urban/rural designation.

Within the PPA Excel Workbook, the first column for all **summary data** sheets is always the sub-national unit (such as region) and the second column is always health facility type. Note that each **summary data** sheet has a slightly different (or perhaps dramatically different) set of health facility types represented. This is to be expected; the fact that each **raw data** source categorizes health facility types differently is the reason for the health sector and level standardization process described in section 0.5, which is applied next in the **mapped data** sheet.

Each **summary data** sheet corresponds to at least one metric, and optionally may correspond to more than one metric if a single **raw data** source supplies data for more than one metric. In this case, one could optionally store all **summary data** pertaining to a single **raw data** source in a single **summary data** sheet by simply adding additional columns for additional metrics. For example, this might apply if a data source on lab records provided data on microscopy as well as Xpert.

If the scope of the PPA is national, then the number of rows in each **summary data** sheet should equal the number of health facility types represented. Optionally, in the case of a sub-national PPA, the **summary data** sheet will segment the data according to sub-national unit, such as region, and each sub-national unit will have a row for each facility type. In that case, each **summary data** sheet must reference the same set of sub-national units. If running a sub-national PPA, it is useful to also include a line for the national level data in the **summary data** sheets as well. This is shown in the PPA Excel Workbook.

Step 3: Organize Data and Calculate Metrics

SUMMARY DATA: EXAMPLE

Fig. 13 Copying care seeking data from a report to a worksheet in the PPA Excel Workbook

TABLE 3.3: PERCENT DISTRIBUTION OF HEALTH CONSULTATIONS IN PAST 2 WEEKS BY TYPE OF HEALTH PROVIDER CONSULTED.									
PROVINCE & DISTRICT	HEALTH PROVIDER / CONSULTED								
	PRIVATE DISP/HOSP	PUBLIC DISP/HOSP	RHC/ BHU	HAKEM/ HERBALIST	HOMEOPATH	CHEMIST/ PHARMACY	SAINA/ SAINI	OTHER	
Pakistan	Total	70.28	20.34	3.40	1.47	0.51	3.20	0.54	0.27
	Urban	76.45	19.15	0.36	0.87	0.65	1.95	0.42	0.13
	Rural	66.04	21.15	5.48	1.88	0.41	4.06	0.61	0.37

TABLE 3.3: PERCENT DISTRIBUTION OF HEALTH CONSULTATIONS IN PAST 2 WEEKS BY TYPE OF HEALTH PROVIDER CONSULTED.									
PROVINCE & DISTRICT	HEALTH PROVIDER / CONSULTED								
	PRIVATE DISP/HOSP	PUBLIC DISP/HOSP	RHC/ BHU	HAKEM/ HERBALIST	HOMEOPATH	CHEMIST/ PHARMACY	SAINA/ SAINI	OTHER	
Pakistan	Total	70.28	20.34	3.40	1.47	0.51	3.20	0.54	0.27
	Urban	76.45	19.15	0.36	0.87	0.65	1.95	0.42	0.13
	Rural	66.04	21.15	5.48	1.88	0.41	4.06	0.61	0.37

A	B	C	D	E	F	G	H	I	J	
1	Region	Geograph	PRIVATE DISP/HOSP	PUBLIC DISP/HOSP	RHC/BHU	HAKEM/HERBALIST	HOMEOPATH	CHEMIST/PHARMACY	SAINA/SAINI	Other
2	Pakistan	Total	70.28	20.34	3.4	1.47	0.51	3.2	0.54	0.27
3		Urban	76.45	19.15	0.36	0.87	0.65	1.95	0.42	0.13
4		Rural	66.04	21.15	5.48	1.88	0.41	4.06	0.61	0.37
5	Punjab	Total	77.23	16.95	1.28	1.79	0.48	1.58	0.35	0.34
6		Urban	76.65	20.39	0.05	1.08	0.41	0.97	0.24	0.22
7		Rural	77.55	15.00	1.98	2.19	0.52	1.93	0.42	0.41
8	Islamabad	Total	68.19	30.84	0	0	0.38	0.59	0	
9		Urban	57.34	42.66	0	0	0	0	0	
10		Rural	79.64	18.38	0	0	0	0.78	1.21	0
11	Attock	Total	75.18	21.86	0.71	0.93	0.75	0	0	0.57
12		Urban	73.4	23.91	0	1.25	1.45	0	0	0
13		Rural	77.07	19.68	1.48	0.6	0	0	0	1.17
14	Rawalpindi	Total	74.28	24.84	0.18	0	0.7	0	0	0
15		Urban	78.43	20.42	0	0	1.15	0	0	0
16		Rural	67.93	31.62	0.46	0	0	0	0	0
17	Jehlum	Total	74.45	24.53	0	0	1.01	0	0	0
18		Urban	72.66	27.34	0	0	0	0	0	0
19		Rural	75.44	22.99	0	0	1.57	0	0	0
20										

Not all raw data must be or should be summarized in a **summary data** sheet. One such case is when the PPA does not require the health facility level for a metric, such as for core metrics 6 and 7. The second is the case where the **raw data** does not come with a field for health facility type. In both cases the analysis process can bypass **summary data**, and move directly to **mapped data**. If the PPA team must add new fields to the **raw data**, they can create fields for health sector and level rather than facility type, and aggregate data based on these fields as needed in the **mapped data** sheet (aggregate by sector and level, or by sector only). Sometimes this bypass is required.



TIP

If the team has access to Adobe Acrobat Pro, it will be easier to copy and paste the data from a PDF file into Excel. A feature in Adobe Acrobat Pro called 'Copy With Formatting' will keep the table formatting of the data when it's copied out of the report, and will make the data more structured when it's pasted into Excel.

Step 3: Organize Data and Calculate Metrics

Fig. 14 Example of Health Facility Type (in red boxes)

Place of treatment	Male		Female		Total	
	n	%	n	%	n	%
Public hospital	141	2.8	112	3.2	253	3.0
Private hospital	53	1.1	52	1.5	105	1.2
Public primary health facility	440	8.7	485	13.9	925	10.8
Private practitioner/clinic	283	5.6	284	8.1	567	6.6
Nurse or midwife	168	3.3	213	6.1	381	4.5
Pharmacy, drug store, shops etc	1 575	31.1	1 061	30.4	2 636	30.8
Did not seek treatment	2 404	47.5	1 281	36.7	3 685	43.1
TOTAL	5 064	100	3 488	100	8 552	100

Table 2.19 Choice of health service among those who first seek care												
Where did those that consulted health services first seek care?												
	All	%	Male	%	Female	%	Rural	%	Semi-urban	%	Urban	%
Health centre/PHC	163	14%	61	37%	102	63%	81	50%	8	5%	74	45%
Private hospital	130	11%	56	43%	74	57%	63	48%	4	3%	63	48%
Traditional centre	11	1%	8	73%	3	27%	6	55%	0	0%	5	45%
Chemist	319	28%	146	46%	173	54%	138	43%	5	2%	176	55%
General hospital	419	37%	171	41%	248	59%	158	38%	40	10%	221	53%
Teaching hospital	46	4%	23	50%	23	50%	15	33%	0	0%	31	67%
Mission hospital	42	4%	19	45%	23	55%	30	71%	2	5%	10	24%
Other	9	1%	3	33%	6	67%	2	22%	1	11%	6	67%
Missing	3	0%	1	33%	2	67%	2	67%	0	0%	1	33%
Total	1 142	100%	488	43%	654	57%	495	43%	60	5%	587	51%

For example, this could be the case when calculating core metrics 2 and 4, coverage of TB diagnostics and treatment, if the data on the numerator (number of facilities with a service) comes from a different **raw data** source than the data for the denominator (total number of facilities).

3.3.2 PASTE SUMMARY DATA DIRECTLY INTO THE PPA EXCEL WORKBOOK

If the data source is already in the form of **summary data**, the data can be copied and pasted from the data source into its own **summary data** sheet of the PPA Excel Workbook (Figure 13).

3.4 MAPPED DATA SHEETS

The data mapping presented in this section will support the calculations described in section 0.6 and will mirror the examples in the purple tabs of the PPA Excel Workbook Template.

3.4.1 CREATE HEALTH FACILITY MAPPING SHEET

Step 3: Organize Data and Calculate Metrics

LOOKUP TABLE: EXAMPLE

Fig. 15 Including a list of all facility types from all data sources into the health facility mapping table of the PPA Excel Workbook

Data Source	Health Facility Type	Health Facility Sector	Health Facility Level
[manually enter each data source used in the PPA]	[manually enter unique facility types from each data source]	[manually enter the sector for each facility type, in some cases this will already be provided in the data source]	[manually enter the sector for each facility type, in some cases this will already be provided in the data source]
Health Facility Master List	Barangay health station		
Health Facility Master List	Rural Health Unit		
Health Facility Master List	Infirmary		
Health Facility Master List	Hospital		
Health Facility Master List	Hospital		
Health Facility Master List	Infirmary		
Health Facility Master List	Clinic		
Health Facility Master List	DepEd Clinic		
Health Facility Master List	Provincial Health Office		
Health Facility Master List	Psychiatric Care Facility		

This health facility mapping lookup table will be used as a reference for the rest of the PPA calculations. Standard categorizations are needed to enable the comparison of data across different sources as described in section 0.5.

Start by creating a new sheet in the PPA Excel Workbook, and name it 'Health Facility Mapping.' There should be four columns on this sheet. The first column should be titled 'data source', and will include a list of all the data sources used in the PPA that will require mapping. The second column should be titled 'health facility type'. This is where each of the health facility types that were included in each data source will be listed. There will be multiple health facility types per data source. The third and fourth columns are where each of the health facility types in the second column will be manually categorized by sector and level.

To complete the first two columns, go through each of the data sources and find the list of health facility types used by the source. When the data source is a PDF report, the list of health facilities should be listed in the report (Figure 14). These health facility types should be included in column two of the **facility mapping lookup table**, with the corresponding name of the data source included in column one.

Once this step is complete, two of the four columns of the **facility mapping lookup table** will be filled out (see Figure 15).

The next step is to go through each of the health facility types and categorize them by health sector and level (e.g., level 0, level 1, level 2) as defined in section 0.5. Many raw data sources will already provide a categorization of whether the health facility type is in the public or private sector.

Unless the analysis team has reason to believe that the existing categorization is incorrect, these sector categorizations should be used for the PPA. The team will need to determine whether to classify NGO facilities as public or private. It can also be useful to differentiate private sector facilities between informal and formal providers. Typically, informal providers are those with little or no formal medical training, such as traditional healers and drug shops.

The process of categorizing health facility types is an important step of the PPA process. The rest of the visuals and metrics calculations will be based on the categorizations completed here. Thus, it is important to make sure that those who know the health system well are consulted when categorizing each of the facilities.

After this step is completed, the **facility mapping lookup table** should be completely filled in for all of the relevant data sources (Figure 16). An example of the **facility mapping lookup table** is included in the PPA Excel Workbook Template.



TIP

It can be useful to go through the process of creating a **facility mapping lookup table** with a group. If there is someone in the PPA team who knows the country health system well, they are important to include in this process.

Step 3: Organize Data and Calculate Metrics

LOOKUP TABLE: EXAMPLE

Fig. 16 Completed Health Facility Mapping Lookup Table

Data Source	Health Facility Type	Health Facility Sector	Health Facility Level
[manually enter each data source used in the PPA]	[manually enter unique facility types from each data source]	[manually enter the sector for each facility type, in some cases this will already be provided in the data source]	[manually enter the sector for each facility type, in some cases this will already be provided in the data source]
Health Facility Master List	Barangay health station	Public	Level 0
Health Facility Master List	Rural Health Unit	Public	Level 1
Health Facility Master List	Infirmary	Public	Level 1
Health Facility Master List	Hospital	Public	Level 2
Health Facility Master List	Hospital	Private	Level 2
Health Facility Master List	Infirmary	Private	Level 1
Health Facility Master List	Clinic	Private	Level 1
Health Facility Master List	DepEd Clinic	Public	Level 1
Health Facility Master List	Provincial Health Office	Public	Level 1
Health Facility Master List	Psychiatric Care Facility	Private	Level 1
Demographic and Health Survey	Regional hospital/medical center	Public	Level 2
Demographic and Health Survey	Provincial hospital	Public	Level 2
Demographic and Health Survey	District hospital	Public	Level 2
Demographic and Health Survey	Municipal hospital	Public	Level 2
Demographic and Health Survey	RHII/Irruan Health Center	Public	Level 1

3.4.2 ADD HEALTH SECTORS AND LEVELS TO SUMMARY DATA SHEETS

After creating the **facility mapping lookup table**, go back to the **summary data** sheets. Add a column for health sector and a column for level. Each facility type represented in a **summary data** sheet should be assigned to the health sector and level designated in the **lookup table**. Alternatively, columns for health sector and level may be added to the **raw data**, and each health facility captured in the **raw data** assigned a sector and level. (Some data sources may already have a field for health sector). Consult the PPA Excel Workbook Template for examples.

Either method will enable the PPA team to aggregate the data by health sector and level. It is up to the team to determine which method will best suit their needs. Considerations include: What fields are already available in the **raw data**? What variables have already been summarized in existing **summary data**? Has the PPA team decided to skip the **summary data** step and move straight to **mapped data**?

3.4.3 CREATING MAPPED DATA SHEETS

This section guides the team through the process of turning **summary data** into **mapped data**, aggregated by health sector and level. Each core metric in the PPA should have a mapped data sheet, organized according to the sectors and levels that were created in section 3.4.1. When the analysis team is undertaking a sub-national level PPA, these **mapped data** sheets will present data aggregated at the sub-national area of the analysis (e.g., region or province).

To start, create a new (blank) sheet for each of the core metrics in the PPA, label them by the name of the metric, and change the color to **purple**. The team should have one sheet each for:

- Number of health facilities
- Place of initial care seeking
- Coverage of TB diagnostic services
- Coverage of TB treatment services
- Location of notification by sector
- Treatment outcomes

Note that the metrics 'access to diagnosis at initial care seeking' and 'access to treatment at initial care seeking' do not have their own **mapped data** sheet. These metrics are calculated within the final **master PPA data** sheet.

After creating a blank sheet for each of the PPA core metrics, the PPA team will need to develop the basic structure of each sheet.

Step 3: Organize Data and Calculate Metrics

MAPPED DATA SHEET: EXAMPLE			
Fig. 17 Mapped Data Sheet Structure			
Region	Health Facility Sector	Health Facility Level	
[manual entry]	[copied from Health Facility Mapping]	[copied from Health Facility Mapping]	
Region A	Private	Level 1	
Region A	Private	Level 2	
Region A	Public	Level 0	
Region A	Public	Level 1	
Region A	Public	Level 2	
Region B	Private	Level 1	
Region B	Private	Level 2	
Region B	Public	Level 0	
Region B	Public	Level 1	
Region B	Public	Level 2	
Region C	Private	Level 1	
Region C	Private	Level 2	
Region C	Public	Level 0	
Region C	Public	Level 1	
Region C	Public	Level 2	

The sheets for the metrics, “number of health facilities,” “place of initial care seeking,” “coverage of TB diagnostic services,” and “coverage of TB treatment services,” will all be similar because they will always involve aggregating data by health sector and level. The structure for “location of notification by sector” and “treatment outcomes” will be the same as the other metrics only if the country has individual patient notification data available by health facility type and the PPA team wishes to aggregate these data by level in addition to health sector. If the team is instead using data from the WHO Global TB database, the structure for these final two metrics will be different. For “location of notification by sector,” data will be aggregated by health sector only, and for “treatment outcomes” data may or may not be split by sector at all.

Generally, the structure of these sheets will be comparable to that of the **summary data** sheets that were put together earlier in section 3.3.2. The difference is that there will be no column for facility type. Each row will correspond to a given health sector and level (and possibly region or other sub-national unit). The first column, *region*, should include all the regions that are identified in the scope of the PPA. The second and third columns are *health facility sector* and *health facility level*. The fourth column presents the value for the metric of interest.

For example, in the “number of health facilities” **mapped data** sheet, the value in this column represents the number of health facilities at a specified level and sector of the health system. This value is calculated by referencing the **summary data** sheet for the health facility master list. Alternatively, if the **raw data** has been mapped to sectors and levels, this value can be calculated by referencing the **raw data** sheet directly (see example in Figure 17). Next the process of creating a mapped data sheet for each core metric is presented.

As mentioned earlier, the structure for ‘location of notification’ and ‘treatment outcomes’ will be similar to the other metrics only if the team was able to access a TB Surveillance database which provided notification and outcomes records by facility type. If not, these two metrics will need to be reported at a higher level than sector and level. Most commonly, the ‘location of notification’ metric is included in the PPA aggregated only by public and private sector (rather than by level) and only at the national level. The ‘treatment outcomes’ metric is most commonly reported only as *treatment successful* or *treatment unsuccessful* (using standard WHO definitions). Both of these metrics are also calculated with a denominator of the total incidence in the country, so they will also include a portion of the total incidence that are not notified or treated, i.e. ‘missing’ cases.

3.4.3.1 Number of Health Facilities Metric

As described earlier, an HFML will provide a list of every health facility in the country, listed by name. It also often includes more detailed information about the facility, such as facility owner, director, or address. As described in section 3.4.1, the team should have already gone through an HFML and categorized all the health facilities in the list into different types, and then further categorized them by sector and level of the health system.

This information should be put into the health facility **mapped data sheet** and will be used as a reference for the diagnostic and treatment coverage metrics (Figure 18). If the PPA team has summarized health facilities by type in a **summary data sheet**, then this can be rolled up to health facility and level in a **mapped data sheet**.

Step 3: Organize Data and Calculate Metrics

MAPPED DATA SHEET: EXAMPLE			
Fig. 18 Number of Health Facilities Sheet			
Region	Health Facility Sector	Health Facility Level	Number of Health Facilities
[manual entry]	[copied from Health Facility Mapping]	[copied from Health Facility Mapping]	[SUMIFS formula using the data in the Summary-HFML sheet]
Region A	Private	Level 1	158
Region A	Private	Level 2	8
Region A	Public	Level 0	464
Region A	Public	Level 1	140
Region A	Public	Level 2	21
Region B	Private	Level 1	0
Region B	Private	Level 2	18
Region B	Public	Level 0	793
Region B	Public	Level 1	97
Region B	Public	Level 2	40
Region C	Private	Level 1	0
Region C	Private	Level 2	23
Region C	Public	Level 0	698
Region C	Public	Level 1	92
Region C	Public	Level 2	26

3.4.3.2 Place of Initial Care Seeking

This metric is the proportion of patients who seek care at each sector and level combination (e.g., public level 1, public level 2, private level 1). The two key components of this metric are:

- **Numerator:** Number of patients initiating care for illness or symptoms (preferably TB symptoms) in a specific facility level and sector.
- **Denominator:** Total number of patients initiating care for illness or symptoms (preferably TB symptoms) in the survey population.

The **summary data** sheet for care seeking is likely to already include the data as a percent rather than as an absolute number of patients. In this case, in order to move the data into the **mapped data** sheet, the only step do one more layer of aggregation from health facility type to health sector and level.

If the scope of the PPA is national, then the denominator of the metric should be total care seeking patients at the national level, and all the care seeking proportions among each sector and level should add up to 100 percent. In some cases, there may be a small share of patients who were missing a response or who listed their care seeking as other, which can be omitted from the analysis if the PPA team does not have a reasonable assumption for where the patients might have sought care. If the scope of the PPA is sub-national, then within each sub-national unit (i.e., region, province), the care seeking proportions among each sector and level should add up to 100 percent.

3.4.3.3 Diagnostic and Treatment Coverage Metric

For the diagnostic and treatment coverage metrics, two data points are needed to make the calculations. The first data point is the total number of health facilities belonging to each health sector and level, which should be counted by referencing the relevant **raw data source** (health facility master list, lab records, etc.). This will provide the denominator. The second data point, the numerator, is the number of facilities with a specific type of diagnosis available (e.g. smear) or treatment service available, among facilities of the specified health sector and level. This should also be counted from the relevant **raw data source**. The value shown in the **mapped data** sheet will be the percent of health facilities of each health sector and level that have a given diagnosis or treatment service available (see example in Figure 19). Alternatively, if **summary data** is available for coverage of TB services according to health facility type, then that data needs to be rolled up to a **mapped data** sheet where it is presented according to health sector and level.

Within the **mapped data** sheets for diagnostic and treatment coverage, create the columns for region, facility sector, and facility level. The next columns capture the data points needed for the coverage calculation. As highlighted in section 0.6, the denominator for the coverage metrics is the number of health facilities in a given region at each sector/level.

Step 3: Organize Data and Calculate Metrics

MAPPED DATA SHEET: EXAMPLE

Fig. 19 Structure of the data columns in diagnostic and treatment coverage mapped data sheets

Region	Health Facility Sector	Health Facility Level	Number of Facilities (N)	Number of Facilities with DOTS (N)	Percent of Facilities with DOTS (%)
[manual entry]	[copied from Health Facility Mapping]	[copied from Health Facility Mapping]	[SUMIFS formula using the data in the Summary - HFML sheet]	[SUMIFS formula using the data in the Summary - TB Services sheet]	[calculated: E/D]
Region A	Private	Level 1			
Region A	Private	Level 2			
Region A	Public	Level 0			
Region A	Public	Level 1			
Region A	Public	Level 2			
Region B	Private	Level 1			
Region B	Private	Level 2			
Region B	Public	Level 0			
Region B	Public	Level 1			
Region B	Public	Level 2			

The columns that capture the data in this sheet include:

1. The first data column to be brought in is the total number of facilities. This data can be pulled in from the **mapped data** sheet for number of health facilities.
2. The next columns capture the number of health facilities at each level and sector that have the TB service of interest (e.g., microscopy). These data points may be pulled in from the respective **summary data** sheets that include the number of health facilities with diagnostic and treatment services, or counted directly from **raw data** sheets.
3. Finally, the remaining columns calculate the coverage of each facility level and sector with TB services by dividing the number of facilities with a given TB service by the total number of facilities in that sector and level.

The team will need to complete this process for each type of TB diagnostic and treatment service for which data is available. Once completed, the **mapped data** sheet should show TB services coverage for each sector and level category included in the PPA and across all regions (Figure 20).

MAPPED DATA SHEET: EXAMPLE

Fig. 20 Calculating coverage of TB services by region, sector and level

Health Facility Sector	Health Facility Level	Number of Facilities (N)	Number of Facilities with DOTS (N)	Percent of Facilities with DOTS (%)
[copied from Health Facility Mapping]	[copied from Health Facility Mapping]	[SUMIFS formula using the data in the Summary - HFML sheet]	[SUMIFS formula using the data in the Summary - TB Services sheet]	[calculated: E/D]
Private	Level 1	158	93	59%
Private	Level 2	8	7	88%
Public	Level 0	464	54	12%
Public	Level 1	140	74	53%
Public	Level 2	21	18	86%
Private	Level 1	0	0	0%
Private	Level 2	18	16	89%
Public	Level 0	793	92	12%
Public	Level 1	97	57	59%
Public	Level 2	40	30	75%

Step 3: Organize Data and Calculate Metrics

Start this process by creating a new sheet in the PPA Excel Workbook and calling it '**Master PPA Data**.' The first three columns of this sheet should list the regions and health facility sector and levels within each region.

After that, there will be one column for each core PPA metric. Starting after health facility sector, create a separate column for:

- Number of health facilities
- Place of initial care seeking
- Percent of facilities with <diagnostic service> (include one column for each type of diagnostic service)
- Access to diagnosis at initial care seeking
- Percent of facilities with <treatment service> (include one column for each type of treatment service)
- Access to treatment services at initial care seeking
- Location of treatment location – only available for national level
- Treatment outcome – only available for national level

Each of these metrics should have a corresponding **mapped data** sheet, except for 'diagnostic access at initial care seeking' and 'treatment access at initial care seeking' which will be calculated directly in the **master PPA data** sheet. The data from each calculations sheet should be pulled into the **master PPA data** sheet. The PPA Excel Workbook Template provides an example of how the master PPA data sheet should be structured (Figure 21).

The metrics for access to diagnosis and access to treatment at point of initial care seeking are calculated by multiplying the percent of patients who sought care at a given health level and sector by the coverage of diagnosis or treatment services at that level and sector. See columns I and L from the **master PPA data** of the PPA Excel Workbook Template.

In some cases, a PPA will include coverage metrics for more than one diagnostic or treatment service. To calculate the access to diagnosis and access to treatment metrics, there are several approaches the team could take:

1. One approach is to take the most widely used TB service (e.g., microscopy or DOTS) and multiply coverage of that service by initial care seeking patterns.

2. The resulting metric would then provide an estimation of how likely it would be for a patient to access a facility with that specific service upon initial care seeking.

3. A second approach, which is only possible with access to an HFML and a national TB lab or a treatment facility database, is to calculate a metric for the coverage of health facilities by ANY TB service, rather than by a specific service. This coverage metric can then be used along with initial care seeking patterns to estimate the likelihood that a patient would receive access to any TB service at initial care seeking.

Once all the metrics have been brought into the **master PPA data** sheet from their respective **mapped data** sheets, and the team has calculated the metrics for access to diagnosis and treatment at initial care seeking, the **master PPA data** sheet will be complete. All of the organization, cleaning, and calculation steps are finished, and it's time to move onto visualization and communication of the analysis.

3.6 STEP 3 – CHECKLIST

During Step 3, each of the following activities should be completed:

STEP 3 CHECKLIST

- Create the PPA Excel Workbook
- Load raw data sources into the PPA Excel Workbook that will need to be summarized
- Complete any cleaning or coding as necessary on the raw data
- Summarize any raw data sources within the PPA Excel workbook into summary data sources
- Load pre-existing **summary data** into the PPA Excel Workbook
- Create a **facility mapping lookup table**
- Map **summary data** to health sector and level in a **mapped data** sheet for each applicable metric, according to the mapping scheme defined in the **facility mapping lookup table**
- If and where applicable, create a **mapped data** sheet directly from **raw data**
- Create one **master PPA data** sheet containing data on all PPA metrics

Step 4: Visualize Results

MASTER DATA SHEET: EXAMPLE

Fig. 21 BBA Master Sheet Structures

The next step of the PPA process is to create the summary visuals that start to tell a story about the possible care seeking journeys of TB patients within the health system. There are several common design elements to a PPA visual, and a variety of tools that can be used to create these visuals. These design elements and tools are described in the following sections.

4.1 PPA VISUAL DESIGN

Figure 22 shows an example of a common design of the PPA visual. The PPA visual brings together summary charts of each of the core metrics and aligns them across the sectors and level of the health system to allow for easy comparison between the different metrics. It is helpful to imagine the PPA visual as a grid. Each column of the grid represents one of the core metrics described earlier in this document. Each row of the grid represents a health system sector and level.

Read from left to right, the PPA begins by showing the number of health facilities in each sector and level of the health system, which is included for additional context.

Column 1 shows the proportion of TB patients that initiate care at each sector and level.

Column 2 shows the diagnostic coverage at each sector and level; that is, the proportion of facilities at each sector and level that are equipped with a diagnostic tool (in this example, microscopy).

Column 3 shows the estimated proportion of all TB patients that access a facility with a diagnostic tool at initial care seeking. For each health sector and level, the proportion of patients that seek care at that sector and level (Column 1) is multiplied by the proportion of facilities at that sector and level that are equipped with a TB diagnostic tool (Column 2) to obtain the estimated proportion of TB patients that access a TB diagnostic at that sector and level. In this example, 10 percent of all TB patients seek care at Private Level 2 facilities and 78 percent of Private Level 2 facilities are equipped with a TB diagnostic tool. So, the estimated proportion of all TB patients that access TB diagnosis at initial care seeking in Private Level 2 facilities is 10 percent times 78 percent, which equals approximately 8 percent.

TIP
Tables
available

TIP

Tableau is data visualization software available at www.tableau.com. A free 14-day trial of the commercial desktop version is available here: <https://www.tableau.com/products/desktop>.

Alternatively, a fully free public version is available here: <https://public.tableau.com/s/>. However, data published on Tableau Public is made available for [public consumption](#), so confidential data should not be used with this version.

Finally, the Tableau Foundation makes Tableau licenses available for qualified non-profit partners; to learn more, [click here](#).

Similarly, the estimated proportion of all TB patients that access TB diagnosis at Private Level 1 is approximately 12 percent (or 33 percent times 36 percent). Column 3 shows the total estimated proportion of TB patients that access diagnosis at initial care seeking in each sector: 20 percent in the private sector (12 percent plus 8 percent), and 24 percent in the public sector (which is calculated in the same manner as the private sector).

Column 4 shows the coverage of treatment services in each sector and level; that is, the proportion of facilities in each sector and level that offer TB treatment.

Column 5 shows the estimated proportion of patients that access treatment at initial care seeking. Column 5 is the sum of the products of Columns 1 and 4 just as Column 3 is the sum of the products of Columns 1 and 2

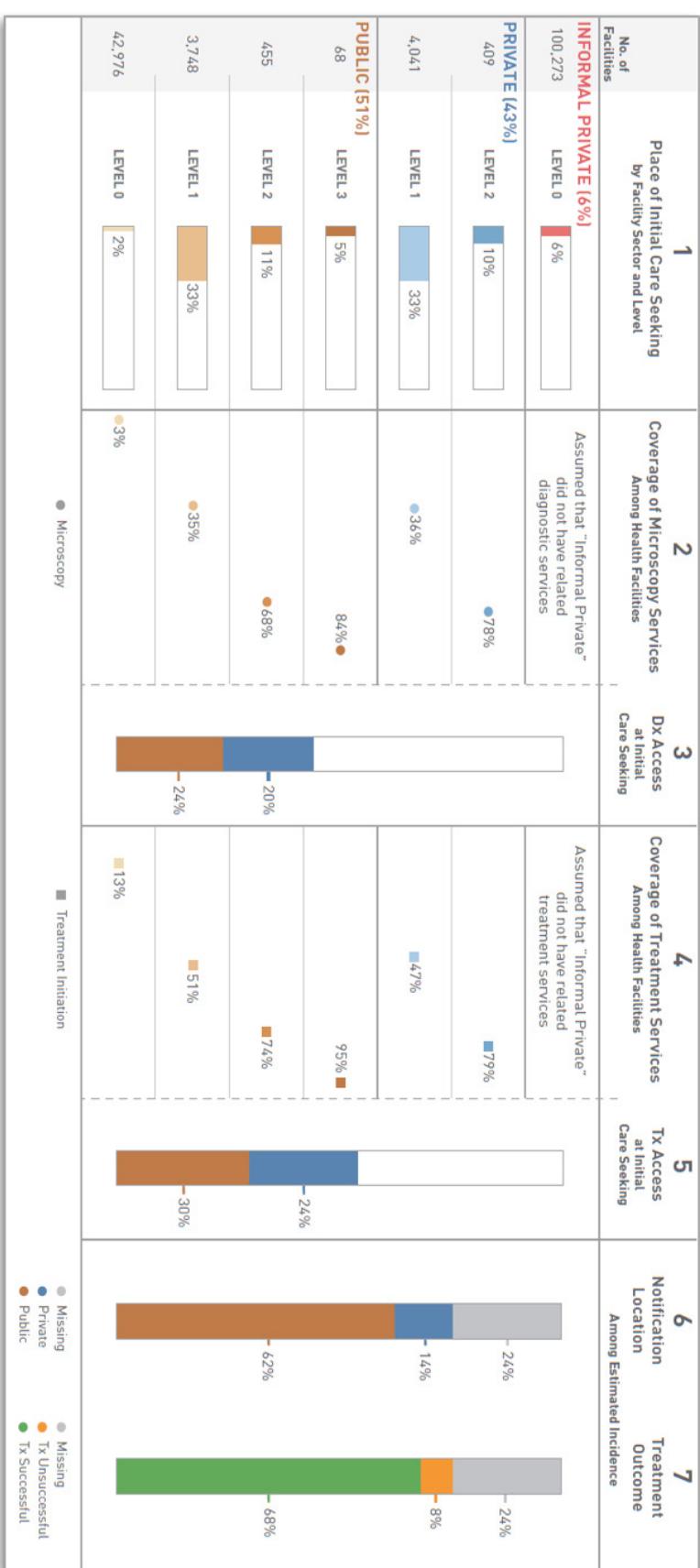
Finally, the PPA finishes with Columns 6 and 7, which show notification by location and treatment outcome of those patients who were notified to the NTP (both shown as a proportion of the estimated incident TB cases at the national level).

Read from bottom to top, the rows of the PPA show the different levels of the health system as described in section 0.5 and created as part of the health facility mapping exercise described in section 3.4.1.

Step 4: Visualize Results

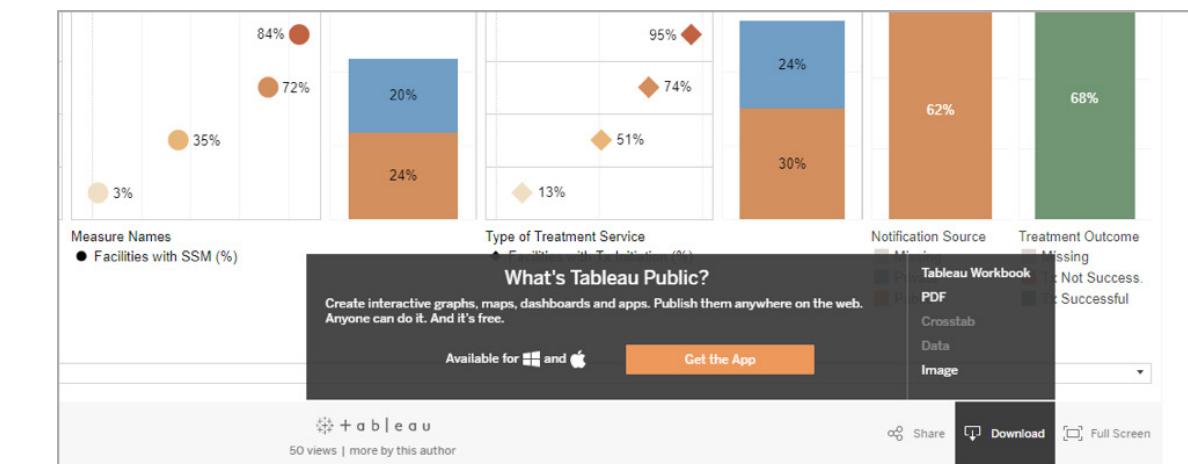
Patient Pathway Description: The PPA visual shows several steps along a patient's journey for TB care. Read from left to right, the PPA begins by showing the number of health facilities in each sector and level of the health system, included for additional context. Column 1 shows the proportion of facilities at each sector and level that initiate care at each sector and level [e.g., Level 0 = microscopy, Level 1 = treatment initiation]. Column 2 shows the proportion of facilities at each sector and level that are equipped with a diagnostic tool (in this example, microscopy). Column 3 shows the estimated proportion of all TB patients that access a facility with a diagnostic tool at initial care seeking. For each health sector and level, the proportion of patients that seek care at that sector and level (Column 1) is multiplied by the proportion of facilities at that sector and level that are equipped with a TB diagnostic tool (Column 2) to obtain the estimated proportion of patients that access diagnosis at initial care seeking in each sector. The total estimated proportion of TB patients that access diagnosis at initial care seeking in each sector and level that offer TB treatment. Column 4 shows the estimated proportion of patients that access treatment at initial care seeking. Column 5 shows the estimated proportion of patients that access treatment at initial care seeking. Column 6 shows the sum of the products of Columns 1 and 4 just as Column 3 is the sum of the products of Columns 1 and 2. The PPA finishes with Columns 6 and 7, showing notification by location and treatment outcomes among patients notified to the NTP (both as a proportion of the estimated incident TB cases at the national level).

Fig. 22 Example of Final PPA Visual



Step 4: Visualize Results

Fig 23 Download the example Tableau workbook



Typically, the PPA visual starts with the lowest level of the public sector on the bottom (e.g., Level 0). Moving up, each row then shows a higher level of the public sector, until it reaches the highest level of the public sector (typically Level 3). After the public sector, the PPA visual starts over with the levels of the private sector, again starting with the lowest level at the bottom and moving up towards the highest levels of the health system in the private sector. Finally, when data are included on the informal private sector. This information is included as the top row of the PPA visual.

The metrics for access to diagnosis and treatment at initial care seeking and location of notification and treatment outcome are included as vertical bars, since they summarize all of the health system sectors and levels.

When creating the PPA visual, it is helpful to use a common color scheme for each of the sectors and levels. Typically, the public sector is represented by brown colors, with each level being represented by a different shade of brown. Similarly, the private sector is represented by blue colors, with each level being represented by a different shade of blue. Finally, the informal private sector is represented as a shade of red to differentiate it from both the private and public sector. Typically, the shades of the colors are lighter among the lower-level health facilities and darker among the higher-level health facilities.

4.2 TOOLS FOR VISUALIZATION

Figure 22 was created in Adobe Illustrator and shows an example of a national-level pathway.

However, if the scope of the PPA is sub-national, it can be time consuming to create each of the sub-national pathways in Illustrator.

Another alternative to Illustrator is to use Tableau, which can load the entire [master PPA data](#) sheet into the software and can easily create PPA visuals for each sub-national unit much more quickly. An example of a PPA created in Tableau is available [here](#).

Once the Demo PPA has been opened, using the link above, click the **Download** button on the bottom right side of the demo visual to download the Tableau workbook (Figure 23). After downloading, the workbook can be opened in Tableau to see how the visual was structured. Tableau allows users to create visualizations based off an underlying data set. Users can load the data set into the software and then make visuals using that data set, without altering the underlying data. In this example, the underlying data set is a PPA Excel Workbook similar in structure to the one the PPA team will have created as a part of this user guide.

4.3 STEP 4 – CHECKLIST

During Step 4, the following activity should be completed:

STEP 4 CHECKLIST

- Create a PPA visual for each sub-national level outlined in the scope of the PPA

Step 5: Interpret Results and Apply Them to Planning and Strategy

5

Once the PPA data have been compiled and the visuals are complete, the final step is to work through interpreting the PPA results for priority setting and program planning. The PPA can provide useful insights into how TB services are aligned with patient care seeking patterns and where there might be areas for further system improvement to better serve patients where they seek care. Translation of the results into programmatic action will require broad dissemination of the findings, and stakeholder engagement to interpret the findings and prioritize.

5.1 SUMMARIZE AND EVALUATE THE PPA PROCESS

Prior to interpreting the results of the PPA, the analysis team should provide a summary of the PPA process and an evaluation of the strengths and weaknesses of the analysis. In the process of summarizing the PPA, it is useful to:

- Make a list of the data sources used for each PPA metric
- Explain the strengths and weaknesses of each data source
- Create a table that explains the process of health facility mapping across all the data sources

The quality of data available for the PPA will vary for each country. The PPA team should consider the quality of the data sources used to create the PPA and how that quality may impact their interpretation of the results.

For example:

- Is the team comfortable with the type of data sources that were available? What are the weaknesses? What are the strengths?
- Based on the available data, where should the team be cautious when interpreting the results? Where does the team feel most comfortable with the results?
 - » Does the team feel comfortable with using the PPA results at the sub-national level?
- If the data sources only provide quality results at the national level, are there areas where better data should be collected to improve the PPA for sub-national use in the future?

Beginning to think about these questions before starting the process of interpreting the results will help ensure that the team appropriately uses the PPA when moving forward in strategy and planning exercises.

The PPA should be interpreted alongside other quantitative and qualitative evidence; e.g., epidemiological reviews, TB inventory studies, and routine program surveillance. Having these data and analytical reports available during discussions of the PPA results can strengthen the validity of interpretation.

5.2 INTERPRETING A PPA

The results of the PPA are now ready for interpretation and application to local priority-setting and planning activities. The approach to disseminating and taking action based on the results of the PPA will differ in each country.

However, it is recommended that existing programmatic meetings, supervisory visits, and planning timelines be used to incorporate the PPA into routine programmatic assessments and standard planning cycles.

Translation of the results into programmatic action will require both broad dissemination of the findings as well as stakeholder engagement to identify and prioritize interventions to address gaps along the patient pathway.

Step 5: Interpret Results and Apply Them to Planning and Strategy

It is important to share the results with a wide variety of stakeholders, as each may interpret the results differently based on their role in TB control. To best apply the findings to programmatic action, various expertise from across the health system will be beneficial; i.e. monitoring and evaluation specialists, health systems planners, laboratory network planners, TB care experts, and political decision-makers.

The PPA is based on the idea that TB systems should be patient-centered and as much as possible be aligned with where patients prefer to seek care. Each column of the PPA visual can be reviewed and discussed in turn to unpack the results and inform discussions of relative priorities for action. Some considerations, by column include:

Initial care seeking. Patients tend to initiate care at health facilities in locations and with types of providers that they prefer. The varied distribution of care initiation across different types of health facilities is partly due to the availability of providers as well as to the acceptability and affordability of providers to patients. The distribution of initial care seeking can be compared to the distribution of the source of TB notification to determine if there is significant mismatch. For example, if 50% of care initiation is in the private sector, but only 10% of case notifications come from the private sector, there may be a significant dropout along the care seeking pathway after care initiation in the private sector. Alternatively, a significant proportion of these patients may remain under care in the private sector but are never notified to the NTP.

Diagnostic coverage. The availability of TB diagnostics at different facilities and sectors should be interpreted within the context of the diagnostic pathway planned for the country of interest. For instance, a country may plan for the diagnosis of TB to be based primarily in primary health care clinics (L1) or the plan may call for patients to be referred to the district hospital (L2) for TB diagnosis. Given a particular diagnostic approach, one can determine if there are major gaps in the availability of TB diagnostics at the particular facility level and sector.



TIP

For example, upon completion of PPAs at national level and for each of its 47 counties, Kenya did the following:

1. updated county-specific TB profiles to highlight the local gaps in the patient pathway; alerting county-level decision-makers about the priorities for action;
2. equipped a joint external monitoring mission team with the PPA results to inform their field-based assessment;
3. interpreted the pathways during a meeting of the county laboratory coordinators; brainstorming solutions to the diagnostic gaps and sharing best practices from counties with fewer gaps;
4. introduced the PPA findings to the steering committee preparing the Global Fund application; and
5. incorporated the findings into a revised national strategic plan.

Diagnostic access at initial care seeking.

Diagnostic access is the product of care initiation and diagnostic coverage. If diagnostic access is low, there might be a mismatch between where patients seek care and where diagnostics are available. Identifying where this happens can help the program provide more diagnostic capacity in highly-accessed facility levels and for preferred provider types.

The diagnostic coverage and diagnostic access indicators provide information on where the greatest gain in case-detection can be made. When applied at the sub-national level, this analysis can identify areas with lower diagnostic coverage or access. Information on diagnostic coverage and access can be combined with information on population size, prevalence or incidence of TB, and current case notification to identify the priority areas of interventions to improve access to TB diagnosis. Additional information on feasibility of patient or specimen referral and capabilities in health facilities can help the program select the best way to improve access to TB diagnosis.

Step 5: Interpret Results and Apply Them to Planning and Strategy

Table 5 Questions to Guide Interpretation of the PPA Results

Interpretation Questions		
PPA Metrics to Consider	Basic Review of Findings	Understanding Programmatic Implications
Place of Initial Care Seeking	Where are patients initiating their care seeking journey? Is it mostly in the private sector or mostly in the public sector? Do they initiate care mostly in lower level (i.e., level 0 or level 1) health facilities or do they bypass primary care and initiate their care seeking primarily at level 2/3 hospitals?	Looking at the distribution of places for initial care seeking, are there levels of the health system or sectors (e.g. private) that are not relatively well supported by the NTP? Are there interventions that could better engage them? If there is a high proportion of initial care seeking occurring at level 0, what are some possible reasons? Does it relate to the drug/pharmaceutical regulations in the country? Are there ways to engage this level in referral or specimen capture?
Place of Initial Care Seeking	How does care seeking differ by sub-national unit? Are there differences in public and private sector care seeking by sub-national unit? Are there any sub-national units where one level of the health system seems to be more widely used than average?	Does initial care seeking suggest more active screening for TB within patient populations? Which kind of screening strategies could be implemented at each level? Who can be involved to implement the screening strategies?
Place of Initial Care Seeking, Diagnostic Coverage	How well do TB diagnostics align with patient care seeking preferences? If patients prefer to seek care in the private sector, how well covered are private sector facilities with TB services?	Are there certain levels or sectors that should be targeted with expansion of diagnostic or treatment services; e.g. are there priority regions / areas for private sector engagement or hospital engagement?
Place of Initial Care Seeking, Treatment Coverage	How well do TB treatment services align with patient care seeking preferences? How well does the coverage of TB treatment services align with TB diagnostic services? Does it appear that patients would have to move between facilities for diagnosis and treatment? What is the driver of this movement?	If diagnostic services are not well aligned with where patients seek care, how strong are the referral systems between different health facility levels and sectors? Are there areas that should be particularly targeted with improved referral systems, including remote specimen capture, specimen transport, and patient support for patient referral?
Access to Diagnosis and Treatment at Initial Care Seeking	How does access to TB services compare across sub-national units? Are there some regions that appear to be better aligned than others? What have these regions done well?	Are there "best practice" examples where the diagnostic gap is small, e.g. where specimen transport systems exist, that warrant scale-up?
Place of Initial Care Seeking, Location of Notification	Assuming that the initial care seeking patterns of patients is representative of the whole TB population, how does this compare to the population of TB patients who are notified to the national TB program? Does a greater share of patients seek care in the private sector than are notified in the private sector (as percentage of incident cases)? Does a greater share of patients seek care in the public sector than are notified in the public sector (as a percentage of incident cases)? Is there significant proportion of patients receiving TB treatment at L2/3 hospital levels? Is there any relationship between place of treatment and notification/report of complicated TB cases?	After reviewing the PPA metrics, are there areas that would benefit from greater involvement or engagement of communities in support of TB control (e.g., TB suspect referral or community based treatment monitoring)? Can treatment be decentralized / ensured where patients initiate care; i.e. where they prefer to seek care?
	If the notification proportion is very low, does the country have a mandatory notification regulation for TB? How effective is the enforcement of this regulation?	How can the PPA results inform action when used other analysis, such as prevalence survey findings? What are the national and sub-national priorities for action that emerge? How does the cost for spending for TB control compare to where the greatest needed is as shown by patient care seeking? How might funding allocations be shifted to align with patient-centered priorities? Are there regulatory barriers to better aligning diagnosis or treatment with where patients seek care? How can these be addressed?

Step 5: Interpret Results and Apply Them to Planning and Strategy

Treatment coverage. Treatment coverage should be interpreted within the context of the treatment pathway planned for the country of interest. For instance, a country may plan for TB treatment to be based primarily in primary health care clinics (L1) or the plan may call for patients to be referred to the district hospital (L2) for TB treatment. Given a particular treatment approach, one can determine if there are major gaps in the availability of TB treatment at different facility levels and sectors. In addition, one can compare diagnostic coverage to treatment coverage to determine if there is a major misalignment between where TB diagnoses are made and where TB treatment is available. The TB program should aim to reduce this misalignment, such that treatment is available as soon as a TB diagnosis is made.

Treatment access. Treatment access is a product of initial care seeking and treatment coverage. If treatment access is low, there may be a mismatch between where patients initially seek care and where TB treatment is available. This provides opportunities to examine whether the availability of TB treatment can be aligned more closely with where patients initially seek care. To do so, it is important to determine whether or not it is necessary to concurrently improve diagnostic access, given that TB treatment follows diagnosis. Expanded diagnostic access is often the first step in improving treatment access.

Table 5 outlines some useful questions to consider when reviewing the PPA data. Using these questions as guidance, the PPA team should start to formulate a list of important results from the PPA. These results should then be discussed among the PPA team, partners and other stakeholders.

5.3 APPLYING RESULTS TO PLANNING AND STRATEGY

What follows are three hypothetical examples of how PPA results could be used for planning and strategy exercises, depending on the quality of data included in the analysis and what the results showed about the program.

In one situation, one of the technical partners of the NTP is using the results to work with regional TB control teams.

By conducting the PPA at the sub-national level, the national team is able to work with regional TB control leaders to identify where there are potential alignment issues between care seeking and service availability. Further, the team is using the PPA to provide supporting evidence for the claim that knowledge and engagement with private providers is insufficient given the level of care seeking in this sector. It has identified another program partner who is engaging in mapping of the private sector provider network. Using the PPA, the team can help identify where (geographically) this mapping is most important for TB control. The team will also be using this exercise to make real-time updates to the PPA as new data is collected.

In a second context, a country who implemented the PPA was able to map sub-national results and gain a detailed picture of service availability in both the public and the private sectors. Because of previous efforts to build a strong TB lab database, the country had robust information on the diagnostic networks in both the public and private sector. Conducting the PPA aligned this data with care seeking information and gave the NTP a picture of where diagnostics might be placed to better meet patients where they are. This country is working on bringing this information to regional lab managers who are responsible for TB labs throughout the country to engage in planning exercises for future diagnostic expansion.

In a third example, an NTP had limited data to complete a PPA. They completed a prevalence survey six years ago, which included one question about where patients sought care for TB-related symptoms. However, the survey was only powered at the national level. The NTP has not yet implemented a national health facility master list, but did conduct a survey of health facilities in a national survey completed four years earlier. In addition, the NTP manages a database of TB microscopy labs for any public sector health facilities and some of the private sector health facilities. Using these data sources, this country has completed a PPA at the national level.

In this final example, the data the country has available did not provide the level of granularity that makes a sub-national pathway useful or reliable.

Step 5: Interpret Results and Apply Them to Planning and Strategy

However, by going through the PPA exercise and developing a national-level pathway, this country can use the PPA to diagnose where new data should be collected which could improve the analysis, and ultimately, strategy and planning work, moving forward. With this in mind, the country will use the PPA framework as a reference as it plans the questionnaire of their new DHS survey, ensuring the survey team asks all the participants who experienced a cough where they sought care. This information will give the survey team the granularity necessary to understand care relevant care seeking patterns at the sub-national level. They will also advocate to the Ministry of Health to start building a health facility master list which maps the health facilities across the country.

5.4 STEP 5 – CHECKLIST

During Step 5, each of the following activities should be completed:

STEP 5 CHECKLIST

- Summarize and evaluate the PPA process
- Create a list of key results based on the interpretation of the PPA
- Identify potential programmatic implications based on the results of the PPA
- Create a plan for engaging relevant stakeholders with the results of the PPA

Step 6: Appendix

6.1 COLLECTING AND SUMMARIZING DHS DATA

DHS data are some of the most widely available data sources for care seeking information. Surveys have been conducted in more than 90 countries—in many countries, they have been completed over several years. In addition to their widespread survey implementation, the DHS program has also created a standardized method of delivering raw data for use in analysis that goes further than what is published in the final DHS report. Unfortunately for the PPA, the DHS reports do not often include information on the variables related to care seeking, so the only way to get this information is to download the raw data and gather the care seeking data manually. Because the DHS raw data files are large and complex, an R script has been provided to allow PPA users to easily pull out several of the care seeking variables very easily for use in the PPA.



TIP

Before starting this process, it will be useful to download R and RStudio. Both are available for free:

R: <https://cran.r-project.org/>

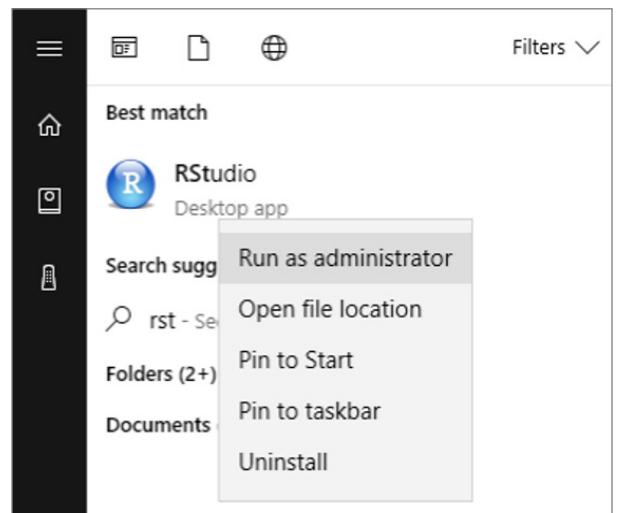
RStudio: <https://www.rstudio.com/products/rstudio/download/#download>

The R script provided here requires the user to be an administrator on their machine. If you are not the administrator of your machine, you can run R as an administrator when you open the software (Figure 24)

The rest of this section will explain the steps needed to access the raw data files from DHS and structure the data to use the R script for DHS care seeking data.

1. The first step in the process of accessing DHS data is to register on the DHS website. A link to the registration page is here: <https://dhsprogram.com/data/new-user-registration.cfm>.
2. After successfully registering, you should navigate to the login page and create a username and password to log in.
3. Once you have logged in, you will need to create a new project before you can request any data sets. Projects require information about the title of the project, the researchers who should be included on the project, and a short description of the project. The project description should describe the patient pathway analysis and any additional potential ways you might use the care seeking information.
4. Once you've created your project, you can then request data sources to add to that project. You will be notified that access was granted via email. Access is typically granted within one to two days.

Fig. 24 Run R as Admin



Step 6: Appendix

Fig 25 Beginning of *analyzeDHS.R* File

```

1 #analyzeDHSdata.R
2
3
4 # use this script to create a careseeking analysis data table from a DHS dataset.
5
6 ##### INSTALLING PACKAGES #####
7
8 # This script requires the attachment of the following libraries: dplyr, foreign, tidyr,
9 # and data.table. If you do not have these libraries installed, uncomment
10 # (remove the pound signs from) the lines below and run the commands to to install them.
11 # You can check if they are installed by running the command -- library(dplyr) --
12
13 # install.packages("dplyr")
14 # install.packages("foreign")
15 # install.packages("tidyr")
16 # install.packages("data.table")
17
18 ##### WORKING DIRECTORY #####
19
20 # using the command below, set your working directory to the
21 # folder containing both this file, and the dhs_functions.R script. The path to the file
22 # must be in quotes and contain only forward slashes. This will be the folder in which
23 # the DHS careseeking .cvs files created by the script will be stored.
24
25 setwd("C:/Dropbox (Linksbridge)/Anna Summer Internship Files")

```

5. After receiving the notification that access was granted, sign back into your DHS account. You will likely be navigated to a page that asks you to select a project you want to work on. Click the project title that you just created (in step 3) and you will be sent to a page where you can request new data. There will be an option to 'Download a single survey'. Click the dropdown menu and select the country for which you'd like to access the raw data file.
6. You will be sent to a page that shows all of the past DHS data sets for that country. For the purposes of the PPA, download the most recent version of the 'Standard DHS'. Find this version and click the Download link.
7. This link will send you to a page that lists the different types of survey data sets and file types by which you can access the data. For the most common care seeking variables, you will want to download the individual recode data set, in the file type 'stata dataset (.dta)'. Click the check box next to this data set and then click the 'Process selected files for download' button at the bottom of the page.
8. On the next page will be a link that says 'Click here to download the files you selected in one zipped file'. Clicking this link will start the download process. The files will be downloaded into a zipped folder. Extract all of the files from the zipped folder into a separate folder in your PPA folder and title this folder 'DHS dataset'. This is the data source folder which you will reference in the R script.

Step 6: Appendix

Fig. 26 Entering the Input Data for the *analyzeDHS.R* Script

```

27 ##### INSTRUCTIONS #####
28
29 # As described in the DHS Careseeking Analysis walkthrough, there are six variables that
30 # you must change to run this script on your data. They include:
31 #
32 #     infilename: The full path to the file storing your downloaded DHS data (which must be in stata format
33 # with a .dta file extension)
34 #     example: "~/Documents/DHS Data/Tanzania/tzir63dt/TZIR63FL"
35 #     note: Do NOT include the file extension (ie: ".dta")
36 #
37 #     country: The name of the country you are analyzing
38 #     example: "Tanzania"
39 #
40 #     careseeking_var_list: A list of the careseeking variables you want to generate files for
41 #     example: c("h46a_1", "h44a_1", "v829", "v842")
42 #
43 #     careseeking_var_names: The names of the careseeking variables you are using
44 #     example: c("fever", "diarrhea", "hiv.test", "antenatal.hiv.test")
45 #     note: make sure they are in the same order as they are listed in careseeking_var_list
46 #
47 #     analysis_var_list: The variables you want to analyze your careseeking variables with
48 #     example: c("v190", "v024", "v025")
49 #
50 #     analysis_var_names: The variables you want to analyze your careseeking variables with
51 #     example: c("wealth_quintile", "region", "residence")
52 #     note: make sure they are in the same order as they are listed in analysis_var_list
53 #
54 # The examples described above are the defaults in the assignment commands below. Modify these
55 # assignments (by changing the text after the <- ) before running the file.
56
57 infilename <- "C:/Dropbox (Linksbridge)/Anna Summer Internship Files/DHS Data/Kenya/keir70dt/KEIR70FL"
58 country <- "Kenya"
59 careseeking_var_list <- c("h46a_1", "h44a_1")
60 careseeking_var_names <- c("fever", "diarrhea")
61 analysis_var_list <- c("v024", "v190", "v025")
62 analysis_var_names <- c("Region", "wealth_quintile", "Residence")

```

9. The folder that you've just downloaded will contain seven files: a DCT file, a DO file, a Word document, a DTA file, a FRQ file, a FRW file and a MAP file.
 - a. The DTA file is the raw data that is going to be analyzed.
 - b. Both the FRQ and the FRW files show frequency statistics about all of the variables included in the data set. The FRQ file shows unweighted frequencies and the FRW file shows weighted frequencies. These can be used as reference to confirm your analysis is correct.
 - c. The MAP file shows descriptive information about each of the variables in the data set, such as the length of the data, the data type, and the possible responses to any variable question.
 - d. The Word Document includes additional background documentation about the data set, including a list of all the variables included.
10. As mentioned in section 3.4.3.2, there are five common care seeking variables in the DHS data set. Four of them are included in the individual recode file you have just downloaded and the fifth is typically included in the summary report if it is available.
11. The four care seeking variables included in the individual recode file are:
 - a. Place first sought treatment for diarrhea (*h44a*)
 - b. Place first sought treatment for fever (*h46a*)
 - c. Place of HIV test (general) (*v829*)
 - d. Place of antenatal HIV test (*v842*)
12. Each of the variables are given a variable ID which is used in the data file, rather than the full variable name. These variable IDs are included in parentheses next to each variable in step 10. You will need to note these variable IDs for use in the R code.
13. Once you have downloaded the data set and reviewed the file above, you'll then need to navigate to the GitHub page that houses the DHS Care Seeking script. You can access the R script for DHS care seeking data here: <https://github.com/linksbridge>.
14. Click on the repo titled 'DHS-Care-Seeking.' There will be two R files included:
 - a. *dhs_functions.R*
 - b. *analyzeDHS.R*
15. Download both of these files and save them in your PPA folder.
16. Once you have downloaded these files, open the *analyzeDHS.R* file. As long as both files are saved in the same folder, you will not need to open the *dhs_functions.R* file.
17. At the top of the *analyzeDHS.R* file, there will be a set of instructions.
18. The first set of instructions directs you to install the necessary R packages to run the analysis. If you do not have these R packages already installed on your machine, you will need to uncomment lines 13-16 and install the four packages listed there. If you already have those packages installed, you do not need to reinstall them and the libraries will be loaded automatically with the code (Figure 25).

Step 6: Appendix

Fig 27 Output of analyzeDHS.R script

category	category_values	PUBLIC SECTOR: Government hospital	PUBLIC SECTOR: Government health center	PUBLIC SECTOR: Government dispensary
Facility Total	Facility Total	0.1578	0.2038	0.3412
Region	Coast	0.1771	0.089	0.4612
Region	North Eastern	0.2194	0.1327	0.2941
Region	Eastern	0.1269	0.2188	0.3432
Region	Central	0.2768	0.1954	0.2719
Region	Rift Valley	0.1953	0.155	0.3769
Region	Western	0.124	0.3109	0.284
Region	Nyanza	0.0881	0.2348	0.3977
Region	Nairobi	0.1522	0.2679	0.1371
Residence	Urban	0.251	0.1901	0.168
Residence	Rural	0.1088	0.211	0.4323
Wealth_Quintile	Poorest	0.0955	0.1931	0.4964
Wealth_Quintile	Poorer	0.1284	0.2529	0.4027
Wealth_Quintile	Middle	0.1579	0.229	0.2772
Wealth_Quintile	Richer	0.1765	0.2296	0.2534
Wealth_Quintile	Richest	0.2972	0.0816	0.1521

19. The next step (as outlined on line 25 of the script) is to set your working directory. This should be the folder where you want the final files to be exported to. Note that when you paste the working directory into this line of code, the whole directory should be pasted within the quotation marks (don't delete them). Additionally, if you are working on a PC, the backslashes in the file path need to be changed to forward slashes, otherwise R won't recognize the directory. (See Figure 25).
20. Once you have installed the necessary packages and set your working directory, you will need to give the script some information about what data you'd like to analyze. Lines 27-54 provide some background on the six inputs that the user needs to provide in order for the script to run. Enter the required information next to the inputs on lines 56-61. Make sure to keep the quotation marks around each of the pieces of input information that you include (Figure 26).
 - a. Infilename > The full path to the file storing your downloaded DHS data (which must be in Stata format with a .dta file extension). Note that you need to include the file name as well, but do not include the .dta at the end of the file.
 - b. Country > The name of the country you are analyzing.

Step 6: Appendix

21. Once you have included all the necessary input information, you can run the code by clicking **source** in the upper right hand side of the R Studio screen.
22. The script should automatically run. Based on the inputs you gave it, it will create a separate Excel file for each care seeking variable, saved in the folder you put in your working directory.
23. Once you open this file, all the care seeking data from that variable will be available and summarized across each of your cross-analysis variables (Figure 27). The data should now be ready to import into your PPA Excel Workbook sheet as a **summary data** sheet.

- c. Careseeking_var_list > A list of the care seeking variables you want to generate files for. These are the variable IDs that were described in step 10. Only include the variable IDs for the care seeking information you'd like to generate.
- d. Careseeking_var_names > These are the names you'd like to use for the care seeking variables. They can be whatever you'd like, but they should be descriptive of the content in the variable (e.g., fever or hiv_test).
- e. Analysis_var_list > This is where you would include any other variables for which you'd like to cross analyze the care seeking patterns. This might include region, wealth quintile, or urban/rural. Some common variable IDs that might be of interest are:
 - i. V024 – Region
 - ii. V025 – Type of place of residence (e.g., urban/rural)
 - iii. V106 – Highest educational level
 - iv. V190 – Wealth index
- f. Analysis_var_names > These are what you would like to name the variables for cross-analysis. For example, you might want to name the region, residence, or education. They need to be in the same order as the variable IDs included in the *analysis_var_list*.

References

1. Subbaraman R, Nathavitharana RR, Satyanarayana S, et al. The Tuberculosis Cascade of Care in India's Public Sector: A Systematic Review and Meta-analysis. *PLoS Med.* 2016;13(10):1-38. doi:10.1371/journal.pmed.1002149.
2. Nadoo P, Theron G, Rangaka MX, et al. The South African Tuberculosis Care Cascade: Estimation of Losses and Methodological Challenges. *J Infect Dis.* 2017;2016 (Supp).
3. Uys PW, Warren RM, van Helden PD. A threshold value for the time delay to TB diagnosis. *PLoS One.* 2007;2(8). doi:10.1371/journal.pone.0000757.
4. Golub JE, Bur S, Cronin WA, et al. Delayed tuberculosis diagnosis and tuberculosis transmission. *Int J Tuberc Lung Dis.* 2006;10(March 2004):24-30.
5. Kenya Ministry of Health. Kenya Master Health Facility List. 2017. <http://kmhfl.health.go.ke/#/home>. Accessed January 1, 2017.
6. The United Republic of Tanzania - Ministry of Health. Health Facility Registry. *Heal Facil Regist.* 2017. <http://hfrportal.ehealth.go.tz/>. Accessed January 1, 2017.
7. Republic of Philippines Ministry of Health. National Health Facility Registry 2.0. *Heal Facil List.* 2013. <http://nhfr.doh.gov.ph/rfacilities2list.php>. Accessed January 1, 2017.
8. World Bank, World Health Organisation (WHO), USAID. *Guide to Producing National Health Accounts.*; 2003. http://www.who.int/health-accounts/documentation/publication_nha_guide/en/.
9. OECD, Eurostat, World Health Organisation (WHO). *A System of Health Accounts.*; 2011. <http://www.who.int/entity/health-accounts/documentation/sha2011.pdf?ua=1>.
10. Pakistan Bureau of Statistics. *Pakistan National Health Accounts*, 2011-2012. Islamabad; 2011.
11. World. *Tuberculosis Prevalence Surveys: A Handbook*. Geneva; 2011. http://www.who.int/tb/advisory_bodies/impact_measurement_taskforce/resources_documents/thelimebook/en/.
12. Ministry of Health Republic of Indonesia. Indonesia Tuberculosis Prevalence Survey. 2015;(June).
13. Federal Republic of Nigeria. *Report First National TB Prevalence Survey.*; 2012.
14. The Federal Democratic Republic of Ethiopia Ministry of Health. *Household Health Service Utilization and Expenditure Survey.*; 2014.
15. Ministry of Health - Government of Kenya. *Kenya Household Health Expenditure and Utilisation Survey 2013*. Nairobi; 2013. www.healthpolicyproject.com/pubs/745_KHHUESReportJanuary.pdf.
16. Philippine Statistics Authority, ICF International. Philippines National Demographic and Health Survey 2013. 2014:353.
17. The World Bank. Living Standards Measurement Survey. *LSMS*. <http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,menuPK:3359053~pagePK:64168427~piPK:64168435~theSitePK:3358997,00.html>. Accessed January 1, 2017.
18. Nigerian National Bureau of Statistics, Nigeria Federal Ministry of Agriculture and Rural Development, The World Bank. *LSMS-Integrated Surveys on Agriculture General Household Survey Panel.*; 2014.
19. Pakistan National TB Control Programme. National Reference Laboratory Database. 2017. <http://mdrtbntp.gov.pk/nrl/login.php>. Accessed January 1, 2017.
20. World Health Organization. *Service Availability and Readiness Assessment (SARA) - Reference Manual*. Geneva; 2015. http://www.who.int/healthinfo/systems/sara_reference_manual/en/.
21. Ethiopian Public Health Institute, Federal Ministry of Health, World Health Organization. *Services Availability and Readiness Assessment*. Addis Ababa; 2017.
22. Health KM of. *Kenya Service Availability and Readiness Assessment Mapping*. Nairobi; 2013.
23. World Health Organization (WHO). Global TB Database. 2016. <http://who.int/tb/country/data/download/en/>.
24. Pakistan Bureau of Statistics. *Pakistan National Health Accounts*, 2009-2010. Islamabad; 2009.