FIND
Diagnosis for all

Forecasting diagnostic testing demand for geospatial analysis

JANUARY 2024

Table of Contents

About this guide

Imagine a healthcare system where every person with a disease has access to the diagnostic tests they need, when they need them, regardless of where they live.

In order to achieve this, healthcare providers need to understand current and future diagnostic testing demand to be able to adequately allocate and distribute the needed resources. To support providers in their planning, this guide summarizes various methods for calculating diagnostic testing demand for tuberculosis (TB), human immunodeficiency virus (HIV) and human papilloma virus (HPV) from different data types, and shows simple approaches to forecast future demand. The aim is to provide an overview of simple and most commonly used methods that can be applied without any expertise in statistical modelling or software.

The methods and approaches presented in this guide are applicable to any demand forecasting project, but we have specifically included considerations and approaches that are critical for geospatial analysis and diagnostic network optimization, which often adds a layer of complexity to the demand forecasting process as we not only need to forecast how much but also where future testing demand will occur.

The guide is intended to inform Ministry of Health officials, programme managers, laboratory specialists, technical partners, and donors who aim to analyse, interpret and forecast diagnostic testing demand. This guide is also relevant for specialists conducting a geospatial analysis and optimization which requires testing demand.

Acknowledgements

The development of the guide was led by Manuela Rehr (independent consultant for FIND) with contributions from Heidi Albert (FIND, South Africa), Sarah Girdwood (FIND, South Africa) and Marie Brunetti (FIND, Switzerland).

Ministries of Health from multiple countries and their partners have conducted demand forecasting within geospatial network analysis and optimization projects. Their experience and work have informed the compilation and development of concepts and methods described in this guide.

9 November 2023

Abbreviations

ANC	Antenatal care
ART	Antiretroviral therapy
BL	Baseline
DHS	Demographic health survey
DNO	Diagnostic network optimization
EID	Early infant diagnosis
FLQ	Fluoroquinolone
GLI	Global Laboratory Initiative
HIV	Human immunodeficiency virus
HPV	Human papilloma virus
INH	Isoniazid
LF LAM	Lateral flow lipoarabinomannan
LRTI	Lower respiratory tract infection
M&E	Monitoring & evaluation
MoH	Ministry of Health
MTB	*Mycobacterium tuberculosis*
mWRD	WHO-recommended molecular diagnostics
N/A	Not available
NCD	Non-communicable diseases
NRL	National reference laboratory
NSP	National strategic plan
NTP	National TB programme
OPD	Outpatient department
PBFW	Pregnant and breastfeeding women
PLHIV	People living with HIV
PMTCT	Prevention of mother-to-child transmission
POC	Point of care
PPA	Patient pathway analysis
RIF	Rifampicin
SARA	Service availability and readiness assessment
SM	Smear microscopy
SRS	Specimen referral system
TB	Tuberculosis
LAMP	Loop-mediated isothermal amplification

VL	Viral load
WHO	World Health Organization

1. Introduction

1.1 What is diagnostic testing demand?

In the context of this guide, we define “testing demand” as the total number of diagnostic or screening test volumes. We differentiate between:

- Historical demand, which is the demand that has occurred in previous years and includes the baseline demand
- Baseline demand, which is the demand that has occurred in the most recent year
- Future demand, which is the demand expected in the short-, medium- or long-term future.

1.2 Why is forecasting testing demand important?

Imagine a healthcare system where every person with a disease has access to the diagnostic tests they need, when they need them, regardless of where they live.

In order to achieve this, healthcare providers need to understand current and future diagnostic testing demand to be able to adequately allocate and distribute resources, such as test kits, reagents and diagnostic platforms, and consequently plan for and provide required funding.

Geospatial methods for diagnostic network optimization (DNO)[1] has become a critically important tool over the past years and helps healthcare providers and policymakers design diagnostic networks that are both equitable and efficient by making optimal decisions about resource distribution and funding.

However, one of the key requirements of DNO is an accurate forecasting of future demand for diagnostic tests. This is not a simple task, as it makes the sometimes already complex demand forecasting process even more complex as it requires determining not only how much demand is expected, but also where it is expected.

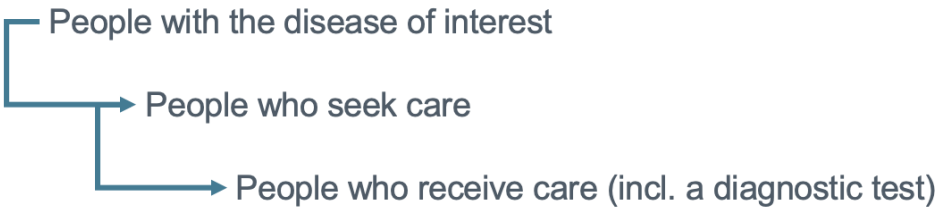
1.3 Why this guide?

To overcome this challenge, we summarize and present various methods and approaches applicable to any demand forecasting project, but with specific considerations and approaches that are critical for geospatial analysis and DNO. Our goal is to help planners and managers develop the insights they need to optimize diagnostic networks and resource allocation and ensure that everyone has access to the diagnostic tests they need.

1.4 Factors that impact diagnostic demand

Diagnostic testing demand analysis and forecasting has to consider the cascade of care, which is critical for choosing the correct method, make correct assumptions and select suitable base datasets for demand forecasting.

[1] FIND “Diagnostic Network Design and Optimization” 2022. Available online [https://www.finddx.org/tools-and-resources/access-and-implementation/diagnostic-network-design-and-optimization/]



The number of people with the disease provides the initial estimate of the potential pool of individuals who might require diagnostic testing. If the objectives of the demand forecasting are to calculate the demand that "should" be realized if all people with disease sought and received care, then there are appropriate methods and base data choices to calculate this, which we will introduce later in this guide.

However, the typical objective of demand forecasting is to determine the realistic future demand that is expected routinely at the health facilities. This "type of demand" has to account for two factors, including the health-seeking behaviour as often a significant portion of individuals affected by the disease might not seek care.

Secondly, not everyone who sought care might actually receive care, again, depending on health system performance and context. Not accounting for health seeking and care receiving can lead to an overestimation of future demand and both factors are therefore critical to account for when choosing a method and base dataset.

The various elements that can influence healthcare seeking and healthcare reception (called "barriers to care") include health service provider and patient/consumer dimensions and are typically categorized into affordability, availability, geographic accessibility, knowledge and acceptability (Figure 1.1).

Figure 1.1: Barriers to care and their impact on healthcare (and diagnostic testing demand). With modifications from Peters et al. [2] and Jacobs et al [3].

[2] Peters DH, Garg A, Bloom G, Walker DG, Brieger WR, Rahman MH. Poverty and access to healthcare in developing countries. Ann N Y Acad Sci. 2008;1136:161-71. doi: 10.1196/annals.1425.011.

[3] Jacobs B, Ir P, Bigdeli M, Annear PL, Van Damme W. Addressing access barriers to health services: an analytical framework for selecting appropriate interventions in low-income Asian countries. Health Policy Plan. 2012 Jul;27(4):288-300. doi: 10.1093/heapol/czr038.

Some of those factors could be of high relevance for demand forecasting, especially for methods and approaches that rely on a more detailed analysis of past health system operations and policies, and account for their anticipated changes in the future.

For example, suppose the Ministry of Health (MoH) in a country has identified the availability of diagnostic instruments as a major barrier to appropriate healthcare as it was observed that a large number of patients were not tested due to very frequent instrument breakdowns. In their upcoming National Strategic Plan (NSP), they therefore planned to set up a regular service, maintenance and repair contract with a qualified company. It is very plausible to assume that this intervention will lead to an increase in diagnostic testing in the future, and appropriate methods can be chosen to adequately account for this future health system performance.

What every MoH essentially does is removing existing barriers to care to improve access to and utilization of the healthcare system.

These planned interventions, which are typically outlined in the NSP, are expected to increase demand (at least in the short-term) and are designed to reach a certain, optimal target.

This optimal target can be either a situation in which all people with the disease seek and receive care (as illustrated in Figure 1.2), or it can be an interim step that was considered feasible within the time frame of the NSP. This concept and relationship is important to understand to adequately design the demand forecast, select appropriate methods and datasets, and we will continue to refer back to it throughout this guide.

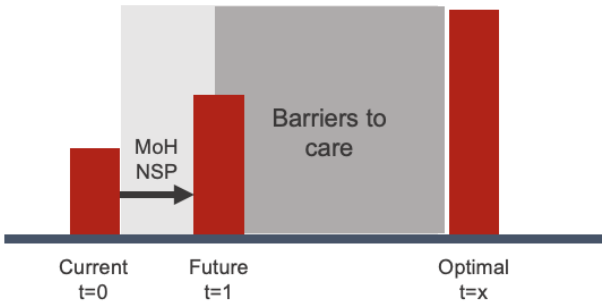


Figure 1.2: The link between demand and planned NSP interventions.

Analysing and forecasting demand for geospatial analysis can add another layer of complexity to the project because the aforementioned factors of disease burden, barriers to care, as well as future planned interventions in the NSP, normally vary geographically (Figure 1.3).

In geospatial analyses, demand is ideally required at facility level, however, depending on the base dataset, this may not be feasible either due to methodological constraints or time limitations. Therefore, we explain for each demand method introduced in this guide any relevant considerations and approaches to ensure the forecasted demand is suitable for geospatial analysis. These are of course not relevant for those who aim to determine demand, for example at national-level aggregate value only.

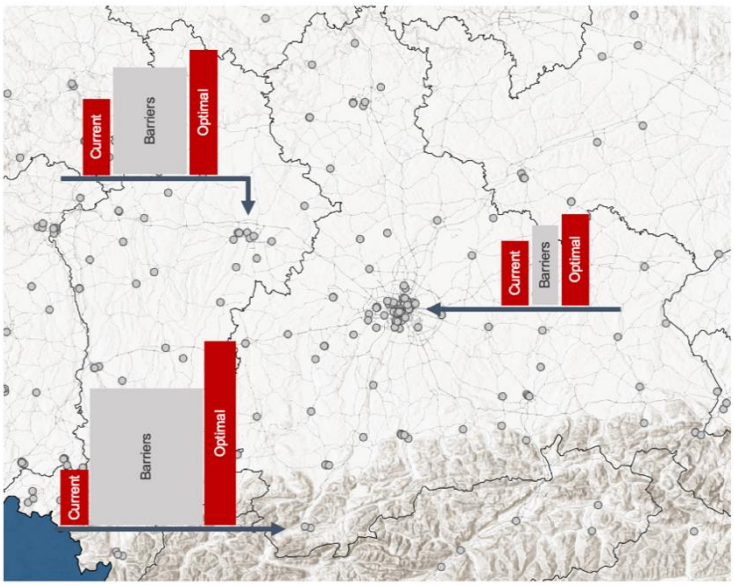


Figure 1.3: Geospatial variations of current demand, barriers to care and future, optimal demand.

1.5 The process of demand forecasting

The process of demand forecasting for diagnostic testing can vary depending on the specific situation, but it normally includes the following steps:

Define objectives and scope

Determine the objectives and scope, outlining which test types and platforms should be covered, the required level of accuracy and, if geospatial analysis is planned, which level of spatial resolution is required (i.e. facility-level or small-area aggregates such as sub-districts or wards).

Identify factors that influence demand

These can include healthcare policies and interventions, changes of disease burden or demographics, and other relevant variables. A review of current policies, testing guidelines and the National Strategic Plan is important in this process, and guiding questions for this review are outlined in chapter 8.1.

Review the availability and quality of data

Collect required data and review completeness, timeliness and relevance; respective methods for data review and cleaning steps are outlined in chapter 2. If desired datasets are not suitable, alternative datasets, methods and/or supplementary data can be considered, as described in chapter 4 and 5.

Chose an appropriate method

Chose an appropriate method for demand forecasting, depending on objectives and scope, factors that influence demand and data availability. Forecasting methods as well as their use cases are described in chapter 6. Guiding questions for the method choice are described in chapter 7.

Forecast demand and refine results

After developing the first set of results, it is important to discuss preliminary findings with stakeholders from all relevant departments, such as laboratory and clinical experts, as well as managers to obtain expert opinion. It might be required to adjust and fine-tune assumptions and calculation methods afterwards.

2. Data preparation

This section describes standard data review and cleaning procedures, relevant for demand calculation and forecasting. These can be applied to all datasets, but the focus is on routine health data, which are the most commonly used data for demand analysis.

2.1 Data formats and timeframes

The utilized datasets should ideally be formatted and prepared as follows:

- Data should be disaggregated by facility and month to avoid that data errors and unusual patterns are averaged out and "disappear" in the annual- or geographic aggregate.
- The observation period should ideally encompass several previous years to be able to understand the impact of past system functionality and policies in the past, and be able to account for health system shocks, such as the COVID-19 pandemic. The minimum observation period should be one year to account for any seasonal fluctuations.
- Data could be reviewed in table formats; however, it is often easier to understand patterns, identify unusual observations, trends and/or geographic clustering if data are visualized in graphs and simple geographic maps.

2.2 Data review and cleaning procedures

2.2.1 Data completeness

Missing data and incomplete datasets are one of the most frequently occurring problems of routine health data. Data could be missing from recording at facility registers or missing from reporting to the next higher level(s); typical reasons include human error, software or hardware malfunctions, interrupted connectivity and incomplete data upload, or a combination thereof. Not adequately recognizing and accounting for missing data could lead to inaccurate demand forecasts, as the below example illustrates.

Case study 2.1

A health facility originally reported 3,994 and 3,150 tests for 2021 and 2022, respectively, which suggested a decline of demand (Fig.2.1A). A more detailed analysis of monthly data revealed however that two months in 2021 and five months in 2022 were missing (Fig.2.1B).

The data analysis team in this case retrieved the missing data from the source and the complete statistics are illustrated in the second row (Fig.2.1C and D). Not surprisingly, the annual totals changed substantially after missing data were included, leading to a much higher testing demand, and, in this case, reverted the annual trend from decreasing to increasing, which illustrates the importance of reviewing and accounting for missing data.

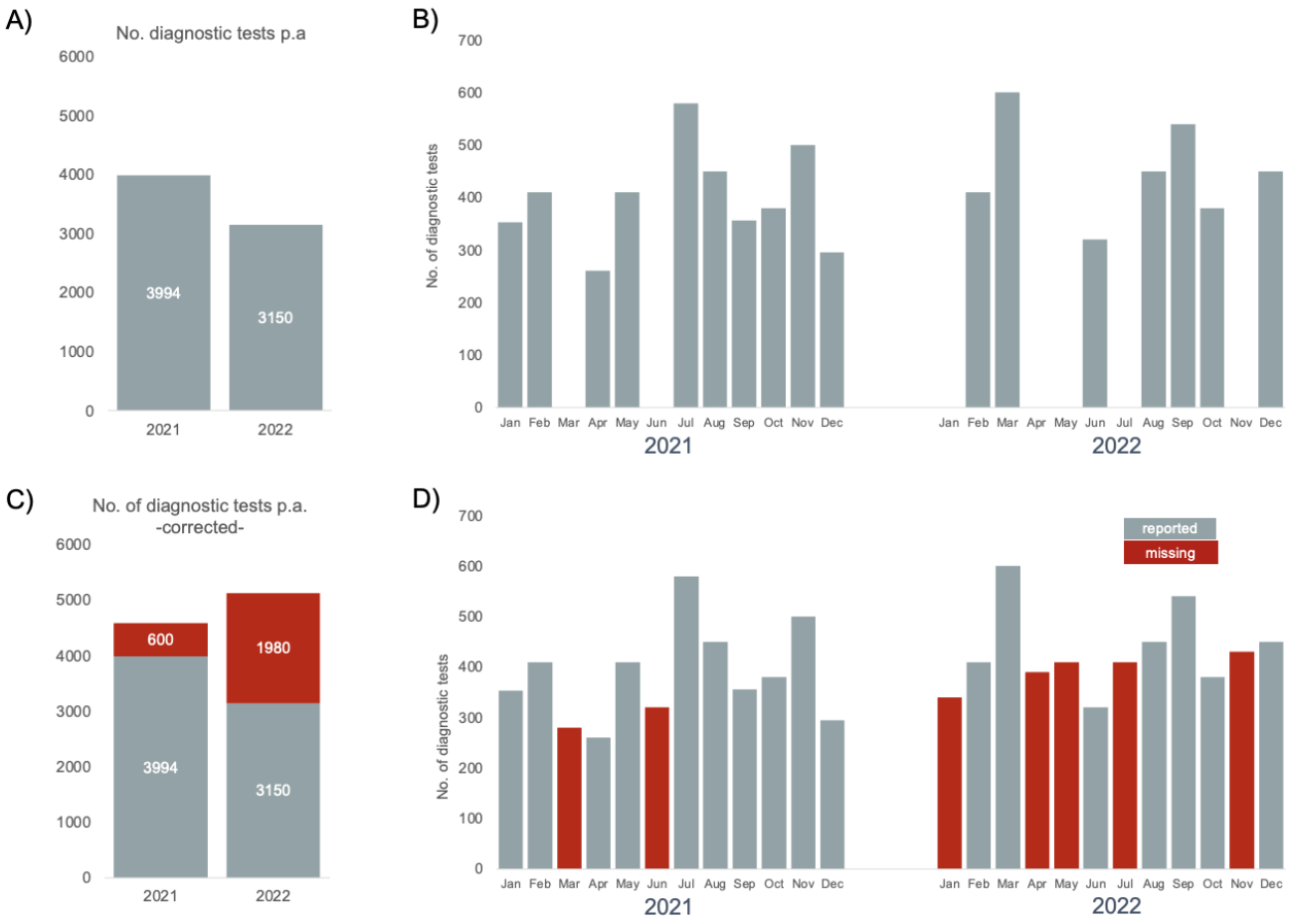


Figure 2.1: Missing data analysis and correction. A) Raw data as aggregate B) raw data disaggregated by month C) raw data (grey) and newly retrieved data previously missing (red) as annual aggregates and D) disaggregated by month.

As in case study 2.1, the most obvious sign of missing data are empty cells in the reporting tables. However, missing data that are often not so easily visible, e.g. if partial data are reported each month and no blank cells occur. In this case, missing data could be identified through examining the data for unusual patterns, calculating and comparing key indicators and/or cross-referencing the data with other datasets.

An additional source of information to identify incomplete datasets are data quality assessments or studies, such as a TB inventory study, which are often conducted by MoHs.

A stepwise approach should be taken to identify missing data:

- Examine the dataset and identifying any inconsistencies or patterns that indicate missing data. Look for empty cells, placeholders (e.g. “N/A”), low outliers or any other unusual pattern.
- Calculate summary statistics for key variables in the dataset, such as mean, median, or calculate the basic indicators (for example the percentage of positive tests by facility). If there are variables with significantly fewer observations or outlying indicator results, it may indicate missing data.

- Create visualizations, such as histograms, box plots, or scatter plots, to explore the distribution, relationships and periodicity of variables. Missing data may be visible as gaps in the plots or as unusual patterns.
- Check for geographical clustering of missing data; possibly, the majority of missing data could be observed from facilities in only a few districts or sub-districts, which is an important information to consider and account for when forecasting demand (chapter 4).
- Cross-reference the data with other datasets, for example, compare the number of tests conducted in the laboratories the reagent consumption in the same period.
- Consult the owner of the data. Respective specialists from the MoH typically know the level of completeness of the routine data and can guide next steps. It is also recommended to directly enquire with those relevant specialists if there have been any non-routine events in the past that could have led to missing data, including but not limited to IT problems with the digital data system.

The first step is to try and quantify missing data by calculating or estimating the number and percentage of missing data by facility and year. Depending on those results, it might be better to consider other, alternative datasets as the quality of demand forecasts from an incomplete dataset will be substantially diminished.

How to deal with missing data

Simple, non-statistical methods for dealing with missing data include the following below. It is important to note that those methods should be considered with great care to not practically invent or inflate demand. If a large quantity of data are missing, it is strongly recommended to rather consider an alternative dataset.

- Following the review of missing data or unusual patterns, first confirm that these observations are really due to missing data and do not reflect an actual low-demand situation: for example, what appears to be a missing value could be caused by service interruptions or reflect low demand during the festive season. This knowledge determines the subsequent approaches when accounting for missing data and estimating demand.
- Try to retrieve missing data from the affected health facilities or responsible persons; consider a different data source, e.g. instead of extracting data from the national database, more complete data could be retrieved by extracting data instead from the facility data systems directly.
- Exclude missing data from analysis, for example, if missing data are mostly observed in facilities that are located in a few districts or sub-districts, those areas could be excluded from analysis. However, this approach should be used cautiously to avoid losing valuable information; especially for geospatial analysis and diagnostic network optimization, not including a geographic area can lead to biased result.
- Imputation of missing data by replacing missing values with plausible substitutes. There are various simple imputation methods available, such as using the mean, median or mode to replace the missing values. It is always helpful to compare the dataset with other years, to get a better understand of the expected value and ensure possible seasonality fluctuations are considered.

- Missing data from individual facilities could be inferred from facilities with similar characteristics (type, instrument size, services offered) in the same area; this method does not provide good confidence though and should be considered with care.

Zero reporting

Zero reporting refers to the practice of reporting zero values in routine data systems, if no patients have been seen, investigated or diagnosed or no tests have been conducted in the period of observation. If zero reporting is consistently used, it helps to differentiate missing data from a situation where no demand occurred at a facility.

However, sometimes, zero reporting is not (or is inconsistently) used, which can have major implications for estimated demand. For example, if zero reporting is not used, a blank cell might not necessarily be missing data but could also reflect a situation in which the facility had zero demand for very real reasons. But if this zero-demand situation is interpreted as missing value and imputed, the demand would be artificially inflated. Conversely, sometimes staff that compile data reports fill blank cells with zeros, which can consequently lead to a serious underestimation of demand. In order to account correctly for missing data, it is essential to clarify if and how zero reporting is used in the dataset at hand.

Case study 2.2

This case study illustrates the approach to identify and account for missing data, including subsequent demand forecasting. A hospital reported routine mWRD testing statistics for a period of 4 years between 2019 and 2022 (Fig. 2.2). For simplicity, we assume that there was no COVID-19 pandemic. The data review revealed six time points with lower-than-average testing volumes, at two of which, data were completely missing. The subsequent assessment revealed different causes and subsequently resulted in different approaches:

- Low demand in Jan–Jun 2019: This was the start-up period of the laboratory, i.e. the months following the installation of a new mWRD instrument. The laboratory experienced some technical challenges during this phase, which all could be resolved. These are not partially missing data but a real lower testing volume; because the situation will not re-occur in the future, it is appropriate to exclude this time from demand analysis.
- Low demand in Apr–May 2020, 2021 and 2022: This was also a real low demand situation as it was festive season. Because it was Ramadan (which shifts every year by 10 days), the low demand pattern in the subsequent years changes and only affects the month of April in 2022. Because this situation will occur every year, the analyst team kept the data as they are in the demand forecast.
- Low demand in Nov–Dec 2020: this was also clarified as an actual low-demand situation, here, the diagnostic instrument broke down and services were not available. Instruments may always have technical issues; therefore, this was considered routine and not corrected.
- Sep–Nov 2022: this was determined as real missing value and the analyst team decided to replace the missing value with the average of the year.

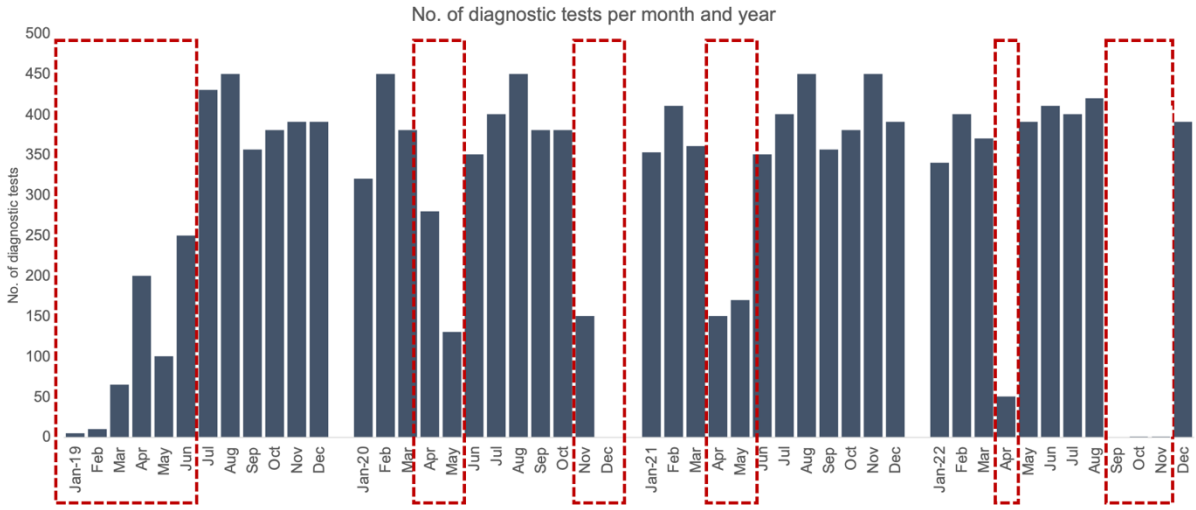


Figure 2.2: Data review and identification of low-demand times and missing data.

2.2.2 Timeliness

When reviewing the timeliness of a routine health dataset, there are several key aspects to check, including data collection and reporting frequency, reporting and processing time as well as monitoring and feedback mechanisms.

For the purpose of demand analysis, the most relevant questions to assess timeliness are:

- What was the date the respective dataset been generated and last updated? Is it the most up-to-date version?
- Are there any further updates expected? Depending on data collection and reporting frequency, as well as the electronic data system, past data could be still uploaded into the database at a later timepoint, which might be the case if network connectivity is not always available. Consequently, the currently available dataset might be still incomplete.
- Are the data the final and officially validated data by the MoH? In most countries, routine data are first quality reviewed and validated before declared final and official. Due to the expected higher data quality after the final validation step, it might be worthwhile to wait for this quality review before commencing analysis.

2.2.3 Internal and external consistency

Consistency checks refer to the review of coherence and logical consistency within one dataset (internal consistency) and between datasets (external consistency) to verify accuracy and reliability of the dataset.

The following aspects should be reviewed when checking consistency:

- Calculate summary statistics for key variables in the dataset and check for consistency: for example, the sum of the number of tests from all facilities should match the total reported number of tests (internal consistency). Or, the facilities reporting testing volumes in the laboratory statistics dataset should be identical to the facilities on the laboratory inventory dataset (external consistency).

- Verify that data values fall within reasonable ranges. For example, all facilities report to have between 1 and 1,000 patients on ART, but one facility reports more than 10,000 patients on ART. This data point might be a mistake and should be further reviewed (internal consistency).
- Conduct logic checks. If a facility has been labelled as “closed” in the master facility list, there should not be subsequent records of patients or tests from this facility (internal/external consistency). Or, the number of test results for rifampicin (RIF) resistance on Xpert MTB/RIF cannot be higher than the number of test results for MTB positive.
- Check for duplicates, for example, duplicate facilities or duplicate patient records. (internal consistency).
- Compare the values, trends or distributions between datasets. For example, if the number of tests reported over a specific time period increases, one would expect that the reagent consumption also increases over the same time period (external consistency).

If discrepancies are identified, either within or across dataset, these should be clarified to determine if they are genuine errors, possibly missing data, or if there are more systemic factors, contributing to the differences, as below examples illustrate.

This root cause analysis might reveal that facilities, which are still reporting data although labelled as “closed” because the master facility list is outdated and does not capture that the facility has been re-opened. This observation has important implications for the entire analysis and guides the correct action: here, it would be either update or discard the status variable in the master facility list. At the same time, considering that the status variable is not up-to-date might reveal missing data, if facilities without data were previously incorrectly assumed to be closed.

If too many discrepancies are discovered, it might be better to consider other, alternative datasets as the quality of demand forecasts from such datasets will be substantially diminished.

2.2.4 Relevance of data: health policies and changes

In the context of this guide, the relevancy of data refers to whether or not historical data are actually representative for the future situation, for which the demand should be forecasted.

For example, using historical data from the years that were heavily impacted by the COVID-19 pandemic might be of very limited relevance as all health systems worldwide experienced a disruptive event, leading to temporarily reduced health seeking and availability of health services in most countries. Demand forecasted from years with disruptive events that are not expected anymore in the future, would not provide an adequate basis for the future demand.

However, there are other, less severe events, that could render historical data as irrelevant, as the following case study shows.

Case study 2.3

Figure 2.3 shows the 4-year historical testing volumes of a district, which the analyst chose as period to use for demand forecasting. The context review revealed that the majority of community health worker contracts were discontinued in Q1-2022 in the entire country, which

led to a reduction of historical testing volumes from Q2-2022 onward. The situation was ongoing for some time and was not expected to change anymore.

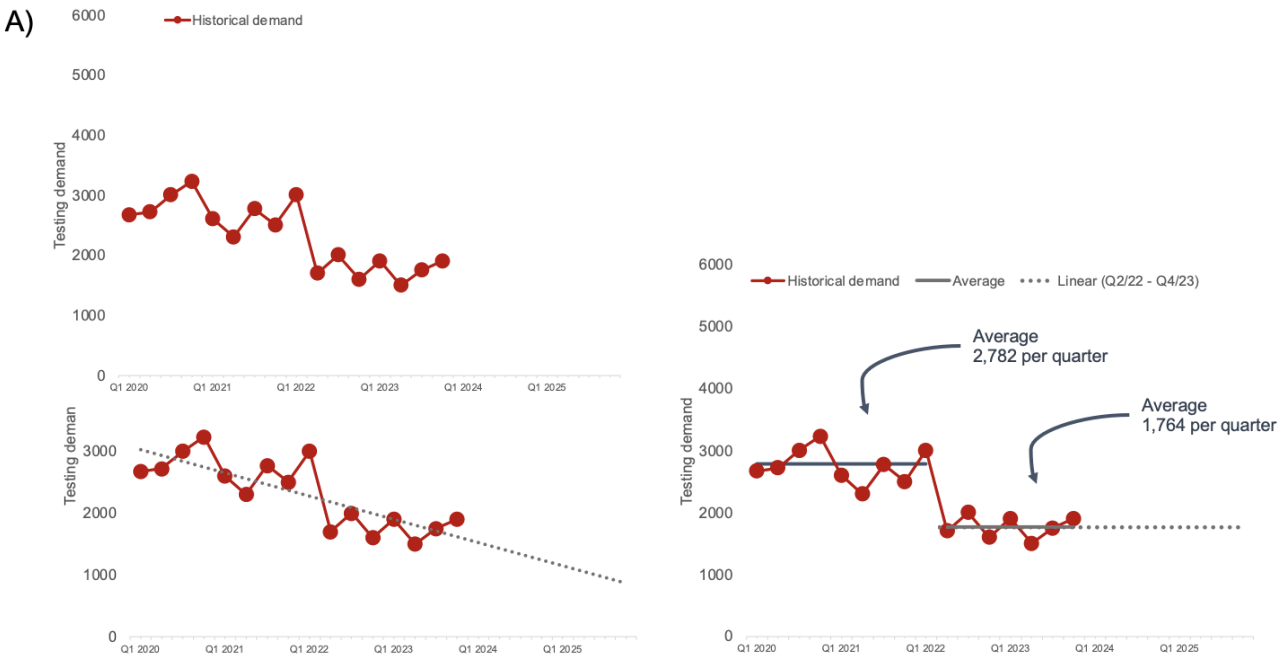


Figure 2.3: Relevance of historical data for future demand. A) Raw data B) left: application of a linear trend across the entire time period. Right: application of a linear trend across the relevant time period.

In this situation, the relevant time period to use for demand forecasting is data from Q2-2022 onward, which can be considered the “new normal” from now on. The importance of excluding the non-relevant time period from demand forecasting is illustrated in Fig.2.3B: if for example a linear trend is fitted to include the time during which community health workers were still employed, the demand trend would substantially decline in the future.

If instead only the relevant period is included (the new normal), the linear trend would be constant and much higher than in the first example.

2.2.5 Relevance of data: geospatial distribution

If demand is forecasted for the purpose of geospatial analysis and diagnostic network optimization, ensuring the precise location of forecasted demand is critical. To achieve high accuracy, it is important to forecast demand using a dataset that is both geographically representative and appropriate; the following factors should be considered when reviewing the geospatial relevance of a dataset.

Testing location versus actual patient location

Depending on the design of the monitoring and evaluation (M&E) system, there might be difference between the real demand location and the location of where the test is registered and reported. This might be the case if specimens for testing are collected at the patients’

origin facility but transported, tested and recorded only at the laboratory. Here, the real demand location would be the patients' origin facility and historical laboratory testing volumes only represent a clustered and aggregated version of demand in the area.

Notification versus testing versus actual patient location

Secondly, problems might arise when alternate data need to be used, for instance, when people with TB notification data have to be used because actual TB testing volume data are not available at good quality. In some contexts, TB notification facilities are not the same as testing facilities or the patients' origin facilities as TB notification is only conducted at a sub-set of all facilities providing TB services. Also in this situation, the geospatial distribution of notification data (and consequently the demand forecasted from those) are not representative, and thus with limited relevance for actual demand locations.

Current versus future service locations

Healthcare facilities offering specimen collection, referral and/or testing services (i.e. demand locations) might change in the future. If the service locations in the future differ notably from the current locations, historical demand data might not be geospatially relevant for future demand locations. A typical example of this is future diagnostic network expansion due to the integration of new facilities into the specimen referral network, which is expected to increase demand at new locations and could also decrease demand from existing locations.

In each of these examples, reviewing geospatial distribution and relevance is crucial, as discrepancies between registered, reported, and actual demand locations can skew analyses and results. Should the review reveal that the initially chosen dataset lacks relevance, the standard approach would be to adjust the demand forecasting method, including considering alternative datasets.

Private-sector health services

In many countries, the private sector provides services to a substantial number of patients, and therefore, it should be clarified if the historical health system data include private-sector facilities or not. Often, diagnostic tests, diagnosed or treated patients are only incompletely reported from the private sector to the MoH, if at all.

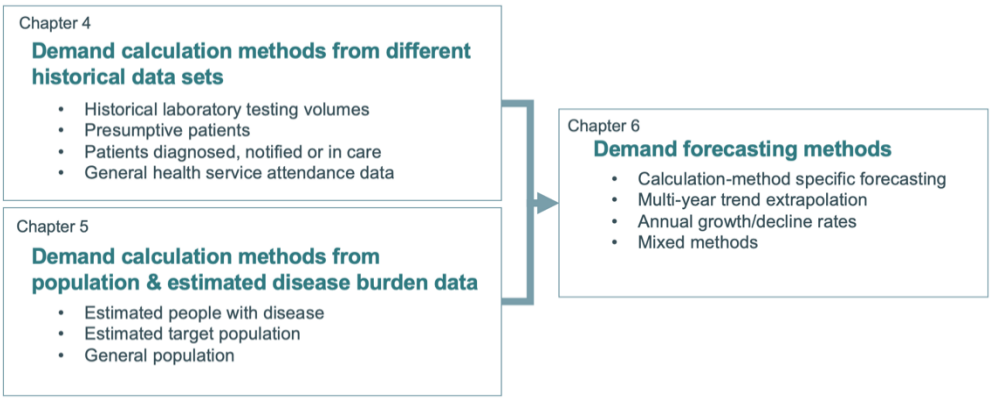
If demand should be forecasted for the entire country, but historical data only include a fraction of private-sector facilities, these data are not representative for the situation for which demand should be forecasted. It should be considered to use a different approach and base dataset instead.

3. Overview of diagnostic demand methods

Most intuitively, diagnostic testing demand is calculated using historical data of testing volumes. However, diagnostic testing demand can also be calculated from other types of data, such as people with presumed, diagnosed or notified with disease or general health service attendance data.

In this guide, we categorized this approach into “demand methods from historical datasets”, as they are all based on routine health service data from the past year(s). The respective calculation methods will be introduced in chapter 4, where we explain, for example, how to calculate the number of HIV viral load (VL) tests from the number of patients enrolled into ART.

But demand can also be calculated from population data or estimated patients with disease. These methods entirely use estimates and are therefore independent of past health system functioning and coverage. We categorized these methods into “methods from population and disease burden estimates” and will explain in chapter 5 how to calculate, for example, the number of TB tests from the WHO-estimated TB incidence.



Both approaches have advantages and disadvantages, and which method type to choose often depends on availability of quality data, the scope and objectives of the analysis as well as the level of detail required (read more in chapter 7). However, the key differences between the two types of data lie in the fact that one relies on health system data, while the other does not, which has implications for analysis and findings, as outlined in the table below.

<table>
<tr><th colspan="2">Demand from historical facility data</th></tr>
<tr><th>Advantages</th><th>Disadvantages</th></tr>
<tr><td>Context-specific, real-world data that account for disease burden, health-seeking behaviour and health system factors (such as availability and accessibility) at once
Strongly recommended to determine the baseline demand
Easy and relatively reliable for forecasting short-term demand in a minimally changing context in the future, for example when no major health system interventions are planned and no other major changes of population or disease burden are expected</td><td>Routine health data might have quality limitations, especially the risk of missing data is high that can have a large impact on forecasted demand
Demand forecasting from these data require a good understanding of past events in health system operations, policies or disruptive events such as reagent stock-outs
If large changes are going to be made to the healthcare system, such as large expansion of service locations, or substantial algorithm changes, forecasting demand from historical data might not be representative anymore for the future, consequently, forecasted demand might be inaccurate.</td></tr>
<tr><th colspan="2">Demand from population and disease burden estimates</th></tr>
<tr><th>Advantages</th><th>Disadvantages</th></tr>
<tr><td>Allows a more accurate calculation of testing demand if health seeking and/or health receiving would change or be less of a bottleneck in the future
Allows the consideration of population- and disease burden changes
Recommended to be used to determine a benchmark in the long-term future, regardless of other methods for interim steps
Quick to calculate, especially if high-level aggregates and assumptions are made</td><td>Population and disease burden are typically not available at facility level, which requires additional distribution approaches at the cost of accuracy and loss of geographical granularity
Potential challenges with disease burden data can be high, depending on the source of data (e.g. survey, surveillance system, prediction models)
High-level aggregates and assumptions lead to reduced accuracy and loss of granularity compared to demand calculated from historical data</td></tr>
</table>

Once demand has been calculated, different methods can be applied to forecast future demand, which will be discussed in chapter 6.

4. Demand calculation methods from historical data

4.1 TB demand calculation methods from historical data

4.1.1 Demand from historical laboratory testing volumes

Calculation	Demand = total number of diagnostic tests conducted in period of observation
Required data	Number of diagnostic tests conducted and/or referred
Stratification as required	By test technology mWRD (GeneXpert, Truenat, TB LAMP), smear microcopy (SM), others, as relevant for analysis design By facility (or small geographical aggregate) By month (or at least by quarter)
Time period	At least the most recent year Ideally include 3-4 years of historic data for better interpretation
Advantages	This is the easiest and most realistic method to determine demand as no further assumptions about disease burden, health-seeking behaviour or the impact of health system factors have to be made
Disadvantages	Routine health service data might be incomplete or have other quality problems that are sometimes difficult to identify. In contrast to mWRD data, SM data are often not available in digital format by facility but only digitized at higher level

Calculating demand from historical testing volumes is the most accurate and simple method[4].

Good to know: location of testing versus location of sample taking

If testing demand is required for a geospatial analysis, it is necessary to attribute demand to a health facility or testing location. Depending on the recording and reporting system, supplementary data might be required to correctly attribute demand to the origin facility because laboratory registers typically include the number of tests conducted, which need to be ideally stratified by origin facility to be able to correctly attribute demand to the origin facilities. If those are however not available from the laboratory register, supplementary data are required that sum the number of specimens referred by origin facility. These could be for example specimen referral data (chapter 8.2).

Should referral data also not be available, but it is at least known which facilities refer specimen, an alternative approach could be to redistribute the total number of tests from the laboratory records back to origin facilities, using proxy data (such as TB notification data, OPD attendance or population) or by making generic assumptions. As this leads to inaccuracies, this approach should be carefully balanced with advantages and disadvantages of using an alternative base dataset.

[4] Please refer to the section with “Definitions” for the difference between testing volumes and demand.

Good to know: technology-specific testing demand

Most often, the key interest of TB diagnostic demand forecasting are the number of mWRD tests. The total diagnostic testing demand would be the sum of all diagnostic tests conducted using all the different technologies. But carefully considering the diagnostic algorithm, policies and standard practices is essential to accurately calculate mWRD demand:

- There might be a notable amount of unnecessary double testing, which is conducted not in line with the guidelines, whereby healthcare workers order mWRD and SM for diagnosis in parallel. This would inflate the total diagnostic demand and should therefore be accounted for, if it is expected that double testing frequently occurs and will be resolved in the future. The same applies for other tests, such as the urinary LF LAM, which is recommended by WHO to be accompanied by an mWRD test, if possible.
- SM laboratory statistics normally record the number of smears, whereby one patient might have more than one sample investigated, depending on the diagnostic algorithm. In this case, the total number of smears has to be divided by the average number of slides per patient, the latter can typically be provided by the national reference laboratory (NRL).
- Pay extra attention to the quality of SM data if these are used: while mWRD data are often digitally collected and are therefore of good quality, SM data are still mostly manually recorded and reported, with implications for data quality.
- Account for unsuccessful tests if those are not included in the historic testing volumes.

4.1.2 TB demand from historical people with presumed TB

Calculation	(1) Demand = number of people with presumed TB* percentage tested with diagnostic test (or diagnostic test of interest) Or (2) Demand = number of people with presumed TB* average number of tests per person with presumed TB
Required data	Clinic data: number of people with presumed TB in time period and percentage tested with a diagnostic test (or diagnostic test of interest) or average number of tests per person with presumed TB
Stratification as required	By test technology mWRD (GeneXpert, Truenat, TB LAMP), SSM, others, as relevant for analysis By facility (or small geographical aggregate) By month (or at least by quarter)
Time period	At least the most recent year Ideally include 3–4 years of historical data for better interpretation
Advantages	This is also an easy and realistic method as no further assumptions about disease burden or health-seeking behaviour have to be made, but the percentage that are tested (i.e. the health system performance) have to be accounted for. If data are of good quality, people with presumed TB are more representative for the real demand than testing volumes.

	In future demand projections, this method offers the opportunity to estimate demand if the percentage of patients tested is increased, representing a good link to future planned interventions.
Disadvantages	If generic assumptions about the percentage of patients that were tested have to be made, this method loses accuracy. Similarly, if national-level data are used, spatial variation differences are lost. Presumptive TB registers are not frequently used in routine services and if they are, they are often only available as hard copy at facility-level but not digitally for all facilities; routine health data might be incomplete.

Using people with presumed TB volumes to calculate testing demand could be an interesting dataset as the number of people with presumed TB is not influenced by laboratory service availability, e.g. a low number of recorded tests conducted during reagent stock-outs.

The variation in formula (2) might be specifically useful if demand is required at facility-level, for example for geospatial analysis, but testing volumes are not available at facility level.

Presumptive TB data availability and quality limitations frequently occur: the number of people with presumed TB is typically not part of the routine recording and reporting system from facility to national level and respective registers are not necessarily widely used. If they are, records are often kept as hard copy registers, without aggregate reporting to the next higher levels, which might limit the accessibility of data. Lastly, presumptive TB registers are sometimes not well kept, which could lead to an underestimation of demand. If none of those aspects are of concern, using presumptive TB data to estimate demand is highly recommended.

Calculations and modifications

If precise patient data on the type of test used to investigate for TB are not available from the presumptive TB register, assumptions have to be made for the percentage of patients that were tested with a diagnostic test (or diagnostic test of interest).

Alternatively, the average number of tests per person with presumed TB can be used. Both can be based for example on the national aggregate or average, or generically assumed by a subject matter expert such as a laboratory or clinical specialist from the national TB programme (NTP).

Of note, if national aggregates or assumptions are used, the result becomes less accurate and essentially averages out geospatial differences in screening and testing efforts. This limitation can be reduced if regional or district-level aggregates if available and used, which would allow to include at least some geospatial differences.

4.1.3 TB demand from historical people notified with TB

Calculation	Demand = number of people notified with TB* average number of diagnostic tests (or diagnostic test of interest) conducted per person notified with TB

Required data	Clinical data: number of people notified with TB Laboratory data: national total number of diagnostic tests (or diagnostic test of interest) conducted
Stratification as required	By test technology mWRD (GeneXpert, Truenat, TB LAMP), SM, others, as relevant for analysis By facility (or small geographical aggregate) By month (or at least by quarter)
Time period	At least the most recent year Ideally include 3–4 years of historic data for better interpretation
Advantages	Still a relatively simple method, that does not require assumptions about health-seeking behaviour or health system functioning. Notification data are typically of good quality.
Disadvantages	Using the national average number of tests leads to inaccuracy as individual facility-level testing efforts cannot be accounted for anymore. A disadvantage could occur if the patients are notified at a different location than their origin facility or the facility that collects and sends the specimen, which would represent a geographic clustering of demand.

This method is especially useful if testing demand is required at facility level (for example for a geospatial analysis), but the actual facility-level testing volumes are not available. If instead the number of people notified with TB is available at facility level, it only requires the national-level sum of diagnostic tests to calculate the average number of tests per person notified with TB in the same year, and with this, estimate the individual facility demand.

If the demand of interest is the mWRD testing demand, the total sum of mWRD diagnostic tests per country can be also obtained from publicly available WHO TB data.[5] For other test types, national-level sums have to be obtained from the NTP.

Multiplying the number of people notified with TB with the average number of tests conducted is less accurate compared to previous method as it cannot account anymore for the individual facility testing efforts. These differences will be lost with this method and it should be reviewed if, and to what extent, this will impact the geospatial analysis.

Watch out: notification data might not be geographically representative for testing demand

It is critically important to review whether TB notification data are a clustered version of testing demand: in some countries, a large number (if not all) of facilities investigate patients for TB, collect and refer specimens for testing, while notification of a person with TB formally occurs only at a sub-set of facilities. In a situation like this, the number of people notified with disease is essentially a clustered proxy for the actual testing demand.

[5] WHO Global TB Report. Laboratory data set Variable: m_wrd_tests_performed. Available from: https://www.who.int/teams/global-tuberculosis-programme/data

It can help to generate simple maps to visually inspect for any clustering, it is however not mandatory: if the number of notifying facilities is notably smaller than the number of facilities that investigate for TB, collect and refer specimen, this is a good enough indication for clustering. Reviewing alternative datasets is recommended in this situation as clustered demand data are of limited use for geospatial network optimization analysis.

A (less than ideal) alternative is to reassign the number of people notified with TB (and hence estimated demand) back to their origin facilities, for example by using other proxy data (such as OPD data) or by making generic assumptions. This requires knowledge about precise facility linkages as well as the additional proxy data or reasonable assumptions. Because this approach leads to additional inaccuracies, it should be only considered with utmost care.

4.1.4 TB demand from historical general health service attendance

Calculation	(1) Demand = number of OPD consultations * percentage symptomatic with presumed TB * percentage identified * percentage tested with diagnostic test (or diagnostic test of interest) Or (2) Demand = number of OPD consultations * average number of TB tests per OPD consultation
Required data	Clinical data: number of outpatient department (OPD) consultations (ideally new consultations, excluding follow-up consultations) Estimates for percentage symptomatic with presumed TB, percentage identified and percentage tested with diagnostic test (or diagnostic test of interest)
Stratification as required	By test technology mWRD (GeneXpert, Truenat, TB LAMP), SM, others, as relevant for analysis By facility (or small geographical aggregate such as sub-district) By month (or at least by quarter)
Time period	At least the most recent year Ideally include 3-4 years of historic data for better interpretation
Advantages	OPD data at facility-level are often easily available as many countries have digital databases (such as DHIS2) established. This method does not have to make assumptions about health-seeking behaviour, and offers the opportunity to integrate future improvements in the cascade of care at OPD into future demand forecasting.
Disadvantages	Expert opinion is required to estimate the percentage of OPD patients that successfully went through the TB patient cascade of screening, identification, sample collection and laboratory testing, as these values are typically not routinely collected data; this reduces the level of accuracy and/or increases complexity.

Basing future demand forecasting on OPD data might be suitable if for example historical TB-specific data are geospatially not representative for the future demand situation. However, if

the future demand situation still has a comparable coverage as a more limited TB healthcare system, this method might not be suitable.

Details about required data

The following information sources and proxy data could support the various assumptions about the patient pathway at OPD: depending on the recording and reporting system, for example in DHIS2, the number of OPD consultations might be already recorded by main morbidity categories, e.g. lower respiratory tract infections (LRTI), which are a good start to estimate the number of patients with symptoms; a further refinement is however required to estimate the percentage of people with presumed TB among patients with respiratory symptoms.

Of note though, the differentiation into upper and lower respiratory tract infection recording categories might sometimes not be very accurate though and expert advice should be sought before using those morbidity categories. If patients present with more than one condition at OPD, often only one morbidity is recorded in the data, which can be either the condition that triggered the visit or the most severe condition. This could lead to an under recording of respiratory tract infections and consequently lead to an underestimation of demand if morbidity categories are used. Expert opinion should be sought to carefully assess if any patients would be missed if morbidity categories are used.

Datasets that could help determining the various assumptions about the patient care cascade at OPD (formula 1) are dedicated surveys, assessments or research. These include, but are not limited to, individual country household health access and/or utilization surveys, WHO Service Availability and Readiness Assessment (SARA),[6] Demographic Health Surveys (DHS),[7] TB Patient Pathway Analysis (PPA)[8] or a dedicated TB patient cascade of care analysis. If available, subnational estimates should be used to account for geographical variations of receiving TB-specific healthcare.

Modifications

The calculation can be simplified by making only one assumption about the average number of tests conducted per OPD consultation (formula 2). The average number of diagnostic tests per OPD consultation can be determined from the total, national-level sum of diagnostic tests, divided by the total number of OPD consultations. Individual facility demand is then calculated by multiplying the OPD consultations with that average number of tests. In doing this, OPD data essentially become a facility distribution key for the sum of diagnostic tests conducted, which can be considered if TB-specific data are not available at facility level; as before, this approach cannot account for individual facility-level testing efforts anymore. Of note, if national aggregates or assumptions are used, the result becomes less accurate and essentially averages out geospatial differences in screening and testing efforts. This limitation can be

[6] WHO Service availability and readiness assessment (SARA). Available online [https://www.who.int/data/data-collection-tools/service-availability-and-readiness-assessment-(sara)]

[7] The DHS Program. Available online: [https://www.dhsprogram.com/]

[8] StopTB Partnership TB Patient Pathway Guide. Available online [http://www.stoptb.org/assets/documents/global/awards/tbreach/TB_Patient%20Pathways%20Guide.pdf]

reduced if regional or district-level aggregates if available and used, which would allow to include at least some geospatial differences.

Depending on data and information available and the level of detail anticipated, the above formula (1) can be even further broken down into smaller, more granular steps, such as the percentage symptomatic, percentage screened, percentage correctly identified, percentage samples taken, percentage samples received, percentage samples referred, percentage samples tested). However, his approach becomes very complex very rapidly, and the more assumptions have to be made without actual data, the less accurate the calculation becomes.

4.2 HIV demand calculation methods from historical data

4.2.1 EID demand from historical laboratory testing volumes

Calculation	Demand = total number of EID tests conducted in time period
Required data	Laboratory data: number of tests conducted
Stratification as required	By technology: POC, conventional methods, as relevant for analysis By type: number of first, second, third EID test, as relevant for analysis By facility (or small geographical aggregate) By month (or at least by quarter)
Time period	At least the most recent year Ideally include 3-4 years of historic data for better interpretation
Advantages	An easy and realistic method to determine demand as no further assumptions about disease burden, health-seeking behaviour and/or the impact of health system factors have to be made.
Disadvantages	Routine health service data might be incomplete or have other quality problems that are sometimes difficult to identify.

Depending on the programmatic context and the purpose of the demand forecast, historical EID testing volumes might have to be stratified by technology categories, such as point-of-care (POC) devices and conventional platform testing. This depends essentially on national policies and algorithms for the use of each of those platforms.

In this regard, it might be useful to explore underlying factors that might have impacted the proportional use of platforms in the past: for example, a shortage of reagents for POC platforms might have reduced the total number of EIDs conducted on POC and increased the number of tests conducted on conventional platforms. This proportional change might not reflect a normal, routine situation and should be accounted for, either by adjusting the period of analysis to include the years prior to the reagent shortage, by making assumptions or by choosing an alternative base dataset all together.

Good to know: location of testing versus location of sample taking

If demand is required at the level of the original facility, for example for geospatial analysis, supplementary datasets might be needed because the laboratory data might not include

information on origin facility. In this case, supplementary referral data could be used for reattribution of demand to origin facility (chapter 8.2).

Should referral data also not be available, but it is at least known which facilities refer specimens, an alternative approach could be to calculate demand from patient data as illustrated in section 4.2.2, or redistribute the total number of tests from the laboratory back to origin facilities, using proxy data (such women on ART) or by making generic assumptions. Redistribution by proxy data can lead to inaccuracies, and this approach should be carefully balanced with advantages and disadvantages of using an alternative base dataset.

In-depth analysis

If data are available, it might be useful to conduct a sub-analysis by algorithm, and consider the number of first (at birth), second and third EID. This allows a better interpretation of demand and might be helpful of future demand forecasting for which the analyst could model a strengthened algorithm implementation in the future.

4.2.2 EID demand from historical patient data

Calculation	(1) Demand = sum of tests for the number of HIV-exposed infants, considering percentage of infants tested at each step of the testing algorithm and percentage of confirmatory testing at each step Or (2) Demand = number of HIV-exposed infants * average number of EID tests per infant
Required data	Clinical data: number of HIV-exposed infants Laboratory data: national sum number of EID tests conducted Performance data: percentage infants tested at each step and percentage of positive tests among EID at birth and follow-on testing that require confirmatory testing
Stratification as required	By technology: POC, conventional methods, as relevant for analysis By facility (or small geographical aggregate such as sub-district) By month (or at least by quarter)
Time period	At least the most recent year Ideally include 3–4 years of historic data for better interpretation
Advantage of method	This method is relatively simple to execute if high-level assumptions or national aggregates are used. It offers opportunity for detailed analysis, although it does become more complex the more granularity is added. Compared to a method based on historical laboratory testing data, these can be easier linked to programmatic activities as it is based on patients in care.
Disadvantage of method	Routine health service data might be incomplete or have other quality problems that are sometimes difficult to identify.

This is a simple way to calculate demand from patient data.

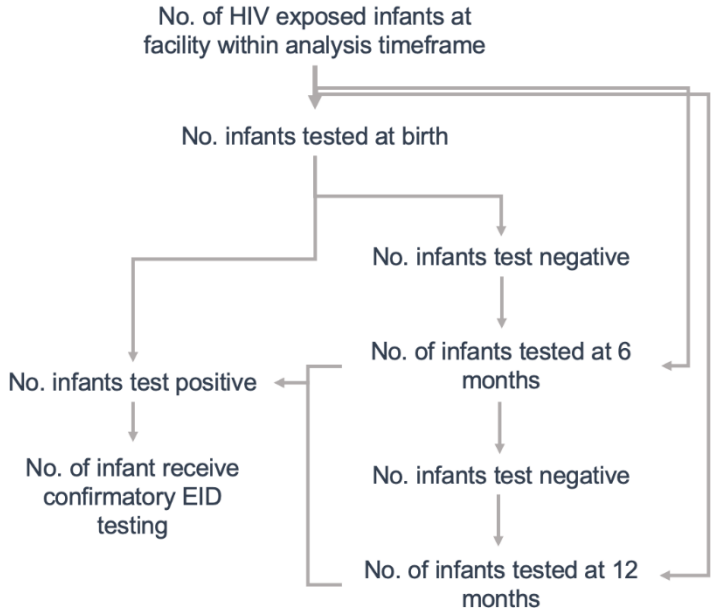
While using national averages or performance aggregates (i.e. percentage tested for initial- and confirmatory testing) introduce inaccuracies, as geographic differences or facility testing efforts (or algorithm implementation) are simply averaged out, these methods offer notable advantages. They could be used if demand is required at facility level (e.g. for a geospatial analysis), but actual testing volumes are not available at facility level and only as national sum. Another major advantage is that later demand forecasting can be easily linked to programmatic targets and/or clinical interventions, such as a strengthening of algorithm application.

Furthermore, the lack of geospatial detail can be reduced if regional or district-level aggregates are available and used, which would allow to include at least some geographical differences.

Calculations and modifications

The number of HIV-exposed infants might be available from prevention-of-mother-to-child transmission (PMTCT) HIV programme data or antenatal care (ANC) programme, depending on the context. If the actual number of HIV exposed infants per facility are not available, the number of women enrolled in a the PMTCT programme could be used as alternative. The simplest approach would be to assume that, on average, one woman in PMTCT = one infant, but other assumptions can be made based on context. It is important to review if PMTCT data are representative enough for the testing demand, e.g. if a large number of EID tests originates from patients who are not formally enrolled and registered in PMTCT, these data might be of limited use.

Formula 1 can be used to calculate demand following the testing algorithm and use either national-level performance data or assumptions for each step, as illustrated in an example testing algorithm below. Starting from the number of HIV-exposed infants, testing demand can be calculated for each step by applying (assumptions or national aggregate data for) the percentage tested at birth, percentage tested negative/positive, percentage re-tested at 6 months, percentage received confirmatory testing, and considering that some infants might not be tested at birth but only at 6 months of age etc. The testing algorithm illustrated here must be adjusted to the national algorithm.



Forecasting diagnostic testing demand for geospatial analysis

Using this algorithm-based calculation allows also for an adjustment based on entry age, which might be necessary if some infants are not tested at birth but only at a later age (such as 6 or 12 months as in the example algorithm above).

Overall, the approach in formula 1 is complex but it does however offer major advantages if later demand forecasting should be linked to programmatic activities that target, for example, a strengthened testing algorithm implementation, and specifically targets the improvement of, for example, the increase of infants tested at birth.

Alternatively, the number of exposed infants (or mothers in PMTCT) could be multiplied with the national EID coverage, which is a standard indicator. Of note, the EID coverage calculates only the percentage of HIV-exposed infants born during the reporting period who received a virological HIV test within 2 months (and 12 months) of birth and does not account for repeat testing, which would also have to be accounted for.

The national average of EID tests per infant in formula (2) can be calculated using the total national sum of EID tests, divided by the total number of HIV exposed infants in the same time period. This may be stratified by type of EID, i.e. first (at birth), second and third test, if required for the demand analysis. This simple approach may be used if demand is required at facility-level and the only available data are the number of HIV exposed infants. Applying the national average is a simplification though, which cannot account anymore for facility-level difference in testing efforts and/or algorithm implementation but could be considered if no better approaches are possible.

4.2.3 VL demand from historical laboratory testing volumes

Calculation	Demand = total number of VL tests conducted in time period
Required data	Laboratory data: number of HIV VL tests conducted
Stratification as required	By technology: POC, conventional methods, as relevant for analysis By patient group: for priority groups for POC VL, for routine follow-up of non-priority groups, as relevant for analysis By facility (or small geographical aggregate) By month (or at least by quarter)
Time period	At least the most recent year Ideally include 3–4 years of historic data for better interpretation
Advantages	An easy and realistic method to determine demand as no further assumptions about disease burden, health-seeking behaviour and/or the impact of health system factors have to be made.
Disadvantages	Often, demand needs to be forecasted separately for priority patient groups and all other patients, which might not necessarily be recorded in the laboratory register and thus requires additional approaches. Routine health service data might be incomplete or have other quality problems that are sometimes difficult to identify.

Very often, demand for VL testing is required separately for different patient groups, e.g. when VL for priority group patients (such as pregnant and breastfeeding women (PBFW) and/or children) are conducted on POC instruments, while all other, routine follow-up HIV VL testing for non-priority groups is conducted on conventional platforms. Ministries of Health might have different scale-up plans for VL testing services for the different platforms, and if that is the case, demand forecasting should be conducted in patient strata.

Historical testing volumes from the laboratory might not record the patient categories (e.g. priority vs. routine). In this case, it might be feasible to assume that tests conducted on a POC device belong to a patient from priority groups, however, this should be carefully reviewed: a laboratory with both, conventional and POC platforms, might simply assign VL specimen to the next available platform, or the platform for which reagents are available, irrespective of patient category. If this is a concern, a method based on patient data should be used instead (see section 4.2.4)

Good to know: origin facility

Demand for geospatial analysis typically requires an attribution of demand to the origin facility. But depending on the recording and reporting system, the laboratory data might not record the origin facility of a specimen. In this situation, alternative or supplementary data are required that provide information about the number of specimens referred by origin facility, such as historical specimen referral data. Alternatively, historical patient data might be used if available at the facility-level.

4.2.4 VL demand from historical patient data

Calculation	(1) Demand = number of patients receiving ART * percentage of patients tested (refine formula considering the national testing schedules for different patient groups, see text for details) Or (2) Demand = number of patients receiving ART * average number of VL tests per patient
Required data	Clinical data: number of patients on ART (newly and previously enrolled) Laboratory data: national sum number of VL tests conducted, or Performance data: (national level) percentage of patients tested as per algorithm
Stratification as required	By patient group as per national guidelines: priority group patient eligible for POC VL and non-priority group patient By age group: if VL schedule depends on the age of the patient By facility (or small geographical aggregate such as sub-district) By month (or at least by quarter)
Time period	At least the most recent year Ideally include 3–4 years of historic data for better interpretation
Advantages	This method is relatively simple to execute if high-level assumptions or national aggregates are used. It offers opportunity for detailed analysis, although it does become more complex the more granularity is added. Compared with methods based on historical laboratory testing data, this method can be more easily linked to programmatic activities as it is based on patients in care.
Disadvantages	Using the national average or performance aggregates for VL testing is a simplification, which lacks geographic resolution. Routine health service data might be incomplete or have other quality problems that are sometimes difficult to identify.

Using national averages or national aggregates of performance values introduce inaccuracies, as geographic differences or facility testing efforts (or algorithm implementation) are simply averaged out. However, these methods offer notable advantages, as they could be used if demand is required at facility level (e.g. for a geospatial analysis), but actual testing volumes are not available at facility level and only as national sum. Of note, the lack of geospatial detail can be reduced if regional- or district-level aggregates are available and used, which would allow the inclusion of at least some geographical differences.

A notable advantage of the above approaches is that they allow a strong linkage of forecasted demand with planned interventions in the future (chapter 6), such as improvement of performance in applying the testing algorithm.

Calculations and modifications

In formula (1) the national algorithm is applied; here, we only show a simple formula, which will have to be adjusted in accordance with national testing guidelines to consider the frequency of VL testing depending on patient age, priority group criteria and date of ART initiation. Additional considerations for repeat testing for patients with non-suppressed previous VL results should be integrated as well and the calculations, similar to the example used in the EID testing schedule in chapter 4.2.3.

Formula (2) is a very simple approach, whereby the average number of tests can be calculated from the total national sum of VL divided by the total number of patients on ART at the end of the reporting year. Stratifications can be made by priority group patients and patients for routine VL, if those data are available. This approach is suitable where testing demand is required at facility-level but actual testing volumes are not available at this level. Of note, here, facility differences in testing efforts and performance are averaged out and the granularity of geospatial analysis is reduced.

4.3 HPV demand calculation methods from historical data

4.3.1 HPV demand from historical laboratory testing volumes

Calculation	Demand = total number of HPV tests conducted in time period
Required data	Laboratory data: number of tests conducted
Stratification as required	By test type and technology: molecular methods (POC, conventional methods) and non-molecular methods (such as Pap smear), as relevant for analysis By facility (or small geographical aggregate) By month (or at least by quarter)
Time period	At least the most recent year Ideally include 3–4 years of historic data for better interpretation
Advantages	An easy and realistic method to determine demand as no further assumptions about population, screening and testing efforts have to be made.
Disadvantages	Routine health service data might be incomplete or have other quality problems that are sometimes difficult to identify.

Depending on the programmatic context, screening and testing guidelines, and the purpose of the demand forecast, historical HPV testing volumes might have to be stratified by technology categories, such as molecular testing (on POC devices and/or conventional platforms) and non-molecular laboratory tests (such as Pap smear).

Good to know: origin facility

Demand for geospatial analysis typically requires an attribution of demand to the origin facility. However, depending on the recording and reporting system, the laboratory data might not record the origin facility of a specimen. In this situation, alternative or supplementary data are required that provide information about the number of specimens referred by origin facility,

such as historical specimen referral data. Alternatively, historical screening data might be used if available at facility-level.

4.3.2 HPV demand from historical screening data

Calculation	(1) Demand = number of people screened and tested for HPV Or (2) Demand = number of people screened for cervical cancer * average number of HPV tests per screened person
Required data	Clinic data: number of people screened and tested for HPV or number of people screened for cervical cancer and percentage of people tested or the average number of HPV tests per screened person
Stratification as required	By test type and technology: molecular methods (POC, conventional methods) and non-molecular methods (such as Pap smear), as relevant for analysis By facility (or small geographical aggregate such as sub-district) By month (or at least by quarter)
Time period	At least the most recent year Ideally include 3-4 years of historic data for better interpretation
Advantages	This is a relatively easy method to determine demand for HPV testing. Compared to a method based on historical laboratory testing data, results from this method can be easier linked to programmatic activities
Disadvantages	If the national-level value for the percentage of women tested is used, it is a simplification but lacks geographic resolution. Routine health service data might be incomplete or have other quality problems that are sometimes difficult to identify.

As these methods are based on individual person data, they offer the advantage that later demand forecasts can be linked to programmatic and strategic activities, such as an expansion of a screening programme or an increase in percentage of people tested among those who attended screening. Both methods require that screening and testing activities are systematically recorded (ideally at facility-level) and are digitized, which might not necessarily be the case, depending on the country and context.

Calculations and modifications

The approach in formula (2) might be useful if the only available information are the number of people screened and the total number of tests, but the demand is required at facility-level for geospatial analysis. Here, the number of screened people essentially become a distribution key for the annual aggregate of historical testing demand but is not a demand calculation method per se. The resulting lack of geographic differences in screening and testing efforts

can be minimized if regional or district-level aggregates are available and used, which would allow to include at least some geographical differences.

4.3.3 HPV demand from historical patient data

Calculation	Demand = number of eligible patients * percentage screened and tested for HPV
Required data	Clinical data: number of eligible target patients in the relevant age group Performance data: percentage of patients screened and tested
Stratification as required	By test type and technology: molecular methods (POC, conventional methods) and non-molecular methods (such as Pap smear), as relevant for analysis By facility (or small geographical aggregate such as sub-district) By month (or at least by quarter)
Time period	At least the most recent year Ideally include 3–4 years of historic data for better interpretation
Advantages	This is a relatively easy method to determine demand for HPV testing. This method can be easily linked to programmatic activities as it is based on patients.
Disadvantages	Using the national performance aggregates for the percentage of patients screened and tested is a simplification, which lacks geographic resolution. Routine health service data might be incomplete or have other quality problems that are sometimes difficult to identify.

The calculations presented here follow the same concept as introduced in chapter 4.3.2 except that here they are based on specific patient groups (and data) that are already in care for other conditions. This may include for example the number of women in the eligible age groups who are on ART, or the number of women in the eligible age group attending OPD services.

While we list here ARV and OPD services as examples, the approach can be extended to any patient characteristics as required by the context. Other target group considerations could be for example women who are enrolled in the PMTCT programme or who attend ANC services, regardless of HIV status.

As this method is based on patient data, it has the advantage that later demand forecasts can be linked to programmatic and strategic activities, such as an expansion of the screening programme within existing health services, with or without a specific target group focus.

5. Demand calculation from population and disease burden data

5.1 TB demand from population and disease burden

Calculations	(1) Demand = number of people * estimated disease burden * percentage seeking care * percentage receiving care (including a diagnostic test)) Or (2) Demand = estimated (or targeted) number of people with TB * average number of tests per patient
Required data	Population data Estimated disease burden (see specifications below) Estimates for percentage of people seeking care and percentage receiving care (including a diagnostic test)
Stratification	By test technology mWRD (GeneXpert, Truenat, TB LAMP), SM, others, as relevant for analysis By geographic area Optional: by patient type or risk factors (for example PLHIV, children etc.)
Time period	Most recent year
Advantages	This method is relatively simple to conduct if publicly available data are used and high-level assumptions are made. In later demand forecasting, these methods have the advantage that they can be tailored to a set NSP target, which are often expressed as the percentage of estimated incidence cases that should be detected in the future, or the targeted number of patients.
Disadvantages	Assumptions and the use of national aggregates lead to a higher degree of uncertainty and lack geographic differences in testing efforts. Attributing population to single health facilities requires GIS software skills, hence this method is therefore better used with small area aggregates, such as districts or sub-districts. If at all, subnational estimates for incidence are typically not available, which consequently limits the consideration of a geographically different disease burden.

If good quality historical facility data are available and suitable for the analysis, these methods would not be the first choice for calculation baseline demand, although they are technically possible. However, these methods might be helpful to overcome challenges with incomplete or otherwise low quality routine data. Another use-case would be to forecast demand for a situation in which the coverage of current services are notably increased, and/or when demand forecasting should be linked to NSP targets that are based on a targeted number of patients to be diagnosed in the future.

Data

Population: Population data should include the number of people by geographic area, such as administrative level 1, 2 or 3 (for example region, district, sub-district), which are typically data that can be retrieved from the most recent census.

Of note, if demand is required for geospatial analysis and optimization, all data should ideally be facility-based or at the very least aggregated at small geographic area (such as ward, town or community); higher-level administrative-level aggregates might not suitable for geospatial analysis, especially if the areas are relatively large, which has to be carefully reviewed. Sometimes, health facilities do have data on population estimates in their catchment area, which could be used instead, if the forecasting method was suitable and data are of good quality.

Alternatively, population estimates by facility could be made using a geospatial analysis software, such as QGIS or ArcGIS. Publicly available raster population data [9] can be used to determine the population in defined catchment areas of health facilities. There are various options for determining facility catchment areas, e.g. creating a buffer zone or calculating road-network based distance/time areas using various geospatial algorithms. However, this method requires advanced GIS software skills. An additional limitation of using GIS approaches is that these often assume that people seek care at their closest facilities, which is not necessarily the case.

Disease burden: One measure for disease burden that is easily accessible is the estimates of TB incidence, which are routinely reported by WHO[10] and developed in consultation with countries, considering either recent TB prevalence surveys, notification data, inventory studies and/or expert opinions. WHO reports the respective method used for each country [11] and it is recommended to review the method used, its reliability and the low and high bound estimates [12].

These data and information sources used by WHO can also be used directly if desired, considering the methodology and consequently suitability. Very often, national TB programmes have also conducted their own epidemiological modelling of disease burden, which should be considered as the first choice. The optimal approach would be to directly use the targeted number of people with TB in the future (in formula 2), which NTPs have typically set and which are already informed by considering individual country disease burden and contextual factors.

Of note, if demand is required at subnational level (for example for geospatial network analysis and optimization), disease burden data would be ideally stratified by smaller geographic area to be able to account for geographic differences in demand. However, estimates of TB disease burden are typically only national-level aggregates, which can still be used if it is acceptable for the analysis that geospatial differences in disease burden are not accounted for anymore.

---

[9] For example, worldpop.org

[10] WHO TB Country Profiles. Available online https://worldhealthorg.shinyapps.io/tb_profiles

[11] WHO Global TB Report

[12] WHO Global TB Report Technical appendices https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022/technical-appendices

Depending on the choice of input data for disease burden, additional tests would have to be considered for example, for patients that are smear positive during treatment follow-up and require an mWRD test for the purpose of Rif-susceptibility testing.

Seeking and receiving TB care: All methods that determine demand from population data need to account for the proportion of people seeking care and receiving care to avoid overestimation. Generally, these figures are estimates and useful sources of information can be individual country household health access and/or utilization surveys, WHO Service Availability and Readiness Assessment (SARA) [13], Demographic Health Surveys (DHS) [14], TB Patient Pathway Analysis (PPA) [15] or a dedicated TB patient cascade of care analysis, which might be available from the TB programme and/or scientific literature. Alternatively, informed assumptions based on expert opinion can be made.

A frequently used alternative approach is shown in formula (2) where an estimated number of patients is multiplied by the average number of tests per patient or the number of patients with the number-needed-to-test to find one patient. We see that number-needed-to-test is often calculated as number-needed-to-test = 100 / (test positivity in percent). In essence, the number-needed-to-test also just determines the average number of tests to be conducted per patient, but here, based on assumptions about test positivity.

If mWRD testing demand is calculated, the mWRD test positivity is used to determine the number-needed-to-test. If the total testing demand should be calculated, the test positivity needs to include microscopy positivity (per patient and not per smear).

It should be noted though, that using the formula above, the number-needed-to-test calculates the number of tests required to detect one bacteriologically confirmed case, while realistically there will be patients who were and will be clinically diagnosed, either without a test, or based on clinical grounds with a negative test result. The number and proportion of patients who will fall into these categories has to be accounted for in the demand forecasting.

Number-needed-to-test (or percentage test positivity) values can be obtained from historical routine data as baseline. For future demand forecasting, it can be replaced by generic assumptions or by findings from previous internal research projects and/or published literature. Number-needed-to-test values from the latter two sources have the advantage that they are typically based on more extensive data analysis and modelling and might therefore provide more advanced values than simply using a generic assumption or test positivity percentage from historic laboratory statistics.

Of note: Private sector

These methods determine demand at population level and do not consider sector-specific demand. If a large number of patients attend the private sector but the demand calculations are only required for the public sector, this must be accounted for by including the percentage of patients seeking care in the public sector.

---

[13] WHO Service availability and readiness assessment (SARA). Available online [https://www.who.int/data/data-collection-tools/service-availability-and-readiness-assessment-(sara)]

[14] The DHS Program. Available online: [https://www.dhsprogram.com/]

[15] StopTB Partnership TB Patient Pathway Guide. Available online [http://www.stoptb.org/assets/documents/global/awards/tbreach/TB_Patient%20Pathways%20Guide.pdf]

5.2 TB demand using the GLI/WHO forecasting tool

This tool[16] calculates the quantities of diagnostic tests required, either based on past consumption or based on a country's historical epidemiological data, including but not limited to the number of people notified with TB previously. For the latter approach, the number of tests required are automatically calculated for the key patient groups, PLHIV, HIV-negative adults, children and people at risk for drug-resistant TB.

After entering people notified with TB previously, other epidemiological data as well as assumptions, the tool calculates essentially the baseline test requirements. Assumptions for the future could be made by increasing the percentage coverage of testing in respective patient groups.

Alternatively, instead of using the historic number of people notified with TB, the underlying calculation methods from this tool can also be applied to the targeted number of future people with TB to forecast future demand.

Screenshot of the GLI-WHO forecasting tool for Xpert MTB/RIF test calculation results:

Xpert MTB/RIF

Two methods are proposed for calculating the expected number of tests to be performed during the period.

Values in red should be entered or adjusted according to actual country data and practices, when possible.

No. of tests performed last year	120000

Method 1: Consumption-based forecasting
The expected number of tests is based on the average daily use per instrument module.

Number of instrument modules in the country	50
Number of working days per year	220
Average number of Xpert MTB/RIF tests performed per module per day	4
Number of Xpert MTB/RIF Ultra tests required for the period	**44,000**

Method 2: Morbidity-based forecasting
The estimated number of tests calculated is based on epidemiological data (notifications) and planned algorithms. The calculated sum is the target to be reached when algorithms are implemented according to the indicated coverage and assumptions. The default coverage below is for an Xpert-for-all algorithm (100% coverage for all patient groups).

Patient population	Coverage of testing	Number of Xpert MTB/RIF tests required for the period
People living with HIV (PLHIV) with signs and symptoms of TB	90%	**450,000**
Children with signs and symptoms of TB	100%	**46,800**
People at risk of drug-resistant TB	90%	**13,600**
Previously untreated HIV-negative adults with signs and symptoms of TB	90%	**298,100**
New smear-positive TB cases receiving Xpert for rifampicin testing purposes (excluding those receiving it as the initial diagnostic test)	100%	**46,600**
al number of Xpert MTB/RIF Ultra cartridges required for the per		**855,100**

5.3 HIV VL demand from population and disease burden

Calculation	(1) Demand = number of people * estimated HIV prevalence * percentage receiving ART * percentage of patients tested (refine formula considering the national testing schedules for different patient groups) or

[16] WHO/GLI. 2021. "Planning and budgeting tool for TB and drug resistant TB testing: calculation, version 2." Available online [https://www.who.int/publications/i/item/WHO-UCN-TB-2021.8] Accessed 23 Sept. 2022.

	(2) Demand = number of people * estimated HIV prevalence * percentage receiving ART * average number of VL tests per patient Whereby: The (number of people * estimated HIV prevalence) can be replaced with the estimated number of PLHIV in country
Required data	Population data ART coverage (percentage of PLHIV receiving ART among estimated eligible PLHIV) Estimated HIV prevalence Or the estimated number of PLHIV
Stratification	By technology: POC or conventional platforms By patient group as per national guidelines: priority group patient eligible for POC VL and non-priority group patients By age group: if VL schedule depends on the age of the patient Any other relevant patient criteria By facility (or small geographical aggregate such as sub-district)
Time period	Most recent year
Advantages	The approach in formula (1) can be almost entirely run on publicly available data and is very simple and easy to conduct. This method is very suitable for forecasting optimal demand, i.e. whereby all (or a certain percentage) of estimated PLHIV would seek and receive services.
Disadvantages	Overall, this method lacks accuracy and granularity as it is based on various high-level assumptions and, as all population-based methods, is best done at small-area aggregates or requires GIS software skills to attribute population to facilities.

The major advantage of this method is that it can be almost entirely calculated using standard HIV data, estimates or indicators routinely collected and reported from an MoH, WHO[17] or UNAIDS.[18] While technically possible, this method would not be the first choice for calculating baseline demand: given the generally high quality of recording and reporting from HIV programmes, historical facility data are likely much more suitable.

However, this method is better used if optimal demand is forecasted, and/or forecast for a situation in which the coverage of current HIV services are notably increased.

In contrast to population and disease burden methods for TB, estimated HIV prevalence data are often also available at subnational level, which allows a more refined calculation of demand considering some geographic differences, for example at district-level.

If demand needs to be attributed to individual health facilities, for example for a geospatial analysis and optimization, this could be achieved as well. This does however require GIS skills

[17] WHO Global HIV Program [https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv-programme]

[18] UNAIDS [https://www.unaids.org/]

and will require accepting the limitation that simpler GIS approaches typically assume that people seek care at their geographically closest facility (also refer to chapter 5.1 for more details). This is already not necessarily the case for general healthcare; if HIV is very stigmatized in a country, it is even less likely to be true, as PLHIV might prefer to seek HIV care in facilities far away from their own neighbourhood.

Calculations and modifications

The percentage of patients tested as well (formula 1) as the average number of tests (formula 2) could be retrieved as national-level performance data using routine data. Using values disaggregated by administrative area (regions, districts, sub-districts) would allow to account for geographical variations in testing efforts and performance, which will be lost to some extent if national-level values are used.

Instead of using routine data, the percentage of patients tested as well as the average number of tests could be also based on expert opinion.

For the first approach, it is recommended to refine the formula in line with the VL testing schedule, considering age groups, if the patient was newly initiated on ART, or if the patient is in a priority group with shorter times between routine VL tests. Additional considerations for repeat testing for patients with non-suppressed previous VL results should be integrated as well.

5.4 HPV demand from the population

Calculations	Demand = number of people * percentage eligible for cervical cancer screening * percentage screened each year * percentage tested for HPV
Required data	Population data Estimated percentage of eligible for screening, percentage screened and percentage tested each year
Stratification	By geographic area Target group definitions and data (age group, or HIV-positive, or others as per national guidelines) By test type and technology: molecular methods (POC, conventional methods) and non-molecular methods (such as Pap smear), as relevant for analysis
Time period	Most recent year
Advantages	This method is very simple and quick to conduct if publicly available data are used and high-level assumptions are made.
Disadvantages	Assumptions and the use of national-level data lead to a higher degree of uncertainty. Attributing population to single health facilities requires GIS software skills, hence this method is better used with small area aggregates, such as districts or sub-districts.

Population data should include the number of people by geographic area, such as administrative levels 1, 2 or 3 (for example region, district, sub-district), which are typically data that can be retrieved from the most recent census.

If eligibility criteria are demographic characteristics, it might be easiest to retrieve those directly: for example, if the national guidelines determined that all women in a certain age group are eligible for cervical cancer screening, these data should be used directly as age and gender stratification of population data are typically easily available.

Please refer to section 5.1 for additional considerations about geospatial population distribution and data analysis.

For the percentage of eligible persons who are screened and the percentage who are tested each year, the simplest approach would be to use national aggregate values. Alternatively, assumptions based on expert opinions can be made.

Of note, if national aggregates or assumptions are used, the result becomes less accurate and essentially averages out geospatial differences in screening and testing efforts. This limitation can be reduced if regional or district-level aggregates are available and used, which would allow to include at least some geospatial differences.

6. Methods for forecasting future demand

6.1. Calculation method-specific forecasting

Method	Modifying input factors for previously introduced demand calculation methods
Suitable for	All methods described in chapters 4 and 5.
Required data	As described in the respective methods in chapters 4 and 5
Advantages	These methods are easy to link with planned interventions, which are expected to have an impact on future demand.
Disadvantages	For facility-level demand, it might be necessary to group facilities into smaller area aggregates; otherwise, this method will be very tedious (refer to section 6.6 for details). Also, grouping leads to a loss of granularity and detail.

In chapters 4 and 5, we have introduced various methods to calculate demand using different base data, such as historical testing volumes, people notified with disease or patients in care, as well as general OPD attendance, population or estimated incidence data.

For each method, calculation formulas were provided, and the most sensible approach to future demand forecasting would be to simply modify the formula factors according to future expectations. This approach is illustrated in the following case studies.

Case study 6.1

In this country, the PMTCT coverage was assessed to be high on average; however, in 12 of the 36 districts in the country, HIV-positive pregnant and breastfeeding women (PBFW) received ART and were registered but only a minority of eligible infants were tested with EID.

Reports from routine supervision visits revealed that the main bottleneck was the knowledge of health facility staff, consequently, the MoH planned retraining in those 12 districts and wanted to forecast future EID testing demand for the period after the trainings. In a situation like this, using the number of HIV-exposed infants in the PMTCT programme in affected districts was expected to provide better evidence for future demand than historical testing volumes, as it could be linked to the programmatic interventions. Hence method 4.2.2(1) was chosen.

Annual demand forecasts were aligned with planned training schedules, i.e. the first four districts would be trained in year 1 and it was calculated that this would increase the demand by 14,600 EID tests in the first year, based on the number of HIV-exposed infants and accounting for those already tested in the previous reporting year. Subsequent years were calculated accordingly.

In this particular country example, it was assumed that the number of HIV-exposed infants would remain constant in the future and no other events or interventions would impact EID demand. Of course, these assumptions could be replaced by additionally accounting for an increase or decrease in total numbers of infants.

Fig. 6.1: Example of using PMTCT data for EID demand forecasting.

Case study 6.2

The country in this case study wanted to forecast demand for future TB testing on mWRD. In the most recent year, 180,540 mWRD tests were reported by the laboratory in the historical testing volume dataset. In the future, the country planned to substantially increase demand through strengthened TB symptom screening and increased testing of people with presumed TB with mWRD, through clinical training.

The country therefore chose to project future demand from OPD attendance data (method 4.1.4), as it would be very difficult to make accurate forecasts from historical laboratory testing volumes in a way that is linked to the expected substantial change in policy and clinical protocol application.

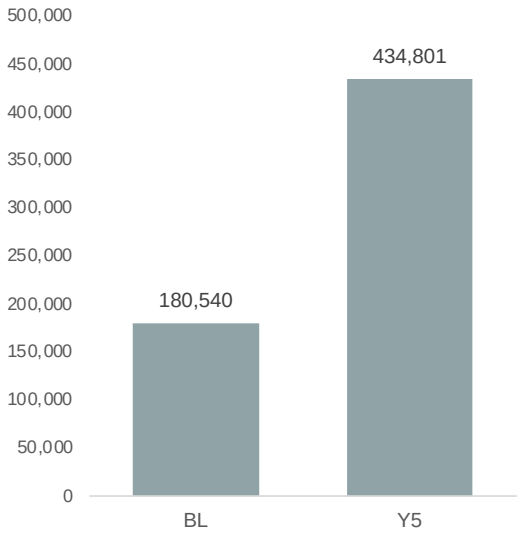


Fig. 6.2: Example of using OPD data for demand forecasting.

Utilizing morbidity data from the DHIS2 database, data from previous surveys and estimates from experts and considering actual mWRD testing volumes, the baseline performance values were estimated to be 34% of patients presenting with respiratory symptoms at an OPD. Of those, 50% were actually screened for TB; of those, 15% were found to be presumptive with TB, of whom 60% received an mWRD.

In the future, it was then expected that the screening efforts would increase from 50% to the target of 85%. The percentage of people with presumed TB tested with an mWRD would increase from 60% to 85%, which was the target. All other values were considered constant. It was therefore forecasted that the mWRD testing demand would increase from 180,540 to 434,801 tests in the final year, once all trainings were completed.

The approaches both case study countries took were relatively simple and allowed the considerations of future interventions. However, they certainly had limitations, as can be seen from the various high-level assumptions that were made in the forecasting approaches. These can have a significant impact on forecasted demand, and the stronger the underlying evidence for the assumptions, the better the results.

Information about effect or impact of future policies and plans could be, for example, evidence from early-implementing areas, pilot projects, published literature, supervision or assessment reports and, most often, expert opinion.

However, there is no one-size-fits-all approach and whether or not a chosen approach is valid has to be carefully reviewed considering available data and information.

6.2. Multi-year trend extrapolation

Method	Past demand trend is extrapolated into the future
Required data	Multiple years of historical demand
Suitable for	All methods in chapter 4
Advantages	This method is simple, quick and generic, meaning that it can be conducted without much specialized knowledge or modelling of future policies and their impact on demand. However, it is also possible to consider future policies to some extent, which can be used to fine-tune the method
Disadvantages	Not considering a policy analysis in the multi-year trend extrapolation can lead to an over-simplification and thus inaccuracies in forecasted demand. Before extrapolating a past trend into the future, it is essential to understand what influenced the demand in previous years to ensure it is valid to extrapolate the history into the future. This will require extra effort.

In this method, historical health system data are analysed for past trends, which are then projected into the future to determine future testing demand. Demand data should be disaggregated by month or at least by quarter; ideally, several years of historical data are plotted.

Excel is a good starter tool for trend extrapolation,[19] which can be done by fitting a trendline to data by choosing one of the most common trend types (e.g. linear, exponential), as illustrated in Figure 6.3. Demand for future years can be calculated through the trendline equation or by using simply the FORECAST functions in Excel.

Simply forecasting the past trend without consideration of future health policies and plans is a relatively generic approach and most suitable for a context where the demand did not show and is not expected to show any major fluctuations or short-term changes in the past or in the future. The assumption that the past trend continues in the future has to be somewhat applicable, which might not necessarily be the case if larger interventions are planned, as the case studies below illustrate.

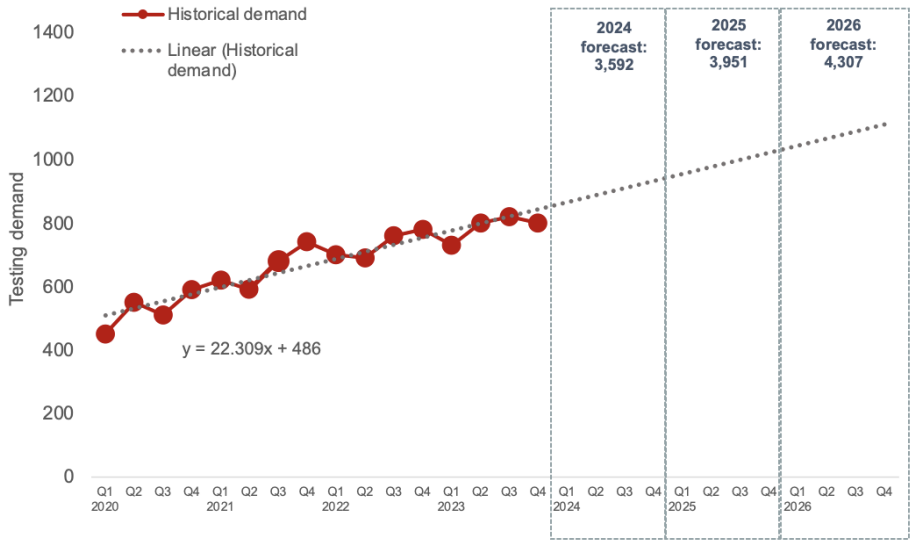


Fig. 6.3 Example graph of multi-year trend extrapolation of historical data conducted in Excel, applying a linear trend.

Case study 6.3

This case study shows mWRD TB testing demand between 2019 and 2023 (Fig. 6.4). The past trend analysis revealed that in Q2-2020, demand decreased significantly, which was caused by the COVID-19 pandemic. The demand remained low during the subsequent quarters but reached pre-COVID-19 levels of demand again from Q2-2021 onwards. However, the data also showed a large increase in testing demand from Q3-2022 onwards, which coincided with the beginning of laboratory network expansion activities in the country, when the NTP simultaneously equipped new facilities with mWRD instruments and expanded the specimen referral system.

In this situation, trend extrapolation is conditional to the expectations for the future, and requires expert opinion: if, for example, the expansion of the laboratory network was completed at the end of 2023, extrapolation of the previous trend should not be used as the

[19] Microsoft “Create a forecast in Excel for Windows” available online https://support.microsoft.com/en-au/office/create-a-forecast-in-excel-for-windows-22c500da-6da7-45e5-bfdc-60a7062329fd#bkmk_calculation

observed trend is unlikely to continue (at least not at this pace) in the future. Here, it would be better to use expert opinion and assume the demand would stabilize somewhere near end-2023 observations or simply choose another method.

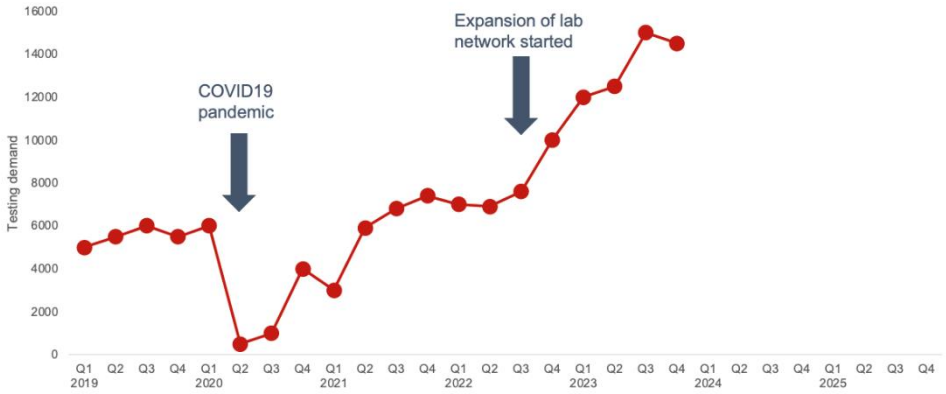


Fig. 6.4: Analysis of the past trend of testing demand in a country.

If in contrast the laboratory network expansion would continue at a comparable scale in subsequent quarters, it could be feasible to assume a continuation of increasing demand for the coming quarters. The demand might reach a plateau once this process is completed. Also in this this case, expert option for future developments should be considered, rather than generically fitting a trendline.

This case study illustrates the importance of understanding past events that have influenced demand and consider future planned policies to ensure the method is chosen and used correctly. It also shows that the considerations of those aspects into trend forecasting can be relatively simple without much requirement of very detailed calculations.

6.3 Annual growth/decline rates

Method	Future demand is forecasted by applying an annual growth or decline rate, starting from baseline demand
Required data	Baseline demand (most recent year)
Suitable for	Methods in chapters 4 and 5
Advantages	This method is simple and quick; it can be executed generically without in-depth knowledge about past or future events. It only requires one year of historical data (the most recent year)
Disadvantages	Not considering future events and their impact on demand can lead to an over-simplification and thus inaccuracies in forecasted demand.

This method is useful if multi-year, historical data are not available or cannot be used, for example, in cases where major policy changes in the past occurred (or are expected in the future) and the assumption that the past trend will continue in the future is not valid. This method is also very useful if specific targets are set to be reached in the future and demand for interim years is estimated as proportional increase until the target is reached.

In the simplest version, a generic assumption about annual growth or decline is made and applied to baseline demand. The assumed proportional change for the future years are typically generic, qualitative assumptions. In the example illustrated in Fig.6.5, analysts simply assumed that the demand will grow by 5% each year.

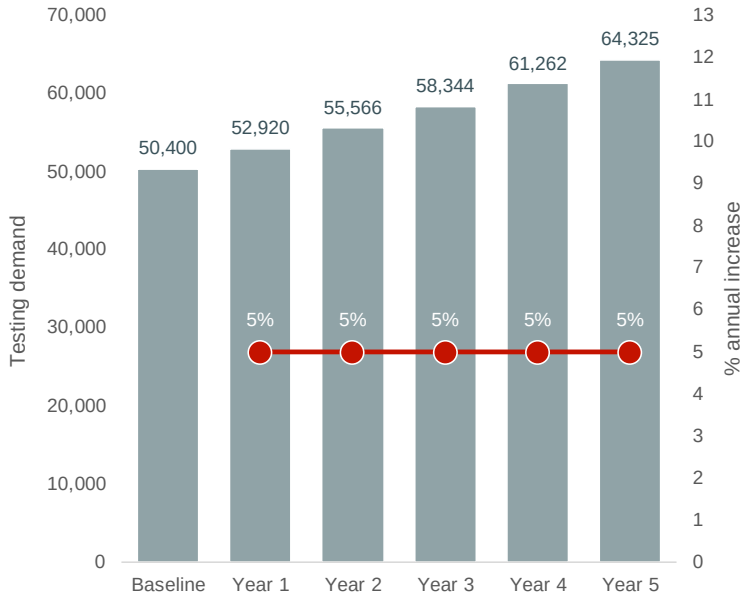


Fig. 6.5: Example of using a constant annual growth rate of 5% starting from the baseline demand.

Certainly, the applied annual growth or decline rates do not have to be constant but vary each year, depending on contextual assumptions.

The advantage of this method is the straightforward approach, which is quick to conduct; however, it is not very accurate, unless additional information about future policies and plans is considered to refine growth/decline rate values and/or determine the future target.

To improve accuracy, assumptions about growth/decline rates or future targets could be linked to future policies and plans. Information about the effect or impact of future policies and plans could be, for example, evidence from early-implementing areas, pilot projects, published literature and, most often, expert opinion. The following two case studies illustrate practical examples of such a tailored annual growth or decline rate forecasting.

Case study 6.4

This country aimed to transfer the majority of TB testing from smear microscopy to mWRD testing. At baseline, 60% of the total testing volume was conducted with mWRD, 40% were tested with SM (Fig.6.6). Within the next four years, the mWRD testing target was set to 80% of the total testing volume.

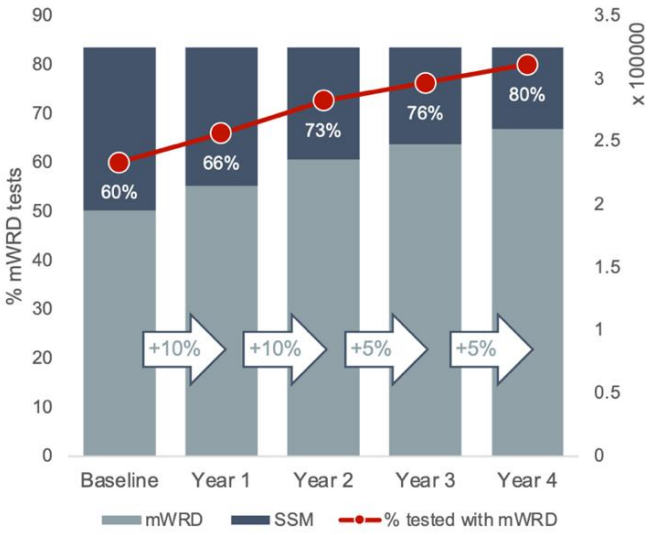


Fig. 6.6: Demand forecasting using a policy-informed annual growth rate with a set target.

To reach this target, the appropriate interventions were planned (here for example clinical trainings) and in line with those, it was simply assumed that the demand would increase by 10% in the first two years and by 5% in years 3 and 4, eventually reaching the set target.

This example illustrates the simplicity of the method, as only high-level assumptions are made, based on reaching set targets. Expert opinion is however crucial to ensure that adequate interventions are designed for the country to actually reach the set targets.

Case study 6.5

This country conducted an assessment of routine VL testing services at randomly selected facilities, which showed that there was a notable amount of unnecessary repeat testing of patients on ART due to incorrect application of the testing algorithm. The report estimated that about 8% of the most recent year's testing volumes was due to unnecessary testing.

Therefore, the HIV programme planned a large-scale retraining intervention to reduce to unnecessary repeat testing and demand was forecasted for the implementation timeframe. Future targets were calculated based on the assessment report findings of 8% unnecessary testing at baseline, i.e. the baseline demand would be reduced from 800,000 to 736,000 in year 5. The country then planned most trainings in year 1 and 2, and simply assumed a decrease of 3% for the first two years, followed by another two years of additional trainings, assuming a 1% annual decrease (Fig.6.7).

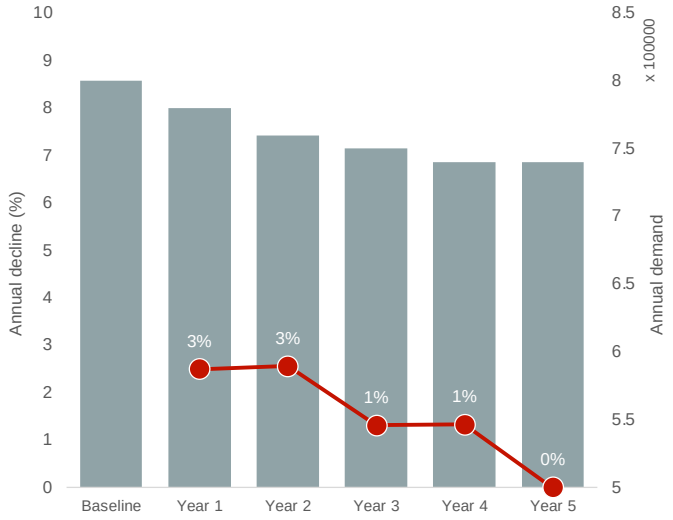


Fig. 6.7: Demand forecasting using a policy-informed annual decline rate with a target.

6.4 Including future policies and plans into forecasting

In previous sections, we introduced methods and approaches for future demand forecasting, most of which can be generically calculated. However, all of them would benefit substantially in terms of accuracy if future planned strategies, policies and interventions were considered, as illustrated in the case studies.

For the purpose of understandability, previous case studies always assumed that just one intervention would take place in the future and no other changes would occur that could simultaneously affect demand. This is however not necessarily a realistic situation as MoHs typically plan several interventions over the periods of their strategic plans, many of which are expected to differentially impact future demand (refer to chapter 1.2).

Considering all future policies and plans could however become a very time- and effort-intensive approach. In addition, it would be wrong to assume that the level of accuracy would continuously increase the more detailed policies and interventions are considered. On the contrary, each policy and intervention consideration requires additional assumptions, and the more assumptions that have to be made, the less precise the overall result might be. It is therefore important to find and apply a very balanced approach.

A good place to start with the selection of relevant future policies and interventions is with the respective disease-specific NSPs (refer to chapter 1.2). NSPs outline the essential future interventions at higher levels which is typically a sufficient level of detail for demand forecasting.

Examples of future NSP policies and interventions include but are not limited to:

- Increasing the percentage of TB diagnostic tests conducted with mWRD to defined targets (or alternatively, increase the percentage of people notified with TB tested with mWRD)
- Strengthening the application of the national testing algorithm and/or clinical procedures
- Changing the diagnostic algorithm or eligibility criteria, or the introduction of chest X-rays and/or population-level screening activities for TB
- Expansion of the specimen referral system for defined areas or facilities
- Introducing a new testing technology
- Strengthening health system performances, such as addressing a reagent shortage or ensuing that all mWRD instruments are repaired (both of which might have led to reduced demand)

6.5 Mixed methods

We have previously only presented examples of single methods used for demand forecasting, while in real life, the various methods are often combined, either for the purpose of simplicity and/or increasing accuracy.

Below are typical example situations in which it might be adequate or necessary to combine different forecasting methods.

- **Forecast demand for geographical areas with a planned intervention and without the intervention separately by applying different methods.** If an intervention is planned in selected districts only, the demand linked to the intervention can be calculated specifically for those districts, while for other districts not affected by the intervention a generic method could be used (for example a generic growth rate or trend extrapolation).

- **Forecast demand for selected facilities with a planned intervention and without the intervention separately by applying different methods.** An intervention could target for example only hospitals or only low-performing facilities (or any other type of characteristics), in which case demand could be forecasted for those specific facilities, while other, generic methods could be applied to all other facilities that are not affected by the intervention.

- **In cases where there is high uncertainty in historical data for selected geographical areas or facilities, different methods may be used.** In some cases, the historical data on testing volumes for a country are of good quality and allow for multi-year trend forecasting, but there are some areas of the country where the data are of lower quality (or are not suitable for trend extrapolation for other reasons). In this case, different methods might be considered for those areas, including assumptions fully based on expert opinion, generic growth/decline rates, or even calculating demand from a different data type, for example using patient data instead of historical testing volumes.

- **Combining methods using historical facility data with methods using population and disease burden data.** Demand can be forecasted from population/disease-burden data or historical facility data, as described in chapters 4 and 5. These methods can be also combined and used for demand forecasting for different years in the future. For example, a country might plan only small interventions in the first 2 years of an NSP, such as linking 10% of new health facilities with the mWRD specimen referral system. For the new sites, demand can be forecasted using historical testing volumes from smear microscopy data. By year 5 however, the country might plan a very large network expansion, clinical training etc. In this case, it might be more suitable to use a method based on population and estimated disease burden data to project this long-term demand, as historical facility data are not representative anymore for this future expansion.

These are just few examples for situations in which it might be sensible to combine methods; however, the best approach depends on the specific context. The combination of methods might introduce a bias, which should be reviewed beforehand. The decision whether to combine methods is a complex one that should be made on a case-by-case basis as there is no one-size-fits-all approach.

6.6. Grouping for forecasting

If demand is forecasted for geospatial analysis and optimization, it would be ideal if demand is attributed to individual facilities to deliver precise results. However, some of the previously introduced methods would become quite time-intensive if applied to all individual facilities. For example, conducting a trend analysis and trend-based demand forecasting for all health facilities individually is typically not feasible. One approach could be to apply the same mathematical operations to all facilities at once; however, this would lose some level of detail in the results.

Alternatively, facilities can be combined in several groups, where the same mathematical operations and/or the same assumptions can be applied to one group of facilities, but can differ for other groups (NB. this should be done without actually merging the individual data as the facility attribution of demand is still required for geospatial analysis). Key requirements are that facilities in a group are comparable and that applying the same operations to the group is valid.

Very often, forecasting by facility group is implied already by the analysis design. An example of this approach was already shown in case study 6.1, in which all facilities in selected districts were planned to be retrained and were therefore grouped, while different forecasting was applied for all other facilities.

Grouping by administrative level and/or by intervention are just two examples, other options include but are not limited to:

- By administrative level (region, districts, sub-districts)
- By future intervention: affected or not affected by the specific intervention(s)
- By area characteristics: urban–rural, disease burden, health service coverage, easy-to-reach or hard-to-reach

- By type of facility, such as primary, secondary or tertiary care, or private versus public, or similar
- By role in the system: diagnostic facility, treatment facility, or similar

Before conducting a grouped analysis, it is recommended to check the validity of this approach, at least qualitatively at a high level. Especially when working with past trends, it is recommended to plot the historical demand data for individual facilities in respective groups to review the validity of this approach. This will reveal if there are outliers to the group trend, or if the observed trend is in fact only reflecting the trend of the busiest facility in the group. The analyst has to decide whether it would be valid to assume the overall group trend will apply to all facilities in the future, or whether it might be more appropriate to further split facilities by other characteristics.

7. Deciding on the appropriate method

We have introduced numerous different methods for demand forecasting. But how to decide which method is the best for a specific context and analysis? While there is no one-size-fits-all approach, the choice of the appropriate method can be guided by three main questions:

What level of detail is needed?

Previously introduced methods include simple approaches that use publicly available, national-level data as well as high-level assumptions. But we have also introduced methods that are very complex, are based on historical facility data and include a notable amount of future policy consideration. If these types of granular results are not needed, or if there is reason to believe that the required detailed assumptions might introduce an unknown uncertainty, it is often better to use simpler methods that require fewer assumptions.

Which high-quality data are available?

In real life, the availability of high-quality data is often the main decision-making criteria. Typically, we use historical testing volume data as the first choice to forecast future demand. Sometimes, however, these data might be not available at sufficient quality, or may not suitable for the analysis objective, as we have illustrated in several examples. In this case, there might be no other choice than to use a different base dataset and/or select a different method to forecast demand.

How representative are the base data and the method for the future situation?

If the objective is to forecast demand for a greatly expanded network, for example a network in which many new facilities are connected to the specimen referral system, the use of historical data from only previously connected facilities might be of limited use as demand would have to be forecasted for facilities which did not have any previous demand recorded. In this situation, historical data are not representative for the future and instead it might be more appropriate to choose a method that calculates demand from population and disease burden data or general OPD attendance data.

Which future interventions should be considered?

Sometimes, the future planned interventions already suggest an appropriate method by design. For example, if an intervention's target strengthened clinical procedures or increased testing coverage (such as in case studies 6.1 or 6.2), using historical patient data and respective calculation methods would be more suitable than using historical laboratory testing volumes.

8. Appendix

8.1 Guiding questions for a policy and guideline review

This section lists selected guiding questions for a policy and future strategy review to inform demand forecasting. The list of questions is not exhaustive, but focusses on frequently relevant topics; it should be adjusted depending on the local context and objectives of the planned analysis.

Targeted patient groups

- Eligibility criteria: which patients are currently tested with the diagnostic test of interest?
    - For example: is testing used for all people with presumed TB as an initial diagnostic test or only for certain risk groups?
    - If only risk groups are tested, what are the clinical/demographic criteria?
- Location: are the eligibility criteria practically implemented in all facilities and in all geographic regions or do they differ?
    - Example: some countries allow the use of mWRD TB testing as initial diagnostic test only for walk-in patients at facilities with an instrument on-site, while non-mWRD facilities can refer specimens only from selected risk groups, such as people at risk for drug-resistant TB.
- Did the eligibility criteria or geography change during the period of observation or are they planned to be changed in the future?
        - For example, is a country planning to allow all facilities to use mWRD as an initial diagnostic test for all people with presumed TB (rather than for risk groups only)?
    - If yes, what was/will be changed? If yes, how was it/will it be implemented, i.e. all facilities switch at once or gradual change, for example by district?
- Have there been expanded or reduced screening activities, for example intensified or active case finding, contact tracing activities or policies?

Availability

- Have any changes occurred with regards to availability of diagnostic services during the period of observation? For example:
    - Has the diagnostic network been expanded or reduced, for example in numbers of diagnostic facilities and/or diagnostic instruments in the network? Specify location and installation/removal date.
    - Have any instruments been relocated and/or repurposed for, including but not limited to, COVID-19 testing?
    - Have there been notable service interruptions for diagnostic testing, including but not limited to reagent stock-outs, broken instruments or malfunctional IT, COVID-19 lockdowns, strikes, natural disasters, security issues? If yes, specify location, time and duration. Is it expected that these interruptions will continue in the future or are interventions planned to reduce them?
    - If these events occurred, what were the mitigation strategies, including but not limited to: rerouting specimens for diagnostic testing to other facilities, use of alternative diagnostic tests instead, or the use of clinical diagnosis only?

- Have there been or will there be any major changes in the workforce, that directly or indirectly affect services, such as available workforce of laboratory staff, nurses or doctors or community healthcare workers? When and where did this occur in the past, or when is it expected in the future?

Accessibility

- Is there a dedicated specimen referral system (SRS)?
    - What is the geographic coverage of the SRS? Are all regions/districts/facilities covered or only a selection? If only a selection is covered, specify.
    - Does the SRS typically function well or have there been service interruption times/locations of the SRS? If there were service interruptions of the SRS, when and where did they occur and for how long? What happened (or typically happens) if the SRS is not functioning, for example, are specimens referred to a neighbouring facility or not tested at all anymore?
    - Was the SRS implemented during the period of observation? If yes, describe the place and time.
- Are there plans to expand or reduce the SRS in the future? If yes, where and how?
- Instead of specimen referral, are patients often or sometimes referred for diagnostic testing as well?
    - How often does this occur, where and what is expected to happen in the future?

Knowledge

- Have there been or will there be any large-scale workforce trainings or retrainings that impact demand, for example retraining on signs and symptoms of a disease, algorithms, testing procedures?
- Have there been or will there be any public awareness-raising activities, e.g. on social media, local events, targeted facility campaigns?

Disruptive events

- What was the impact of the COVID-19 pandemic on health services, including diagnostic testing demand for non-COVID-19 diseases in country?
    - Were health services, including general and/or disease-specific health services reduced or completely interrupted? If yes, what was the timeframe and were there geographic differences, such as a lockdown in some regions? Are services now restored to pre-COVID-19 levels?
    - Was reduced health-seeking behaviour observed; if yes, has health-seeking behaviour returned to the pre-COVID-19 situation?
- Have there been any other shocks to the healthcare system, for example, natural disasters, strikes or conflicts?

8.2 Supplementary data for demand forecasting

Specimen referral linkages

- A list of facilities indicating current specimen referral linkages for the diagnostic test of interest; ideally one row = one origin-destination link. There might have been changes to the origin-destination links over multiple years that were covered in the retrospective analysis, and it is important to time-match these two datasets. Ideally, the number of specimens referred are included. This dataset can guide the re-allocation of centrally recorded demand or forecasted demand

Diagnostic service interruption

- A list of health service reduction/interruption events, including diagnostic services but also all health services, for example in case of natural disasters, strikes or COVID-19 lockdowns. These data describe the availability of diagnostic services, which impacts the testing volumes registered at facilities and are therefore important if an in-depth demand forecast is required.