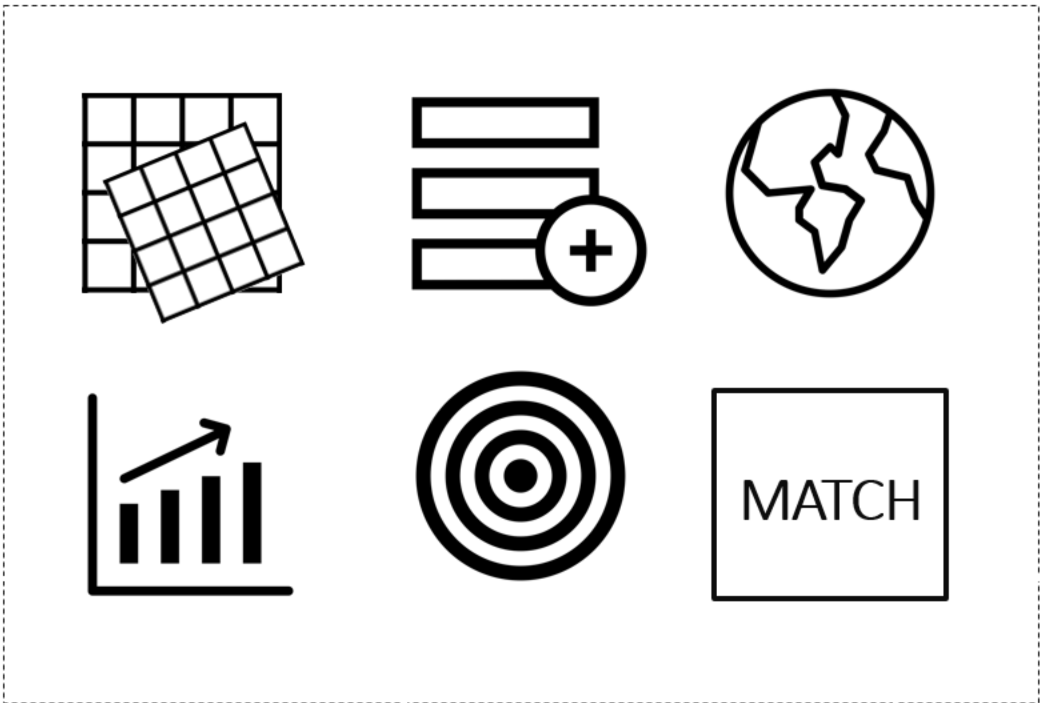
CASE
CENTRE FOR APPLIED SPATIAL EPIDEMIOLOGY



The MATCH

Manual

Amsterdam, October 2018

KIT Royal Tropical Institute

This manual is an initiative of the Stop TB partnership, funded by the Global Fund to Fight AIDS, Tuberculosis, and Malaria, and written by the Centre for Applied Spatial Epidemiology (CASE) of KIT Royal Tropical Institute.

This manual was been prepared by:
Mirjam Bakker
Ente Rood
Christina Mergenthaler
Lucie Blok
Margo van Gurp
Masja Straetemans
Sandra Alba

All documents and annexes referred to in this manual can be found at:

https://www.kit.nl/case

With the support of:

MATCH ///

Mapping and

Analysis for

Tailored disease

Control &

Health system strengthening)

Contents

1 Introduction to this manual

Globally, disease programs are increasingly recognising the value of sub-nationally disaggregated data to plan, monitor and evaluate locally-tailored disease interventions. To achieve the ambitious United Nations' SDG targets to end the epidemics of HIV, Tuberculosis (TB) and Malaria by 2030, a paradigm shift is required, one which moves from national strategies to locally-tailored interventions. To enable this change, the availability, quality and use of health data needs to be strengthened and improved at both the national and subnational levels. It is important to build, together with partners, long-term capacity to utilize data to plan and monitor interventions embedding the use and mapping of subnational data into routine practice, increasing ownership and accountability of policy makers. By linking data to policy objectives, combining and simplifying complex data to enable access and stimulate use and building capacity for data management and analytical techniques, we aim to strengthen capacities to reduce the burden of disease across the globe.

The increased availability of tools and software applications to strengthen health information systems through digital recording and reporting of routinely collected data have resulted in a rapid expansion in the availability and access of health program data. This trend requires the development of new methods and tools to optimally utilize these data to inform decision making and program planning. To make better use of routinely collected health program data, the Centre for Applied Spatial Epidemiology of KIT Royal Tropical Institute developed an innovative spatial data processing and analysis framework using Geographic Information Systems (GIS) to integrate routine health surveillance data with other data sources to inform health policy and planning. The approach, known as Mapping and Analysis for Tailored disease Control and Health system strengthening (MATCH), aims to support disease programs in using subnational data for more differentiated programme planning.

The MATCH Approach employs multiple sources of geographically, temporally and demographically disaggregated data with the aim of identifying people with health ailments who are missed by the health system throughout the pathway of care at subnational level. The approach utilizes multiple sources of geographic, temporal and demographic disaggregate data with the aim to identify subnational variations and gaps in the delivery and access to healthcare. The MATCH Approach takes advantage of the considerable amount of existing, yet uncollated subnational data which potentially describe the people who do not have access to health care, including who they are, where they are located, explanations for why they are missed, and indications of how to find and treat them. The cornerstone of this approach, as outlined in this manual, is to move away from one-size-fits-all health strategies towards locally-tailored interventions which recognise and address the local context of disease. MATCH maps and spatially analyses routinely collected health data and triangulates these with other relevant publicly available data to create a better understanding of health disparities at the subnational level. The results of these analyses provide insight into subnational variations and gaps in health care delivery and demand.

The MATCH Approach is done through:

1. Strengthening capacities to use publically available and user friendly tools to effectively analyse subnational data of multiple sources to understand and interpret the main programmatic gaps and challenges.
2. Increasing critical thinking and appraisal of the available data and analytical results to guide decision making on planning, monitoring and evaluation of health (data -> information -> knowledge -> policy).
3. Enabling routine utilisation of data and to determine which steps need to be taken to collect more useful subnational data when it is found to be unavailable.

Target audience

This manual is written for health program managers, monitoring and evaluation (M&E) specialists and data managers who are interested in mapping and analysing their subnational program data. The first two chapters describe the general approach and the analytical framework used by MATCH. Next, concise descriptions and examples of the data management tasks required for the analysis and stepwise instructions on how to conduct the spatial analysis are provided. The final chapter provides a few general principles and suggestions for sustainable implementation of this approach.

2 The MATCH Approach

The MATCH Approach employs multiple sources of geographically, temporally and demographically disaggregated data and builds on the idea that:

- making use of available information will lead to a better understanding of local gaps in order to design differentiated interventions.
- differentiated, subnational interventions will lead to a more efficient use of resources, to more effective targeted interventions and ultimately to improved case detection and treatment.

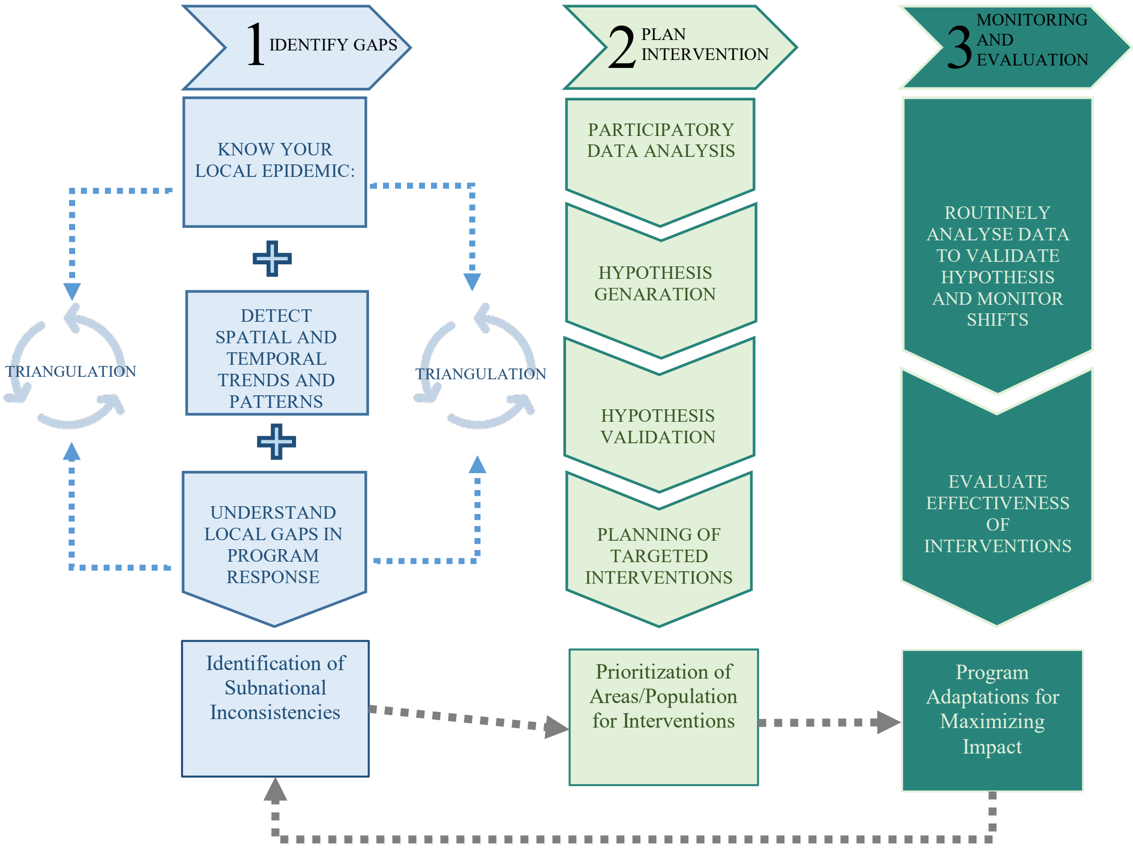
MATCH employs existing data, makes it available and accessible for usage, builds capacity to analyse these data and supports the development of subnational differentiated responses using the following logic: (see also figure 1 below):

(1) Identifying gaps (figure 1, step 1)**:** Multiple sources of social, demographic, program and health data are linked based on their shared geographic coverage and stored in a single standardized spatial database. The MATCH analytical framework (section 2.1) is then used to identify heterogeneities in spatial and temporal epidemic trends, accounting for local variations and gaps in program performance and service delivery.

(2) **Planning interventions** (figure 1, step 2)**:** The outputs of the MATCH analysis are then validated by experts knowledgeable about local health programs and contexts and translated into context specific hypotheses to where and why people might not have access to the care they need. Based on the results of the analysis and validation, locally tailored interventions and corresponding locally differentiated responses to mitigate these are formulated.

(3) **Monitoring and evaluation** (figure 1, step 3)**:** Following reprogramming and the strategy implementation, the MATCH analysis is routinely conducted to evaluate the effect of interventions and to verify and adjust interventions.

Figure 1. **Schematic overview of the MATCH Approach, which illustrates the components of each step.**



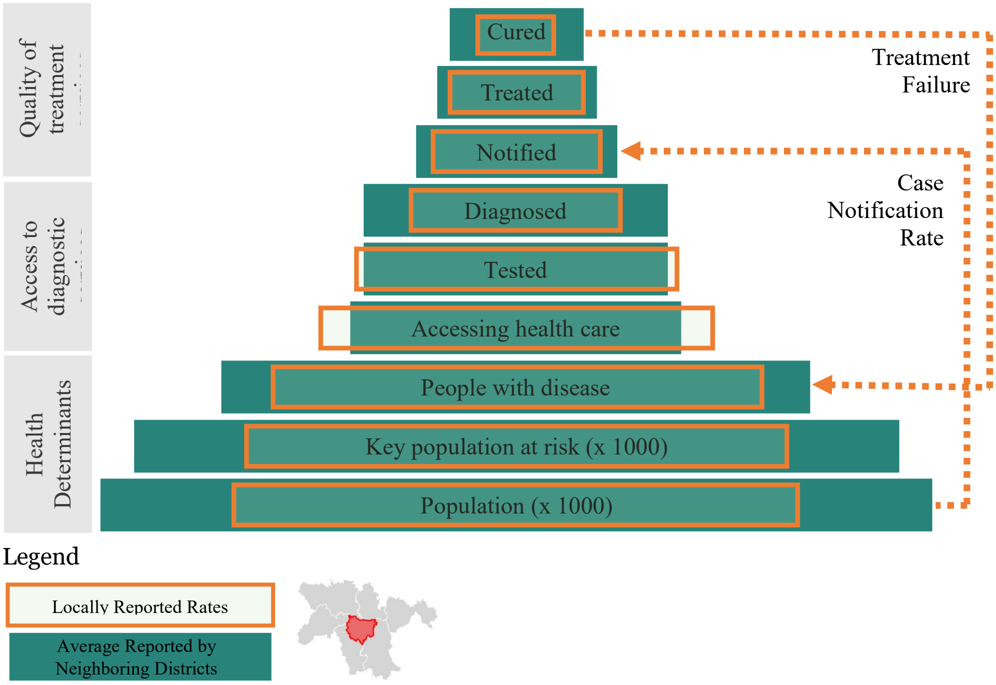
2.1 MATCH analytical framework

The MATCH approach combines the conceptual framework known as the patient pathway (pyramid structure, figure 2) with quantitative spatial statistical methods. This analytical framework is used to identify gaps in patient flow through the pathway of care. A schematic model of the patient pathway along this cascade is shown in figure 2. The analytical framework differs from other monitoring and evaluation tools in three ways. First, it performs data analysis using subnational data aggregated at the smallest available geographic unit, depending on the resolution of the available routinely collected data. Using subnational data allows one to visualise and quantify variations in health disparities notified by health programs. These variations can be used to investigate local anomalies in which areas deviate from their surrounding areas – anomalies which then need to be investigated.

Second, data which are commonly unavailable or not used by health programs – such as estimates of the subnational population at risk, laboratory data, or environmental and ecological land use factors – can be included in the analysis by harmonizing health program data with external data sources. The use of GIS and spatial data (e.g. districts/counties/health facility catchment populations) allows the integration of data collected across varying levels of aggregation – based on their shared geographies – into a single, unified spatial database.

Third, subnational variations are explicitly quantified and accounted for by using spatial analysis and spatial statistics to detect spatial patterns in the cascade. It therefore allows one to assess the geographic scale at which socio-economic, epidemiological and health system processes operate and are linked. For example, small-scale clusters of TB transmission might correspond with local clusters of poverty which in turn coincide with larger geographic clusters of low health system coverage. In the same way, an area with low notifications coinciding with areas reporting low screening rates or low volumes of diagnostic testing would raise suspicion regarding the low number of patients being notified. Such findings provide insight into the effectiveness of health systems to detect and diagnose disease and provide important information for future planning. These unique aspects of the MATCH framework are sequentially addressed in the data analysis.

Figure 2. Schematic model showing the pathway of a patient along the cascade of care. Along each step of the cascade a certain proportion of patients is expected not to proceed to the next step due to barriers in access and program responses. In the MATCH Approach, locally observed rates (orange) are compared to regional rates (green) to account for geographic associations between nearby areas and identify spatial anomalies; for example, due to patient movements or high transmission between areas. Drop-out rates can be calculated between various levels of the pyramid to inform where bottlenecks occur along the patient pathway.



3 MATCH data management

3.1 Data Requirements

To use existing data to identify geographic and temporal trends, it is essential to simplify, harmonize and integrate the data obtained from multiple sources into one coherent and unified source. Data derived from various sources should be validated using fit-for-purpose criteria (i.e. derived data is assured of its suitability for its end-use context) and should be geographically and temporally consistent to allow integration and analysis. Standardized data formats, processing and management procedures will need to be developed.

To improve the quality, access and use of data, it is preferable to upload available data into cloud based SQL data bases (box 1).

> **BOX 1. CLOUD BASED DATA STORAGE**
>
> There are numerous advantages of using a cloud based storage over local archiving systems:
>
> - All data are available from a single location to different users with different access rights.
> - There are no issues with versioning of data tables generated by various users as these are centrally managed and archived.
> - Consistent formatting of data according to predefined standards
> - The use of a relational database, where different data tables are linked using unique identifiers for each row in a table, reduces (removes) redundancy of data elements which are used in different data-tables (e.g. single data tables for each data type)
> - The use of schemas (standardized data containers analogous to a folder) within the database allows to manage access rights to different users.

3.2 Data types and sources

The data collected and collated for the subnational mapping and analysis of epidemiological and program responses include five broad categories:

SURVEILLANCE DATA:	Data regarding disease notification rates, incidence or prevalence estimates, laboratory testing and treatment outcomes
POPULATION DATA:	Estimates of age stratified population Gridded population data of the world by WorldPop[1]
RISK FACTORS:	Socioeconomic, environmental or ecological risk factors.
HEALTH SYSTEMS DATA:	Access to care, health and diagnostic facilities, health care utilization, insurance data, Survey Provision Assessment (SPA) and Service Availability and Readiness Assessment (SARA).
SPATIAL DATA:	Subnational administrative boundaries (polygon) Health facilities (point) Road network data (line)

3.3 Data preparation and management

Data should be processed and stored in different thematic tables according to the thematic data categories. Data tables are formatted as long tables in which single records (e.g. rows/tuples) represent a single geographic reporting unit at one time interval (Quarter, Year). Therefore, each record can be identified by a combination of two or three variables depending on the frequency of reporting (quarterly vs annually): 1) the unique geographic unit of reporting (BMU, for example), 2) the year of reporting,

[1] www.worldpop.org

and 3) the quarter of the report. In each thematic table, different data elements (indicators, reported values) are represented as single variables. Using data formatted into a long table has the advantage that it can be easily queried (e.g. selections and aggregations) by geographic area and time period.

3.4 Record Linkage

To map the properties (attributes) of different spatial features, tabular data collected from different sources needs to be linked to their respective geographic (spatial) units in their respective rows. As the reporting units encountered in different M&E frameworks, surveys and surveillance networks commonly do not correspond to administrative geographical units in a country, the linking of the two can be subjected to the following issues:

A. Health reporting units do not match administrative units contained in spatial data files

B. Linking based on a common identifier is not possible

Linkage problems due to mismatching identifiers (e.g. reporting unit names vs geographical unit names) can be solved by manual assessment and linkage. This procedure, however, is an enormous undertaking with datasets containing more than 100 records. Deterministic linkage using a matching algorithm or probabilistic linkage should then be considered.

3.5 Data Quality

Data quality assessments should be performed to ascertain that the data is fit-for-purpose. This means that the indicators used to assess geographic variation in the epidemiology, program response, and the geographic scale and resolution used to make these assessments are appropriate to reflect real world processes and therefore can be used to inform actions. As any data used in this process is a result of data collection and processing activities which inherently contain a certain degree of error, it is important to assess the completeness and internal (single source/variable) and external (across sources / variables) consistencies of various data elements used in the analysis. The following data quality elements will be assessed against the quality standards determined by the purpose of use:

1. ACCURACY: The data are internally consistent. This means that data values fall within the range of expectation and that no apparent outliers are present in the data. Spatial accuracy is assessed by comparing the geographic units in a map to the area covered by a reporting unit. For example, all reporting health facilities should fall within the districts where they are reported.
2. TIMELINESS: The data should be recorded and reported in a timely fashion. For example, predictions for 2017 can only be made when data for the preceding period, including 2016, are available.
3. COMPLETENESS: The data used to calculate indicators and map outcomes for different geographic areas should be based on a complete set of reports including all reporting units contained by that area. Population based notification rates should include the reports of cases across all ages.
4. COHERENCE: The data (spatial and non-spatial) should be coherent between sources. For example the number of individuals tested positive by a laboratory cannot exceed the number notified cases within an area.
5. PRECISION: The number of cases notified should reflect the true number of patients accessing healthcare. External quality assurance of diagnostic services can be used to assess the expected numbers of false diagnosis.
6. CURRENCY: The data should reflect the current situation. For example, when NTP data are mapped per district in a certain year, the district spatial data should reflect the actual administrative divisions for that year (e.g. do not use a district boundaries file for 2000 to map district case notifications in 2017).

7. SCALE: The scale (or resolution) of your outputs should be appropriate to answer operational questions. For example, to show subnational variations of TB case notification rates you would like to use a map of administrative areas which allow you to assess spatial patterns (rule of thumb > 30). NOTE: Using a too high resolution can result in the use of areas with very small populations and unstable rate estimations as a result.

A full description of the data management steps and standards are provided in the MATCH data management plan.

4 MATCH analysis – TB example

4.1 Know your local TB epidemic

Traditionally, national TB programs have collected and reported case notification rates nationally to monitor changes in program coverage and disease burden over time. Variations in TB case notification rates reflect variations in disease detection, TB transmission within different population groups and effect of treatment. Notification rates are inherently different across different geographic regions within a country. A first step in understanding the underlying causes which lead TB case notifications to vary across regions is to determine and describe these geographic variations and trends in case notification rates.

Step 1: How do the reported rates of TB vary over space and time?

1. What is the distribution of case notification rates (CNR) sub-nationally?
2. How consistent is the CNR over time and across geographic areas? Are the subnational CNRs related to each other in space?
3. Are there areas identifiable which deviate from the general pattern (inconsistencies)?

After variations in disease have been mapped and analysed, geographic trends and inconsistencies can be identified. These findings will determine various scenarios which could explain the observed trends and outcomes. Ideally one would like to have an estimate of the local burden for each location. TB prevalence surveys often provide a number of regional estimates, but not at the required subnational level. However, combining estimates of national prevalence, stratified by province, or risk groups (age, sex, urban-rural), with geographical distribution of other TB risk factors can provide insight into the expected local TB burden. These estimates are triangulated with reported rates to get a first insight into where possibly most TB cases are being missed.

Step 2: Do the subnational case notification rates reflect the true local status of the TB epidemic, including understanding of the distribution of TB risk factors and key populations?

1. What is the national or provincial estimated TB burden?
2. How variable is the TB burden over time and across geographic areas?
3. What is known about the difference in TB burden among specific populations (age, sex, urban/rural, key populations)?
4. What is the estimated subnational distribution of the burden based on the demographic composition (including presence of key populations at risk of TB) of each district's population?
5. How do the prevalence of TB risk factors and key populations vary across geographic areas?
6. How consistent are case notification rates with expected TB burden locally? And what does this tell us about areas with possibly a high number of missed people with TB?

4.2 Understand the local health care response

For a complete picture of the reported rates, it is important to triangulate these with program input and outcome indicators which will further allow one to ascertain the different scenarios that may be causing the observed variations in TB case notification. Using geographic maps has the added benefit of providing additional information, allowing one to interpret these variations in a wider geographic context.

Step 3: What is the relation between the case notification rates, the estimated burden and the diagnostic coverage sub-nationally?

1. What is the diagnostic coverage of past and current program interventions (diagnostic centres, lab efforts, ACF efforts, access to diagnostic services)?
2. How consistent are the case notification rates, burden and diagnostic coverage?
3. To what degree are existing TB services functional and used? For example, are there any populations in the district completely missing out on care due to poor quality of services?
4. Do the program efforts match the need of the different populations? Geographic coverage of key populations?

Once the geographic patterns of case notification rates, burden and program delivery have been mapped and compared spatially, it is possible to establish broad ideas regarding the effectiveness of a program to detect TB cases and classify areas according to their expected burden and program performance. An additional layer in this pathway is the presence of risk factors and key populations who face particular barriers in accessing care and which require different strategies of case detection and treatment.

Once geographic areas have been identified in which TB cases are likely significantly underreported, the next step is to do a program analysis which helps in understanding whether it is for example an issue of low quality diagnostics, or poor access of certain risk populations to care.

The next steps focus on mapping various stages along the pathway of care to identify bottlenecks in health care access, diagnosis and treatment.

Step 4: Which areas and population groups are likely to have limited access to TB care?

1. What is the geographic coverage of health facilities providing TB services (public and private)? For example, are there any populations in the district that are completely missing out on care due to total lack of (or access to) services?
2. To what degree are existing TB services functional and used? For example, are there any populations in the district completely missing out on care due to poor quality of services?
3. What is known about general health service utilization across different populations and parts of the country (geographic, financial and social barriers)?
4. Mapping of outputs and results of various health facilities
5. Do the program efforts geographically match the need of the different key populations?
6. Which additional program efforts are implemented? Active case finding activities for specific population groups.

Step 5: Mapping the pathway of care: what is the pathway from case finding to successful treatment?

1. What/how large are the numbers of people with TB which are not being reported and what is the time between the point of infection and testing?
2. What/how large are the numbers of people with TB which are not being reported and what is the time between testing and treatment outcomes?
3. How consistent are the quality of diagnosis, percentage bacteriologically confirmed, and treatment outcomes over time and geographic area?

5 Develop locally tailored approaches

To improve the delivery and access to primary care services, disease programs need to prioritize actions to places where people are most likely being missed. It is important to note that low coverage of primary care is usually caused by a combination of factors and these are likely to differ in different parts of the country and for different population groups. Therefore solutions which are expected to have a high impact will vary across countries and epidemiological settings. While the mapping and analysis of data following the guidance above provides valuable clues to which population groups should be prioritized for additional interventions and where these are located, it is important to collect additional information to fully understand why these groups of patients are not reached by the national program. Discussions with patients, providers and other stakeholders, and service quality assessments may provide valuable information to guide a tailored response.

5.1 Geographical locations and population groups to be prioritized

The MATCH Approach provides insight into the geographical areas and population groups that are most likely to be missed by the health system. One may find that case notification is much lower than the estimated burden in only specific geographical locations. This would call for prioritizing these underreporting areas for intensified efforts to increase coverage and/or quality of services. The first step will be to verify the results with people with knowledge of the local context and practices to validate possible causes for low access to health care. For instance, in situations where the notification rate proportionate to the estimated burden is particularly low in rural areas one should focus interventions specifically on barriers for rural populations to access effective diagnosis and care. On the other hand, finding a disproportionate high number of missed patients in urban areas warrants a deeper understanding of appropriate ways to overcome urban access barriers. There are also instances in which the problem seems more generalized across all geographic regions, but in which case notification is found to be particularly low during certain times of the year. This indicates the existence of time-bound problems (e.g. seasonal inaccessibility of services, seasonal migration, unrest, strikes, interrupted reporting or lab-supplies). Lastly the mapping exercise and analysis could lead to the conclusion that the general population is reasonably well covered by the health system in place, but specific key populations or risk groups might still not be reached.

5.2 Barriers in accessing services

Barriers to accessing services are often a combination of geographical barriers (distance to services, impassable terrain) financial barriers (cost of consultation, diagnosis, treatment, transport, loss of income), and inconvenience (opening hours, waiting times). In the case of geographic problems, increased decentralization and coverage of services can sometimes be readily achieved through expanding the number of clinics that offer TB diagnosis and treatment services. This may not be feasible in remote rural areas with low population density without jeopardizing the quality of diagnosis. In such situations, community based services, mobile outreach, or transportation services for patients or diagnostic samples are potential solutions. Costs of diagnosis and waiting times have successfully been assessed by provision of vouchers and fast-lane services for people referred with symptoms of disease. User friendly services may be designed to prevent loss of income for daily labourers through more inconvenient opening hours and locations of services. Barriers to accessing services may also be defined by social, cultural, or stigma related factors. The mapping of service coverage and notification among risk groups and key populations will provide clues as to which groups may have high prevalence but are still missing out on services. Analysis and mapping of the notification rate data of the general population – disaggregated by age and sex – and comparing these with the expected burden may provide clues to segments of the general population being disadvantaged (male vs female; elderly, children or other age groups).

5.3 Quality of diagnosis

Some situations may show that the number of people screened or samples collected and tested is high relative to the number of patients reported and treated in a district. This could be indicative of a low quality of diagnosis. In this case it is important to evaluate diagnostic performance such as the diagnostic error rates of (molecular) laboratory tests. A review of internal and external quality assurance reports will assist in better understanding lab quality issues. Another potential cause of low quality diagnosis is transportation of samples with long duration under unfavourable conditions.

A low rate of disease relative to the number of people tested is not in all cases a sign of poor quality diagnosis. It could signal that though many people are screened and tested the programme focuses on the wrong population groups (with low prevalence). This highlights the importance of assessing geographic patterns of disease reporting to the spread of risk factors.

5.4 Non reporting leading to low case notification

If the mapping of private sector providers, analysis of sale of drugs, and pathway of care analysis indicate a substantial number of patients being diagnosed and possibly treated in the private sector, it is possible that a large number of patients are being diagnosed in the private sector. If these patients are not flagged to the national program, this contributes to the low case notification rate. Intensified public private partnership should then be pursued. In other situations the analysis of disaggregate BMU-level reporting over time may lead to identification of incomplete reporting by certain BMUs. Under a paper-based reporting system reports may have been lost or delayed. Similarly, problems have been observed with inflexibilities in electronic reporting systems, which may lead to long delays and incomplete notification. These kind of issues warrant follow up and troubleshooting by the health program supervisors at all levels.

EXAMPLE OF MATCH IN BANGLADESH:

TB case notification gaps – indications of TB under detection.

In Bangladesh the MATCH Approach was implemented using district-level TB case notification rate data, which were integrated with socioeconomic data from the 2013 census as well and routine laboratory data. During a participatory data analysis workshop the NTP, together with their implementing partners (BRAC, ICDDR,B, MSH), collaboratively mapped and analysed these data using the MATCH framework.

In the central-southern region (yellow highlighted districts), TB case notification rates were considerably lower than the country average and also compared to the CNR in directly adjacent districts. Analysis and triangulation with other program components showed that:

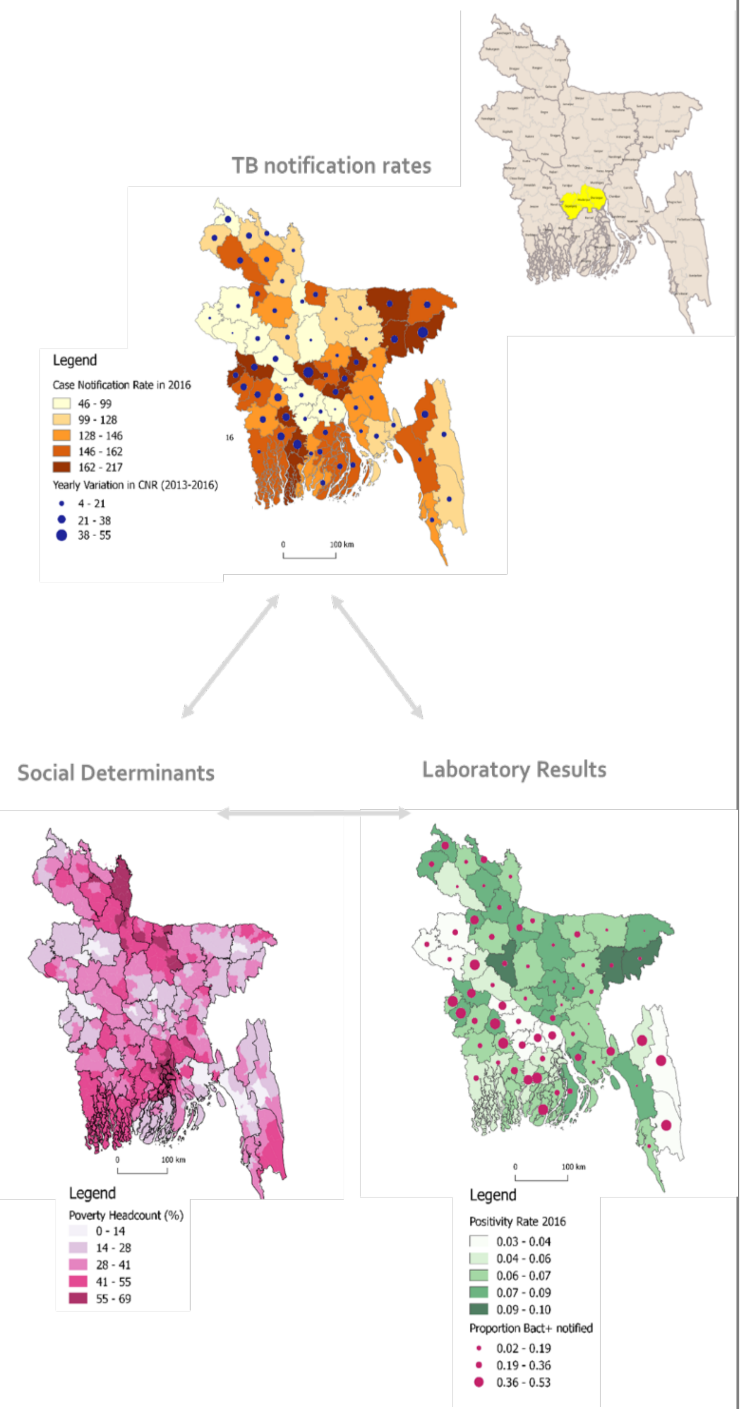
- Test rate is low but the proportion of bac+ patients among all notified is comparatively high;
- Positivity rate of tests performed is low;
- The poverty rate in these areas is relatively high as compared to the country average.

These findings led to the following hypotheses:

- The information on socioeconomic status and risk factors suggest no reason to believe the actual burden in this area is significantly lower than in the neighboring districts;
- Low test rates and low positivity rates suggests many patients are not reaching quality diagnosis;
- Low coverage of microscopic facilities indicate poor coverage of diagnostic services;
- Above average percentage of bacteriologically confirmed patients among the low number of notified patients suggests that also many smear negative and extra-pulmonary patients might be missed.

Interventions to be considered in these areas after verification of hypotheses:

- Need to increase presumptive case finding through improved screening in facilities and community;
- Improve coverage of Xpert testing for all microscopy negative and X-ray for B- presumptive cases;
- Conduct more supervisory visits to find out the root cause of lower notification rate over the years;
- Check completeness of notification system.



ANNEX 1 ///

Mapping and analytical methods

GIS offers several options which can be used to analyze subnational data, ranging from basic visualizations to complex geo processing techniques. These techniques should be carefully chosen depending on the availability and type of data and the purpose of the analysis.

Thematic mapping

CHOROPLETH MAPS

Data: choropleth maps are best made with numerical data such as proportions and rates, but not counts.

Choropleth mapping is a useful tool to explore your data and identify geographical patterns. It shows you the geographical variation in your data by dividing the data into groups with each its own color, this makes it easy to identify geographical areas which belong to the same group. The data can be grouped in several ways depending on the data characteristics:

QUANTILES:	Divides the data in to 4 equally sized groups, the first group contains the 25% lowest values and the last group contains the 25% highest values.
NATURAL BREAKS:	Creates classes of data with minimal variance within each class and maximum variance between classes.
PRETTY BREAKS:	Creates classes of which the boundaries are round numbers.
EQUAL INTERVAL:	Creates classes that have the same number of units (e.g.: 0-10, 10-20, 20-30 etc.)
STANDARD DEVIATION:	Creates classes based on the standard deviation from the mean.
CATEGORICAL:	Mainly used for categorical data and not for numerical data. Each class contains observations belonging to the same category.

The type of grouping to apply and the number of groups to create should be based on several considerations:

- Conceptually: how is the data often presented? Are there standardized cut-offs?
- Range of data: data with a very small range is sometimes better visualized using few categories.
- Number of administrative units: tune the number of categories to the number of geographical areas to avoid having too much or not enough observations within a group.

Proportional symbols

Data: proportional symbols are best used for numerical data such as proportions, rates and counts.

A second method of visualizing your numerical data is with proportional symbols. As the name suggest this way of visualizing data creates a symbol which size is proportional to the value it represents, hence a large symbol indicates a high value and a small symbol indicates a low value. Proportional symbols allow you to instantly see where larger number of occurrences are.

BOX 1. OVERLAYING MAPS

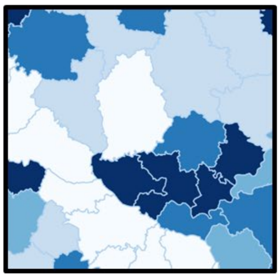
Overlaying two maps can be very useful in identifying local inconsistencies or patterns in data.
For example, you can map a disease rate using a choropleth map and overlay it with proportional symbols showing the number of people tested for that disease.

Heat map

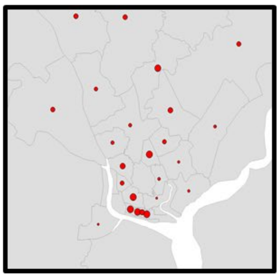
Data: a heat map is made with raster data. Essentially, raster data is made of pixels which are assigned a specific value.

A heat map is similar to a choropleth map in that it shows the geographical distribution of your data. However - unlike a choropleth map - a heat map does not group data within administrative boundaries. It looks at the intensity of occurrence on each point.

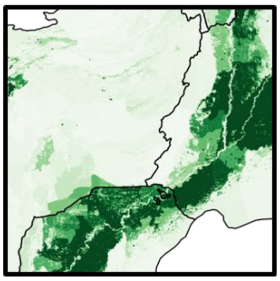
a) Choropleth map



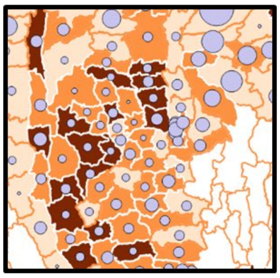
b) Proportional symbols



c) Heat map



d) Overlay of choropleth map with proportional symbols



Geo spatial techniques

SMOOTHING

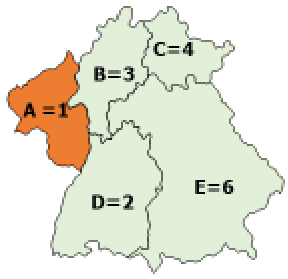
Data: smoothing is often done to identify geographical units that have disease rates which deviate substantially from its surroundings, indicative of under or over reporting or to reduce random irregularities in data.

Data smoothing is a method in which original data are transformed by taking the moving average of a set of data points which are closely related. In spatial statistics such relatedness is measured by geographical distance or adjacency. We distinguish two types of spatial smooths with different application and reasoning:

1. Weighted average of an area and its neighboring areas
2. Weighted average of only neighboring areas (spatial lag)

BOX 2 Smoothing Moving Averages

Data smoothing is a method in which original data are transformed by taking the moving average of a set of data points which are closely related. In spatial statistics such relatedness is measured by geographical distance or adjacency.



We distinguish two types of spatial smooths with different application and reasoning:

1. Weighted average of an area and its neighbouring areas
2. Weighted average of only neighboring areas (spatial lag)

The first type of smoothing is used to get more stable estimates: small denominators can lead to instable rate estimations. Furthermore, patients may cross borders to find health care, which is not reflected in population estimates (this is especially true for urban areas). In addition visualizing smoothed values could help in finding a spatial pattern.

The second type of smoothing is used to find discrepancies in values between an area and its neighbors. The two maps should always displayed together.

The weights applied to each adjacent area are represented by Matrix W which gives specific weights of each area *i* relative to its neighbor *j*.

The matrix weights of each area (rows) relative to a neighbor (columns), can reflect the inverse distance between the centroids of two areas, or it can simply reflect whether two areas share a common border. In the example here the length of the shared border was used. To ensure that the weighting sums back to 1 the rows are standardized (e.g. sum to one).

neighbor

local	A	B	C	D	E
A	0	0.7	0	0.3	0
B	0.4	0	0.3	0.1	0.2
C	0	0.5	0	0	0.5
D	0.2	0.1	0	0	0.7
E	0	0.3	0.2	0.5	0

An example of the calculation of the weighted neighboring average for area A is given below.

$$A_{smooth} = \frac{((W_{AB} * B) + (W_{AD} * D))}{(W_{AB} + W_{AD})} = \frac{(0.7 * 3 + 0.3 * 2)}{(0.7 + 0.3)} = 2.7$$

The first type of smoothing is used to get more stable estimates: small denominators can lead to instable rate estimations. Furthermore, patients may cross borders to find health care, which is not reflected in population estimates (this is especially true for urban areas). In addition, visualizing smoothed values could help in finding a spatial pattern.

The second type of smoothing is used to find discrepancies in values between an area and its neighbors. This map should always be displayed alongside a map with the original (i.e. unsmoothed) values.

It is important to take into account that some areas are more closely related than others, for example because they share a longer border or are closer in distance. To account for this we apply weights. The weights that we apply to each adjacent area are represented in by a matrix which gives specific weights for each area *i* relative to its neighbor *j*. The matrix weights of each are (rows) relative to a neighbor (columns), can reflect the inverse distance between the centroids of two areas, or it can simply reflect whether two areas share a common border. In the example here, the length of the shared border was used. To ensure that the weighting sums to 1, the rows are standardized.

SPATIAL AUTOCORRELATION

Data: Measures of spatial autocorrelation are used to quantify whether observed geographic patterns in data are statistically significant and hence should be interpreted as tangible spatial variation reflecting real world differences between geographic areas.

Spatial autocorrelation implies that observations are geographically clustered and thus cannot be treated as independent data points. This dependency indicates that disease rates are more similar in geographically adjacent areas than in geographically distant areas. For example, neighbouring areas might have similar tuberculosis case notification rates due to similarities in underlying TB risk factors, TB health care infrastructure or demographic characteristics associated to TB. There are two types of spatial autocorrelation analysis:

1. Global patterns: Global Moran's I
2. Local patterns: Local Indicators of Spatial Autocorrelation

The first one measures the overall spatial autocorrelation of your variable of interest and is expressed as a value ranging between -1 (perfect dispersion) and +1 (perfect clustering), with 0 being completely random.

The second calculates a similar statistic, but then on a local level, for each spatial unit. This allows us to identify 'hotspots' – clustering of geographical units with similar values - and 'cold spots' – areas which are significantly different from their neighbours.

2-STEP FLOATING CATCHMENT METHOD (2SFCA)

This method is used to analyse access to health care or diagnostic services. Access to health care is often regarded of great influence on the burden and of disease and healthcare seeking behaviour. Therefore, analysing the access to healthcare can support in program planning. The 2SFCA method aims to determine the following things:

1. Determining the population within a given catchment of a health care provider (e.g. 30 minute drive or 5km distance)
2. Determining the health care provider within the catchment of a population centre.

The first uses the 'supplier', in this case the health care provider, as a starting point whereas the latter uses the 'consumer', in this case the population that wants to access health care, as a starting point. The outcome of both (often a provider-to-population ratio) can be summed of to calculate the overall accessibility of health care.

ANNEX 2 ///

Geographical analysis & data triangulation

The following sections provide a stepwise description of the MATCH approach. Each step starts with a description of the purpose and leading questions to be addressed. Then a list of data requirements and a non-exhaustive list of potential outcomes are provided (both maps and non-spatial charts). Finally a reference to practical exercises using QGIS[2] and resources is provided.

The approach is illustrated using the example of Bangladesh. These data were kindly shared by Dr. Rouseli Haq of the Bangladesh National TB Program. The maps shown were produced following a workshop in Dhaka on the MATCH approach, together with the M&E office of the Bangladesh NTP (20-24 August 2017).

Box 3: Data Management Specifications For Bangladesh NTP Data Presented In The Manual

1. Total number of first administrative units (adm1): 7
2. Total number of second administrative units (adm2): 64
3. Particularly helpful was the fact that the NTP's number adm1 and adm2 reporting units match the available shapefile units. This means that in both the NTP data and adm2 shapefile, there are 64 units, and in both the NTP data and adm1 shapefile, there are 7 units. This avoided the need to resolve any discrepancies between the sources by doing research online or consulting the national bureau of statistics.
4. Some spelling differences were found between the shapefile administrative units and NTP reporting units, which is normal and was easy to resolve based on similarities in spelling. Some shapefiles additionally had a variable with alternative spellings of administrative units, which helped to resolve any differences with the NTP data.
5. Shapefiles used were not corrupted and boundaries were up to date for the data analyzed.
6. Some data were only available at the admin 1 level, which can be expected, but there was also a large amount of data available at the admin 2 level, making possible more informative maps.
7. The NTP was able to provide a large amount of data, including lab (testing) data, EQA data, notifications both age and sex disaggregated, census population data both age and sex disaggregated, and treatment outcomes. Other sources of data were widely available: DHS, MICS, bureau of statistics data, as well as statistical yearbooks.
8. Due to these favourable factors, Bangladesh was an ideal country to use as an example for this manual.

Know your local epidemic

Step 1: Study case notification rates over space and time

In step 1 routinely collected notification data is studied over space and time to identify patterns and inconsistencies to those patterns (areas deviating from what is expected).

National TB programs collect and report national TB case notification figures to monitor changes in program coverage and TB burden over time. TB case notification rate is a useful indicator for local epidemic situations for a number of reasons, with the most obvious being that TB case notifications

[2] QGIS (Quantum GIS) has been found to be a very useful software for this work as it is open-source and free (FOSS). QGIS can be downloaded here: download.qgis.org.

are the most routinely collected data by TB programs, often available at the lowest administrative level of reporting. It is available on a quarterly basis, and while there are routine reviews ensuring that late quarter 4 and early quarter 1 notifications are appropriately allocated, this is the data that is generally least delayed, in comparison to laboratory DST results and treatment outcomes. Quarterly case notification data also indicate the level of completeness in terms of the proportion of expected reporting BMUs and private providers actually reporting to the national TB program, and identification of specific BMUs or providers which deviate from expected notification levels or not reporting at all. More and more patient based electronic reporting systems are being implemented, allowing different aggregations levels.

Variations in TB notification rates will reflect differences in TB detection, TB transmission within different population groups and TB treatment which are inherently different across different geographic regions. A first step in understanding the underlying causes which lead TB case notification to vary across regions is to determine these geographic variations and trends in TB case notification rates.

In areas with high MDR prevalence, the same approach can be used to identify geographic and temporal patterns and inconsistencies to those patterns of MDR notification rates.

<u>In this step the following questions are suggested to guide analyses:</u>

A. What is the distribution of reported rates of disease subnationally?
B. How consistent are disease rates over time and across geographic areas? Are these geographically related to the neighbouring locations?
C. Where are areas deviating from the general patterns? (look for inconsistencies)

<u>Required data inputs:</u>

- quarterly (drug sensitive and drug resistant) notification data disaggregated by type of TB, age, sex and adm2 level or lower for at least one full year (for geographic trends).
  For a minority of countries, adm1[3] level data may be of sufficient disaggregation to yield a properly detailed subnational analysis to inform policy, i.e. countries where there are a relatively large number of adm1 units or in smaller countries, in which the small size of the country means that any further subdivisions will cover a small population or geographic area comparable to adm2 level.
- for time trends at least 3 years of data but preferably 5 years
- population size disaggregated by age and sex for the same subnational level as the notification data
- polygon shapefile showing the required subnational units (adm1, adm2 or lower)

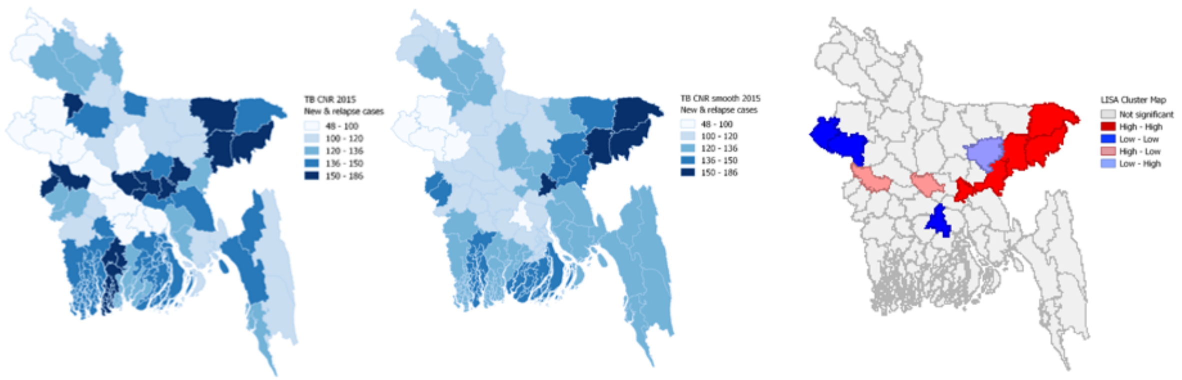
<u>Suggested outputs:</u>

1. graph showing subnational quarterly trends in case notification rates for the past years for different types of TB (All Forms, new cases, new B+, EPTB)
2. maps showing subnational annual case notification rates for the most recent complete year for different types of TB (All Forms, new and relapse cases, new B+, EPTB)

---

[3] Adm1 level refers to the highest level of administrative division in a country (usually regional or province level). Adm2 refers to the nest level of administrative division (often district level)

3. map showing subnational spatially smoothed notification rates with smoothed rates based on, for example, the average rate of the district and its adjacent districts (see box 2 for an explanation of spatial smoothing using moving averages.).
4. map showing subnational notification rates significantly lower or higher than their neighbours
5. map showing the notification hot and cold spots (Local Indicators of Spatial Autocorrelation[4])
6. map showing subnational areas with significantly changing notification rates over the past years
7. map showing the seasonality index (quarterly variation)
8. map showing subnational secular trends
9. map showing subnational paediatric disease rates
10. map showing subnational child/adult notification ratio
11. map showing subnational male/female notification ratio
12. map showing subnational MDR notification rates
13. map showing subnational percentages of MDR confirmation among newly detected patients with DST performed



A: Notification rates of new and relapse patients in 2015 by 100,000 population. Categories defined by equal counts (quantiles)

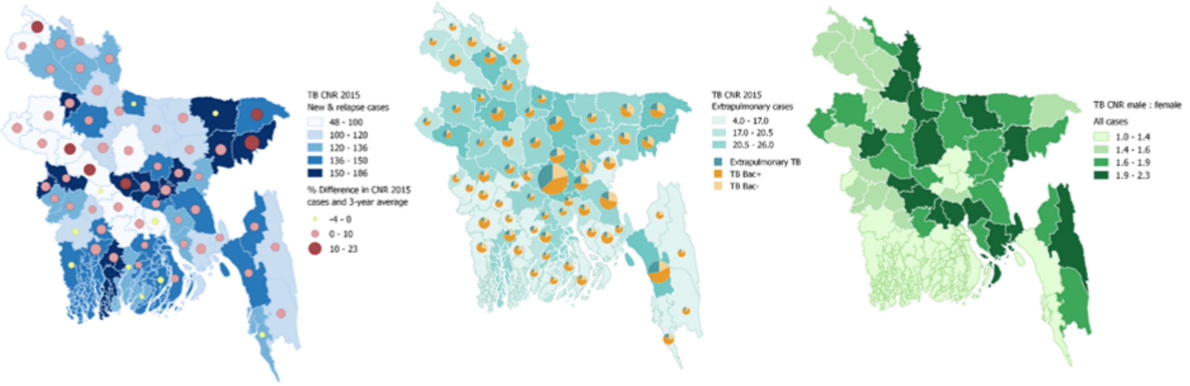
B: Smoothed notification rates of new and relapse patients in 2015 by 100,000 population. Same categories used as in map A. Smoothed values based on averaging the area value and its neigbouring values (no weights applied).

C: Hot and cold spots of notification rates of new and relapse patients in 2015 using local Moran's I (LISA). Areas deviating from its neighbours are also indicated.

The maps show large subnational variations in CNR. Subnational CNRs show a clear spatially clustered pattern: areas with low or high CNRs are clustered. Some areas deviate from this pattern and have lower or higher CNRs compared to their neighbouring areas. Those areas with lower CNR compared to their neighbouring areas are especially interesting when looking for missed TB cases. These areas need further investigation in terms of presence of key populations and diagnostic coverage.

Figure 1. Example outputs of analysing annual TB case notification rates in Bangladesh.

[4] https://en.wikipedia.org/wiki/Indicators_of_spatial_association



A: Notification rates of new and relapse patients in 2015 by 100,000 population and the % difference between the CNR in 2015 and the 3-year average (2013-2015).

B: EPTB notification rates in 2015 by 100,000 population and types and distribution of new TB patients

C: Male female ratio of all forms CNR of 2015

The maps provide different pieces of information. Map A suggests the CNRs are relatively constant over time. This gives higher confidence that the observed spatial pattern is consistent. Map B shows a different spatial pattern for the types of TB with more EPTB and relatively less bacteriologically confirmed cases in the northern districts. More male than female TB patients are notified, and the extent of this difference varies greatly. In the south-west and central regions the ratio is closest to 1 (map C).

Figure 2. Example output of analysing various types of notification data in Bangladesh.

Methods:

- Exploratory data analysis in Excel or DHIS2 through histograms, trends
- Mapping notifications rates according to predefined thresholds (e.g. programmatic targets, Natural means for natural groupings) using choropleth visualization.
- Mapping proportional notifications according to different types using diagram maps (i.e age disaggregation, relapse vs new)
- Mapping ranked notifications according to quantile groups (IQR, Percentile or Box, range).
- Smoothing through spatially weighted averages (see box 2).
- Moran's I to find spatial correlations for example between a district and its neighbouring districts
- Hotspot analysis to identify significant geographical hot and cold spots using local Moran's I[5].

Interpretation:

After variations in TB CNR have been mapped and analysed, geographic trends and geographic inconsistencies can be identified. These findings will determine various scenarios which could explain the observed trends and outcomes. For example, a district reporting very low case notification rates as compared to its direct surrounding would raise questions regarding the access, referral and diagnosis of TB. On the other hand a region of contiguous districts which consistently report high rates (also called a hotspot) can be hypothesized to reflect a high TB burden associated with the presence of a well-functioning program and high case detection.

Step 2: Investigate whether the reported subnational disease rates are likely to reflect the true local status of the TB epidemic, including understanding the distribution of TB risk factors and key

[5] Anselin (1995), Local Indicators of Spatial Association—LISA

populations.

The TB case notification rate can be interpreted as a product of on the one hand the effectiveness of the TB program and on the other hand the underlying TB burden. The TB program represents the efforts invested in detecting and treating TB patients (including generating awareness, access to care, case finding activities, and case holding etc). The TB burden reflects the number of TB patients in need of care. Low notification rates do not necessarily mean a low burden, but could also reflect a weak program. The reverse scenario is also possible. A high notification rate does not exclude the possibility that still large numbers of TB patients are missed. Based on this logic, notification rates can therefore only be properly interpreted in combination with information on the TB burden and the TB program.

Ideally one would like to have an estimate of the local burden for each location. TB prevalence surveys are designed to provide a national estimate of the TB prevalence in a country, but often also provide estimates for one level below national (e.g. region). The surveys usually also provide TB prevalence estimates for subpopulations such as males, females, different age groups and urban versus rural populations, or HIV-TB co-infections.

Combining what we know of the national prevalence, prevalence by province, demographic and rural/urban differences in prevalence, geographical differences in comorbidities such as HIV and diabetes, and distribution of other risk factors (smoking, malnutrition, elderly, socio-economic status etc) and key populations (miners, prisoners etc), can tell a lot about the expected local TB burden. In step 2 these data are compared to CNRs to get a first insight in where possibly most TB cases are being missed.

A similar exercise can be done for comparing estimated multi drug resistant (MDR) TB burden with MDR notification rates by region or district, making use of drug sensitivity survey data.

<u>In this step the following questions are suggested to guide analyses:</u>

A. What is the national or provincial estimated TB burden?
B. How variable is the TB burden over time and across geographic areas?
C. What is known about difference in TB burden among specific populations (e.g. age, sex, urban/rural, key populations)?
D. How do the prevalence of TB risk factors and key populations vary across geographic areas?
E. What is the estimated subnational distribution of the burden based on the demographic composition (including presence of KPs) of each district’s population?
F. Which locations are expected to have significantly higher or lower TB burden than the national average?
G. How consistent are CNRs with expected TB burden locally? Where are areas with possibly a high number of missed cases?

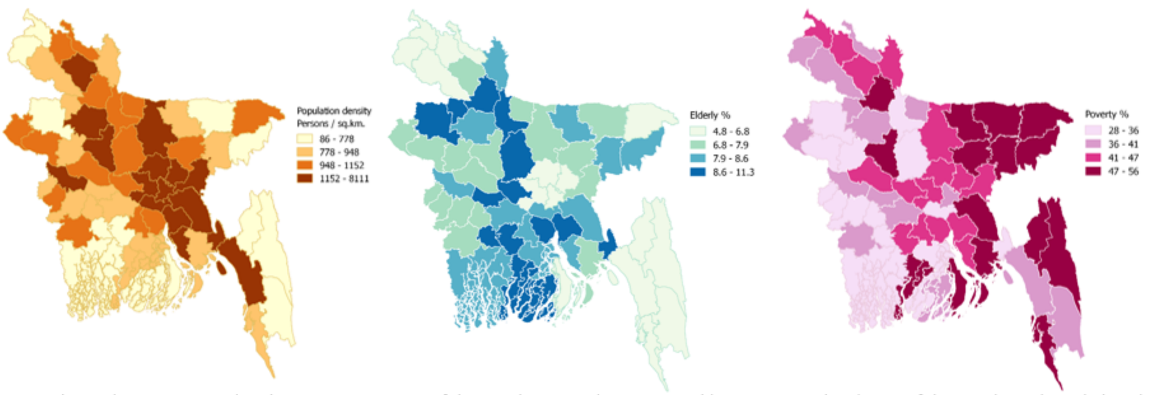
<u>Required data inputs:</u>

- results of most recent TB prevalence survey with subnational estimates and by age, sex, urban, HIV categories
- population size disaggregated by age, sex and urban/rural areas for adm2 or lower level
- subnational case detection survey data
- HIV prevalence estimates at subnational level
- subnational estimates of prevalence of risk factors: HIV, diabetes, urbanization, smoking habits, malnutrition, socio-economic status (wealth, literacy), elderly age

- location and size of key populations: migrants, miners, prisoners, urban poor, people who use drugs, homeless population
- results of most recent drug resistance survey with any subnational estimates

Suggested outputs:

1. map showing the results of the latest TB prevalence survey at the lowest level available
2. map showing subnational (preferably adm2) estimates of the TB prevalence based on demographic characteristics of the adm2 area (sex, age, HIV, urban/rural)
3. maps showing distribution of indicators/proxies of TB burden (risk factors, comorbidities), e.g. HIV prevalence, diabetes, population density, proportion of urban population, socio-economic status (wealth literacy), smoking habits, prevalence of malnutrition, elderly age.
4. map showing TB/HIV coinfection (% TB patients co-infected with HIV in combination with the HIV testing rates and/or with HIV prevalence)
5. maps showing locations of key populations such as migrants, miners, prisoners, urban poor, injecting drug users (IDU), homeless population



A: Population density measured as the average population in each km²

B: % of the population aged 65 years or older

C: % headcount of the population living below the national poverty line

The maps depict some known risk factors for TB. If there is local evidence that the TB prevalence increases with higher population density, and that prevalence of TB is higher among elderly and among the poor, these maps could be indicative of where to expect higher TB prevalence. Comparing these with the maps of the first example output could give indications of where TB patients may be missed.

Figure 3. Example output of mapping risk factors for TB in Bangladesh.

Methods:

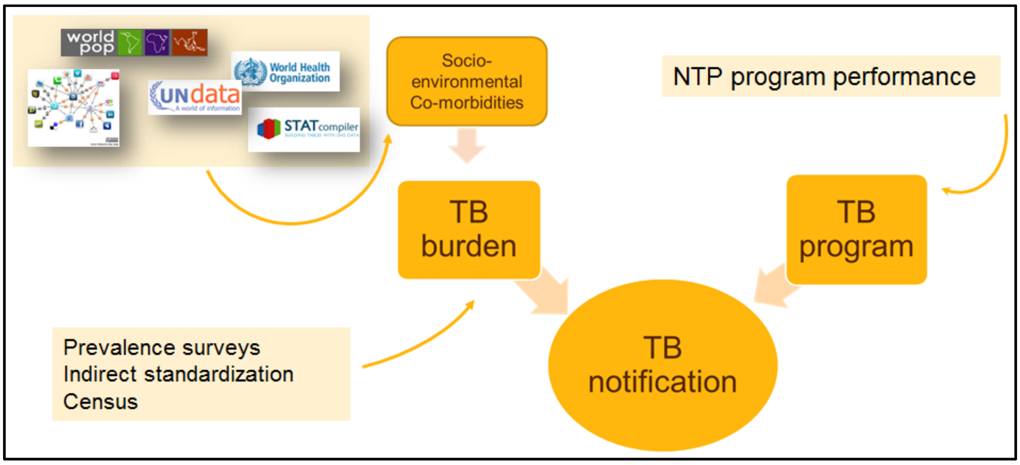
- Comparative (multi-indicator) mapping of TB case notification data and TB prevalence data to assess possible missed cases. Missed TB cases are expressed as the gap between the prevalence and the case notifications.
- Comparative (multi-indicator) mapping of TB case notification data and at risk population data to assess possible missed cases. Missed TB cases are expressed as the gap between the risk populations and the case notifications.
- Comparative (multi-indicator) mapping of treatment coverage (still commonly referred to as case detection) data relative to the case notification rate.

Interpretation:

In a perfect surveillance system all new and relapse cases of TB would be recognized, have access to health care , would be diagnosed with 100% specificity and sensitivity and reporting would be flawless. In this case the case notification rate would be exactly the same as the incidence rate as all cases would be detected. However, when the case notification rate is lower than the true incidence rate, some TB cases are being missed due to deficiencies in case recognition, access to health care, TB diagnosis or TB case reporting. Because the incidence is hypothetically linearly proportional to the prevalence rate, the proportion of cases notified out of the prevalence rate or population at risk should be the same throughout a country. Variations in these indicators could therefore be indicative of TB cases being missed.

Step 3: What are the relationships between the case notification rates, the estimated burden and program diagnostic coverage subnationally?

In step 3 the variety in CNRs will be further explored by adding a third layer: programmatic efforts to diagnose TB patients. As discussed earlier notification rates can only be properly interpreted in combination with information on the TB burden and the diagnostic coverage (see figure below). The diagnostic coverage of TB patients, can be studied through programmatic factors, such as the number of presumptive TB patients identified and tested over the total population (Testing Rate), the proportion of bacteriologically positive TB patients identified among persons tested for TB, and the proportion of bacteriologically positive TB patients among TB cases notified. Combining testing rates with the proportion positive can be informative on the diagnostic coverage, for example a low testing rate in combination with a high percentage positive can be indicative for low diagnostic efforts where only the most severe TB patients are tested. An area with high testing rates among the population and relatively low positivity rate may indicate a higher diagnostic coverage for instance through the use of a more inclusive definition of presumptive TB. Overlaying these diagnostic efforts with case notification rates and expected TB burden provides further insights in where individuals with presumptive TB are potentially missed.
Examples of areas of interest are those with geographically inconsistent[6] low CNRs, in combination with an expected high TB burden and low TB efforts.



[6] Having spatial inconsistencies in which there are administrative unit(s) with lower CNRs compared to their neighbors with higher CNRs, implies that further investigation may be warranted to determine whether the low CNRs are caused by a lower burden, by less efforts to identify cases, or reduced access to quality diagnosis.

Figure 4. Schematic relationship between TB notification, TB burden and TB program.

Using geographic maps has the added benefit that maps make it possible to interpret these variations in a wider geographic context.

Triangulation of data is key in this step and results are not always leading to straightforward conclusions. This is an explorative analysis and often leads to further questions to be answered. Studying time trends is important to exclude errors in data.

Caution is required though. For example an area with a low CNR and high testing rate with low positivity (indicative for good programme efforts) could still miss patients when the programme is only targeting people who are accessing health services and is missing out on a very important target population. Knowledge on the expected TB burden should shed light on this.
In interpreting lab data it is also important to consider the distribution of diagnostic facilities. If for instance district A reports a very high case notification rate but relatively low laboratory testing rates, suspicion regarding the quality of the data would be raised. However, if the same mapping shows that a directly adjacent district B is reporting proportionally high testing rates this could suggest that testing for district A might be conducted in district B.

<u>In this step the following questions are suggested to guide analyses:</u>

A. What is the intensity of past and current diagnostic efforts (diagnostic coverage, lab efforts, ACF efforts)?
B. How consistent are the CNR, burden and diagnostic coverage? (compare CNRs with diagnostic efforts, and burden with diagnostic efforts at subnational geographic level and for specific key population groups)

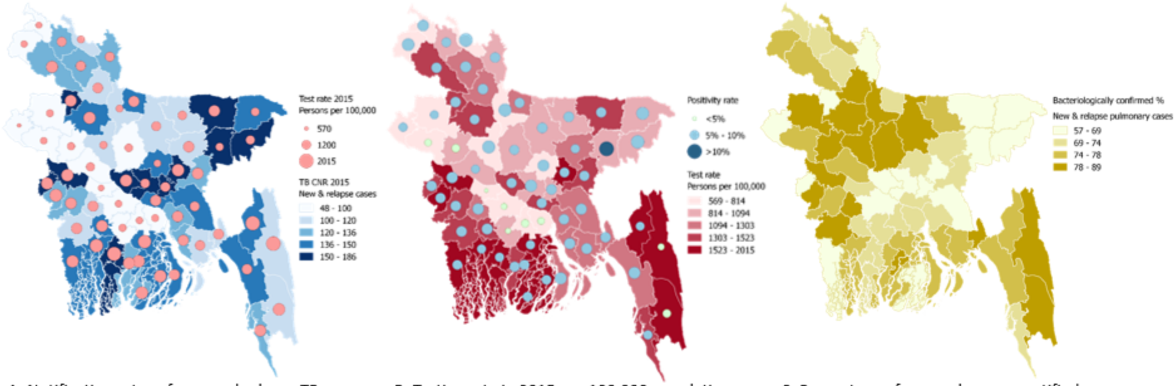
<u>Required data inputs:</u>

- laboratory data per laboratory or adm2 level: number persons identified with signs/symptoms suggestive of TB (presumptive TB), numbers of persons tested for TB (number of slides read), number bacteriologically positive (separate by microscopy and GeneXpert where possible)
- number of diagnostic facilities per adm2 level
- population density
- results of active case finding activities (number of contacts investigated, number of PLHIV screened for TB etc) by subnational (adm2) level and test positivity rates

<u>Suggested outputs:</u>

1. map showing the subnational testing rate (per 100,000 pop) in combination with the proportion positive
2. map showing the % bacteriologically confirmed out of all new patients
3. map showing the number of slides read per laboratory (categorized into 2 classes) and the percentage positivity
4. map showing facility density per subnational unit and in combination with population density (# diagnostic facilities and treatment facilities/population)
5. maps showing results of active case finding activities among general population and among populations at special risk of TB (e.g. contact investigation, testing PLHIV)
6. maps showing case notification amongst specific risk/key populations (notification of

TB among PLHIV, notification TB in miners, prisoners, or other specific groups)



A: Notification rates of new and relapse TB patients in 2015 by 100,000 population and the testing rate (i.e. number of persons tested for TB per 100,000 population)

B: Testing rate in 2015 per 100,000 population and the proportion with a positive test result (i.e. positivity rate)

C: Percentage of new pulmonary notified cases with bacteriological confirmation

Map A clearly shows that in areas with low CNR in 2015 fewer people were tested for TB compared to other districts. If these areas have high positivity rates this could be indicative of a too low testing rate (map B). Inversely low positivity rates in areas with low testing rates and low CNR could be indicative of a true low burden, but could also be the result of low sensitivity of referral and testing algorithms or insufficient quality of the diagnostic test. Where smear microscopy is the main diagnostic to confirm TB, a high percentage of bacteriological confirmation among patients notified may suggest patients with smear negative TB are being missed (map C).

Figure 5. Example output of analysing diagnostic coverage in relation to the case notification rate in Bangladesh.

Methods:

- Triangulation of data analysed in steps 1 and 2 with TB laboratory data by overlaying maps, for example choropleth maps (burden or notification) with symbology mapping (testing rates) to find spatial correlations.
- Looking for geographic and temporal trends by comparing these overlaid maps (described in the point above) over time, i.e. mapping annual data three or more years in a row.

Interpretation

As outlined in figure 9, the case notification rate is a product of the actual TB epidemiology (burden) and the efforts of the TB program to detect and treat patients. Good programme functioning would result in higher detection, treatment success and lower transmission. This would result in an initial rise of the case notification rate followed by a decline due to successful stop of transmission. Therefore case notification rates should be triangulated with screening efforts (testing rate) and the results of laboratory testing as this allows to identify areas where testing and diagnosis are relatively low and could be indicative of under detection of TB

Understand gaps in the local TB response

Step 4: Identify areas with limited access to tb care

Now that we have a better idea of where TB patients are possibly missed, the next steps focus on mapping various stages along the pathway of care to identify bottlenecks in health care access, diagnosis and treatment. Two levels of analysis can be identified: (1) identify subnational areas or clusters of areas where discrepancies between TB risk, burden , notification and detection efforts exist and (2) within those subnational areas where TB under notification is likely, assess health care delivery (also in relation to specific risk groups, local context analysis).

This will allow to ascertain different scenarios for what may cause the observed variations in TB case notification.

In this step the following questions are suggested to guide analyses:

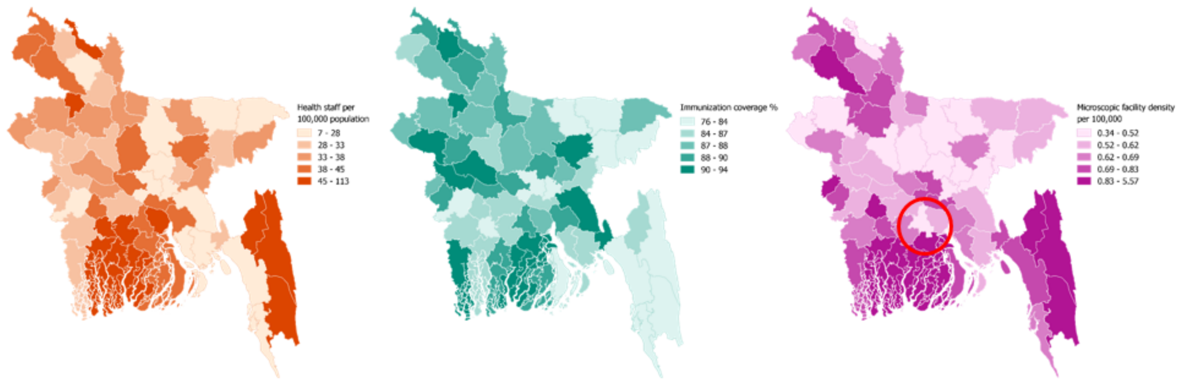
A. What is the geographic coverage of health facilities providing TB services (public and private)? For example, are there any populations in the district completely missing out on care due to total lack of (access to) services?
B. To what degree are these TB services functional and used? For example, are there any populations in the district completely missing out on care due to poor quality of services?
C. Do the program efforts match the need of the different populations? Match to geographic coverage of key populations

Required data inputs:

- Geographic point data of the location of diagnostic facilities and TB service facilities (geographic coordinates).
- list of services present at each diagnostic facility (microscope, GeneXpert, x-ray and functionality), type of facility (private, public), staff availability and capability
- location of treatment facilities (geographic coordinates)
- notification and laboratory data for each facility
- subnational % of notified cases coming from the private sector (PPM data)
- spatial line vector data of country road network

Suggested outputs:

1. map showing coverage of TB diagnostic facilities (microscopy, Xpert, x-ray) – public and private
2. map showing coverage of functional microscopes
3. map showing staff availability (lab staff, TB workers) – trainings
4. map showing coverage of TB treatment facilities
5. maps showing proxies for access to care: immunization coverage, U5 mortality
6. map showing health insurance coverage
7. map showing usage of general care services
8. map showing usage of TB services
9. map of network weighted travel distances to the nearest TB diagnostic and treatment facility (network analysis)
10. map of population and distance weighted coverage of TB services (2SFCA)



A: Number of health staff per 100,000 population

B: Vaccination coverage for all basic vaccinations of children aged 12-23 months in percentage

C: Microscopic facility density per 100,000 population

Maps A to C show indicators of access to care. Low number of health staff per 100,000 population (A) and low immunization coverage (B) are all suggesting poor access to health care. Facility density (map C) shows great variation between districts. The marked district in map C was marked as a cold spot in the first example output and showed low testing rates in the fourth example output and low facility density. This could be one of the reasons why TB patients are being missed.

Figure 6. Example output of analyzing access to care in relation to population in Bangladesh.

Methods:

- Two step floating catchment area[7] analysis combining high resolution population data with point locations of TB facilities
- Network analysis[8] to calculate travel distances to the nearest facility and delineate facility catchment areas.

<u>Interpretation</u>

Subnational variation in TB case notification and inconsistencies with the presence of risk groups and TB detection efforts are expected to be indicative of the missing of TB cases. This does not per se mean that those cases are actually not finding their way into care or are not being diagnosed. Mapping access indicators and the involvement of the private sector will allow to assess whether access to care is likely preventing people with TB to be diagnosed and find treatment. Access to TB diagnostic services can be visualized by mapping different access indicators ranging from physical access (presence, distance to care) to availability (types of diagnostic services and staffing), appropriateness (quality and performance) and financial barriers (public, cost recovery or private). Careful assessment of (small scale) access variation will allow to identify key barriers which can result in the missing of TB cases.

Step 5: Mapping the pathway of care, from case finding to successful treatment

Analysis of programmatic gaps is key to understand not only how TB cases are being found, and diagnosed, but also how the program is able to treat patients and to break future transmission. Even when TB case detection is functioning according to standards, deficiencies in the treatment and follow-up of patients will lead to continued transmission preventing to stop the TB epidemic.

---

[7] 2SFCA: http://www.niu.edu/landform/papers/JHAP741_e2sfca.pdf

[8] Network analysis: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3511293/

Therefore it is key to analyse the pathway of care and to identify discrepancies in the proportion of individuals which successfully find their way into TB care resulting a successful treatment.
People at high risk of TB are commonly not equally distributed across different geographic areas. Likewise program efforts to detect and treat patients are not equally accessible and performing at the same level. Therefore bottlenecks in the pathway of care will vary across areas depending on the local context. Knowledge about the context specific bottlenecks in the patient pathway process will help to evaluate which interventions could be appropriate within specific settings.
<u>In this step the following questions are recommended to guide analyses:</u>

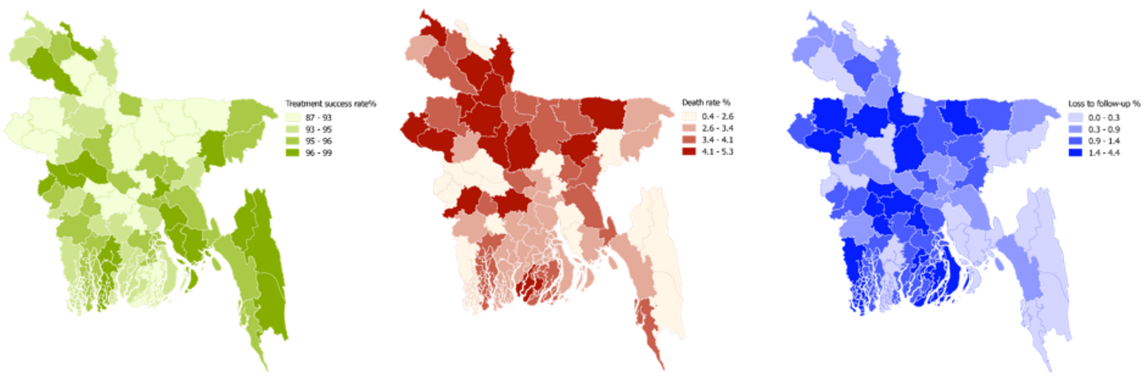
A. What/how large are the losses and delays between the point of infection and testing?
B. Are there locations with limited access and utilization of general health services (e.g. low vaccination coverage, low coverage with antenatal care)?
C. Are there locations where low quality of diagnosis leads to missed cases?
D. What/how large are the losses and delays between testing and treatment outcomes?

<u>Required data inputs</u>:

- Numbers screened for TB symptoms
- Numbers tested for TB, distinguished by diagnostic method (microscopy vs Xpert)
- Numbers bacteriologically confirmed for TB and total TB notifications
- Treatment outcome data
- Laboratory EQA data

<u>Suggested outputs:</u>

1. maps showing testing rates and test positivity rates
2. maps showing quality of diagnostics: External Quality Assurance results of microscopy (high false positives, high false negatives)
3. maps comparing diagnostic rates and notification rates
4. maps showing patients lost to follow up
5. maps treatment outcomes (treatment success rate, default, failures, death)



A: Treatment success rate (TSR) of the TB patient cohort starting treatment in 2014

B: Percentage of cases notified who died after treatment was initiated in 2014

C: Percentage of cases notified who were lost to follow-up after treatment was initiated in 2014

Maps A to C show indicators of program performance and effectiveness which could reflect issues with late case detection and referral, low adherence or treatment initiation. TSR could be indicative for the quality of the TB program. Some of the earlier identified districts with low CNR and poor diagnostic coverage also show low TSR (map A). Higher death rates (map B) in absence of high HIV prevalence could indicate ineffective treatment and continuous transmission. Loss to follow-up (map C) could indicate low access to treatment services, both physical as well as financial and could reflect lowered ability to detect, diagnose and treat cases of TB

Figure 7. Example output of analysing programme effectiveness for TB cohort starting treatment in 2014.

<u>Methods:</u>
Calculate the loss to follow up at each stage of the pathway to care, i.e.:

- number symptomatic for TB / number screened for TB
- number tested for TB / number symptomatic for TB
- number with a diagnosis of TB / number tested for TB
- number initiating treatment for TB / number diagnosed with TB
- number receiving additional lab work (culture and DST) / number eligible according to diagnostic algorithm
- number with successful treatment outcomes / number initiated on TB treatment, etc.

Interpretation

Analysis of the TB patient care-seeking and service delivery can help programs to identify gaps to be prioritized for targeted planning. Identifying the health systems-related challenges and obstacles along the patient pathway is key to identify which interventions are expected to be most effective and where these should be targeted. Combining the outcomes of the TB epidemic assessment with those of the TB service delivery (access) and patient pathway will inform strategic planning and decision making prospectively and should be part of routine planning.