

## **Methodology, Sampling, Interviewer Training**

### *Contents*

*Methodology*

*Sampling*

*Survey Examples*

*Coding*

*Interviewer Training Issues*

## **1. Methodology**

The general procedures in survey research require the following steps, which we will apply to the Tool to Estimate Patients' Costs:

1) Decide on research question and goals of the study

The research question here is: What are the costs for TB patients before / during diagnosis and during treatment? The answer to this question will then also give us an idea whether TB services are affordable to patients.

2) Decide on the population relevant to the study

The survey aims to have a close look at the poor affected by TB. Since TB patients are usually found among the low-income groups, the population should contain a larger proportion of poor people than what is found in the overall population of a country. Inclusion and exclusion criteria need to be decided upon, such as whether only new patients, re-treatment, MDR or a mixture shall be studied; whether specifically defaulters are of interest; whether unregistered patients shall be included or patients who also/only visit private sector facilities, etc.

3) Decide on method used to collect the data

This will be done through administering a questionnaire in face to face interviews (see Guidelines for Adaptation).

4) Write research protocol and submit to ethics committee for ethical clearance

5) Develop a sampling strategy

This section on methodology will be followed by a section on sampling.

6) Adapt questionnaire and codes

See the Guidelines for Translating and Adapting the provided questionnaire.

7) Conduct interviews

8) Enter and analyze data

We recommend to analyze **median** and mean values for costs.

You can use any statistical package. We used **Epi Info**, which can be downloaded for free on the CDC website.

We have prepared a data entry template with the questions of the questionnaire for Epi Info. Epi Info will be the mechanism that stores and analyzes your data and subsequent data set, so it is crucial that the template is suitable for the context in which data is being collected. This template uses the generic Tool questionnaire, and there is a data entry field for each question and possible answer. Each completed questionnaire is entered into the database (using Enter Data), and saved there for cleaning and later analysis (using Analyze Data). Epi Info stores the entered data in a format that can also be used in Excel and Access. Codes for each answer are already included in the template.

**To make changes to the layout of the data entry template, select 'Edit View' while in the Enter Data section of the program, or simply go to 'Make View', open the Tool template called "Kitui", and make changes directly to the template. It is recommended to first make a copy of the file and work in the copy, because Epi Info stores changes immediately and later revisions may then not be possible anymore.**

It is strongly recommended that a statistician familiar with Epi Info be consulted before and during data entry with the Tool template to minimize errors. Furthermore, the '**Epi Info for Beginners**' manual is recommended to give an overview of how to use this free software: "Epi Info Beginner's Manual with Exercises" by the Great Lakes Epidemiology Center Community Based Research Training, Updated: 02/2004.

## 9) Report results

See the Guidelines on Interpretation of Results and Suggestions for Interventions based on Results.

## 2. Sampling

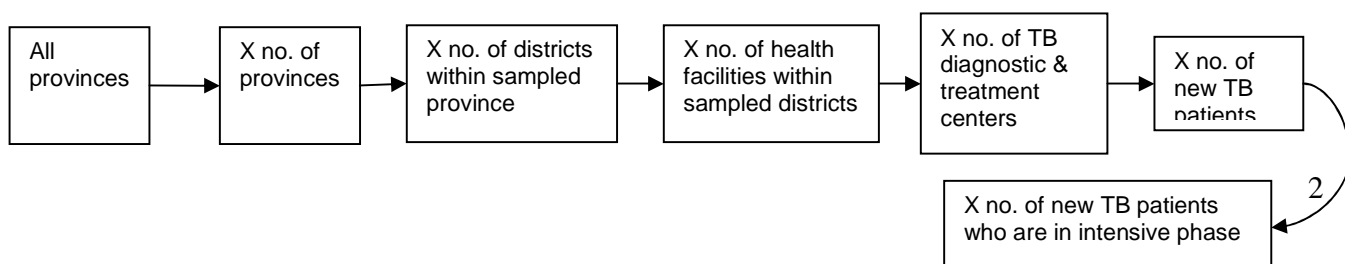
This survey will provide information on a subset of a population and, if appropriately selected, can provide the basis for making inferences about the population. The sample is selected from a sampling frame, which is the total of the sampling units. Advice on sampling may be sought from the national statistics office.

Points to consider when estimating sample size are:

- main indicators to be measured
- number of sites / mean costs
- hypothesized proportions
- population to be studied (new / retreatment / MDR patients / defaulters etc.)

### **Possible Scenario:**

The Tool is used to survey only new patients with at least one month of treatment completed. TB patients will be interviewed at health facilities. The unit of analysis is the TB patient; sampling units are health facilities; secondary units are districts and primary units are provinces. The sampling frame is the TB register of a diagnostic and treatment facility. A small survey to generate baseline data without large sample sizes is conducted. One province is purposively selected. Within this province, 5 districts are sampled, and in each district there are 16 diagnostic and treatment facilities, a total of 80 facilities. If we want to interview 200 patients, 50 per facility, we need to sample 4 facilities. Statistical software programs such as SPSS or STATA can help generate a sample.



### **Purposive Sampling**

It is conceivable to draw a purposive / convenience sample if a certain district or province is of particular interest. Inferences about the studied population are then however limited to this particular district/province. An initial decision should be made as to whether the aim of the survey is to establish a representative result/baseline or to get a basic idea about the current situation in a certain area. Representativeness of a sample size can be assessed through the total number of TB patients at a point in time (not the number of new cases!).

Patients can be sampled as they come in on clinic days, until a certain number is reached, or they can be sampled by going through the register and taking a random sample. Depending on the size of the health facility and the number of patients on treatment, the first method might be more practicable in small clinics, whereas the latter might be easier in big clinics. The type of clinic days need to be taken into account to avoid bias. For example, if those patients in intensive phase might attend the health facility on different days than those in continuation phase. Patients can be also interviewed at home, but this may introduce problems of confidentiality and social desirability bias<sup>1</sup>. Our questionnaire is flexible about which methods are chosen.

### **Non-response bias**

When patients are interviewed, it is important to record non-responses to minimize bias in data interpretation, including those who refused an interview and those who cannot be interviewed because of language barriers (such as migrants or minorities).

## **3. Target group, inclusion and exclusion criteria**

The thoughts below may help in defining inclusion and exclusion criteria before the questionnaire is adapted to local circumstances and according to the objectives of the survey.

- **Timing of interview:** The questionnaire is designed to interview TB patients who have completed at least one month of TB treatment. A good balance between recall bias and cost experience needs to be found.
- **Target group:** In assessing the financial burden of poor patients, the target group could be narrowed down to poor TB patients, targeted through purposively sampling poor districts. Since costs due to an HIV co-infection are to be taken into account and can be significant, the sampling of the patients may take the prevalence of HIV in the region into account. Here, purposive sampling may also be applied. The questionnaire in its generic form is tailored towards new patients, but it can be adapted to interview re-treatment and even MDR patients. It is recommended to choose only one or two of these sub-categories of TB patients, in order limit variance in the data and produce meaningful results.
- **Place of interview:** Patients can be either interviewed at the health facility or at home. Both have advantages and disadvantages. Our generic questionnaire leaves this open, as well as how far the patient has advanced into treatment.

## **4. Examples of small surveys on patient costs**

Jackson et al (2006). *Poverty and the economic effects of TB in rural China*. Int J Tuberc Lung Dis 10(10):1104-1110.

---

<sup>1</sup> **Social desirability bias** is a term used in scientific research to describe the tendency of respondents to reply in a manner that will be viewed favorably by others. This will generally take the form of overreporting "good" behavior or underreporting "bad" behavior.

Four counties were selected on the grounds of being part of a World Bank financed project. Cases were chosen from economically productive age group (25-60), restricted to new cases until the target of 40 patients per county was reached (160 cases).

Kemp et al (2007). *Can Malawi's poor afford free tuberculosis services? Patient and household costs associated with a tuberculosis diagnosis in Lilongwe*. Bulletin of the World Health Organization 85(5), 580-587.

Nhlema Simwaka et al (2007). *Developing a socio-economic measure to monitor access to tuberculosis services in urban Lilongwe, Malawi*. Int J Tuberc Lung Dis 11(1): 65-71.

"Five health centers in urban Lilongwe were selected where there is the highest burden of TB cases in Malawi. Only patients in the intensive phase of treatment were selected above 16 years of age. At each treatment centre a list of new pulmonary TB patients in the intensive phase of treatment was drawn up in chronological order and numbered. The number of patients systematically sampled from this list was proportional to the total number of TB patients in the intensive treatment phase at the centre."

"For the cross-sectional survey, a sample size of 179 TB patients was calculated based on the assumption that the poverty rate within urban Lilongwe is 38%, that the rate of poverty among TB patients is at least 15% higher than in the general population. Confidence interval 95%. Treatment registers of five urban DOT centers were used to identify every third new adult patient in the intensive phase of treatment."

Kamolratanakul et al (1999). *Economic impact of tuberculosis at the household level*. Int J tuberc Lung Dis 3(7), 596-602.

"Socio-economic data for tuberculosis patients were collected in a cross-sectional survey performed at 16 health care facilities in Thailand in 1996/97. The study was conducted at two district hospitals, one provincial hospital and one referral center within each of the country's four administrative regions. The study sites for each region were determined by random sampling. At each study site, a cluster sample of all adult tuberculosis patients who completed treatment between August 96 and February 97 was subjected to in-depth interviews using a structured questionnaire." The total sample size was 687 patients. The distribution of these 673 patients grossly corresponded to the distribution of the general population.

## 5. Coding

Each question in the questionnaire needs to receive a code. This code serves to give a name to the variable which is measured and will be used to identify the question during data analysis. The version of the questionnaire for the analyzer should include these codes. Following are examples of codes:

<b>Where do you take your medicine?</b>	1. Health facility / hospital 2. Home 3. Community 4. Workplace	<b>DOTHC</b> <b>DOTHOME</b> <b>DOTCOM</b> <b>DOTWP</b>
---	--	---

How often do you travel to the health facility / hospital for taking your medicine? <b>NRVISITDOT</b>		Times / week
How much does one of these visits cost on average? <b>COSTVISITDOT</b>		
1. transport	2. food	3. fees
4. tests: sputum:	xray:	other tests:
5. TB medicines	6. Other medicines	7. Food supplements
		8. Other
Total:		
Have you ever stopped working due to TB? <b>STOPWORKTB</b>		1. Yes      2. No
If yes, for how long? <b>TIMESTOPWORKTB</b>		1. one month    2. 2-3 months    3. 4-5 months 4. more than 6

## 6. Training of Interviewers – important issues

- a. Depending on how far the patient has progressed with treatment, it might be difficult for him/her to recall cost items. The interviewer should make it as easy as possible for the patient to recall by using local methods of time structuring; Interviewers should be given examples how to prompt responses regarding time and types of costs.
- b. Interviewers need to be instructed about indicator definitions, such as types of costs, what is meant by cost of food, cost of travel and cost of accommodation, what is included and what is excluded and how they can help patients recalling items by prompting. This will help to ensure consistency in interviews and prompting by interviewers.
- c. Interviewers need to be sensitized on the different phases (intensive, continuation) and types of TB treatment (hospitalization, different forms of DOT) and associated costs (sputum conversion test, follow up test, medicine collection etc.), to avoid double counting costs. It also needs to be clear to the interviewers what counts as TB drugs and what are additional drugs that are prescribed/bought.
- d. Interviewers should be familiar with the concept of guardian costs; a guardian is someone who accompanies the patient to the health facility/hospital or other visits because the patient cannot go by himself. The guardian incurs direct and indirect costs. If the guardian lives in the same household as the patient, the cost to the household is much higher if guardian costs are included.
- e. Difficult concepts such as willingness to pay, coping and guardian costs need to be explained in detail with clear instructions on the intent behind the questions to ensure that interviewers are able to explain these concepts to patients.
- i. It may be complicated for some patients to address questions that refer to their situation “prior to their diagnosis” vs. “post diagnosis” as they may not be clear when or where their diagnosis was made. In addition there may be other big changes in their life that coincide with their illness. It is important to teach interviewers how to tease apart issues (if any) so that changes in income etc. are truly caused by their TB, and not because of some other unrelated event.
- f. Interviewers should be informed about the nature of TB, what their participation means for their own health and how they can protect themselves. Depending on which kind of patients are interviewed (new, re-treatment or MDR patients), and how far the patient is into treatment, risks to

interviewer health differ. For example, patients who are in their first month of treatment might still be infectious. The interviewer needs to be aware of that and knowledgeable about infection control measures; i.e. conducting the interview outside or in a well-ventilated room.

## **References:**

Cochran (1977). *Sampling Techniques*. New York: John Wiley & Sons.

Meier, Brudney & Bohte (2008). *Applied Statistics for Public and Nonprofit Administration*. Boston, MA: Wadsworth Publishing.

World Bank (2007). *Analyzing Health Equity using Household Survey Data*. Available under <http://siteresources.worldbank.org/INTPAH/Resources/Publications/459843-1195594469249/HealthEquityFINAL.pdf>