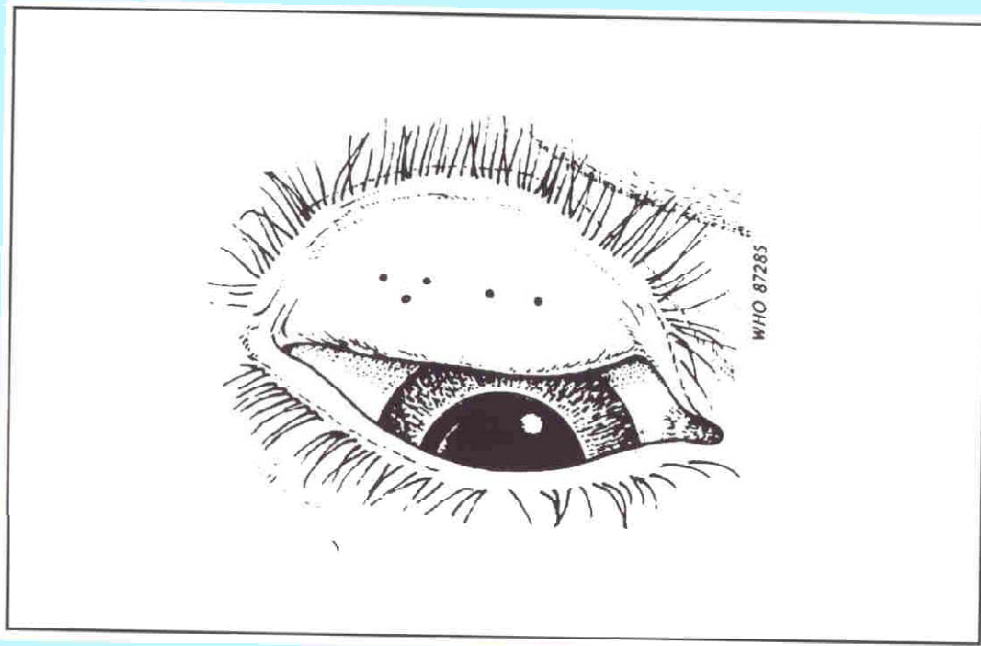# World Health Organization

TRACHOMA EPIDEMIOLOGIC SURVEY PROTOCOL

WHO 87285

14. THE PREVALENCE SURVEY

Introduction

This second section tells how to gather sufficient data for preparation and evaluation of a trachoma control component within a national programme for prevention of blindness. It is not a detailed description of a rigorous scientific survey on trachoma; a statistician/ epidemiologist should be consulted for that purpose.

Before a trachoma control programme is initiated, all the necessary data must be available. Such data can be obtained either from hospital records or from the records of mobile units, or by interviewing health personnel working at the peripheral level. Yet data on clinical consultations (number of cases of acute follicular trachoma, number of surgical operations on the eyelid) lead to underestimation of the scope of the problem; many sufferers do not have recourse to health services and are therefore never counted. This is why a small-scale **survey** in the general population is often necessary.

A. Principles of a prevalence survey

The most accurate way of finding out the trachoma situation in a community is to undertake a prevalence survey. This means that the proportion of cases of disease and complications is assessed at a given moment (cross-sectional) in terms of **prevalence**, in relation to the total examined population under study.

1. It is important that there be a clear **objective/purpose** of the survey. In most instances it will be to find out the prevalence of active inflammatory disease (TF and/or TI) or ensuing complications (TT). These are the variables that will determine the need for medication or surgical treatment.

2. There should be a well-defined **target population** for the survey. As a general rule for simplified trachoma surveys, it is recommended to consider only **rural** populations, and only communities with less than 5000 inhabitants. Those are the settings where trachoma is most likely to be a significant cause of visual loss.

3.   Most surveys require a **sampling** of the population concerned, i.e., only a portion, determined according to a defined procedure, will be examined, in order to save on work and cost. The common principle in all samples is **randomization**, implying that villages, households or individuals are selected at random, to be representative of the whole population concerned.

The ideal is to have a simple random sample, drawn from a list of all the people in the study area. This is not practical, however, and groups of households or individuals are therefore often examined, which makes the survey work easier and more effective. Each group is called a **cluster**.

When cluster sampling is used, the sample is determined in two or more steps (multistage). For the present study, the first step will be to sample **villages**. The second step is to select **households** in the chosen villages. All the individuals belonging to the identified households should then be included in the sample.

4.   For trachoma assessment, it is usually enough to estimate the prevalence of disease and complications in the total population of administrative or health care units; in the present manual, the assessment refers to the **district** level.

B.   Selecting the sample

The choice of the sample is governed by certain conditions:

1.   REPRESENTATIVENESS

The villages or communities selected must faithfully reflect (represent) all the villages or communities in the survey zone. This means that, if communities are chosen because they are near a dispensary and easily accessible, the survey will not show the real epidemiological situation of trachoma in the study area.

The people to be examined in each village or neighbourhood should be representative of the local poulation. If the village is very small, all inhabitants should be examined. Otherwise, people should be chosen from all social classes and from all parts of the village.

The results of the survey can be extrapolated to the whole poulation of the area **only if the sample is deemed representative of the population from which it is drawn**.

2. THE SAMPLING FRAME

This refers to the setting and information needed to be able to select correctly a sample in a given area.

A **census** for the district is needed, listing all **villages** with their populations. If the census is old (more than five years), it needs to be updated, making use of more recent information as far as possible (recent health surveys, school records, population increases, etc.). A recent map of the entire district, with the location of each village, is very useful.

Sometimes the census enumeration areas combine several villages and, therefore, information on the population of each village is not available. In this case, census enumeration areas should be used in place of villages.

A complete **listing of all households** in selected villages may be found from a recent census or other registration system. These data may also be available from other sources: for example, immunization coverage or agricultural development programmes. Detailed maps of the area are also useful. Care should be taken to be sure that the listing of households is complete and includes all sections of the village. If a complete listing is not available, it will be necessary to create one by walking through the village and writing down each household.

3. SAMPLE SIZE

Trachoma is an infectious disease, and it will thus be more common in certain families, neighbourhoods and villages; therefore, individuals within selected clusters tend to resemble each other and present a certain homogeneity. This homogeneity should be taken into consideration when determining the sample size.[1]

---

[1] Thus, if it is intended to examine more people per village, the sample size should be further increased to compensate for the "cluster effect". The rate of homogeneity (roh) is the statistic used to measure homogeneity.

The size of the total sample to be selected will depend on the "expected" prevalence of the sign in question, on the degree of accuracy required of the estimate, and on the number of individuals to be examined in each cluster.

The larger the acceptable confidence interval, the smaller the sample size. Thus, it may be helpful to confine the examinations to groups at high risk for trachoma. For example, it may be decided to confine the sample for the estimate of TF and/or TI to children under the age of 10. This group could be expected to have more active, inflammatory cases. For assessing signs of trichiasis or corneal opacity, on the other hand, the sample could be confined to women over the age of 15. The "key" prevalences for operational decisions are: TF >20% and TI >5% in the 0-9 years age group; TT >1% in women over 15 years of age (see further pages 39-40).

Example

It should be noted that the following section is based on certain assumptions and simplified approaches. This model gives reasonably accurate estimates, as a basis for operational decisions; but if a more sophisticated or large-scale assessment is planned, a statistician/epidemiologist should be consulted for a detailed sampling.

```
╔═══════════════╗
║    Stage 1    ║
╚═══════════════╝
```

Choice of villages from which the clusters are selected. First of all, a list must be made of all the villages or census enumeration areas in the survey area, giving the precise or estimated number of inhabitants of each (the sampling frame). Communities are selected with "probability proportional to size", so that communities of equal population will have the same chance of being chosen to provide a cluster, while communities with larger populations will have a proportionally greater chance.[1] An example of this sampling procedure is given below; it also takes into account the grouping of villages into categories of known or suspected trachoma-endemic communities (implicit stratification).

---

[1] Only villages with a population of less than 5000 inhabitants may be considered.

District "X" contains 60 villages with a total population of 100 000. A sample of 4000 persons composed of 20 clusters should be drawn from the total population. This is based on the following considerations:

(i)   To identify the least common sign, TT, with reasonable accuracy (see section D), at least 50 women aged 15 years or more should be examined in each village (cluster).

(ii) At least 20 clusters should be included in the sample (see section D).

(iii) A field team may easily examine 200 persons/day.

> **Step 1**

Prepare a complete list of villages in the district, with the known or estimated pouplation in each village. This is the sampling frame.

| SAMPLING FRAME ||
Village	Population
**1	4000
*2	1000
3	1200
*4	450
|	|
|	|
|	|
58	13 000
**59	700
60	7000

Thus: 60 villages, 100 000 population

Eliminate villages of more than 5000 population, for simplicity of field work.

Step 2

Now **group** the villages into **three** categories of known or estimated endemicity. Mark this in the above table as follows:

(i) Villages with **STRONG** evidence of severe disease, i.e., previous surveys, visiting health teams' reports, etc.: *mark these villages with two asterisks, or similar, next to their number.*

(ii) **SOME** evidence of trachoma, but probably little or mild disease, as judged by anecdotal or few patients' evidence: *mark these villages with one asterisk, next to their number.*

(iii) **NO** evidence of trachoma in the village: *no mark needed.*

Step 3

To identify the 20 clusters needed, a table is set up (see below):

- First the category of village endemicity (strong, some, none, as from the previous table)

- Column 1 with the identification of each village

- Column 2 with the total population of that village

- Column 3 with cumulative population, adding each village population to the previous ones

- Column 4 will contain the identified clusters, as per instructions.

Category	1 Locality	2 Population	3 Cumulative population	4 Identified clusters
STRONG evidence of prevalent and severe trachoma	1 5 11 13 17 26 30 31 39 43 48 54 59	4 000 1 000 1 500 4 500 600 2 100 500 1 000 1 600 350 650 3 000 4 900	4 000 5 000 6 500 11 000 11 600 13 700 14 200 15 200 16 800 17 150 17 800 20 800 25 700	3 392 = Cluster 1 7 392 = Cluster 2 11 392 = Cluster 3 15 392 = Cluster 4 19 392 = Cluster 5 23 392 = Cluster 6
SOME evidence of trachoma but little or mild disease	2 4 etc. 24 villages	1 000 450 etc.	etc.	Clusters 7-19 (75 392 = Cluster 19)
NO evidence of trachoma	3 7 14 etc. 21 villages	1 200 1 100 700 800	76 200 77 400 78 500 79 200 80 000	79 392 = Cluster 20
Random number = 3392 Sampling interval = 4000				

NOTE: *Two villages of 13 000 and 7000 inhabitants were excluded (Nos 58 and 60).*

Step 4

Calculating the sampling interval

Divide the total cumulative population (80 000) by the requisite number of clusters (20) to calculate the sampling interval (SI):

$$SI = \frac{80\ 000}{20} = 4000$$

Step 5

Drawing a random number and identification of the first cluster

Use a random numbers table or a pocket calculator with a random numbers function (or the serial numbers of bank notes) to choose a random number between 1 and the sampling interval (4000). For this example, the number is 3392. It can be seen from the "Cumulative population" column that 3392 is less than 4000, the population of the first village. The first cluster is therefore taken from the first village. Report it in column 4 of the table.

Step 6

Identification of remaining clusters

To find out from which village the second cluster should be taken, add the sampling interval (4000) to the number chosen at random (3392); this makes 7392, which falls between 6500 and 11 000 in the "Cumulated population" column; the second cluster therefore comes from village 13. Repeat this operation to identify the other villages. Two or more clusters could come from a single community if its population were greater than the sampling interval (SI).

This **first** stage of sampling identifies the **villages** in which a cluster of households should be examined.

Stage 2

Selection within those villages of the households to be included in the sample. The cluster consists of **all** the members of these households. This stage consists of defining the clusters and identifying the people in them. From the **full** list of households in a community, certain households are selected to provide the requisite number of subjects. The examiner does not begin his work until **all** the households are identified and numbered, and all their members are registered. Use a **local census** listing and **detailed map** for numbering and choosing the households for examination.

Step 7

Determine the number of households constituting each cluster.

Step 8

At least 50 women aged 15 years or more must be examined in each village to be able to tell if the prevalence of trichiasis is more than 1% (see further section D).

Assuming that women constitute 50% of the total population, and that 50% of women are above 15 years of age, there is a need to examine 200 persons (both sexes and all ages) in each village cluster. This will give the required precision of the prevalence of trichiasis.

Step 9

If we estimate the average size of a household to be eight people, with an average of two women aged 15 years or more, we need to examine, on average:

$$\frac{200}{8} = 25 \text{ households per village}$$

Step 10

Check the total number of households in the selected village cluster.

Then **divide** by the number of estimated needed households (in this case 25). For example:

300 households in the village

Divide by 25 = 12

Thus, **every twelfth household** throughout the village should be examined, with all its members.

Step 11

Decide on the most practical way to select and examine each twelfth household, making use of lists and maps. Steps 4-6 can be applied in principle, i.e., the sampling interval is 12 in this case (300/25) and a random number between 1 and 12 can be drawn to identify the first household on a numbered list of households or a map. The subsequent households are then identified as in Step 6.

Step 12

Define and **follow strictly** rules for:

- abandoned households (another household should be selected at random);

- revisits, if people are absent;

- additional households, to be selected if needed (there must be a minimum of 50 women examined in each cluster of households).

C. Conducting the survey: resources and costs

This kind of fast, simple survey is not usually very expensive. The main expenditure is on transport, equipment and supplies. Additional funds should be allocated to wages and data analysis.

1. TRANSPORT

Transport is needed to ensure access to each of the clusters selected. In many regions, transport is a big problem. If the households in the cluster are scattered, some form of transport must be provided for travel in the cluster (one or more visits to each household). The examinations can be conducted either when the households have been brought together or in the course of home visits. Whereas it is easier to find people in their homes, such an approach takes longer. Either way, when the population is very scattered, it is better to make home visits even in large villages, where only a few (selected) households have to be examined.

2. EQUIPMENT AND SUPPLIES

The trachoma examiner and the team will need simple equipment (loupes and torches) and drugs (tetracycline eye ointment). During the survey, the examiner may have to treat routine medical problems; he should therefore bring antibiotics, antimalarials, aspirin, etc. For systematic gathering of standardized data, the examiner should have a sufficient number of **record forms**. A simple and practical specimen is shown below.

SPECIMEN FORM FOR REGISTRATION OF DATA
FOR THE EPIDEMIOLOGICAL SURVEY OF TRACHOMA

EPIDEMIOLOGICAL SURVEY OF TRACHOMA

Examiner: Mazengo   **Village:** Makawa
Subject examined: Judith Oh   **Household no.:** 31
                                         Sex: F

ENTER ONE OF THE TWO FOLLOWING SYMBOLS FOR EACH SIGN

0 = Absent
1 = Present

	TT	CO	TF	TI	TS
Right eye	0	0	0	0	1
Left eye	0	0	0	0	1

---

3. THE TEAM

The personnel required for field work may simply be the trachoma examiner and an assistant. On occasion, there might be an advance team that would undertake the mapping and census work. The examiner could be accompanied by a driver and one or two assistants who would help with the examinations and registration of results. The assistants could be recruited in the area and could help locate the households. There should also be clerical support for tabulating (or coding) the initial forms and processing the data.

4.   THE TIMETABLE

Another factor that must be considered is the time involved. This depends on the number of health workers operating in the field, the size of the sample, and local geography.

5.   ANALYSIS

Tabulation and "manual" analysis of data need not cost much, setting aside staff costs. If data are processed by computer, then the cost of data-processing must be taken into account.

D.   Data analysis

1.   After the survey, the form must be put together, checked for completeness, and analysed. First, consider the **attendance**, e.g. how many people were actually examined out of all those listed?

Example:

It was envisaged to examine a sample of 900 people in 30 clusters. In fact, only 800 of the total census population of 900 were actually examined.

Thus, attendance was (%): $\frac{800}{900}$ x 100 = 88.9%

As a rule, an attendance rate of **more than 85%** is considered satisfactory. For a more detailed analysis, the attendance for each age group and by sex may be calculated.

2.   To be able to consider the **representativity** of the examined sample, a comparison by age group and sex must be made to the known population structure ("age pyramid") in the area.

3. **Prevalence** is the easiest indicator to calculate.

Example:

If 200 people are examined and 10 have evidence of TT in one or both eyes, the prevalence of trichiasis is (%):

$$\frac{10}{200} \times 100 = 5\%$$

It is always useful to estimate specific prevalence by age, village and district of active trachoma and of trichiasis, as this implies the need for treatment to be provided. For the prevalence of "active trachoma", both TF and TI should be considered, in particular as to the possible need for mass ("blanket") treatment of a community.

As a routine data analysis procedure, proceed as follows:

1. Calculate the prevalence of TF and/or TI by age group, as shown in Table II (page 42). This is referred to as "**active trachoma**"; check the rate for children ≤ 10 years of age and compare to the recommended strategies for treatment (page 16).

2. Now look at the prevalence of TI by the same age groups. Again verify the rate for children ≤ 10 years of age, and compare to the recommendations for treatment (page 17). This is referred to as "**intense trachoma**".

3. Finally, calculate the prevalence of TT in the age group ≥ 15 years; by far, most cases of trichiasis will be in that age group. It may therefore be useful to refer to "**trichiasis**" by age group, for example "2% of the population above 15 years". This facilitates the estimation of surgery needed in other areas, where the composition of population age may be different.

NOTE: *If the above model for sampling is applied, the 95% confidence interval implies the following:*[1]

- *To be 95% sure that the district prevalence rate of TT in women is more than 1%, the "survey prevalence" found in women more than 15 years old should actually be 2.5%.*

- *To be 95% sure that the district prevalence rate of TI in children 0-9 years is more than 5%, the "survey prevalence" should be at least 8%.*

- *To be 95% sure that the district prevalence rate of TF in children 0-9 years is more than 20%, the "survey prevalence" should be at least 25%.*

In addition to the above three points, it is often useful to calculate the TS rate by age group. This may be referred to as **"conjunctival scarring"**, which in children ≤10 years of age gives an idea as to the severity of trachoma in the population and the future need for trichiasis surgery. However, this "scarring" is more useful in estimating trends over time, to measure a **change** in the disease, as not all children with TS will develop trichiasis.

The prevalence of CO, finally, gives an indication of how much visual loss may be caused by trachoma, and it is again useful to calculate this by age group. This may be referred to as **"corneal opacity"**; it can be used to measure change over time of the disease. However, it is a long-term indicator, as the overall CO rate will change very slowly, but there should gradually be less of it, after a few years, if the prevention and treatment of trachoma have been successful in an area.

There is often a difference between **males** and **females** with regard to trachomatous inflammation and complications; usually, females tend to be more affected. The data analysis can, of course, be done also by sex in addition to age, but it is less imperative unless there are very marked differences leading to a change in treatment strategy.

---

[1] This is based on the assumption that at least 50 women aged 15 and over are examined in each of 20 villages. The assumed rate of homogeneity (roh) was 0.8 for TF and TI and 0.4 for TT. In addition, to calculate the finite correction factor it was assumed that there were 60 villages in the district.

NOTE: *The overall prevalence of trachoma is often referred to. It is generally **not** a useful measure of the disease, as it does not tell anything about present or past intensity or complications of trachoma.*

Of course, computer processing of the data makes statistical analysis easier. If no computer is available, then simple tables filled in by hand give information which is just as good on the scale and severity of trachoma. Table II is useful for a detailed analysis of all signs of trachoma by age group and sex. For more detailed analysis, an epidemiologist/statistician should be called in, or reference books on calculation of other variables should be consulted.

TABLE II. SUGGESTED TABLE FOR OVERVIEW OF DISTRIBUTION OF THE KEY SIGNS OF TRACHOMA, BY AGE AND SEX

Age / Signs of trachoma	0-9			10-14			15+			Total		
	Male N(%)	Female N(%)	Total N(%)	Male N(%)	Female N(%)	Total N(%)	Male N(%)	Female N(%)	Total N(%)	Male N(%)	Female N(%)	Total N(%)
Active trachoma (TF+TI)	62(31.6)	67(32.8)	129(32.3)	14(18.2)	17(23.3)	31(20.7)	11(10.0)	18(12.8)	29(11.6)	187(22.7)	102(21.5)	189(23.6)
Intense trachoma inflammation (TI)	19(9.7)	21(10.3)	40(10.0)	2(2.6)	4(5.5)	6(4.0)	0(0.0)	5(3.6)	5(2.0)	21(5.5)	30(7.2)	51(6.4)
Trachomatous trichiasis (TT)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	4(3.6)	16(11.4)	20(8.0)	4(1.0)	16(3.8)	20(2.5)
Trachomatous scarring (TS)	14(7.1)	6(2.9)	20(5.0)	13(16.9)	17(23.3)	30(20.0)	44(40.0)	47(33.6)	91(36.4)	71(18.5)	70(16.8)	141(17.6)
Corneal opacity (CO)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	2(1.8)	10(7.1)	12(4.8)	2(0.5)	10(2.4)	12(1.5)
Trachoma, total	78(39.8)	80(39.2)	158(39.5)	23(29.9)	32(43.8)	55(36.7)	59(53.6)	65(46.4)	124(49.6)	160(41.8)	177(42.4)	337(42.1)
Examined (denominators)	196	204	400	77	73	150	110	140	250	383	417	800