



WORLD HEALTH ORGANIZATION
ORGANISATION MONDIALE DE LA SANTE

DISTR.: LIMITED
DISTR.: LIMITEE
WHO/TB/94.180

52/21

GUIDELINES FOR HIV SURVEILLANCE AMONG TUBERCULOSIS PATIENTS

**Tuberculosis Programme
World Health Organization, Geneva, Switzerland**

and

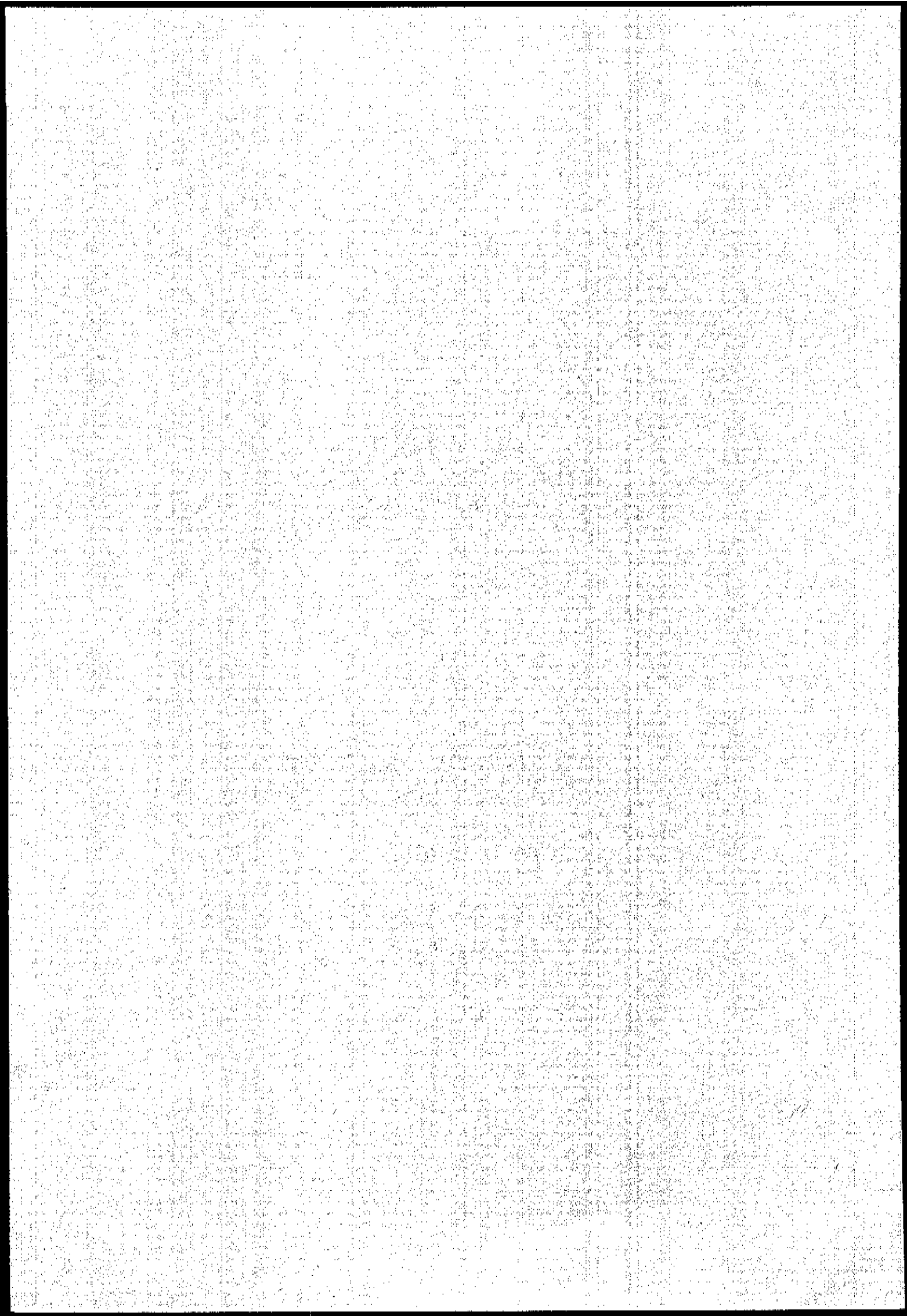
**International Union Against Tuberculosis
and Lung Disease, Paris, France**

This document is not issued to the general public, and all rights are reserved by the World Health Organization (WHO). The document may not be reviewed, abstracted, quoted, reproduced or translated, in part or in whole, without the prior written permission of WHO. No part of this document may be stored in a retrieval system or transmitted in any form or by any means - electronic, mechanical or other - without the prior written permission of WHO.

The views expressed in documents by named authors are solely the responsibility of those authors.

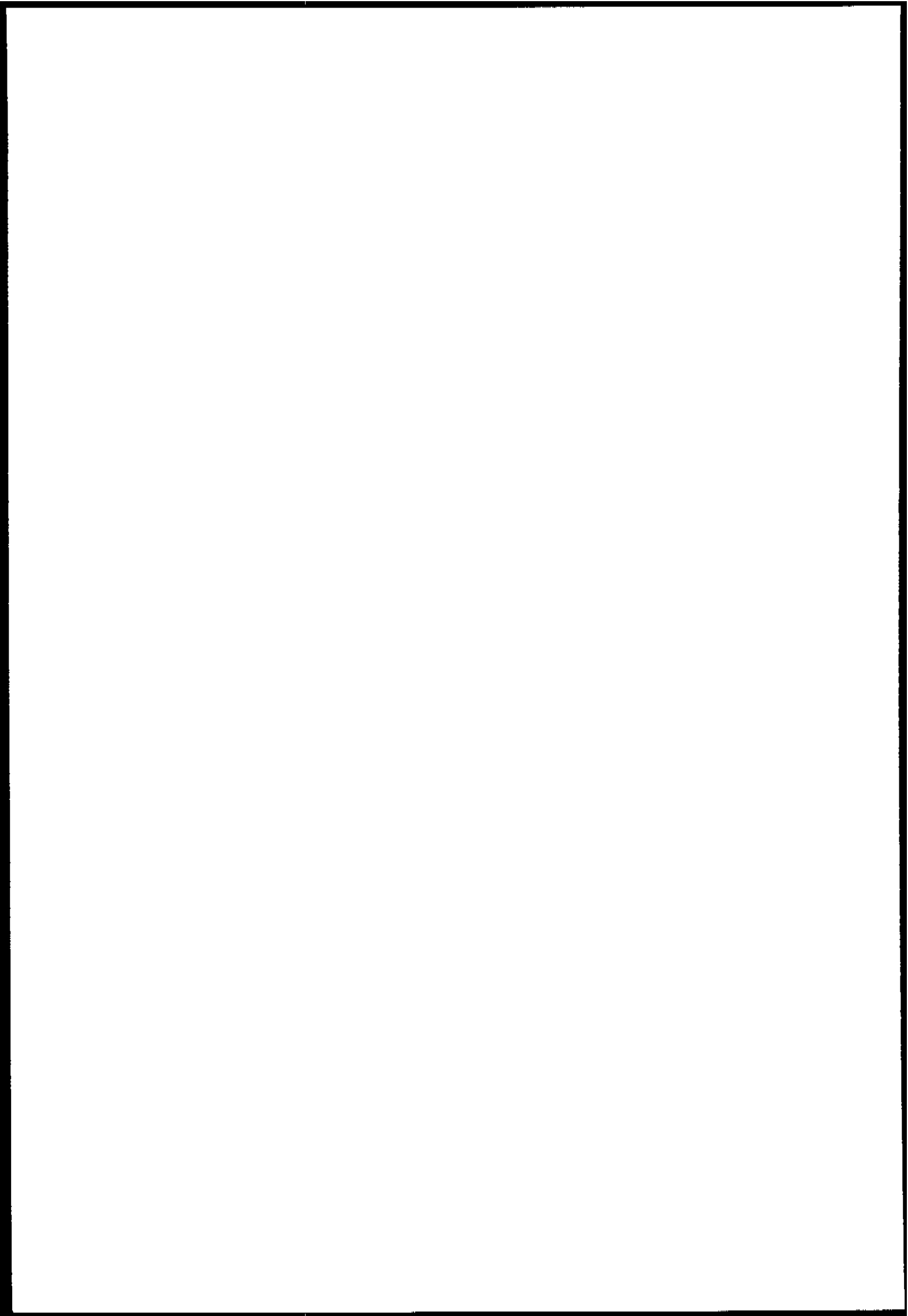
Ce document n'est pas destiné à être distribué au grand public et tous les droits y afférents sont réservés par l'Organisation mondiale de la Santé (OMS). Il ne peut être commenté, résumé, cité, reproduit ou traduit, partiellement ou en totalité, sans une autorisation préalable écrite de l'OMS. Aucune partie ne doit être chargée dans un système de recherche documentaire ou diffusée sous quelque forme ou par quelque moyen que ce soit - électronique, mécanique, ou autre - sans une autorisation préalable écrite de l'OMS.

Les opinions exprimées dans les documents par des auteurs cités nommément n'engagent que lesdits auteurs.



CONTENTS

	Page
1. Background and rationale	3
2. Objectives	4
3. Study population	4
3.1 Eligibility criteria	4
3.2 Exclusion criteria	4
4. Sampling	5
4.1 Sample size	5
4.2 Sampling procedure	6
4.2.1 Representative sampling	6
4.2.2 Sentinel surveillance	7
5. Operational procedures	7
5.1 Data and specimen collection	7
5.2 Serological test	8
6. Data management and analysis	9
7. Recruitment and training of personnel	9
8. Time frame	9
9. Budget	9
10. References	11
Annex 1 - Sample size calculation	
Annex 2 - Cluster sampling	
Annex 3 - Data collection form	
Annex 4 - Attributable risk calculation	



1. Background and Rationale

The pandemic of infection with human immunodeficiency virus (HIV) is having a profound medical, social and economic impact, particularly in developing countries. Among HIV-infected adults, particularly in Africa, tuberculosis is one of the most common opportunistic diseases (1-3). The World Health Organization estimates that globally over 5 million persons have been infected with both HIV and *Mycobacterium tuberculosis* since the beginning of the HIV pandemic, with over 3.7 million in sub-Saharan Africa. These individuals have a 5-10% annual risk (2,4) and an estimated life-time risk of developing clinical tuberculosis of 30% or more (2,5). The result of this increased risk is evident from rapid increases in the incidence of tuberculosis, coincident with the spread of the HIV epidemic in populations where the prevalence of both infections is high (2). In many sub-Saharan African countries, where perhaps one third to one half of the adult population is infected with *M. tuberculosis*, the number of pulmonary tuberculosis cases has more than doubled over the past 5 years (2,6). Similar increases may be expected in countries in Asia where tuberculosis is already a major health problem and where HIV is making rapid inroads.

On the one hand, tuberculosis patients may be selected as one of the sentinel populations for monitoring HIV infection levels and trends in a country or region. On the other hand, it is important for the national tuberculosis control programme to determine and to monitor over time the prevalence of HIV infection in tuberculosis patients, since it is a direct measure of the impact of the HIV epidemic on the tuberculosis problem. It is a useful tool for evaluating the current situation and for predicting future changes in tuberculosis incidence. Data obtained from this HIV surveillance may be used by the programme to guide the choice or changes in therapeutic regimens, to improve management of HIV-infected tuberculosis patients and to organize AIDS information/education aimed at reducing HIV transmission for patients and staff. Where local representativeness can be ensured by the sampling procedure, data may be used to target interventions in those areas of the country most affected by the HIV epidemic.

For surveillance purposes, it is essential to obtain unbiased estimates of HIV prevalence. This is best achieved through unlinked anonymous testing, which consists of testing sera routinely collected for other purposes after having removed all personal identifiers so that test results cannot be linked with the identity of the person tested. Since the method does not require individual patient's consent, it minimizes the participation bias which would be introduced by patients refusing to be tested because they perceive themselves to be at risk of HIV infection. Therefore, unlinked anonymous testing is the recommended procedure to monitor HIV prevalence, provided ethical considerations are carefully taken into account. Since patients are not informed of their test results and thus cannot be appropriately counselled, unlinked anonymous HIV testing cannot achieve the objectives of voluntary confidential HIV testing and counselling. It must therefore be established whether conducting unlinked anonymous HIV surveillance is a priority, particularly in places where counselling services are not available and resources are scarce. In situations where voluntary HIV testing and counselling are offered to patients, sufficient resources must be available to conduct unlinked anonymous testing parallel to voluntary confidential testing and counselling. In all situations where removal of personal identifiers cannot be absolutely ensured, unlinked anonymous testing of tuberculosis patients is not recommended.

2. Objectives

The overall objective of this protocol is to establish standardized procedures for conducting HIV surveillance among tuberculosis patients through unlinked anonymous HIV testing.

The specific objectives of the HIV surveillance are:

- i) to determine HIV prevalence rates among persons with tuberculosis;
- ii) to monitor trends in HIV prevalence over time among persons with tuberculosis.

Depending on the method used, additional objectives may be pursued:

- iii) to monitor HIV prevalence rates separately by age, sex, site of disease or bacteriological status (if the sample size is sufficient);
- iv) to monitor HIV prevalence rates separately by area of residence (if the sample is stratified by area of residence);
- v) to estimate the proportion of tuberculosis cases that could be attributable to HIV infection (this requires the selection of a suitable comparison group that can be considered representative of the general population; see Annex 4).

3. Study population

3.1. Eligibility criteria

- a) Patients with a new diagnosis of tuberculosis (*i.e.*, patients who have never had treatment for tuberculosis, or who have taken antituberculosis drugs for less than four weeks);
- b) Age between 15 and 59;
- c) Bacteriological status: patients with smear-positive pulmonary tuberculosis have a confirmed diagnosis. Thus, they provide the most internationally comparable study population. One option is to include all patients with tuberculosis, whether pulmonary (smear-positive or smear-negative) or extrapulmonary (including meningeal and disseminated tuberculosis), *provided sputum smear-positive patients are identified as such and that results are analyzed separately*. Another option is to include only smear-positive cases.

3.2. Exclusion criteria

- a) relapse cases, *i.e.*, cases previously notified as cured and who are subsequently diagnosed with active disease. *However, these relapse cases may be included provided they are identified as such and that results are analyzed separately*;
- b) other cases, such as patients failing treatment (smear positive patients who remained, or became again, smear positive five months or later after commencing treatment),

returning defaulters (patients who return to treatment after having interrupted treatment for two months or more) or others.

4. Sampling

4.1. Sample size

The sample size necessary to estimate HIV seroprevalence among tuberculosis patients should be calculated so that the obtained result of the study may be within x percentage points of the true value with 95% confidence.

In order to select a minimum sample of tuberculosis patients for study, the following information is necessary:

- 1) The anticipated population proportion (P), *i.e.*, estimated HIV seroprevalence (expressed as a fraction) in tuberculosis patients;
- 2) A confidence level (generally 95%);
- 3) The desirable precision of the estimate (d), *i.e.*, the confidence interval.

The sample size can be calculated by the following formula : $N = 3.84 P(1-P) / d^2$ (see also Annex 1). This formula is valid for simple random sampling or sentinel surveillance.

If the cluster sampling technique is used (see 4.2.1), the sample size calculated above should be multiplied by a number F. This is because patients belonging to the same cluster have a high probability of having common characteristics. This may be the case particularly for HIV infection: patients diagnosed in the same urban hospital may have a higher probability of being HIV-infected than patients diagnosed in a rural centre. This results in less precise estimates than those which would result from simple random sampling with the same number of patients. The number F increases with the heterogeneity of HIV prevalence and decreases with an increasing number of clusters. If 30 clusters are used, it is reasonable to multiply the size calculated with the above formula by 2 to obtain the sample size for cluster sampling.

To monitor trends over time, the sample must be sufficiently large to provide an accurate baseline estimate of prevalence which can be compared with that observed later, and detected differences found to be statistically significant. At a minimum, baseline seroprevalence estimates with a sufficient degree of precision ($\pm 5\%$) should be obtained. When it is planned to conduct surveys at regular time intervals in order to monitor time trends, it is better to calculate the sample size that would be required to compare two consecutive estimates. An example of this calculation is given in Annex 1.

In order to monitor the level and trend of HIV prevalence separately by region (in the case of representative sampling) or by sentinel sites (in the case of sentinel surveillance), a sample size must be calculated for each region or for each site. This will greatly increase the total number of patients tested in the country and may only be done if sufficient resources are available.

4.2. Sampling procedure

Ideally, the procedure should ensure that the sample of patients tested is representative of all tuberculosis patients in the country. Representativeness can only be achieved through a random selection procedure. Where this is not feasible because no sampling frame is available or because of logistical constraints, it is better to test patients under standardized conditions in a limited number of diagnostic centres taken as sentinel sites. One must be aware, however, that a HIV prevalence rate among tuberculosis patients in one sentinel site can rarely be considered as representative of the area where the site is situated. This should be taken into account when combining results from several sentinel sites.

4.2.1. Representative sampling

Sampling of individual tuberculosis patients (e.g. systematically sampling one out of X patients) is usually not feasible for logistical reasons.

If resources permit, an option is to test all eligible patients in each diagnostic centre diagnosed within the same limited period of time. To calculate the length of this period, the sample size is divided by the total number of eligible patients per year in the country. For example, if around 6,000 eligible patients are diagnosed per year and if a sample size of 250 patients is required, the inclusion period will be $6,000/250 = 1/24$ year, i.e. two weeks. In this case, all consecutive eligible patients should be tested during two weeks in all centres, either during the same 2 weeks or in rotation (centres of region 1 during the first 2 weeks, centres of region 2 during the next 2 weeks and so on). If the total time to complete the study exceeds one year however, time trends will be difficult to monitor. The representativeness of this design is ensured by the inclusion of all diagnostic centres and by the use of the same inclusion period for each of them.

Where this option is not feasible, the optimal solution is very often to make up a sample out of several randomly selected groups. Rather than a single person, a "cluster" of persons is the sampling unit. A survey using this technique will allow one to obtain an HIV prevalence estimate for all tuberculosis patients diagnosed in the country. *However, it will not permit comparisons among different clusters of the total population surveyed.*

Each cluster will consist of the same number of eligible patients tested consecutively at a diagnostic centre. The cluster size is obtained by dividing the sample size by the desired number of clusters. Since clusters, and not patients, are the statistical units, a minimum of 30 clusters should be chosen. The optimum number of clusters will depend largely on the homogeneity of the HIV prevalence in the population: if the distribution of HIV infection is expected to be patchy, a few large clusters will give much less information than a large number of small ones. If for logistical reasons, less than 30 clusters are selected, this will result in less precise estimates.

The cluster sampling procedure requires the availability of a sampling frame consisting of a complete list of all diagnostic centres in the country, with the number of tuberculosis patients diagnosed per year in each centre. A diagnostic centre is defined as any health unit providing diagnosis and treatment services to tuberculosis patients (e.g., tuberculosis hospital or clinic, hospital with tuberculosis ward). The selection of clusters is conducted from this list, so that the probability of selecting a cluster in a particular centre is proportional to the size of this centre (number of patients diagnosed in this centre). Detailed procedure for cluster selection is explained in Annex 2.

When there is only one cluster selected in a centre, the first patient to be selected can be the first eligible patient of any suitable month. When there are several clusters, each first patient has

to be randomly selected at different points in time: for example, if there are 3 clusters in a centre and if the survey lasts 6 months, 3 months are selected at random among the 6 (e.g. months 1,2,5), and the first eligible patient in month 1, month 2 and month 5 is selected. Patients are then taken consecutively after the first until the size of the cluster is achieved.

To assess time trends, surveys should be repeated at regular intervals. The same sampling procedure should be conducted again for each survey. If cluster sampling is used, this will lead to a new selection of clusters each time (probably located differently than in the first survey).

4.2.2. Sentinel surveillance

In sentinel surveillance, sentinel sites are not selected at random, but rather according to certain criteria. Major criteria may consist of the following:

- a) Recruitment of patients: one may want to select sites attended by patients with high, medium and low HIV prevalence (e.g. referral hospital, urban centre, rural centre) in order to get some indications on the general situation of the country. Selecting only a referral hospital should be avoided. However, it must be emphasized that representativeness is not achieved by this method.
- b) Willingness of the hospital/clinic staff to participate and cooperate in carrying out HIV seroprevalence surveys.
- c) Conditions which facilitate the collection of blood samples: blood already drawn from patients as part of routine care, availability of cold chain, presence of a laboratory which can perform HIV serological tests.

To obtain separate HIV prevalence estimates for each sentinel site, eligible patients are tested consecutively until the desired sample size is achieved in each site. Alternatively, it is possible to estimate a study period and to test all eligible patients during this period. The length of the period is calculated by dividing the total number of eligible patients diagnosed per year at the site by the sample size. This is particularly applicable when patients tested at several sentinel sites are considered as a single sample (however, combining data from different sentinel sites is not advisable, see 6). For example, if a total of 500 patients are diagnosed per year in the selected sentinel site(s) and if the sample size is 250, the study period will be $500/250 = 1/2$ year, *i.e.*, 6 months, and all patients diagnosed at the site(s) during a 6-month period should be tested.

To assess time trends, the study should be repeated at regular intervals *in the same sites during the same period* (this can be the entire year if patients are tested throughout the year). If additional sites are selected in subsequent periods, it is preferable to assess trends separately for those centres (see 6).

5. Operational procedures

5.1 Data and specimen collection

In the best conditions, blood is already routinely drawn from tuberculosis patients and part of the blood specimen can be taken for HIV serological tests. Alternatively, if blood is not routinely drawn, patients may be asked to voluntarily donate a blood sample for a study on the role of infections such as HIV on tuberculosis.

A unique study number should be attributed to each patient selected. The blood specimen should be labelled with this number only, without any other identifier. The same study number should be recorded on the data collection form of the patient.

The possibility of linking test results with a patient's identity must be excluded at all steps of the study. Therefore, data collected should not only be completely anonymous, but they should also be kept to a minimum to avoid indirect identification of the patient by his/her characteristics. This is particularly important if data are to be analyzed and presented separately by sentinel site.

Basic information to be collected includes sex, age (preferably broad age groups rather than exact age), site of disease (pulmonary/extrapulmonary) and bacteriological status (sputum smear positive/negative). If relapse cases are included in the sample, the status of the patient (new case/relapse) should be recorded also. Information can be abstracted from the medical records. Area of residence is necessary only if separate samples are taken in different regions. For cluster sampling surveys, the identification of the cluster must be recorded. Identification of the study site should be recorded only if absolutely necessary (sentinel surveillance). If surveys are conducted at repeated intervals, the study period should be mentioned without an exact date. An example of a data collection form is given in Annex 3.

When HIV tests are performed in a reference laboratory, data collection forms can be sent to this laboratory together with coded serum specimen. Laboratory test results are recorded on the forms and the forms then sent to the data manager. Alternatively, data collection forms can be sent to the data manager, and laboratory test results (identified by code numbers) are sent separately.

When HIV tests are performed in the centre attended by the patients, it is preferable that recording of demographic/clinical data and recording of HIV test results be performed by two different persons on two separate forms. The two forms are then sent independently to the data manager. Linking of data by code number should be done at the central level.

If, for one reason or another, a blood specimen cannot be obtained from some patients, it would be useful to complete a data collection form for these patients as well, in order to subsequently analyze characteristics of patients not included.

5.2 Serological tests

Specimens are tested for HIV by ELISA or other rapid/simple methods (such as dot immunoassay and agglutination tests) recommended by the National AIDS Programme. The strategy for testing should be based upon the recommendations of WHO and take into account the objective of the test (surveillance) and the HIV seroprevalence of the population under study (8). In brief, as the test is performed for surveillance purposes only and if the general HIV seroprevalence is estimated to be higher than 10%, one ELISA or rapid/simple assay is sufficient; if the HIV seroprevalence is 10% or lower, any positive ELISA or rapid/simple test will need to be retested and found positive by a second ELISA or rapid/simple test based on a different antigen preparation and/or a different test principle to be considered as positive.

6. Data management and analysis

The person responsible for data management and analysis links demographic and clinical data with serological test results. After each testing period, the data can be analyzed by the principal investigator with the help of a statistician or epidemiologist. However, all data collection forms should be carefully checked for errors prior to analysis.

Data analysis and presentation of results depend on study design. With representative sampling, results can be presented for the area surveyed as a whole. Results cannot be presented by region if separate samples (with sufficient sample size) have not been drawn in the different regions. If cluster sampling is used, HIV prevalence data should never be presented by cluster as a cluster is the sampling unit. In sentinel surveillance, results can be presented separately by sentinel site if sample size is sufficient in each site. Combining data from different sentinel sites should only be done with caution, since results can rarely be meaningfully interpreted.

Confidence intervals (CI) around HIV prevalence estimates can be calculated for simple random sampling or sentinel surveillance with the usual formula :

$$CI = \pm 1.96 \sqrt{\frac{P(1-P)}{N}}$$

For cluster sampling, this formula is not appropriate. It is possible to calculate a confidence interval using data tabulated by cluster (see Annex 2); alternatively an approximation of the CI can be obtained by multiplying the CI calculated as for simple random sampling by 1.4 (square root of the multiplier 2 used to calculate sample size for cluster sampling).

To assess trends in HIV-associated tuberculosis, results from surveys repeated at regular time intervals using the same methodology can be compared: an observed difference in prevalence (*i.e.*, from baseline) can be tested for significance by a chi-square test for trend.

7. Recruitment and Training of Personnel

This should be detailed in the study protocol according to the local conditions.

8. Time-frame

This should be decided by the investigator.

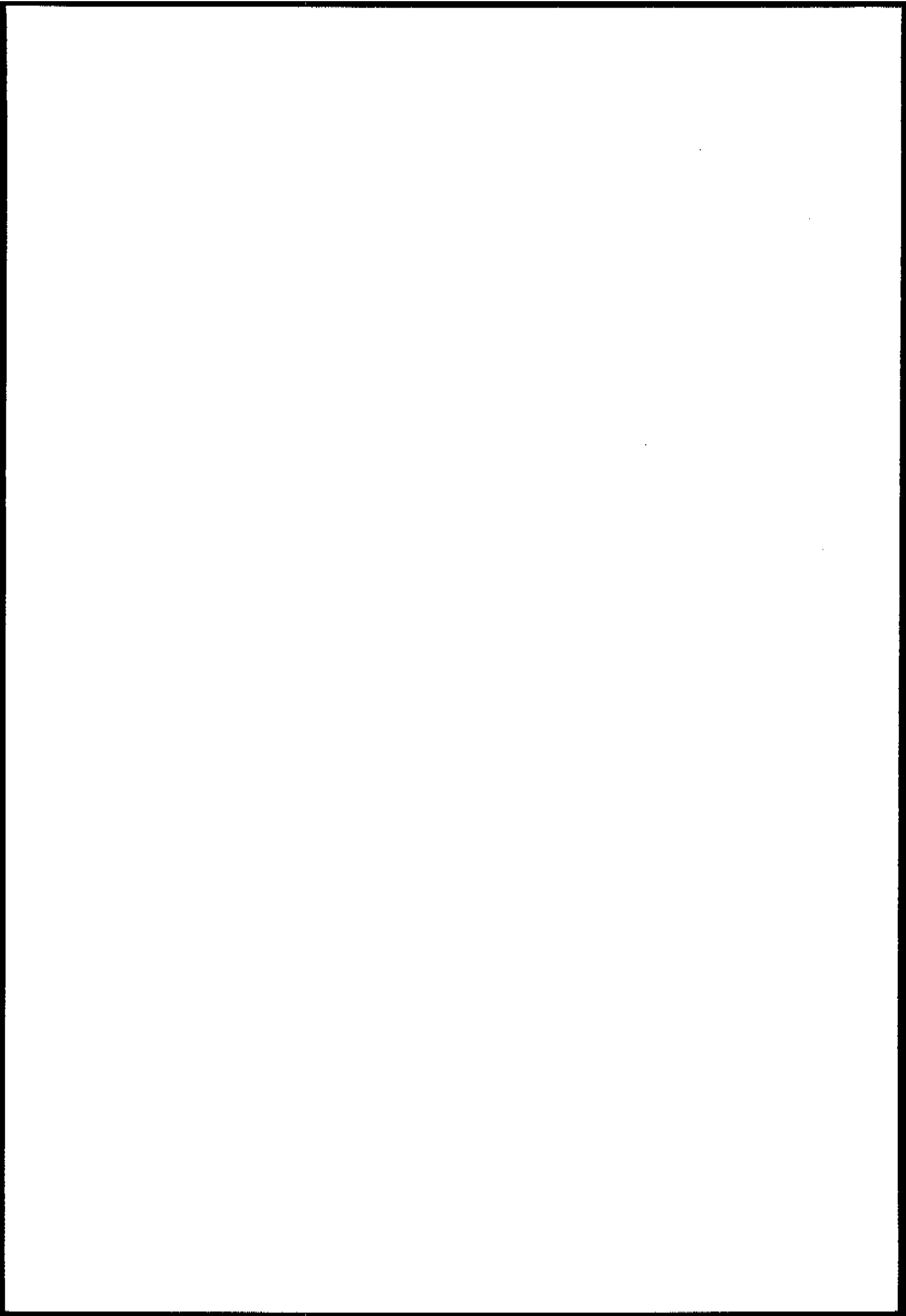
9. Budget (annual)

This is to be developed by the investigator. In brief, it should include salaries for support staff and a data entry clerk; supplies such as HIV test kits; stationery; if not already available, a computer with printer and adequate software for data management; travel expenditures for the investigator(s) to visit sites; if not already available, a vehicle and expenditures for its maintenance. In addition, the budget should also contain the expenses related to setting up a system of monitoring.

The following persons participated in the development of the protocol: from the WHO's Tuberculosis Programme Drs J.P. Narain (currently with the Global Programme on AIDS, SEARO), R. O'Brien, M.C. Raviglione, and V. Schwoebel (temporary adviser); from the International Union against Tuberculosis and Lung Disease Dr H.L. Rieder. Important contribution was offered by Mr H.G. ten Dam, former Scientist at the WHO's Tuberculosis Programme; Dr E. van Praag, Global Programme on AIDS; Drs R.A. Ancelle, J.B. Brunet, E. Couturier, Mrs F. Cazein and C. Six, European Centre for Epidemiological Monitoring of AIDS.

10. References

1. Styblo K. The potential impact of AIDS on the tuberculosis situation in developed and developing countries. *Bull Int Union Tuberc Lung Dis* 1988; 63: 25-8.
2. Narain JP, Raviglione MC, Kochi A. HIV-associated tuberculosis in developing countries: epidemiology and strategies for prevention. *Tuber Lung Dis* 1992; 73:311-21.
3. De Cock KM, Soro B, Coulibaly IM, Lucas SB. Tuberculosis and HIV infection in sub-Saharan Africa. *JAMA* 1992; 268:1581-7.
4. Selwyn PA, Hartel D, Lewis VA et al. A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. *N Engl J Med* 1989; 320; 545-50.
5. Rieder HL, Cauthen GM, Bloch AB et al. Tuberculosis and AIDS - Florida. *Arch Intern Med* 1989; 149: 1268-73.
6. World Health Organization. Treatment of tuberculosis. Guidelines for national programmes. Geneva 1993.
7. Dondero TJ, Pappaioanou M, Curran JW. Monitoring the levels and trends of HIV infection: The Public Health Services HIV Surveillance Programme. *Public Health Reports* 1988; 103: 213-20.
8. World Health Organization. Global Programme on AIDS. Recommendations for the selection and use of HIV antibody tests. *Wkly Epidemiol Rec* 1992; 67: 145-9.



ANNEX 1

Sample size calculation (for simple random sampling; for cluster sampling, see 4.1)

Example 1:

The tuberculosis control programme wishes to evaluate the HIV seroprevalence among patients attending a tuberculosis clinic. How many patients with tuberculosis should be included in the sample so that the HIV seroprevalence may be estimated to within 5 percentage points of the true value with 95% confidence, if it is known that the true seroprevalence is unlikely to exceed 10%?

Solution:

- 1) Anticipated HIV seroprevalence (P): 10%
- 2) Confidence level: 95%
- 3) Absolute precision (d): 5 percentage points (from 5-15%).

From Table below in the column headed "0.10"(10% seroprevalence) and the row headed "5"(5 percentage points), it is found that a sample size of 138 would be needed.

Sample size necessary to estimate P within d absolute percentage points with 95% confidence.

d (%)	Anticipated population proportion (P)				
	0.50	0.40	0.30	0.20	0.10
1	9604	9220	8067	6147	3457
2	2401	2305	2017	1537	864
3	1067	1024	896	683	384
4	600	576	504	384	216
5	384	369	323	246	138
6	267	256	224	171	96
7	196	188	165	125	71
8	150	144	126	96	54
9	119	114	100	76	43
10	96	92	81	61	35

Source: Lwanga SK, Lemeshow S. Sample Size Determination in Health Studies. A Practical Manual. World Health Organization, Geneva, 1991.

Example 2:

The tuberculosis programme wishes to evaluate the HIV seroprevalence among patients attending a tuberculosis clinic annually in order to assess time trends. What would be the minimum sample size in each survey necessary to compare HIV prevalence between year 1 and year 2? The baseline prevalence in year 1: P_1 is expected to be 0.10(10%) and one wants to detect a difference of 0.10 between the two years (prevalence in year 2: $P_2 = 0.20$).

Solution

The formula to calculate sample size with alpha error of 0.05 and power of 90% is :

$$N = 8.6 \frac{P_1(1-P_1) + P_2(1-P_2)}{(P_1 - P_2)^2}$$

Each year 215 patients should be included in the sample.

Source: ten Dam H.G. Surveillance of tuberculosis by means of tuberculin surveys. WHO/TB/85.145

ANNEX 2

Cluster sampling

Cluster selection

Example: A sample size of 360 tuberculosis patients has been calculated after taking into account the effect of cluster sampling. 30 clusters of $360/30 = 12$ patients will have to be selected. The following steps must be taken (see below):

- a) establish the list of the diagnostic centres with their annual number of patients (see below).
- b) calculate the cumulative numbers of patients and record them in an additional column. Cumulative number for second centre will be (number in first centre) + (number in second centre). Cumulative number for third centre will be (cumulative number for second centre) + (number in third centre) and so on. The total number of patients diagnosed in the country is 6,322.
- c) determine the sampling interval: $6,322 / 30 = 211$
- d) select a number between 0 and 211 at random (with a table of random numbers or by using the last digits of a currency note for example). In this case the number selected is 120.
- e) the first cluster is selected using this number 120 : it will be in the first centre because 120 falls between 0 and 246 (number of patients in the first centre).
- f) selection of next clusters is done by adding the sampling interval 211 each time to this first number 120. The next number $(120 + 211) = 331$ falls between 246 and 1,823 (cumulative number of patients for second centre), therefore the 2nd cluster is selected in the 2nd centre. The 3rd number $(331 + 211) = 542$ falls also between 246 and 1,823, the 3rd cluster is therefore selected in the 3rd centre as well.

Name of diagnostic centre	Number of patients diagnosed per year	Cumulative number of patients	Cluster number
A	246	246	1
B	1,577	1,823	2,3,4,5,6,7,8,9
C	468	2,291	10,11
D	340	2,631	12
E	220	2,851	13
F	246	3,097	14,15
G	190	3,287	16
H	1,124	4,411	17,18,19,20,21
I	61	4,472	
J	154	4,626	22
K	139	4,765	23
K	60	4,825	
M	14	4,839	
N	38	4,877	
O	19	4,896	
P	41	4,937	
Q	120	5,057	24
R	455	5,512	25,26
S	51	5,563	
T	26	5,589	
U	199	5,788	27
V	21	5,809	
W	32	5,841	28
X	69	5,910	
Y	6	5,916	
Z	145	6,061	29
AA	129	6,190	
BB	87	6,277	30
CC	10	6,287	
DD	35	6,322	

Cluster sampling

Confidence interval calculation

If cluster selection is done with probability proportional to size (method described above) and if clusters have the same size, a simplified formula for the confidence interval (CI) around the HIV prevalence is :

$$CI = \pm 1.96 \sqrt{\frac{\sum_i (P_i - P)^2}{n(n-1)}}$$

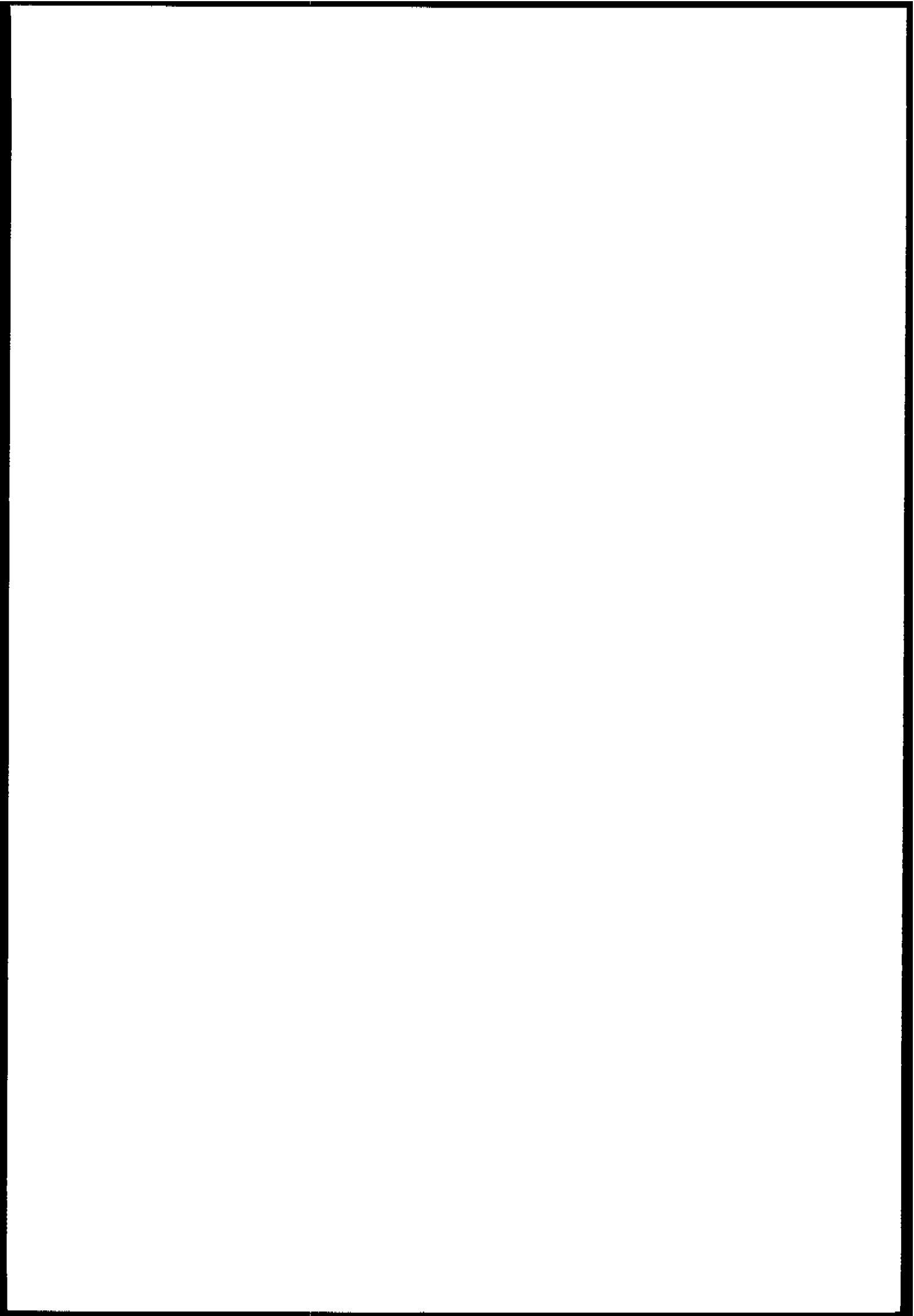
where P is the proportion (HIV prevalence) calculated for the total sample,
 P_i is the proportion calculated in each cluster i
 n is the number of clusters (30)

To calculate the sum of the $(P_i - P)^2$ over all 30 clusters the following table can be used:

Cluster number	P_i	$P_i - P$	$(P_i - P)^2$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			

The total of the last column can then be used in the formula.

Source: ten Dam H.G. Surveillance of tuberculosis by means of tuberculin surveys. WHO/TB/85.145



ANNEX 3

Data collection form
HIV surveillance of tuberculosis patients (unlinked anonymous testing)

TB/HIV surveillance form:

Study site: _____ (if necessary)
Cluster number: _____ (if necessary)
Study period: _____ (if necessary)
Study region _____ (if the sample is stratified by region)

Patient code number: _____

Age group	15 - 19	<input type="checkbox"/>
	20 - 24	<input type="checkbox"/>
	25 - 29	<input type="checkbox"/>
	30 - 34	<input type="checkbox"/>
	35 - 39	<input type="checkbox"/>
	40 - 44	<input type="checkbox"/>
	45 - 49	<input type="checkbox"/>

Sex: male female

If patients with relapse are included:

Patient: new relapse

If patients with all forms of tuberculosis are included:

Site of tuberculosis: pulmonary extra-pulmonary
(check 1 or both)

If pulmonary: sputum smear positive negative

Laboratory form

Study site : _____ (if applicable)

Patient code number _____ (same number as that on the surveillance form)

ELISA 1: Reactive non reactive doubtful not done

If due:

ELISA 2: Reactive non reactive doubtful not done

ANNEX 4

Attributable risk calculation

The attributable risk is defined as the proportion of disease incidence (or disease risk) which can be attributed to a specific exposure. In case of tuberculosis, we want to estimate what proportion of the tuberculosis incidence can be attributed to HIV, both for the exposed (HIV-infected) population and for the general population.

The calculation of an attributable risk requires a comparison group. This comparison should be as representative as possible of the general population with respect to HIV infection. Data on seroprevalence among women attending antenatal clinics or among blood donors from the same area can be used for comparison. However, results should always be taken with caution, since representativeness of these groups is often poor.

The sample of tuberculosis patients serves as a case group, and the general population group (blood donors, antenatal clinic attenders) as a control group. Data are analyzed as a case-control study: the odds ratio (OR) will provide an estimate of the relative risk of tuberculosis for HIV-infected persons as compared to non-infected.

	TB cases	Controls		a . d
HIV (+)	a	b	OR =	-----
HIV (-)	c	d		b . c

The OR is then used to calculate the attributable risk as follows:

1. To determine the proportion of tuberculosis cases attributable to HIV infection among the HIV-infected:

$$1 - 1/OR$$

2. To determine proportion of tuberculosis cases attributable to HIV infection among the overall population:

$$(1 - 1/OR) \cdot (a / a+c)$$

where $a / a+c$ is the prevalence of HIV infection among TB patients

If data are available by age group and sex in both groups, it is better to exclude groups not represented among both groups and to estimate the OR from stratified data using the Mantel-Haenszel summary estimate:

	TB cases	Controls	Total
HIV (+)	a_i	b_i	n_i
HIV (-)	c_i	d_i	n_2
Total	m_1	m_2	T_i

$$\text{Summary odds ratio} = \frac{E(a_i \cdot d_i / T_i)}{E(b_i \cdot c_i / T_i)} \quad (E=\text{summatory})$$