Available online at www.sciencedirect.com
SCIENCE DIRECT®

Microbes and Infection 8 (2006) 721−727

Microbes and Infection

www.elsevier.com/locate/micinf

Original article

# An evaluation of the actual incidence of tuberculosis in French Guiana using a capture-recapture model

Vanina Guernier [a,b,*], Jean-François Guégan [a], Xavier Deparis [c]

[a] UMR 2724 IRD-CNRS, Génétique et Évolution des Maladies Infectieuses, Equipe Évolution des Systèmes Symbiotiques, 911 avenue Agropolis, BP 64501, 34394 Montpellier cedex 05, France
[b] Unité S140 Expertise et Spatialisation des Connaissances en Environnement, IRD, Maison de la Télédétection, 500 rue J-F. Breton, 34093 Montpellier cedex 05, France
[c] Institut Pasteur de la Guyane, BP 6010, 97306 Cayenne, French Guiana

Received 28 April 2005; accepted 7 September 2005
Available online 13 January 2006

**Abstract**

In order to estimate the level of under-reporting and to improve estimates of the incidence of tuberculosis (TB) in the vicinity of Cayenne, French Guiana, we performed capture-recapture analysis from 1996 through 2003. We cross-linked data from the Institut Pasteur, the Département d'Information Médicale of Cayenne Hospital, and the Service de Lutte Anti-Tuberculeuse. The estimate of 381 TB cases obtained after matching those three sources was revised to 425 (95% confidence interval: 407, 453) using the capture-recapture model based on sample coverage. The corresponding average annual incidence was 63.1 TB cases per 100,000 population. The evaluated sensitivity of the compulsory notification system was 35.3%, indicating wide under-notification of TB in the vicinity of Cayenne. The estimated coverage reported by the three sources was fairly accurate (i.e. 85.9%), but not sufficient to evaluate the risk of transmission of TB in the Ile-de-Cayenne (Cayenne and its suburbs).


*Keywords:* Tuberculosis; French Guiana; Disease notification; Transmission; Surveillance; Incidence; Statistical models

## 1. Introduction

Human tuberculosis (TB) is caused by the bacterium *Mycobacterium tuberculosis* (and occasionally by *Mycobacterium bovis*, which is generally more pathogenic for animals). This major human disease, responsible for 1.9 million deaths per year worldwide, has re-emerged as a serious public health threat. Almost two billion people are latently infected with the tubercle bacillus and more than eight million new cases are diagnosed each year [1]. *M. tuberculosis* is transmitted from person to person via the aerosol route through small droplets of saliva and sputum expelled when an infectious patient sneezes or coughs. Active TB infection requires treatment with a combination of three or four antibiotics for at least six months. Failure to complete the full course of drug therapy can lead to the emergence of drug-resistant strains of mycobacteria [2,3]. This so-called acquired resistance, as opposed to primary resistance [4], severely limits effective treatment options. Moreover, it can lead to the maintenance of a pool of potentially resistant bacteria in the environment, thus hindering arrest of the chain of transmission. Likewise, if not

*Abbreviations:* CI, confidence interval; DIM, Département d'Information Médicale; INSEE, National Institute for Statistics and Economic Studies; InVS, Institut de Veille Sanitaire; IPG, Institut Pasteur de la Guyane; LAT, Lutte Anti-tuberculeuse; MDRTB, multidrug-resistant tuberculosis cases; TB, tuberculosis.

* Corresponding author. UMR 2724 IRD-CNRS, Génétique et Évolution des Maladies Infectieuses, Equipe Évolution des Systèmes Symbiotiques, 911 avenue Agropolis, BP 64501, 34394 Montpellier cedex 05, France. Tel.: +33 4 67 54 87 12; fax: +33 4 67 54 87 00.
*E-mail address:* guernier@mpl.ird.fr (V. Guernier).

1286-4579/$ - see front matter 
doi:10.1016/j.micinf.2005.09.009

screened and treated appropriately, persons with unnotified cases are likely to transmit the disease and/or to develop drug-resistant TB [5,6].

These points underscore the need for effective control strategies, specifically for TB. Early identification of cases ensures that treatment is completed and enables close-contact investigations and screening [7]. It also provides data to measure disease incidence, monitor epidemiological trends, detect outbreaks and plan new strategies for prevention and treatment. For these reasons, obligatory notification of all TB cases is particularly useful. Several studies have evaluated the completeness of TB notification over a defined period and within particular geographical areas [8−11], while recent studies often use capture-recapture methodology [12]. In France, TB has been statutorily considered a notifiable disease since 1964. The Health Monitoring Institute (InVS, Institut de Veille Sanitaire), in charge of epidemiological surveillance, reported the highest TB rates in French Guiana and in the suburbs of Paris. But an evaluation of the completeness of notification in French Guiana, comparing culture-positive cases with notified cases from 1994 to 1996, has shown strong evidence of under-notification of TB in that department [13]. For example, in 1998, only six cases were notified to the InVS for the all French Guiana, whereas the Institut Pasteur reported 20 biologically confirmed cases in the Ile-de-Cayenne only (i.e. Cayenne and its suburbs). At present, dissemination of health services in the territory (3 hospitals on the coast and 21 health centers dispersed throughout the territory) and under-reporting of TB by medical practitioners preclude data completeness and analysis of a larger data set for the whole of French Guiana. This is the reason why we considered only TB cases in the population of the Ile-de-Cayenne, where most new cases occur and for which most health information is available. In the present work, we performed capture-recapture analysis to evaluate the actual TB incidence in the Ile-de-Cayenne and to provide estimates of TB incidence for all of French Guiana.

## 2. Materials and methods

### *2.1. Study period*

The analysis was restricted to the period from January 1, 1996 to December 31, 2003, which is the most reliable period in terms of available data sources. TB cases were notified using the date of registry in each monitoring system, thus providing daily data.

### *2.2. Study site and population*

French Guiana is a French overseas department comprising 32,432 square miles. It is located in the northeast of South America, between Surinam and Brazil. The study sector, locally referred to as the Ile-de-Cayenne, is located along the Atlantic coast in the north of French Guiana. This area is made up of three communities, Cayenne (the capital), Rémire-Montjoly and Matoury, and constitutes the main economic pole of attraction with the highest population density in French Guiana. More than 90% of the 157,213 inhabitants of French Guiana (census 1999 from INSEE − National Institute for Statistics and Economic Studies; http://www.insee.fr) are concentrated in the coastal region, a tropical rainforest spreading over 90% of the area, and the Ile-de-Cayenne represents more than 50% of the total population of French Guiana (84,181 inhabitants in 1999).

### *2.3. Data sources*

In order to select patients to be included in our analysis, we used the European consensus definition of a notifiable case of TB [14,15]. A case is defined as any person administered a full course of antituberculosis treatment (at least three antibiotics). This definition excludes prophylaxis TB and atypical mycobacteria, but includes definite cases in which *M. tuberculosis* was confirmed by direct examination and/or culture, as well as clinical cases, non-confirmed in the laboratory but treated based on a clinician's decision. Cases which appeared more than once in a source, and were generally relapsed cases, were included once in our data set, even when the cases were reported several times upon obligatory notification. The earliest record was retained from each duplicate set. The same rule applied to two cases notified as chronic. Cases already treated for TB before January 1, 1996 and again treated after that date were ignored. Since no proper databases yet exist, we censused all cases via surveys conducted in three different health monitoring systems.

The Mycobacterium Laboratory of the Institut Pasteur de la Guyane (IPG) is in charge of the culture of isolates from patients suspicious for TB and forwarded by hospitals or private general practitioners. Cultures are then sent to the Institut Pasteur de la Guadeloupe (French West Indies), where identification of the mycobacterium and molecular analysis are performed. Because genotyping requires the availability of a viable isolate of *M. tuberculosis*, the population for which the culture is positive and for which genotyping is performed constitutes a subset of all cases, excluding clinical cases or extrapulmonary TB for which biopsy specimens are not always available for bacteriology. For this reason, data censused at the IPG are not exhaustive for TB notification.

The Département d'Information Médicale (DIM) of Cayenne Hospital collects and computes medical information on all patients treated in or admitted to the Cayenne Hospital, such as the name, surname, date of birth, address, health problems and results of key medical tests. Diagnoses are recorded using the 10th International Classification of Diseases (ICD-10) coding system. For the needs of the study, a listing of cases was automatically extracted from the DIM general database based on principal as well as secondary diagnosis of TB. When there was doubt concerning the diagnosis, the medical record was inspected for confirmation of a true case of TB.

The Service de Lutte Anti-Tuberculeuse (LAT), a branch of the Conseil Général of Cayenne, i.e. a Department Council, is in charge of follow-up and treatment of TB cases when patients leave the hospital in which they were originally treated in individual isolation rooms during the infectious stage of the disease. The LAT is also in charge of screening surveys conducted among close relatives or neighbors of the patients in order to detect persons at risk of exposure, and to provide post-exposure prophylaxis if necessary. Patients immigrating from surrounding countries where they started a treatment are followed up by the LAT upon their arrival in French Guiana, and they were also included in our data set.

### *2.4. Linkage among registries*

Identifiers common to the three sources were needed to match patients present in more than one source. Personal information (name, sex, date of birth, address and date of registry in the system) was collected for each case and used to detect duplicates. It should be noted that this research proposal, including the collection of data from clinical records, ensures protection of confidentiality according to the code of medical ethics, legislation and research findings. Duplicate cases were identified by examining the name and date of birth, and occasionally a third item when in doubt. The earliest record was retained from each duplicate set, and the registered date was used for all further analyses.

### *2.5. Capture-recapture method*

Population estimates using capture-recapture analysis consist of cross-matching the information from two or more databases in order to identify the number of individuals common to paired lists (matched cases) and then using overlapping information and statistical techniques to evaluate the number of missing individuals and, consecutively, the total population size [16–20]. This method has been used in the epidemiology for over 50 years to improve estimates of disease incidence and deaths [21–23], with increased use during the past 20 years [24–28]. Four assumptions underlie this method [12,28,29]: cases must have equal probabilities of being observed (captured) in any source; cases must be uniquely matched among the various data sources; case ascertainment by sources must be independent; the population under study must be ‘closed’. Two sources are considered independent when the probability of notification of an event in one source is not dependent upon the probability of notification of that event in the other source, and when the probability of notification of an event is not influenced by its characteristics (e.g. age, gender, symptoms, etc.) [30].

When only two data sources are available, the assumption of independence is difficult to justify in epidemiological surveys. For example, in our study, the majority of cases reported from IPG and LAT comes from the Cayenne Hospital where suspicious TB cases are first noted; therefore, dependence should occur between DIM and IPG, and DIM and LAT, and independence between IPG and LAT. But with a multiple source approach (more than two sources), which is the case here, the assumption of independence can be dropped by using specified statistical techniques which take into account interdependence among data sets, the most commonly used being log-linear models and the sample coverage approach [31].

### *2.6. Estimate of the number of TB cases*

Specific software for capture-recapture models was used to estimate the size of the population of French Guiana infected with TB: CARE-1 [31], developed for an S-Plus environment [32] (available at http://chao.stat.nthu.edu.tw/softwareCE.html). Three different statistical methods were tested: Petersen and Chapman estimators, log-linear models and the sample coverage approach. The standard Petersen estimator and the nearly unbiased Chapman estimator, based on any pair of samples, must be used under the assumption of independence, which is unlikely to be met here, but those estimators are useful in preliminary analysis. With the log-linear modeling approach, data are considered as a $2^t$ contingency table (with $t$ the number of sources) with an empty cell, corresponding to the number of cases not identified by any of the lists. Those missing cases are estimated while modeling two-way interactions (and therefore dependence) between sources in various log-linear models. The estimate of missing cases can be derived from the model that best fits the data [33,34]. The sample coverage approach incorporates the correlation bias contained in a three-sample model, taking into account recapture information and sample dependencies. It is used as a measure of the overlap fractions among sources. Three types of estimators are calculated: the population size estimate for independent samples ($\hat{N}$hat-0), a sufficiently high number of sample coverage cases ($\hat{N}$hat) and low sample coverage cases ($\hat{N}$hat-1) (see [31] for details).

## 3. Results

### *3.1. Simple matching of sources*

After cross-linking of records from the IPG, DIM and LAT, 381 TB cases were reported in the Ile-de-Cayenne from January 1, 1996 to December 31, 2003. Among these 381 identified cases, 235 (61.7%) were confirmed by isolation and cell culture of *M. tuberculosis*, 111 (29.1%) were found in only one registry, 132 (34.7%) in only two registries and 138 cases (36.2%) in all three sources of information. The cumulative number of TB cases found in one, two or three registries is illustrated in Fig. 1. During the census, no cases of *M. bovis* infection were identified.

### *3.2. Capture-recapture estimates*

The Petersen and Chapman estimates are in the range of 365–406 (see Table 1 for details), given that the Petersen estimator underestimates the true size when both samples are positively dependent and overestimates the true size in case of negative dependence (see the sample coverage approach for identification of dependent sources).

Results for all log-linear models considered are summarized in Table 2, with corresponding deviances and estimates

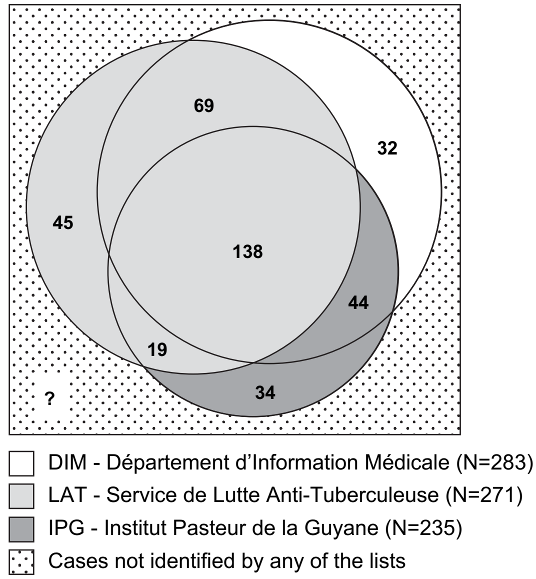


Fig. 1. Three-source combinations of epidemiological reports for the Ile-de-Cayenne for 1996 through 2003 ($N = 381$). The number of cases identified in one, two or three sources is noted.

of the total number of TB cases. The best-fitting model is 12/23 (i.e. dependence arises between sources one and two and sources two and three), which produces an estimate of 462 cases (95% confidence interval (CI): 423, 536). The dependence between sources one and two (IPG and DIM) and between sources two and three (DIM and LAT) was highly intuitive (see Section 2). The degree of ascertainment of monitoring systems (or "sensitivity") was 61.3%, 50.9%, and 58.7%, respectively, for DIM, IPG and LAT, with this model.

Using the third approach, the sample coverage estimate for our data set, i.e. the proportion of cases that were not unique, was 85.9%, and the average of the overlapped cases, i.e. the number of cases found in two lists, excluding the third, was equal to $D = 344$ (Table 3). The sample coverage estimate of 85.9% was sufficiently high to correct for undercount, and estimator N̂hat was used. It gives an estimated population size of 425 (95% CI: 407, 453) using 1000 bootstrap replications. Sensitivities were 66.6% for DIM, 55.3% for IPG and 63.8% for LAT. These data were sufficient to produce a reliable population size estimate of TB cases escaping epidemiological surveillance and health care systems.

Table 1
Population size estimates based on any pair of samples using Petersen and Chapman estimators

| | Petersen | Chapman | se[a] | cil[b] | ciu[c] |
|---|---|---|---|---|---|
| Pair (1,2)[d] | 365 | 365 | 8 | 354 | 384 |
| Pair (1,3) | 406 | 405 | 12 | 386 | 434 |
| Pair (2,3) | 370 | 370 | 6 | 361 | 387 |

[a] se, Asymptotic standard error.
[b] cil, 95% CI lower limit. A log-transformation is used so that the lower limit is always greater than the number ascertained. See [17] for details.
[c] ciu, 95% CI upper limit.
[d] Refers to paired samples (1, Institut Pasteur de la Guyane; 2, Département d'Information Médicale; 3, Service de Lutte Anti-Tuberculeuse).

Table 2
Estimates based on log-linear models

| | dev[a] | df | est | se | cil | ciu |
|---|---|---|---|---|---|---|
| Independent | 39.75 | 3 | 395 | 4 | 389 | 407 |
| 13/2[b] | 38.77 | 2 | 393 | 4 | 387 | 406 |
| 23/1 | 27.47 | 2 | 406 | 7 | 395 | 424 |
| 12/3 | 28.27 | 2 | 403 | 6 | 394 | 419 |
| 12/23 | 1.84 | 1 | 462 | 28 | 423 | 536 |
| 12/13 | 28.20 | 1 | 402 | 7 | 392 | 421 |
| 23/13 | 27.47 | 1 | 406 | 9 | 394 | 430 |
| Symmetry[c] | 32.75 | 4 | 463 | 34 | 419 | 560 |
| Quasi-sy | 17.28 | 2 | 464 | 34 | 419 | 561 |
| Part-qs1 | 0.32 | 1 | 495 | 49 | 431 | 638 |
| Part-qs2 | 11.37 | 1 | 483 | 43 | 427 | 606 |
| Part-qs3 | 15.03 | 1 | 460 | 33 | 417 | 555 |
| Saturated | 0.00 | 0 | 498 | 51 | 433 | 647 |

All possible models derived from a triple sample are considered: independent model, models with one or two interactions which consider local independence only (i.e. 13/1, 23/1, 12/3, 12/23, 12/13, 23/13), models taking heterogeneity into account (quasi-symmetric and partial quasi-symmetric models), and saturated models.
[a] dev, Deviance statistics; df, degree of freedom; est, population size estimate; se, asymptotic standard error; cil, 95% confidence interval lower limit; ciu, 95% confidence interval upper limit; 1, IPG; 2, DIM; 3, LAT.
[b] If dependence arises in samples 1 and 2, the model is denoted as model 12/3. If dependence also appears in samples 1 and 3, the model is noted 12/13, and so on for other interactions.
[c] See [31] for details about the models.

### 3.3. Incidence estimates

Considering the low number of TB cases counted in our survey in the Ile-de-Cayenne (38–70 cases per year), and in order to avoid multiplication of biases, capture-recapture

Table 3
Outputs obtained with the sample coverage approach

| | $M$[a] | $D$ | $\hat{C}$ | est | se | cil | ciu |
|---|---|---|---|---|---|---|---|
| N̂hat-0[d] | 381 | 344 | 0.859 | 401 | 6 | 392 | 415 |
| N̂hat | 381 | 344 | 0.859 | 425 | 11 | 407 | 453 |
| N̂hat-1 | 381 | 344 | 0.859 | 419 | 9 | 405 | 441 |
| | $u1$[b] | $u2$ | $u3$ | $r12$[c] | $r13$ | $r23$ | $r123$ |
| N̂hat-0[d] | 0.59 | 0.71 | 0.68 | 0.10 | -0.01 | 0.08 | 0.06 |
| N̂hat | 0.55 | 0.67 | 0.64 | 0.16 | 0.05 | 0.15 | 0.03 |
| N̂hat-1 | 0.56 | 0.68 | 0.65 | 0.15 | 0.03 | 0.13 | 0.03 |

Three estimators were calculated, taking into account the heterogeneity of redundant information between the three samples (or sources) considered.
[a] $M$, number of cases listed in at least one list; $D$, average number of cases found in two lists excluding the third; $\hat{C}$, sample coverage estimate; est, population size estimate; se, asymptotic standard error; cil, 95% CI lower limit; ciu, 95% CI upper limit.
[b] $u$, Estimated mean probabilities depending on the estimate of $N$ for samples 1, 2 and 3.
[c] $r$, Estimated coefficient of covariation (CCV) depending on the estimate $N$ for different interactions between samples 1, 2 and 3.
[d] Refers to the population size estimate for independent samples (N̂hat-0), sufficiently high sample coverage cases (N̂hat) or low sample coverage cases (N̂hat-1). See [31] for details.

estimates were calculated once with the sum of cases reported from January 1, 1996 to December 31, 2003. The estimated mean incidence was therefore calculated using the estimated number of 425 cases divided by eight years, i.e. 53.1 cases per year for 84,181 inhabitants. Thus, the average annual incidence rate determined by capture-recapture estimates was 63.1 TB cases per 100,000 population in the Ile-de-Cayenne. However, this might represent an underestimation of the actual incidence, since the number of TB cases increased from 1998 in the Ile-de-Cayenne. During the same period, annual incidences per 100,000 population officially reported by the InVS were 24.2 for all of French Guiana, 27.8 for the Ile-de-France (Paris and its suburbs) and 11.3 for metropolitan France.

## 4. Discussion

This is the first study to estimate the true number of TB cases in the Ile-de-Cayenne from 1996 through 2003. Indeed, the disease is recognized to be widespread but is also highly underestimated in French Guiana. Using the capture-recapture method, the estimate of 381 TB cases obtained by simple linking of information from three sources was revised to 425 (95% CI: 407, 453) with the sample coverage approach, which represents an increase of 11.6%. Similar estimates were found using log-linear models. The difference between notified cases and cross-linking results is probably due to the fact that there are only two pulmonary disease specialists in the Ile-de-Cayenne and a general shortage of physicians who do not have enough time for notifications, whereas differences between results from cross-linking and sample coverage might be related to cases that are less likely to be treated, among homeless and non-sedentary populations, for example.

Despite the low declaration of TB cases in French Guiana, the sensitivity of the three combined systems in the Ile-de-Cayenne was quite good, but such cross-linked data are not available, as there is no centralization of data. For this reason, the degree of ascertainment of each monitoring system was evaluated independently. Estimated sensitivity was 35.3% for the notification system compared to sample coverage approach results. Sensitivities of the three monitoring systems varied between 55.3% and 66.6% as a function of the model used. These values are similar to those evaluated in previous studies, for example, on malaria in French armies, with sensitivities between 49.6% and 56.2% [26], and Legionnaire's disease in France, with sensitivities between 58% and 66% [35]. This implies that the sensitivity of health monitoring systems in French Guiana is no worse than that of similar French institutions concerning different diseases.

It is possible to obtain an overall picture of the TB incidence in the whole of French Guiana based on the fact that 61.7% of cases in the Ile-de-Cayenne were culture-confirmed. If the same ratio is applied to the 345 confirmed cases censused by the IPG throughout the region during the same time period, the total number of TB would be about 559, i.e. an average annual incidence rate of 44.4 TB cases per 100,000 population. This value is undoubtedly underestimated, given that the TB incidence has increased in French Guiana over the past five years. Nevertheless, such an estimate is higher than the incidence previously reported by the InVS: 23.8 per 100,000 population per annum on an average for the period 1996–2003, with a maximum of 39.4 in 2000 (http://www.invs.sante.fr). The number of TB cases actually occurring annually since 1996 in French Guiana has been estimated so as to enable comparison between monitoring system estimates and reality (Table 4 and Fig. 2). In 2003, for example, the number of biologically confirmed cases (from IPG) was 66, providing an estimate of $66 \times 100/61.7 = 107$ cases in French Guiana. The population size in French Guiana was evaluated at 176,117 inhabitants in 2003, hypothesizing a linear monotonous increase in the population between the 1990 and 1999 censuses; therefore, the 2003 incidence may be almost 60.8 TB cases per 100,000 population, showing a strong increase in the TB threat over the last few years. Moreover, if the actual number of cases (including missing cases) is presumed to be 11.6% higher, as evaluated by the sample coverage approach, the estimated 2003 incidence would be 67.8 per 100,000 population. Fig. 2 provides a comparison of the degree of ascertainment of the different TB monitoring systems, taking into account the number of cases reported.

When interpreting our findings, certain limitations should be noted. First, the final calculated TB incidence in French Guiana is an approximate result based on different assumptions. However, from a public health point of view, it is important to estimate the number of people who actually have TB and are thus potentially contagious. The aim of our evaluation

Table 4
Comparisons between obligatory notifications of (TB) from governmental institutions and the estimated number of cases that should have actually occurred (1) in the Ile-de-Cayenne and (2) in all of French Guiana from 1996 through 2003

| Year[a] | Ile-de-Cayenne | | French Guiana | |
|---|---|---|---|---|
| | DSDS[b] | IPG × DIM × LAT[c] | InVS[d] | est[e] |
| 1996 | 22 | 38 | 43 | 50 |
| 1997 | 11 | 41 | 24 | 70 |
| 1998 | 4 | 28 | 6 | 55 |
| 1999 | 13 | 35 | 19 | 55 |
| 2000 | 33 | 46 | 66 | 60 |
| 2001 | 34 | 62 | 61 | 81 |
| 2002 | 21 | 61 | 43 | 81 |
| 2003 | 17 | 70 | 37 | 107 |
| Total | 160 | 381 | 299 | 559 |

[a] To homogenize the results between sources and enable comparisons, notifications from DSDS were corrected, excluding two declared atypical cases and revising the year of notification by the year notified at the beginning of treatment.

[b] DSDS, number of TB cases notified by the Département de la Santé et du Développement Social in the Ile-de-Cayenne.

[c] IPG × DIM × LAT, number of matching cases from IPG, DIM and LAT in the Ile-de-Cayenne.

[d] InVS, number of TB notifications from Health Monitoring Institute for all of French Guiana.

[e] est, Estimated number of total cases, considering that confirmed cases represent 61.7% of the total report (without any capture-recapture correction).

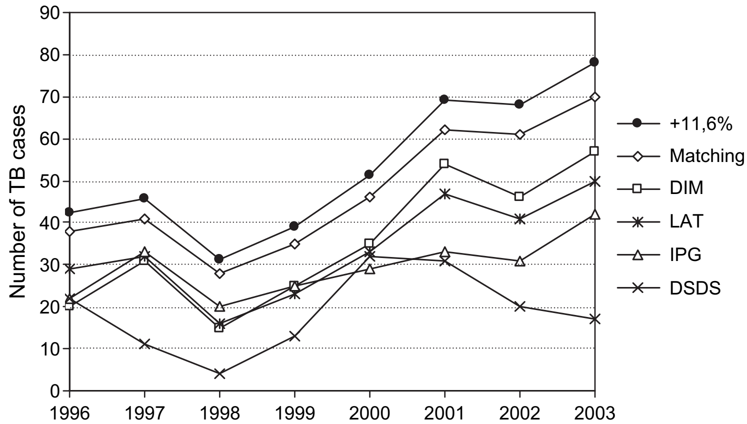


Fig. 2. Evolution of the number of TB cases in the Ile-de-Cayenne between 1996 and 2003 according to the different monitoring systems, i.e. IPG (Institut Pasteur de la Guyane), DIM (Département d'Information Médicale), LAT (Service de Lutte Anti-Tuberculeuse), and the DSDS (Département de la Santé et du Développement Social) which rebroadcast obligatory notifications to the InVS. So-called ''cross-linking'' data corresponding to a simple match between data from the IPG, DIM and LAT are also given. ''+11.6%'' is the estimated number of TB cases per year, taking into account the sample coverage approach, which gives 11.6% more TB cases than after simple matching of the three sources.

was not to provide an exact incidence rate, but to give an order of magnitude, which is already a significant advance compared to biased notified data.

A second limitation is related to assumptions inherent in capture-recapture methods. In this study, the closure assumption, i.e. the constant size of the population during the entire study period, is difficult to evaluate, as is the assumption of equal probabilities of being observed (captured) in any sample. Further research is required to evaluate the bias induced by not taking into account these assumptions and to identify the minimum data sources required using capture-recapture. Nevertheless, in spite of this remark, the combined use of cross-linkage over three sources, along with capture-recapture methods, could help to adjust for underestimates in order to achieve a more accurate view of risks and risk factors.

Finally, we should underline that our estimates do not provide an idea of the proportion of patients reported in the three monitoring systems who are presumed to have at least begun treatment and who have been cured. This rate could also be estimated, since treatment failure strongly affects TB dynamics, with the development of drug-resistant cases, which may remain infectious for long periods of time, thus maintaining the chain of transmission.

## Acknowledgements

We would like to thank the physicians and health care professionals who provided generous help in reporting cases in the Ile-de-Cayenne: Dr. Guillot and Dr. Quintard from the Service de Lutte Anti-Tuberculeuse of Cayenne; Dr. Sobeski from the Département d'Information Médicale of the Cayenne Hospital; Dr. Duval from the Laboratory of the Institut Pasteur de la Guyane; Dr. Rastogi, Dr. Sola and K. Brudey from the Mycobacteria Reference Center of the Institut Pasteur de la Guadeloupe; and Dr. Huguet, Dr. Cottrelle and N. Joly from the Département de la Santé et du Développement Social (DSDS) of French Guiana. We would also like to thank the personnel of the ''Main Entry-Reception'' of Cayenne Hospital, and E. Miatti, of the Médecine B Service of the hospital, who gave of their time. We are grateful to M. Choisy and O. Gimenez for useful discussions. V.G. is supported by a PhD fellowship from the French Ministry of Higher Education and Research, J.-F.G. by both the IRD and the CNRS, and X.D. by the Institut Pasteur.

## References

[1] C. Dye, S. Scheele, P. Dolin, V. Pathania, M.C. Raviglione, Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project, JAMA 282 (1999) 677–686.

[2] K. Neville, A. Bromberg, R. Bromberg, S. Bank, B.A. Hanna, W.N. Ross, The third epidemic: multi-drug resistant tuberculosis, Chest 105 (1994) 45–48.

[3] C.S. Lambregts-van Weezenbeek, J. Veen, Control of drug-resistant tuberculosis, Tuber. Lung Dis. 76 (1995) 455–459.

[4] W.W. Yew, C.H. Chau, Drug-resistant tuberculosis in the 1990s, Eur. Respir. J. 8 (1995) 1184–1192.

[5] World Health Organisation, Anti-Tuberculosis Drug Resistance in the World: The WHO/IUTLD Global Project on Anti-Tuberculosis Drug Resistance Surveillance 1994–7, WHO, Geneva, 1997.

[6] P. Farmer, J.Y. Kim, Community based approaches to the control of multidrug resistant tuberculosis: introducing ''DOTS-plus'', Br. Med. J. 317 (1998) 671–674.

[7] J. Pillaye, A. Clarke, An evaluation of completeness of tuberculosis notification in the United Kingdom, BMC Public Health 3 (2003) 31.

[8] C.D. Sheldon, K. King, H. Cock, P. Wilkinson, N.C. Barnes, Notification of tuberculosis: how many cases are never reported? Thorax 47 (1992) 1015–1018.

[9] G.B. Migliori, A. Spanevello, L. Ballardini, M. Neri, C. Gambarini, M.L. Moro, et al., Validation of the surveillance system for new cases of tuberculosis in a province of northern Italy. Varese Tuberculosis Study Group, Eur. Respir. J. 8 (1995) 1252–1258.

[10] R.M. Washko, T.R. Frieden, Tuberculosis surveillance using death certificate data, New York City, 1992, Public Health Rep. 11 (1996) 251–255.

[11] A.K. Mukerjee, Ascertainment of non-respiratory tuberculosis in five boroughs by comparison of multiple data source, Commun. Dis. Public Health 2 (1999) 143–144.

[12] International Working Group for Disease Monitoring and Forecasting (IWGDMF), Capture-recapture and multiple-record systems estimation II: applications in human diseases, Am. J. Epidemiol. 142 (1995) 1059−1068.

[13] N. Rastogi, L. Schlegel, F. Pfaff, I. Jeanne, C. Magnien, G. Lajoinie, et al., La tuberculose en région Antilles-Guyane. Situation épidémiologique de 1994−1996, BEH 11 (1998) 45−47.

[14] H.L. Rieder, J.M. Watson, M.C. Raviglione, M. Forssbohm, G.B. Migliori, V. Schwoebel, et al., Surveillance of tuberculosis in Europe. Recommendations of a Working Group of the World Health Organization (WHO) and the European Region of the International Union Against Tuberculosis and Lung Disease (IUATLD) for uniform reporting on tuberculosis cases, Eur. Respir. J. 9 (1996) 1097−1104.

[15] V. Schwoebel, H.L. Rieder, J. Watson, M. Raviglione, for the working group for uniform reporting on TB cases in Europe, Surveillance of TB in Europe, Euro. Surveill. 1 (1996) 5−8.

[16] J.N. Darroch, The multiple-recapture census I: estimation of a closed population, Biometrika 64 (1958) 435−447.

[17] G.A.F. Seber, The estimation of animal abundance and related parameters, second ed., Griffin, London, 1982.

[18] J.D. Nichols, K.H. Pollock, Estimation methodology in contemporary small mammal capture-recapture studies, J. Mammal. 64 (1983) 253−260.

[19] A.E. Magurran, Ecological Diversity and Its Measurement, Chapman & Hall, London, 1988.

[20] Mathematical Sciences Education Board, National Research Council, Measuring What Counts: A Policy Brief, National Science Teachers Association, Washington DC, 1993.

[21] T.J. Wittes, W.V. Sidel, A generalisation of the simple capture-recapture model with application to epidemiological research, J. Chronic Dis. 21 (1968) 287−301.

[22] J.T. Wittes, T. Colton, V.W. Sidel, Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources, J. Chronic Dis. 27 (1974) 25−36.

[23] A. Chao, in: P. Armitage, T. Colton (Eds.), Encyclopedia of Biostatistics, vol. 1, Wiley, New York, 1998, pp. 482−486.

[24] E.B. Hook, R.R. Regal, Capture-recapture methods in epidemiology: methods and limitations, Epidemiol. Rev. 17 (1995) 243−264.

[25] P. Bernillon, L. Lievre, J. Pillonel, A. Laporte, D. Costagliola, Groupe d'épidémiologie clinique des CISIH, Estimation de la sous-déclaration des cas de Sida en France par la méthode de capture-recapture, BEH 97 (1997) 19−21.

[26] X. Deparis, B. Pascal, D. Baudon, Evaluation of the completeness of the epidemiological surveillance system for malaria by the capture-recapture system in the French armies in 1994, Trop. Med. Int. Health 2 (1997) 433−439.

[27] D.J. Nanan, F. White, Capture-recapture: reconnaissance of a demographic technique in epidemiology, Chronic Dis. Can. 18 (1997) 144−148.

[28] K. Tocque, M.A. Bellis, N.J. Beeching, P.D. Davies, Capture-recapture as a method of determining the completeness of tuberculosis notifications, Commun. Dis. Public Health 4 (2001) 141−143.

[29] Y. Ding, S.E. Feinberg, Multiple sample estimation of population and census undercount in the presence of matching errors, Surv. Methodol. 22 (1996) 55−64.

[30] J. Iñigo, A. Arce, J.M. Martin-Moreno, R. Herruzo, E. Palenque, F. Chaves, Recent transmission of tuberculosis in Madrid: application of capture-recapture analysis to conventional and molecular epidemiology, Int. J. Epidemiol. 32 (2003) 763−769.

[31] A. Chao, P.K. Tsay, S.H. Lin, W.Y. Shau, D.Y. Chao, Tutorial in biostatistics: the application of capture-recapture models to epidemiological data, Stat. Med. 20 (2001) 3123−3157.

[32] Mathsoft, S-Plus User's Manual, Version 4.0, Mathsoft Inc, Seattle, 1997.

[33] S.E. Fienberg, The multiple-recapture census for closed populations and incomplete 2k contingency tables, Biometrika 59 (1972) 591−603.

[34] E.B. Hook, R.R. Regal, Internal validity analysis: a method for adjusting capture-recapture estimates of prevalence, Am. J. Epidemiol. 142 (1995) 48−52.

[35] A. Infuso, B. Hubert, J. Etienne, Under-reporting of legionnaires' disease in France: the case for more active surveillance, Euro. Surveill. 3 (1998) 48−50.