

Estimation of underreporting of AIDS cases in a Brazilian Northeast metropolis

Abstract

Underreporting is one of the main challenges epidemiologic AIDS surveillance is faced with. The objective of this study was to estimate the level of underreporting of AIDS in adults in Fortaleza in the 2002-2003 period. In addition, the level of underreporting at two referral hospitals (Hospital São José – HSJ, and Hospital Geral de Fortaleza – HGF) was estimated. The study used the capture-recapture method and relied on three secondary databases: SINAN (national disease surveillance), SISCEL (laboratory test control), and SIM (mortality information). SINAN was compared to SISCEL and to SIM. Cases confirmed by SINAN were considered as reported. Cases from the databases were paired using the *RecLink II* software. Subsequently, cases eligible for the capture-recapture method were selected using the Lincoln-Petersen and Chapman estimators. The levels of underreporting were estimated at 33.1% and 14.1% for SISCEL and SIM, respectively. Underreporting for SISCEL was 5.4% at HSJ and 90.5% at HGF, comparing Siscel to Sinan. The study shows a considerable level of underreporting of AIDS cases in Fortaleza, and suggests that SISCEL is an important source of AIDS reporting considering that it allowed detecting levels of underreporting more than twice the estimates derived from SIM and SINAN databases. Considering the findings of the present study, SINAN-AIDS should be compared periodically with all relevant information systems in order to reduce levels of AIDS underreporting.

Keywords: Acquired Immunodeficiency Syndrome. Epidemiology. Disease reporting. Epidemiological surveillance. Epidemiological surveillance services. Information systems. Estimation technique

Valéria Freire GONÇALVES^{1,2}

Lígia Regina Franco Sansigolo KERR²

Rosa Maria Salani MOTA³

João Maurício Araújo MOTA³

¹Núcleo de Epidemiologia da Secretaria da Saúde do Estado do Ceará.

²Departamento de Saúde Comunitária da Universidade Federal do Ceará.

³Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará.

Financial support: Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico - FUNCAP/Processo: Nº. 0578/06

Mailing address: Rua Paulo Morais, 175 aptº 501 – Papicu, Fortaleza, Ceará – CEP: 60.175-175. Fone: (XX85)32627461/91372124 e-mail:val@saude.ce.gov.br

Introduction

In Brazil, the main purpose of AIDS epidemiological surveillance has been to follow up the temporal and spatial development of the disease, the infections and risk behaviors, aimed at guiding prevention and control measures¹. Surveillance develops its activities using universal reporting of AIDS cases that fulfill the criteria established by the Ministry of Health as a reference. The development of AIDS criteria definitions accompanies technological advances and their availability². When AIDS case reporting is relatively adequate and based on the natural history of the disease, it allows monitoring the development of the epidemic retrospectively. It is also important to validate data generated by sentinel surveillance systems, and to guide activities related to prevention and planning of needs in care and disease treatment³.

Nevertheless, some factors are responsible for the questions on the number of AIDS cases reported, both qualitatively and quantitatively. Among those factors, stand out: lack of reporting by medical professionals, especially those in the private network, delay in investigation, besides the low quality of the information collected which enters the National Disease Surveillance System – SINAN. Quite often, the collecting procedure disregards the guidelines of epidemiological surveillance process and adds a bureaucratic and limiting outlook to health services⁴⁻⁵⁻⁶.

The low level of investigations may be confirmed by the number of inconsistencies found in SINAN, as well as by the amount of unknown information, which jeopardizes the epidemiological analysis. Concerning the lack of AIDS records, one of the facts that most contribute to that has been the disorganization of epidemiological surveillance systems, besides the stigma attached to the disease itself⁵. All those factors reduce the usefulness of reporting as a tool to follow up the magnitude and course of the epidemic.

In 2004, aiming at decreasing AIDS underreporting and delayed reporting, the

Ministry of Health established a connection between the SINAN database up to July 2004 and cases reported at SISCEL, that presented a number of TCD4+ lymphocytes under 350 cells p/mm³. Among the cases reported at SISCEL, 50.5% did not exist in SINAN. In the state of Ceará, 15% of the cases reported at SISCEL, but not yet reported at SINAN, were identified⁷. Therefore, in spite of the advances towards better quality and higher comprehensiveness of health information, and implementation of epidemiological surveillance, there is still significant underreporting of this disease in the state of Ceará. These results reinforce the need of routinely using other complementary epidemiological surveillance systems.

The incidence curve of AIDS in Ceará is still ascendant, with 166 (90.2%) municipalities reporting the disease. The epidemiological situation is characterized by poverty, inland localization and predominance of women, but more than 50% of cases are still in the men who have sex with men⁸. In spite of that, the epidemiological surveillance should look for alternatives to improve its performance, appropriately reinforcing identification and reporting of cases, in order to guide disease prevention and control actions. Through the information obtained, epidemiological surveillance should concentrate efforts to develop effective actions that suit the epidemiological reality identified⁹. Therefore, this study aimed to estimate AIDS case underreporting of individuals aged thirteen years or over, living in Fortaleza, from 2002 to 2003, thus contributing to improve the surveillance profile of this disease, in order not only to evaluate the magnitude and course of the epidemic, but also to redirect prevention and control measures.

Method

Study and population design

Cross-sectional, observational, epidemiological study comprising AIDS cases in individuals thirteen years old or over, living

in the city of Fortaleza, diagnosed in 2002 and 2003, and recorded in the databases searched, i.e., SINAN, SISCEL and SIM.

During the study period, six hospitals were treating AIDS cases in the city of Fortaleza. One of them provided outpatient and inpatient care (Hospital São José – HSJ) and another only outpatient care (Hospital Geral de Fortaleza – HGF). These hospitals were reference for AIDS in the state and concentrated the highest number of cases of the disease. Their underreporting was estimated. For two other hospitals, only the underreporting rate was calculated, because there were not enough cases to calculate the estimative. In the two remaining hospitals, it was not possible to assess any parameter, as there were no cases reported in SISCEL.

Data source

Data came from three sources: SINAN, SISCEL and SIM, where reports of AIDS cases may be found. The Ceará State Health Department provided the data from SINAN and SIM, and the Ministry of Health, the data from SISCEL. All data from SINAN were analyzed, with 6,410 AIDS cases reported, 6,007 (93.7%) of them confirmed, and 403 (6.3%) discarded, from 1983 to July 2005. The purpose of working with all this data was the probability of identifying AIDS cases found in SISCEL and SIM, in 2002 and 2003, which had already been diagnosed either prior to the years selected for the study, or after that period. Such cases were excluded using the capture-recapture method. The inclusion criteria for SINAN cases were diagnosis date in 2002 and 2003, and residential address in the city of Fortaleza. This way, of the 1,418 cases diagnosed in SINAN in 2002 and 2003, 576 were excluded, because the address was inland state, 142 were discarded, and 12 duplicated, therefore remaining 688 cases eligible for the study. Regarding SISCEL, inclusion criteria were cases with $TCD4+ > 350 \text{mm}^3$ in 2002 and 2003, with an address in the city. Initially, 767 cases were taken into consideration and from those, 271 were excluded, because the address re-

ported was inland state, 30 because subjects were under 13; 34 cases because the address reported in SINAN source was inland state, 6 because the exams had been requested by other states, and 185 because they had already been diagnosed and reported in SINAN in years other than 2002 and 2003. So, only 241 cases were eligible for the study. Concerning SIM, there were 405 deaths that had occurred in 2002 and 2003, primarily because of AIDS, according to ICD-10, whose codes vary from B20 to B24. Initially the 405 deaths reported in Ceará in 2002 and 2003 were selected. From this number, 175 were excluded because of inland state addresses, 2 because subjects were under 13, 8 because cases were duplicated, 7 with inland addresses in SINAN, and 81 because deaths had already been reported in SINAN in years other than 2002 and 2003. Therefore, only 135 deaths were eligible, 132 from SIM with addresses in Fortaleza, and another 3 with inland addresses on the Death Certificate, but considered as living in Fortaleza, according to SINAN.

Development and validation of databases

For the development and validation of databases, an Excel file was created, with variables selected in order to refine and identify inconsistencies, aiming at preparing data for the pairing of the sources. Information such as birth date, mother's name, and residential address was collected and standardized for all sources, using data from SINAN as reference. The selected variable inconsistencies were corrected through investigation such as: checking the completeness of the data recorded in the available banks, i.e., SINAN, SISCEL, SIM, and investigation in the health units that reported the cases or recorded the deaths. With the complete data, no inconsistencies, and duplicated data excluded, the identifying variables (patient's name, age, birth date and mother's name) were selected at the SINAN/SISCEL and SINAN/SIM databases for the list of cases in the ReLinkII Program.

In the RecLinkII Program, the blocking used initially for the pairing of cases, was the combination of the patient's last surname and first name, using SINAN as a reference source and SISCEL and SIM as comparative sources, i.e., all the cases from SISCEL and SIM checked one by one, among total cases selected at SINAN in 2002 and 2003. After crossing databases, true pairs, doubtful pairs and non-pairs were identified. Doubtful pairs were checked manually with the purpose of defining them as true or non-pairs; the criteria for classifying them as true being the cases in which all the identifying variables agreed, or in the case of similar names with small discrepancies, but with the same birth date or age, and the same mother's name. After those procedures, two new databases were prepared, Excel, SISCEL/SINAN and SIM/SINAN considered as the final bases.

These bases were prepared with the purpose of effectively defining all variables, so that every captured list would be able not only to fulfill the assumptions for the use of the capture and recapture method, but also to estimate underreporting, that is: 1) a closed population, i.e., stable in its size all along the capture period; 2) each captured individual very well defined, to facilitate his/her identification for recapture; 3) independent samples, i.e., the possibility of an individual being in a list should not influence the possibility of him/her being or not in another list. When this method is applied in Epidemiology, it means that each list used, i.e., each information source is called a capture episode⁷.

Statistical analysis

To estimate underreporting of AIDS cases in the city of Fortaleza, in 2002 and 2003, the Lincoln-Petersen Estimator was the capture and recapture method used, obtained by the moment statistical method¹⁰ and the Chapman estimator, which is a modified version of the Lincoln-Petersen estimator. Lincoln-Petersen and Chapman estimators were designed assuming that samples in two

stages are selected without reposition¹¹. The Lincoln-Petersen estimator presupposes the probability of capture to vary according to the source, and independence among captures being defined as:

$$N = \frac{A \times B}{C}$$

In the formula A and B are the captured elements, C is the number of recaptured elements, and N the unknown value, which is being estimated. As the number of recaptured elements grows, N decreases. So, the larger the recaptured number, the lower the estimate of the true value of N.

This is known as the Lincoln-Petersen estimator¹². See Venn's diagram in Figure 1 to better understand the procedure.

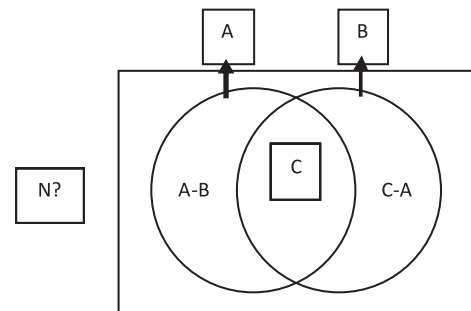


Figure 1. Venn's Diagram for two sources

An overestimate of the N value may occur when the Lincoln-Petersen estimator is used for samples that are not large enough. That is why the Chapman estimator has been the most appropriate for *Hospital Geral de Fortaleza* health unit – HGF, as it allows an approximation of the former estimator, allowing an unbiased N estimate, with independence between capture and recapture¹³. The underreporting evaluation, with confidence interval (CI) of 95% was performed using Excel Office 2000 Program.

Ethical aspects

This research was submitted to the ethical committee for research, from *Universi-*

dade Federal do Ceará, and was approved under number 308/05.

Results

There were 1,604 AIDS cases considered as eligible for the research, distributed in the three sources: 688 in SINAN; 241 in SISCEL and 135 deaths in SIM. Data analysis concerning the three sources, SINAN, SISCEL and SIM has shown that only 10 cases were common to the three of them, 151 cases were in SINAN and SISCEL, 106 in SINAN and SIM, and only one in SISCEL and SIM. Regarding the number of cases found in a single source, 421, 79 and 18 cases were identified in SINAN, SISCEL and SIM, respectively (figure 2).

For the underreporting analyses, using the Lincoln-Petersen estimator for SISCEL and SIM, 1,029 and 801 cases were estimated, respectively. The values were about

the same, using the Chapman estimator for the same sources, i.e., 1,028 and 800 respectively. Taking into account that the number of cases reported in SINAN, at that period, were 688, 341 and 113 cases not been reported in SISCEL and SIM, respectively, when the Lincoln-Petersen estimator was used, and 137, when Chapman was used. Therefore, there was practically no difference between the two estimators used. Underreporting in the period was 33.1% and 14.1%, for SISCEL and SIM respectively, using SINAN as a reference (Table 1).

Considering underreporting for Hospital São José (688 cases reported in SINAN and 324 in SISCEL), 706 cases were assessed by the two estimators, Lincoln-Petersen and Chapman. Underreporting, at this hospital, was 5.4%. Concerning HGF (13 cases in SINAN and 81 in SISCEL), 151 cases were assessed by the Lincoln-Petersen estimator and 137, by Chapman's (Table 2). As already

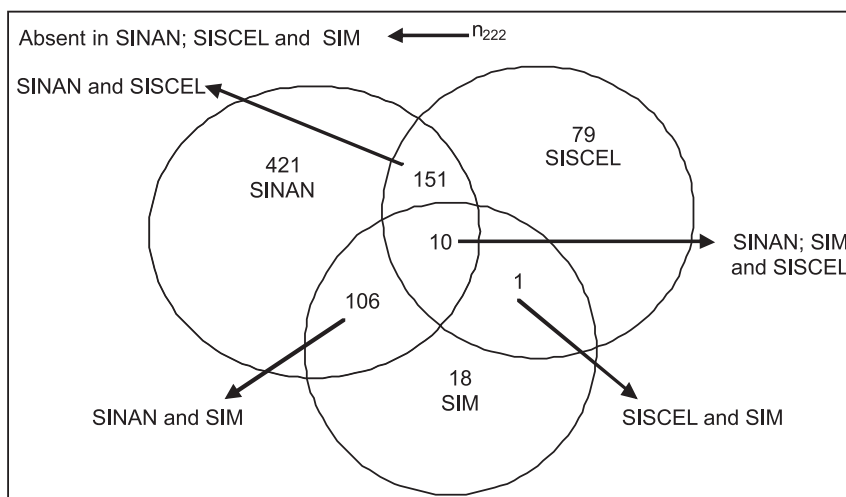


Figure 2. Venn's Diagram – SINAN / SISCEL and SIM

Table 1. Point and interval estimates of AIDS underreporting in SISCEL and SIM, using Lincoln-Petersen and Chapman estimators. Fortaleza, 2002-2003.

Estimator	SINAN / SISCEL		Underreporting % (IC 95%)	SINAN / SIM		Underreporting % (IC 95%)
	N (IC 95%)	SD		N (IC 95%)	SD	
Lincoln e Petersen	1,029 (937 – 1,121)	46.8	33.1 (30.2 – 36.0)	801 (746 - 855)	27.9	14.1 (11,7 - 16,5)
Chapman	1,028 (949 – 1,107)	40.5	33.1 (30.2 - 35.9)	800 (751 - 849)	25.1	14.0 (11.6 - 16.4)

Table 2. Point and interval estimates of AIDS underreporting based on SISCEL database per reporting facility, using Lincoln-Petersen and Chapman estimators. Fortaleza, 2002-2003.

Estimador	HSJ*			HGF**		
	N (IC 95%)	sd	Underreporting % (IC 95%)	N (IC 95%)	sd	Underreporting % (IC 95%)
Lincoln e Petersen	706 (680 - 732)	13.30	5.4 (3.7 - 7.0)	151 (24 - 277)	64.5	91.4 (86.9 - 95.9)
Chapman	706 (683 - 729)	11.55	5.4 (3.7 - 7.0)	137 (64 - 210)	37.3	90.5 (85.6 - 95.4)

* Hospital São José; ** Hospital Geral de Fortaleza

stated, an overestimate of the N value may occur, when the Lincoln-Petersen estimator is used for small samples, as in the case of the HGF. Therefore, an estimated value of 137 cases was considered, with an underreporting of 90.5% (Table 2). In the other health units where it was not possible to estimate AIDS underreporting, due to the small number of cases, only the underreporting rate was calculated, resulting in 83.3% for *Hospital Universitário Walter Cantídio* (18 cases reported in SISCEL and only 3 in SINAN), and 100% for *Hospital Distrital Gonzaga Mota de Messejana* (3 cases in SISCEL and none in SINAN).

Discussion

In spite of the major effort towards reducing AIDS underreporting in Ceará, the present results show important AIDS underreporting in adults, in the city of Fortaleza, in the years of 2002 and 2003.

As the purpose of this work was to assess underreporting, the discussion is going to start with the methods employed. Estimates were carried out using Lincoln-Petersen and Chapman's estimators, and practically no difference between both methods was observed, considering that a large sample was studied. When the methods were used for small samples, such as the evaluation in a health unit with a small number of cases, there was a significant difference between the results obtained.

Concerning the underreporting rate assessed per source, results have shown that underreporting with SISCEL/SINAN was twice higher than SIM/SINAN. Therefore,

SISCEL is an important source for assessing AIDS underreporting. A recent Brazilian study listed the databases from the SINAN, SISCEL and SICLOM systems, through a Reclink system⁷. Nevertheless, this study did not assess underreporting by the capture and recapture method. Even though other studies used the mentioned method, they used other sources, such as the *Sistema de Informações Hospitalares do SUS*, *Sistema de Informações sobre Mortalidade*, hospital patient's files, data from committees of hospital infection control, and from doctor's offices⁸⁻¹⁵.

As to the underreporting rate found, other studies that used the same method and compared other sources came to different results. For example, in a study carried out in six Brazilian cities, the underreporting found varied from 24% to 65% among them, even though they used the same method and compared SIH with SINAN^{7,15}. OLIVEIRA¹⁵ found 68% of underreporting, when studying AIDS underreporting in 1995-1996 in Belo Horizonte, using the capture and recapture method. Another study that can be mentioned is BUCHALLA's⁴, in which underreporting in São Paulo from 1983 to 1986 resulting from the comparison of death certificates, was 15% of non-reported deaths. In a recent study, the state of São Paulo evaluated the number of underreported deaths in SINAN by year, finding rates from 5.2% to 17%, from 1980 to 2005¹⁷. Although these studies show the existence of AIDS underreporting, the comparison between them is limited, because of the diversity of the data sources studied. Nevertheless, despite the high amount of underreporting

presently found in Fortaleza, it is lower than the ones reported in other cities, except for the rate of 24% found in Florianópolis. It should be pointed out that in the two last studies mentioned, just the underreporting rate, not the estimate, was evaluated by probability methods.

In the analysis of underreporting per unit, the two main health units that treat AIDS patients in Fortaleza, HSJ and HGE, presented very different underreporting rates of AIDS cases from the others. This difference may be due to several factors, such as the presence of an Epidemiological Surveillance Unit that has developed experience since the 1980's and is far more demanded by AIDS patients for treatment than the remaining reporting units, in addition to having a central medication dispensing unit located within the facility during the whole study period.

In the two remaining reporting units, *Hospital Universitário Walter Cantídio* and *Hospital Distrital Gonzaga Mota de Messajana*, underreporting rates are alarming, given the hospital units included in this research are considered reference units for

the follow-up and treatment of AIDS.

Conclusion

In spite of all the efforts by the agencies responsible for the epidemiological surveillance in all government spheres, the conclusion is that there still is high AIDS underreporting in Fortaleza. These findings indicate that Ceará state and Fortaleza Health Departments should put some pressure on the professionals working with the diagnoses and treatment of AIDS, so that all confirmed cases are duly reported. Regarding epidemiological surveillance, better quality information should be attained, using all data sources available, mainly SISCEL, which is a useful reference as an AIDS surveillance database; and pairing laboratory data in the service routine should also be implemented. Finally, the present results are expected to contribute to reduce AIDS underreporting in Fortaleza, which may allow for a better understanding of the magnitude of the AIDS epidemics in Fortaleza and in Ceará.

References

1. Ministério da Saúde (BR). Secretária de Vigilância em Saúde. Guia de vigilância epidemiológica. 6. a ed. Brasília; 2005.
2. Ministério da Saúde (BR). Secretaria de Vigilância em Saúde. Programa Nacional de DST e AIDS: critérios de definição de casos de AIDS em adultos e crianças. Brasília, 2004a.
3. Barbosa MTS, Struchiner CJ. Estimativas do número de AIDS: comparação de métodos que corrigem o atraso da notificação. In: Ministério da Saúde (BR). Coordenação Nacional de DST e AIDS. Simpósio satélite: a epidemia de AIDS no Brasil: situação e tendências. Brasília, 1997. p. 15-26.
4. Buchalla CM. A AIDS/SIDA: as estatísticas de mortalidade como fonte de informações. São Paulo: Centro da OMS para Classificação de Doenças em Português; 1990. (Série de Divulgação, n. 6).
5. Ministério da Saúde (BR). Secretaria de Políticas de Saúde. Coordenação de DST e AIDS. Vigilância do HIV no Brasil: novas diretrizes. Brasília; 2002.
6. Carvalho DM. Grandes sistemas nacionais de informações em saúde: revisão e discussão da situação atual. *Inf Epidemiol SUS* 1997; 6(4):7-46.
7. Lucena FFA, Fonseca MGP, Sousa AA, Coef CM. O relacionamento de bancos de dados na implementação da vigilância da AIDS. Relacionamento de dados e vigilância da AIDS. *Cad de Saúde Coletiva*, 14(2):305-312, 2006
8. Ministério da Saúde (BR). A subnotificação de casos de AIDS em municípios brasileiros selecionados: uma aplicação do método de captura-recaptura. *Bol Epidemiol AIDS* 2004b; 18(1): 7- 11.
9. Secretaria da Saúde do Estado (CE). Informe epidemiológico AIDS. Fortaleza, 2005.
10. Ferreira VMB, Portela MC, Vasconcelos MTL. Fatores associados à subnotificação de pacientes com AIDS, no Rio de Janeiro. *Rev. Saúde Pública*, v. 34, n. 2, p. 170-177, São Paulo, 2000.
11. Garthwaite PH, Jolliffe IT, Jones B. Statistical inference. New York: Prentice Hall; 1995.

12. Abuabara MAP, Petrele Júnior M. População aberta: o método de Jolly-Seber. In: Abuabara MAP, Petrele Júnior M. Estimativas de abundância de populações animais: introdução às técnicas de captura-recaptura. Paraná: EDUEM; 1997. cap. 6, p. 131-52.
13. LaPorte RE, McCarty DJ, Tull ES. Counting birds, bees and NCDs. *Lancet* 1992; 339: 494-5.
14. Hook EB, Regal RR. Effect of variation in probability of ascertainment by sources ("variable catchability") upon "capture-recapture" estimates of prevalence. *Am J Epidemiol.* 1993; 137:1148-66.
15. Oliveira MTC. A subnotificação de casos de AIDS em Belo Horizonte, Minas Gerais: uma aplicação da técnica de captura-recaptura [dissertação]. Belo Horizonte: Universidade Federal de Minas Gerais; 2000.
16. Buchalla CM. A Síndrome de Imunodeficiência Adquirida e a mortalidade masculina de 20 a 49 anos, município de São Paulo, 1983-1986. [tese]. São Paulo: Faculdade de Saúde Pública da Universidade de São Paulo; 1993.
17. Secretaria de Estado da Saúde(SP). A Vigilância Epidemiológica da AIDS no Estado de São Paulo (dados até 30/06/2005), *Bol Epidemiol C.R.T. DST/AIDS C.V.E.* 2005; 24(1):25.