Results from a Method for Estimating HIV Incidence Based on the First Cd4 Count among Treatment-Naïve Cases: Brazil, 2004-2013

Célia Landmann Szwarcwald[1]*, **Paulo Roberto Borges de Souza Júnior[1]**, **Ana Roberta Pati Pascom[2]** and **Orlando da Costa Ferreira Júnior[3]**

[1]*Institute of Communication and Information Science and Technology in Health, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil*
[2]*Department of STD, AIDS and Viral Hepatitis, Ministry of Health, Brazil*
[3]*Institute of Biology, Federal University of Rio de Janeiro, Brazil*

Abstract

Background: This paper introduces a method to estimate HIV incidence in Brazil using surveillance data. The interest is to estimate the annual lag (time from infection to reporting) distribution among incident cases in a given year with observations arising from a right truncated version of the distribution.

Methods: For each treatment-naïve HIV case aged 15 years and over reported from 2004 to 2015 we estimated the time since infection based on a statistical model that relates the first CD4 count to time since infection. Under the assumption the lag distribution is expressed by a logistic probability distribution, we estimated HIV incidence as the upper limiting value of the logistic function. Since this approach requires at least eight observations per year, to estimate HIV incidence in recent years (2009-2013), we used linear regression models to estimate the missing observations for these years due to truncation in 2015. Using this approach, HIV incidence was estimated from 2004 to 2013, separately for males and females.

Results: In 2013, HIV incidence among people aged 15 years and over was estimated to be 44827 (95% CI 41143-47987), 32459 men (95% CI: 29775-34642) and 12368 women (95% CI: 11368-13345). Results from 2004-2013 have shown an increase among men and a slow decrease among women. The estimated proportion of cases reported less than one year after infection increased from 24.6% to 35.2%, among men, and from 35.1% to 45.8%, among women. For men who became infected from 2004 to 2013, 35.6% of the cases were not reported by year 2015 and among women, 23.2%.

Conclusion: The delay between HIV infection and diagnosis is of concern. Designing interventions to motivate testing is essential, especially among most-at-risk groups, as the faster HIV infected cases are linked to care, the faster the HIV incidence curve will turn downward.

Keywords: Surveillance; HIV incidence; Estimation; CD4 count; Depletion model; Diagnosis delay; Brazil

Introduction

Over the past two decades, antiretroviral therapy has brought enormous health benefits to people living with HIV/AIDS (PLWHA), with remarkable increase in survival [1]. To ensure effective implementation of prompt access to treatment and other prevention interventions, it is essential to have estimates of HIV incidence for describing the current dynamics of the epidemic [2,3]. Trends in HIV incidence indicate the degree to which HIV transmission is controlled [4], which groups are most at risk for HIV infection and help to identify the emergence of new sub-epidemics in the general population [5].

HIV incidence is therefore the most valuable indicator for epidemiological surveillance, both for planning prevention activities and for monitoring ongoing interventions effectiveness [6]. Despite its relevant role in surveillance, estimating the annual number of new infections remains challenging in many countries [7,8].

Historically, calculation of HIV incidence has been based on reported AIDS cases, using back-calculation model of AIDS incidence, with the assumption that temporal trends in AIDS incidence reflect trends in HIV incidence in the past [9]. However, the expansion of antiretroviral therapy has lengthened the time to the onset of AIDS [10] making inferences about HIV incidence based on AIDS reported cases very limited. More recently, updates of back-calculation models relying on new HIV diagnoses have been applied for HIV incidence estimation [11-15]. One of the difficulties faced is to distinguish the contribution of changes in testing patterns in the trends of HIV reported cases [4].

Another approach for estimating HIV incidence is to measure the rate of seroconversion in a cohort of individuals at risk, not HIV infected, followed over time [16]. Apart from the difficulties in implementing this type of approach, this methodology is subject to selection bias of individuals who agree to participate and who remain in the study, and is affected by changes in risk behavior over time [8,17].

HIV incidence can also be calculated based on the change in HIV prevalence estimated at two points in time [18]. The assumption underlying this methodology is that the number of new infections is the number of prevalent cases at the second point in time minus the number of cases who survived between the two time points. This is the basis of the methodology used by UNAIDS to estimate key HIV indicators such as the number of people living with HIV, new infections, and AIDS deaths [19]. Limitations of this approach stem

from both estimates of prevalence that are subject to errors of sample surveys [20], especially for concentrated epidemics, as well as outdated estimates of mortality due to expansion of antiretroviral therapy and resultant increase in survival [21].

In the late 1990s, laboratory tests were developed to estimate HIV incidence in cross-sectional studies [22]. The algorithms are based on laboratory assays that identify if infections are recent, that is, if they occurred within a certain period of time after HIV infection. The main advantage of this type of approach is the use of a single blood sample collected at one point in time, as in cross-sectional surveys, which does not require follow up of subjects as in cohort studies.

This method has been widely used to estimate the incidence of HIV in several countries and different epidemiological settings [3,23-26], including HIV incidence estimation in two Brazilian cities [27]. However, validation studies have consistently shown that assay-based estimates of HIV incidence vary according to which assay is used to identify recent infections [28]. Recommendations have recently been made to use multi-assay algorithms that include multiple biomarkers, including viral load and CD4 count, to identify recent HIV infections and thus provide more accurate HIV incidence estimates [29].

Lately, methods based on the first CD4 count after HIV diagnosis have been developed to estimate HIV incidence in United Kingdom [4,15], France [11], Brazil [12] and United States [13,14]. The main assumption of these models is that among antiretroviral therapy (ART)-naïve individuals, CD4 cell counts decrease over time [30] and the time since infection can be ascertained by applying an estimated rate of CD4 count decline [31]. Although it is well known a small proportion of HIV cases experiences preservation of CD4 [32], these approaches have the advantage of using a CD4 count back-calculation model to estimate HIV incidence with routinely available data.

In this study, we propose a new method to estimate HIV incidence in Brazil in recent years. The method is based on the first CD4 count after HIV diagnosis among all treatment-naïve HIV infected cases reported to the Ministry of Health in the time-period 2004-2013.

Methods

The information source is Brazil's Laboratory Tests Control System (SISCEL), which is the national laboratory-based information system created to monitor CD4+/CD8+ T lymphocyte counts and HIV viral load. SISCEL is managed centrally by technical staff of the Brazilian Department of Sexually Transmitted Diseases, AIDS and Viral Hepatitis. Data include some characteristics of the patient, such as age, sex, and municipality of residence.

In the present study, we analyzed the SISCEL database after removing patients' identifiers. All treatment-naïve HIV infected cases 15 years or older who underwent a CD4 count for evaluation of treatment indication in the period 2004-2015 were included in the analysis. The project was approved by the Ethics Committee of the Oswaldo Cruz Foundation, Ministry of Health, Brazil.

CD4 depletion model

The method is based on a statistical model proposed in an earlier work by Lodi et al. [31] that relates the first CD4 count to time of HIV infection through a linear mixed model:

$$\sqrt{CD4_1} = b_0 + b_1 t \tag{1}$$

where t is the time from HIV infection to date of first CD4 count, $CD4_1$ is the first CD4 count, and the slope (b_1) and the intercept (b_0)

are random variables following normal distributions. In model (1), the mean values and the standard deviations of the slope and the intercept were estimated separately for combinations of sex, quartile of age at infection, and risk group.

To calculate the time since HIV infection among reported cases in Brazil, we started from the premise that among antiretroviral therapy (ART)-naïve individuals, CD4 cell counts decrease over time [30] and that the time since infection can be estimated by applying a CD4 count depletion model, developed as an adaptation of model [1].

To calculate the intercept (b_0), as SISCEL does not have information about risk group, we adapted model [1] and calculated the intercept by sex and age group (defined by quartiles of age at first CD4 count) using the distribution of risk group among AIDS cases in Brazil to weight the b_0 estimates within each sex and age group. To calculate the slope (b_1), we considered all treatment naïve cases reported to SISCEL with first CD4 count greater or equal than 500 and followed up at least one year before starting antiretroviral treatment. The slope (b_1) was estimated by sex and age group as the ratio between the difference in the square root of the CD4 counts and the time between the first CD4 count and the last CD4 count before treatment (Table 1).

Then, for each treatment-naïve HIV infected case aged 15 years and over reported to SISCEL, we estimated the time since infection (t) based on the linear model coefficients by sex and age group given in Table 1. To account for cases tested in the private sector, we weighted the SISCEL database with weights inversely proportional to the coverage of private health insurance by geographical area of residence [33].

Distribution of time since infection (t) among reported cases

For the analysis, the time since infection was expressed by the number of years from infection to the first CD4 count and the values of the variable t derived from the CD4 count depletion model were aggregated in intervals of years since infection (t<1; 1<=t<2; 2<=t<3; ….; 19<=t<20; t>=20 years). That is, if the square root of the first CD4 count was greater or equal to b_0, we estimated the year of infection as the same as the year of SISCEL reporting (t<1).

Intercept*		Coefficient
Sex	**Age at first CD4 count**	
Males	<=27	23.498
	28-34	23.018
	35-40	22.684
	>=41	22.294
Females	<=27	24.132
	28-34	23.755
	35-40	23.483
	>=41	23.169
Slope**		Coefficient
Sex	**Age at first CD4 count**	
Males	<=27	-1.0373
	28-34	-1.0356
	35-40	-1.0450
	>=41	-1.1505
Females	<=27	-0.9745
	28-34	-0.9698
	35-40	-1.0140
	>=41	-1.0784

* Adapted from the Lodi et al. model [31] for application in Brazil
** Estimated from Brazilian data considering all treatment naïve cases reported to SISCEL with first CD4 count greater or equal than 500 and followed up at least one year before starting antiretroviral treatment

Table 1: Coefficients of the CD4 depletion model applied to SISCEL data.

The annual frequency distribution of the time (years) from infection to the first CD4 count (t) among reported cases from 2004 to 2015 is represented mathematically in Table 2, where $y_{j,u}$ is the number of cases infected in year k and reported in year u=k+j-1, for u ranging from 2004 to 2015 and j from 1 to 21 (representing $\geq$ 20). For each year u, Y_u is the total number of reported cases ($Y_u = \sum_{j=1}^{21} y_{j,u}$).

HIV incidence estimation

In the context of studying SISCEL reported cases, individuals are observed at the time of a subsequent event (reporting) to infection, and at that time, the lag (time between infection and reporting) is ascertained using the CD4 depletion model (Table 1). The annual lag distribution among reported cases from 2004 to 2015 is known. However, to estimate HIV incidence in a given year, interest is to estimate the annual lag distribution among all incident cases of that year, with observations arising from a right truncated version of the distribution.

For the analysis, observations are the cumulative sum of cases reported to SISCEL less than one year after infection, less than two years after infection, and so on. That is, let $X_{k,k+j-1}$ be the total number of cases infected in year k and reported to SISCEL until year u=k+j-1. For example, for year 2004, $X_{2004,2015}$ is the sum of incident cases in 2004 and reported in 2004, 2005,…, 2015 and HIV incidence is the upper limiting value of the cumulative sums.

Let I_k be HIV incidence in year k. Then,

$X_{k, k+j-1}=I_k$. P_k (t<j) for k=2004, …, 2015, j=1, 2, …, L_k, L_k=2015–k+1, where P_k (t<j) is the probability the case is infected in year k and reported to SISCEL less than j years after infection. (2)

Under the assumption the probability of reporting less than j years after infection are expressed by a logistic probability distribution, we estimated HIV incidence as the upper limiting value of the logistic function. For year 2004, separately for males and females, we estimated HIV incidence through an iterative procedure described in the supplementary methodology section.

As SISCEL data are available from 2004 to 2015, the maximum number of observations available for fitting the logistic probability distribution function is 12. To check the possibility of using the same approach for years after 2004, we applied the procedure for 2004 with a smaller number of observations, i.e., truncating the distribution for t<11, t<10, t<9, etc. The results showed the estimated HIV incidence loses accuracy (relative differences greater than 1%) with less than 8 observations, that is, the procedure could only be applied from 2004 to 2008. Therefore, to estimate HIV incidence in recent years (2009-2013), we estimated the missing observations for these years by extrapolating, for each interval of time since infection, the number of cases to be reported in the years 2016 to 2020 based on the 12 year series of reported data.

To this end, with the mathematical notation used in Table 2, for each fixed j varying from 4 to 12, we fitted linear regression models to the observed values $y_{j,u}$ (u=2004, …, 2015) with year of reporting as the independent variable. The predicted observations $\hat{y}_{j,u}$ in years u=2016, …, 2020 correspond to the estimated number of cases infected in year k=u-j+1 and expected to be reported in year u. For example, for year 2009, we estimated the number of cases infected in 2009 and expected to be reported in 2016 ($\hat{y}_{8,2016}$) based on the 12 observations $y_{8,u}$, for u ranging from 2004 to 2015, arranged along line j=8 in Table 2.

After completing the missing data to obtain at least 8 observations per year from 2009 to 2013, we estimated HIV incidence from 2004 to 2013, separately for males and females, fitting the logistic probability distribution to 2005-2013 data with the same standard deviation estimated for 2004, as described in the supplementary methodology section. Confidence interval estimates for HIV incidence were obtained by combining the uncertainties associated with each modeling component, the linear model intercept of equation (A.2) of the supplementary methodology section and the linear regression estimates of the number of cases expected to be reported between 2016 and 2020.

Using the data mentioned above, the method permits the

Intervals of t (years)	j	Year of Reporting											
		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
<1	1	**$y_{1,2004}$**	$y_{1,2005}$	$y_{1,2006}$	$y_{1,2007}$	$y_{1,2008}$	$y_{1,2009}$	$y_{1,2010}$	$y_{1,2011}$	$y_{1,2012}$	$y_{1,2013}$	$y_{1,2014}$	$y_{1,2015}$
$1 \leq t<2$	2	$y_{2,2004}$	**$y_{2,2005}$**	$y_{2,2006}$	$y_{2,2007}$	$y_{2,2008}$	$y_{2,2009}$	$y_{2,2010}$	$y_{2,2011}$	$y_{2,2012}$	$y_{2,2013}$	$y_{2,2014}$	$y_{2,2015}$
$2 \leq t<3$	3	$y_{3,2004}$	$y_{3,2005}$	**$y_{3,2006}$**	$y_{3,2007}$	$y_{3,2008}$	$y_{3,2009}$	$y_{3,2010}$	$y_{3,2011}$	$y_{3,2012}$	$y_{3,2013}$	$y_{3,2014}$	$y_{3,2015}$
$3 \leq t<4$	4	$y_{4,2004}$	$y_{4,2005}$	$y_{4,2006}$	**$y_{4,2007}$**	$y_{4,2008}$	$y_{4,2009}$	$y_{4,2010}$	$y_{4,2011}$	$y_{4,2012}$	$y_{4,2013}$	$y_{4,2014}$	$y_{4,2015}$
$4 \leq t<5$	5	$y_{5,2004}$	$y_{5,2005}$	$y_{5,2006}$	$y_{5,2007}$	**$y_{5,2008}$**	$y_{5,2009}$	$y_{5,2010}$	$y_{5,2011}$	$y_{5,2012}$	$y_{5,2013}$	$y_{5,2014}$	$y_{5,2015}$
$5 \leq t<6$	6	$y_{6,2004}$	$y_{6,2005}$	$y_{6,2006}$	$y_{6,2007}$	$y_{6,2008}$	**$y_{6,2009}$**	$y_{6,2010}$	$y_{6,2011}$	$y_{6,2012}$	$y_{6,2013}$	$y_{6,2014}$	$y_{6,2015}$
$6 \leq t<7$	7	$y_{7,2004}$	$y_{7,2005}$	$y_{7,2006}$	$y_{7,2007}$	$y_{7,2008}$	$y_{7,2009}$	**$y_{7,2010}$**	$y_{7,2011}$	$y_{7,2012}$	$y_{7,2013}$	$y_{7,2014}$	$y_{7,2015}$
$7 \leq t<8$	8	$y_{8,2004}$	$y_{8,2005}$	$y_{8,2006}$	$y_{8,2007}$	$y_{8,2008}$	$y_{8,2009}$	$y_{8,2010}$	**$y_{8,2011}$**	$y_{8,2012}$	$y_{8,2013}$	$y_{8,2014}$	$y_{8,2015}$
$8 \leq t<9$	9	$y_{9,2004}$	$y_{9,2005}$	$y_{9,2006}$	$y_{9,2007}$	$y_{9,2008}$	$y_{9,2009}$	$y_{9,2010}$	$y_{9,2011}$	**$y_{9,2012}$**	$y_{9,2013}$	$y_{9,2014}$	$y_{9,2015}$
$9 \leq t<10$	10	$y_{10,2004}$	$y_{10,2005}$	$y_{10,2006}$	$y_{10,2007}$	$y_{10,2008}$	$y_{10,2009}$	$y_{10,2010}$	$y_{10,2011}$	$y_{10,2012}$	**$y_{10,2013}$**	$y_{10,2014}$	$y_{10,2015}$
$10 \leq t<11$	11	$y_{11,2004}$	$y_{11,2005}$	$y_{11,2006}$	$y_{11,2007}$	$y_{11,2008}$	$y_{11,2009}$	$y_{11,2010}$	$y_{11,2011}$	$y_{11,2012}$	$y_{11,2013}$	**$y_{11,2014}$**	$y_{11,2015}$
$11 \leq t<12$	12	$y_{12,2004}$	$y_{12,2005}$	$y_{12,2006}$	$y_{12,2007}$	$y_{12,2008}$	$y_{12,2009}$	$y_{12,2010}$	$y_{12,2011}$	$y_{12,2012}$	$y_{12,2013}$	$y_{12,2014}$	**$y_{12,2015}$**
$12 \leq t<13$	13	$y_{13,2004}$	$y_{13,2005}$	$y_{13,2006}$	$y_{13,2007}$	$y_{13,2008}$	$y_{13,2009}$	$y_{13,2010}$	$y_{13,2011}$	$y_{13,2012}$	$y_{13,2013}$	$y_{13,2014}$	$y_{13,2015}$
$13 \leq t<14$	14	$y_{14,2004}$	$y_{14,2005}$	$y_{14,2006}$	$y_{14,2007}$	$y_{14,2008}$	$y_{14,2009}$	$y_{14,2010}$	$y_{14,2011}$	$y_{14,2012}$	$y_{14,2013}$	$y_{14,2014}$	$y_{14,2015}$
$14 \leq t<15$	15	$y_{15,2004}$	$y_{15,2005}$	$y_{15,2006}$	$y_{15,2007}$	$y_{15,2008}$	$y_{15,2009}$	$y_{15,2010}$	$y_{15,2011}$	$y_{15,2012}$	$y_{15,2013}$	$y_{15,2014}$	$y_{15,2015}$
$15 \leq t<16$	16	$y_{16,2004}$	$y_{16,2005}$	$y_{16,2006}$	$y_{16,2007}$	$y_{16,2008}$	$y_{16,2009}$	$y_{16,2010}$	$y_{16,2011}$	$y_{16,2012}$	$y_{16,2013}$	$y_{16,2014}$	$y_{16,2015}$
$16 \leq t<17$	17	$y_{17,2004}$	$y_{17,2005}$	$y_{17,2006}$	$y_{17,2007}$	$y_{17,2008}$	$y_{17,2009}$	$y_{17,2010}$	$y_{17,2011}$	$y_{17,2012}$	$y_{17,2013}$	$y_{17,2014}$	$y_{17,2015}$
$17 \leq t<18$	18	$y_{18,2004}$	$y_{18,2005}$	$y_{18,2006}$	$y_{18,2007}$	$y_{18,2008}$	$y_{18,2009}$	$y_{18,2010}$	$y_{18,2011}$	$y_{18,2012}$	$y_{18,2013}$	$y_{18,2014}$	$y_{18,2015}$
$18 \leq t<19$	19	$y_{19,2004}$	$y_{19,2005}$	$y_{19,2006}$	$y_{19,2007}$	$y_{19,2008}$	$y_{19,2009}$	$y_{19,2010}$	$y_{19,2011}$	$y_{19,2012}$	$y_{19,2013}$	$y_{19,2014}$	$y_{19,2015}$
$19 \leq t<20$	20	$y_{20,2004}$	$y_{20,2005}$	$y_{20,2006}$	$y_{20,2007}$	$y_{20,2008}$	$y_{20,2009}$	$y_{20,2010}$	$y_{20,2011}$	$y_{20,2012}$	$y_{20,2013}$	$y_{20,2014}$	$y_{20,2015}$
>=20	21	$y_{21,2004}$	$y_{21,2005}$	$y_{21,2006}$	$y_{21,2007}$	$y_{21,2008}$	$y_{21,2009}$	$y_{21,2010}$	$y_{21,2011}$	$y_{21,2012}$	$y_{21,2013}$	$y_{21,2014}$	$y_{21,2015}$
Total		Y_{2004}	Y_{2005}	Y_{2006}	Y_{2007}	Y_{2008}	Y_{2009}	Y_{2010}	Y_{2011}	Y_{2012}	Y_{2013}	Y_{2014}	Y_{2015}

Table 2: Mathematical representation of the frequency distribution of time (**t**) since infection among reported cases, 2004-2015.

simultaneous estimation of HIV incidence, proportion of cases reported in the same year of HIV infection and proportion of cases reported until 2015, from which we established the changes in testing patterns from 2004 to 2013.

Regarding the trends of HIV incidence, equation (3) with j=1 shows that $I_k = \dfrac{X_{k,k}}{P_k(t<1)}$. Therefore, the trend in HIV incidence is the same as that of the cases diagnosed in the year of HIV infection if the proportion of cases diagnosed in the year of infection remains constant throughout the period.

Results

After weighting SISCEL data to take into account CD4 measures in private laboratories, the total number of reported treatment-naïve HIV infected cases 15 years or older from 2004 to 2015 was 511,328, with 323,055 males and 188,273 females.

For each treatment-naïve case aged 15 year and over reported to SISCEL, we applied the CD4 depletion model by age group and sex (Table 1) and estimated the time between infection and the first CD4 count.

Results of the application of the proposed procedure for men and women HIV infected in year 2004 are presented in Table 3. The comparison between the observed distribution and the probabilities estimated by the logistic probability distribution shows the goodness of fit of the logistic distribution to the data.

In Table 4, we present the number of reported cases from 2004 to 2015, separately for males and females, according to year of reporting and intervals of time since infection ascertained by the depletion model. For each lag interval, the predicted values of the number of cases expected to be reported in years 2016 to 2020 were used to estimate observations originally missing due to truncation in 2015. Exemplifying, the predicted observations $y_{4, 2016}$ $y_{5, 2017}$ $y_{6, 2018}$ $y_{7, 2019}$ and $y_{8, 2020}$, were used to complete the 2013 dataset and estimate HIV incidence in this year.

Estimates of HIV incidence according to sex from 2004 to 2013 are presented in Table 5. In 2013, HIV incidence among people aged 15 years and over was estimated to be 44827 (95% CI 41143-47987),

32459 men (95% CI: 29775-34642) and 12368 women (95% CI: 11368-13345). Based on the Brazilian population data (15 years old and over), HIV incidence rate was 43.5 (95% CI 39.9-46.5) per 100,000 men, 15.9 (95% CI 14.6-17.1) per 100,000 women and 29.4 (95% CI 27.0-31.5) per 100,000 population aged 15 years old and over.

Results also presented in Table 5 showed HIV incidence increased among men and slowly decreased among women in the period 2004-2013. The male-female incidence ratio was 1.53 in 2004 but increased to 2.62 in 2013. The estimated proportion of cases having first CD4 count in less than one year after infection increased from 24.6%, in 2004, to 35.2%, in 2013, among men, and from 35.1% to 45.8%, among women. The estimated average time since infection shortened from 6.1 to 4.6 among men, and from 4.4 to 3.4 among women, from 2004 to 2013. Although men present a faster decline, the average time to SISCEL reporting is always higher among males. Among men who became HIV infected between 2004 and 2013, 35.6% of the cases were not reported to SISCEL by year 2015 and among women, 23.2%. The corresponding proportion for the total number of cases is 31.5%.

Discussion

In this paper, we proposed a practical method for estimation of HIV incidence in Brazil using surveillance data. The method is based on the first CD4 count among treatment-naïve reported cases and a statistical model that relates the first CD4 count to time since HIV infection. One advantage of the approach here proposed is that the method is based on recent trends of new HIV diagnoses using regularly available data.

As SISCEL is based on reimbursement of each lab examination by the Brazilian government, the system coverage of CD4 count exams performed in public laboratories is considered complete. However, SISCEL does not include HIV cases tested in the private sector. To bypass this difficulty, we assumed that the proportion of cases tested in private labs would be the proportion of people who has private health insurance in their area of residence [33]. Using this approach, the proportion of people tested in private labs was estimated as 35% similar to the proportion of cases that are in antiretroviral treatment (100% in the public sector) and are not included in SISCEL [34].

One limitation of SISCEL is that it does not collect information on exposure category, restricting the actual HIV incidence analyses to available variables, such as age group, sex, race, and area of residence

Lag (t) Interval (years)	Males			Females		
	Observed cumulative cases	Observed truncated distribution	Estimated probabilities (logistic distribution)	Observed cumulative cases	Observed proportions	Estimated probabilities (logistic distribution)
t<1	5129	0.2463	0.2507	4771	0.3514	0.3565
t<2	6284	0.3017	0.2984	5717	0.4211	0.4160
t<3	7384	0.3546	0.3510	6515	0.4799	0.4780
t<4	8639	0.4148	0.4075	7440	0.5479	0.5407
t<5	9837	0.4724	0.4665	8247	0.6074	0.6021
t<6	11056	0.5309	0.5265	8987	0.6619	0.6605
t<7	12122	0.5821	0.5857	9707	0.7149	0.7144
t<8	13282	0.6378	0.6426	10300	0.7586	0.7628
t<9	14303	0.6868	0.6957	10828	0.7975	0.8052
t<10	15242	0.7319	0.7440	11318	0.8336	0.8416
t<11	16208	0.7783	0.7871	11751	0.8654	0.8723
t<12	17141	0.8231	0.8246	12191	0.8979	0.8978
HIV Incidence estimate	20825	Sum of square differences	0.00049	13578	Sum of square differences	0.00033
95% CI	20185-21508			13338-13826		

Table 3: Estimates of HIV incidence separately for males and females using the proposed approach Brazil, 2004.

Intervals of t (years)	3 ≤ t<4	4 ≤ t<5	5 ≤ t<6	6 ≤ t<7	7 ≤ t<8	8 ≤ t<9	9 ≤ t<10	10 ≤ t<11	11 ≤ t<12
Year of Reporting					**Males**				
2004	997	978	966	967	868	737	778	665	720
2005	1379	1228	1139	1105	1012	939	917	791	838
2006	1139	1131	1130	1050	956	894	799	827	728
2007	1255	1317	1181	1083	938	874	890	844	760
2008	1226	1198	1094	1093	942	947	909	904	776
2009	1348	1131	1219	1091	1010	941	940	832	834
2010	1321	1212	1233	1066	983	918	914	886	782
2011	1402	1342	1383	1150	1159	954	1012	924	931
2012	1516	1467	1414	1195	1095	1022	906	913	823
2013	1743	1504	1543	1397	1219	1064	938	934	889
2014	1871	1682	1671	1371	1330	1177	989	966	876
2015	2019	1950	1619	1509	1321	1207	1075	902	933
2016	1928	1769	1684	1434	1321	1183	1042	988	927
2017		1834	1743	1474	1360	1215	1060	1007	943
2018			1802	1515	1399	1248	1078	1026	959
2019				1555	1438	1280	1097	1045	975
2020					1476	1312	1115	1064	991
Year of Reporting					**Females**				
2004	898	911	801	799	742	661	575	500	445
2005	1096	965	1026	869	845	748	661	531	457
2006	844	821	858	800	677	676	583	532	502
2007	924	935	882	816	709	669	596	508	455
2008	840	807	746	771	651	575	562	494	487
2009	827	859	740	730	660	567	528	479	424
2010	761	762	763	720	538	556	475	480	458
2011	754	689	672	687	593	563	483	447	388
2012	773	753	701	614	566	528	483	422	394
2013	786	726	745	670	619	555	490	395	430
2014	780	770	696	697	604	567	499	433	417
2015	766	729	656	725	605	543	502	440	441
2016	706	679	627	638	535	498	451	402	408
2017		659	604	623	517	482	438	391	403
2018			582	607	500	466	424	381	397
2019				591	482	450	411	370	392
2020					464	435	398	359	387

*For each interval of time since infection, we estimated the predicted observation (2016-2020) by linear regression models based on the 12 year series of reported data

Table 4: Reported (2004-2015) and predicted observations* (2016-2020) according to intervals of time since infection (t).

Year of Infection	Males							Females						
	Estimated HIV Incidence	95% CI		Average time since infection (years)	% Reported in the same year of infection	Not Reported until 2015		Estimated HIV Incidence	95% CI		Average time since infection (years)	% Reported in the same year of infection	Not Reported until 2015	
		LL	UL			n	%		LL	UL			n	%
2004	20825	20185	21508	6.1	24.6	3685	17.7	13578	13338	13826	4.4	35.1	1387	10.2
2005	21005	20721	21323	5.4	29.6	3613	17.2	14190	14110	14288	3.8	40.3	1465	10.3
2006	22446	21938	22989	5.8	26.3	5493	24.5	13123	12968	13308	4.0	38.9	1799	13.7
2007	23550	23154	24056	5.5	28.8	6209	26.4	12925	12737	13187	3.9	41.0	2082	16.1
2008	24729	24388	25198	5.4	29.6	7600	30.7	12902	12633	13272	3.9	40.9	2576	20.0
2009	26251	25694	26911	5.2	30.7	9096	34.6	12692	12328	13208	3.8	42.1	2998	23.6
2010	27558	26495	28351	5.2	30.6	11013	40.0	12407	12032	13023	3.8	41.8	3573	28.8
2011	29861	28285	30988	5.0	32.0	13132	44.0	12174	11592	12895	3.8	41.9	4113	33.8
2012	31172	28707	32863	5.0	31.9	15526	49.8	11983	11200	12799	3.7	42.3	4667	38.9
2013	32459	29775	34642	4.6	35.2	17050	52.5	12368	11368	13345	3.4	45.8	5102	41.3
∑	**259856**					**92417**	**35.6**	**128341**					**29760**	**23.2**

Table 5: Estimates of HIV incidence and other indicators by year of infection and sex, Brazil, 2004-2013.

(region, state, and municipality). However, as the Department of STD, AIDS and Viral Hepatitis routinely links SISCEL to the National System of Notified HIV/AIDS cases [34], it is possible to obtain exposure category for at least 70% of the SISCEL cases, enabling the estimation of HIV incidence by exposure category with the use of imputation methods [35]. Moreover, any other information available in SISCEL related to the time since infection, such as viral load or patient's condition (symptomatic/asymptomatic) that could be used to refine our estimates [36,37] will be further investigated.

In order to use the CD4 depletion model in Brazil, we adapted the model originally proposed by Lodi et al. [31] and calculated specific CD4 count decline rates by sex and age at diagnosis (first CD4 count) using Brazilian data. In general, the uncertainties associated with HIV incidence estimates derived from this method could be underestimated due to some restrictive assumptions, for example, the mean square root of CD4 count is linearly related to the time since infection, the number of cases to be reported to SISCEL is increased or decreased linearly in future years, etc.

Although the assumption that the square root of CD4 count is linearly related to time since infection has frequently been used in different settings [12,14,15], other types of functions have also been proposed [38,39]. In addition, Rice and collaborators have suggested that besides age at diagnosis, ethnicity and region of birth are significant predictors of rate of CD4 decrease [15]. Sensitivity of the estimates deriving from changes in the CD4 depletion model underlying assumptions will be examined in next studies.

The application of the CD4 depletion model to reported cases made possible to estimate HIV incidence in Brazil and to identify changes in testing patterns. Based on fitting a logistic distribution to the cumulative number of cases classified by intervals of time since infection, the procedure used in this study represents a simplification compared to the previous model [12], since it requires extrapolation of reported data over 5 years only. The proposed approach allowed us to estimate HIV incidence and 95% confidence limits by sex in the period 2004-2013, and the results are consistent with previous findings [12].

For year 2013, our estimate of HIV incidence (44827) is similar to the UNAIDS estimate (44000) for the same year [40], but almost twice the estimate of the Global Burden of Disease (GBD) group (24187) [41]. Reasons for differences in the UNAIDS and Murray et al. estimates have been discussed before [42,43]. In the case of Brazil, the most likely explanation is the influence of the GBD 2013 mortality estimate, with 10217 AIDS deaths [41]. According to the Mortality Information System, 12564 AIDS deaths occurred in 2013, but it is believed that this number is still underestimated due to misclassification of the causes of death [44].

The application of the method in Brazil showed a delay of approximately 4.3 years (51 months) between infection and date of first CD4 count. The delay is a little shorter among women, probably due to the Ministry of Health policy of HIV testing during antenatal care [45].

In France, a similar model was used for Estimating HIV Incidence Based on the estimated time from infection to diagnosis using HIV surveillance data. The results are comparable to those obtained in Brazil: the estimated mean time since HIV infection ranged from 37 to 53 months among men who have sex with men (MSM) and heterosexual men, respectively, with an intermediate delay for women [11]. In the United States of America, HIV cases diagnosed in 2011 had been, on average, infected 5.6 years before their diagnosis [13]. Despite the improvement in shortening delays from infection to first CD4 count from 2004 to 2013 (Table 5), for both men and women, the delay remains excessively long [46].

A growing number of studies have shown evidence that wide and early initiation of antiretroviral therapy can reduce the level of HIV incidence in the population [10,47]. The 'treatment as prevention' (TASP) policy is based on the premise that the rate of new HIV infections can be reduced by increasing frequency of HIV testing to diagnose people living with HIV in early stages and initiate antiretroviral treatment regardless of CD4 count or viral load [46]. In Brazil, difficulty in early detection of HIV infection impairs the benefits of TASP, adopted since 2014, and the spread of HIV transmitted by undiagnosed individuals may actually be the main factor driving the HIV/AIDS epidemic. In fact, this study results showed that there are around 45000 new HIV infections among people aged 15 and over per year in Brazil and only 38% is diagnosed in the same year of infection.

HIV incidence estimates showed an increase among men and a slow decrease among women from 2004 to 2013, with men currently accounting for 72% of new infections. Although we did not analyzed data by exposure category, the most likely explanation is an increasing rate among men who have sex with men (MSM) and a decreasing rate among injection drug users, which was a major contributor to the expansion of heterosexual transmission in the 1990s in Brazil [48].

During more than 30 years of the HIV/AIDS epidemic, Brazil has maintained a concentrated epidemic, with an HIV prevalence rate less than 1% in the general population [39]. Despite earlier concerns about an increase in heterosexual cases and "feminization" of the epidemic, it seems the actual scenario is the predominance of HIV infection among MSM, similar to data from many other countries [49,50]. In a study conducted in two Brazilian cities in 2013 using laboratory tests to distinguish recent from long-term infections, the estimated incidence rate was greater than 1% among MSM [25].

Conclusion

In this paper, we introduced a method to calculate HIV incidence in Brazil. As it is based on the first CD4 count and date of first CD4 count, the approach is applicable to all countries that monitor these data among HIV infected cases. The method permits the simultaneous estimation of HIV incidence, proportion of HIV cases diagnosed in the same year of infection and proportion of undiagnosed cases. The application of the method in Brazil showed a decreasing trend in the time between HIV infection and the first CD4 count in the period 2004-2013 but the delay is still too long (4.3 years). Recent government policies are focused on increasing uptake of HIV testing with expansion of the offer in health units and home-based testing [51]. Therefore, an additional contribution of the proposed method is its potential use to monitor and evaluate the impact of these new HIV testing strategies.

Supplementary Methodology Section

Under the assumption that the probability of reporting less than **j** years after infection is expressed by a logistic probability distribution with mean **μ** and standard deviation **σ**, we have:

(A.1) $F(j; \mu; \sigma) = P(t<j) =$, where , and π is the number pi ~3.1416.

So, the cumulative number of cases reported to SISCEL can be written as:

(A.2) $X_{k, k+j} = I_k \cdot F(j; \mu_k; \sigma_k)$, and

(A.3) , for and .

If we use equation (A.3) with known **λ** (or **σ**), we can estimate HIV incidence in year **k** (I_k) by fitting a linear regression model to the

inverse of the observed values $X_{k,k+j}$ The inverse of the intercept will be the estimated incidence in year **k**.

For year 2004, separately for males and females, we estimated **σ**, I_{2004}, and **θ** by using an iterative procedure with an initial guess for **σ** to generate successive approximations to a solution. For each **σ**, we used the regression model (A.3) to estimate the mean (**μ**) and equation (A.2) to estimate the probabilities of reporting **j** years after infection. The selected **σ** was the one that minimized the sum of squared differences between the observed truncated distribution and the estimated probabilities of reporting less than **j** years after infection by the logistic probability distribution. The estimate of HIV incidence was given by the inverse of the intercept in the linear regression model (A.3) corresponding to the selected **σ**. Confidence intervals for HIV incidence were based on 95% confidence intervals for the linear model intercept.

Ethics Approval and Consent to Participate

The project was approved by the Ethics Committee of the Oswaldo Cruz Foundation, Ministry of Health, Brazil (Protocol 485.175).

Author's Contribution

CLS and PRBSJ designed the research. CLS and OCFJ collaborated in the writing of the manuscript. PRBSJ and ARPP performed the statistical analyses and participated in the interpretation of results. All authors revised the manuscript before submission.

References

1. van Sighem AI, Gras LA, Reiss P, Brinkman K, de Wolf F; ATHENA National Observational Cohort Study (2010) Life expectancy of recently diagnosed asymptomatic HIV-infected patients approaches that of uninfected individuals. AIDS 24: 1527-1535.

2. Stover J, Bollinger L, Izazola JA, Loures L, DeLay P, et al. (2016) What is required to end the AIDS epidemic as a public health threat by 2030? The cost and impact of the fast-track approach. PLoS One 11: e0154893.

3. Aghaizu A, Murphy G, Tosswill J, DeAngelis D, Charlett A, et al (2014) Recent infection testing algorithm (RITA) applied to new HIV diagnoses in England, Wales and Northern Ireland, 2009 to 2011. Euro Surveill 19: 20673.

4. Birrell PJ, Gill ON, Delpech VC, Brown AE, Desai S, et al. (2013) HIV incidence in men who have sex with men in England and Wales 2001-2010: A nationwide population study. Lancet Infect Dis 13: 313-318.

5. Rehle T, Shisana O, Pillay V, Zuma K, Puren A, et al. (2007) National HIV incidence measures--new insights into the South African epidemic. S Afr Med J 97: 194-199.

6. Bätzing-Feigenbaum J, Loschen S, Gohlke-Micknis S, Zimmermann R, Herrmann A, et al. (2008) Country-wide HIV incidence study complementing HIV surveillance in Germany. Euro Surveill, p: 13.

7. Brookmeyer R (2010) Measuring the HIV/AIDS epidemic: Approaches and challenges. Epidemiol Rev 32: 26-37.

8. Morgado MG, Bastos F (2011) Estimates of HIV-1 incidence based on serological methods: A brief methodological review. Cad Saude Publica 27 Suppl 1: S7-S18.

9. Brookmeyer R, Gail MH (1994) AIDS epidemiology: A quantitative approach. Oxford University Press, New York.

10. Montaner JS, Lima VD, Harrigan PR, Lourenço L, Yip B, et al. (2014) Expansion of HAART coverage is associated with sustained decreases in HIV/AIDS morbidity, mortality and HIV transmission: The "HIV Treatment as Prevention" experience in a Canadian setting. PLoS One 9: e87872.

11. Ndawinz JD, Costagliola D, Supervie V (2011) New method for estimating HIV incidence and time from infection to diagnosis using HIV surveillance data: Results for France. AIDS 25: 1905-1913.

12. Szwarcwald CL, Pascom ARP, de Souza Júnior PR (2015) Estimation of the HIV incidence and of the number of people living with HIV/AIDS in Brazil, 2012. J AIDS Clin Res 6: 430.

13. Hall HI, Song R, Szwarcwald CL, Green T (2015) Brief report: Time from infection with the human immunodeficiency virus to diagnosis, United States. J Acquir Immune Defic Syndr 69: 248-251.

14. Song R, Hall HI, Green TA, Szwarcwald CL, Pantazis N (2016) Using CD4 data to estimate HIV incidence, prevalence and percent of undiagnosed infections in the United States. J Acquir Immune Defic Syndr.

15. Rice BD, Elford J, Yin Z, Delpech VC (2012) A new method to assign country of HIV infection among heterosexuals born abroad and diagnosed with HIV. AIDS 26: 1961-1966.

16. Nelson KE, Vlahov D, Solomon L, Cohn S, Muñoz A, et al. (1995) Temporal trends of incident human immunodeficiency virus infection in a cohort of injecting drug users in Baltimore, Md. Arch Intern Med 155: 1305-1311.

17. Harrison L, Dunn DT, Green H, Copas AJ (2009) Modelling the association between patient characteristics and the change over time in a disease measure using observational cohort data. Stat Med 28: 3260-3275.

18. Hallett TB, Zaba B, Todd J, Lopman B, Mwita W, et al. (2008) Estimating incidence from prevalence in generalised HIV epidemics: Methods and validation. PLoS Med 5: e80.

19. Stover J, Brown T, Marston M (2012) Updates to the spectrum/estimation and projection package (EPP) model to estimate HIV trends for adults and children. Sex Transm Infect 88: i11-i16.

20. Marcus U, Hickson F, Weatherburn P, Schmidt AJ; EMIS Network (2012) Prevalence of HIV among MSM in Europe: Comparison of self-reported diagnoses from a large scale internet survey and existing national estimates. BMC Public Health 12: 978.

21. Guibu IA, Barros MB, Donalísio MR, Tayra Â, Alves MC (2011) Survival of AIDS patients in the Southeast and South of Brazil: Analysis of the 1998-1999 cohort. Cad Saude Publica 27: S79-S92.

22. Janssen RS, Satten GA, Stramer SL, Rawal BD, O'Brien TR, et al. (1998) New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes. JAMA 280: 42-48.

23. Hall HI, Song R, Rhodes P, Prejean J, An Q, et al. (2008) Estimation of HIV incidence in the United States. JAMA 300: 520-529.

24. Hu DJ, Vanichseni S, Mock PA, Young NL, Dobbs T, et al. (2003) HIV type 1 incidence estimates by detection of recent infection from a cross-sectional sampling of injection drug users in Bangkok: Use of the IgG capture BED enzyme immunoassay. AIDS Res Hum Retroviruses 19: 727-730.

25. Mermin J, Musinguzi J, Opio A, Kirungi W, Ekwaru JP, et al. (2008) Risk factors for recent HIV infection in Uganda. JAMA 300: 540-549.

26. Garrett NJ, Lattimore S, Gilbart V, Aghaizu A, Mensah G, et al. (2012) The recent infection testing algorithm (RITA) in clinical practice: A survey of HIV clinicians in England and Northern Ireland. HIV Med 13: 444-447.

27. Szwarcwald CL, Ferreira Jr OC, Brito AM, Luhm KR, Ribeiro CEL, et al. (2016) Estimation of HIV incidence in two Brazilian municipalities, 2013. Rev Saude Publica, p: 50.

28. McWalter TA, Welte A (2009) A comparison of biomarker based incidence estimators. PLoS One 4: e7368.

29. Laeyendecker O, Kulich M, Donnell D, Komárek A, Omelka M, et al. (2013) Development of methods for cross-sectional HIV incidence estimation in a large, community randomized trial. PLoS One 8: e78818.

30. Natural History Project Working Group for the Collaboration of Observational HIV Epidemiological Research Europe (COHERE) in EuroCoord (2014) Factors associated with short-term changes in HIV viral load and CD4(+) cell count in antiretroviral-naive individuals. AIDS 28: 1351-1356.

31. Lodi S, Phillips A, Touloumi G, Geskus R, Meyer L, et al. (2011) Time from human immunodeficiency virus seroconversion to reaching CD4+ cell count

thresholds <200, <350 and <500 Cells/mm³: Assessment of need following changes in treatment guidelines. Clin Infect Dis 53: 817-825.

32. Crowell TA, Ganesan A, Berry SA, Deiss RG, Agan BK, et al. (2016) Hospitalizations among HIV controllers and persons with medically controlled HIV in the U.S. military HIV natural history study. J Int AIDS Soc 19: 20524.

33. IBGE - Instituto Brasileiro de Geografia e Estatística (2010) Pesquisa Nacional por Amostra de Domicílios. Um Panorama da saúde no Brasil: acesso e utilização dos serviços, condições de saúde e fatores de risco de proteção à saúde 2008. IBGE, Rio de Janeiro.

34. Ministério da Saúde. Secretaria de Vigilância em Saúde - Departamento de DST, Aids e Hepatites Virais (2014) Boletim Epidemiológico - Aids e DST Ano III - nº 1 - 27ª à 52ª semanas epidemiológicas - julho a dezembro de 2013 Ano III - nº 1 - 01ª à 26ª semanas epidemiológicas - janeiro a junho de 2014. Ministério da Saúde Brasília, DF 84: 1517-1159.

35. Rubin DB (1987) Multiple Imputation for nonresponse in surveys. John Wiley & Sons Inc, New York.

36. DeGruttola V, Lange N, Dafni U (1991) Modeling the progression of HIV infection. Journal of the American Statistical Association 86: 569-577.

37. Touloumi G, Pantazis N, Pillay D, Paraskevis D, Chaix ML, et al. (2013) Impact of HIV-1 subtype on CD4 count at HIV seroconversion, rate of decline and viral load set point in European seroconverter cohorts. Clin Infect Dis 56: 888-897.

38. Yan P, Zhang F, Wand H (2011) Using HIV diagnostic data to estimate HIV incidence: Method and simulation. Statistical Communications in Infectious Diseases.

39. Meijerink H, Wisaksana R, Iskandar S, Heijer M, Ven A, et al. (2014) Injecting drug use is associated with a more rapid CD4 cell decline among treatment naive HIV-positive patients in Indonesia. Journal of the International AIDS Society 17: 18844.

40. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2014) The GAP Report. UNAIDS.

41. Murray CJ, Ortblad KF, Guinovart C, Lim SS, Wolock TM, et al. (2014) Global, regional and national incidence and mortality for HIV, tuberculosis and malaria during 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. Lancet 384: 1005-1070.

42. Hallett TB, Zaba B, Stover J, Brown T, Slaymaker E, et al. (2014) Embracing different approaches to estimating HIV incidence, prevalence and mortality. AIDS 28: S523-532.

43. Supervie V, Archibald CP, Costagliola D, Delpech V, Hall HI, et al. (2015) GBD 2013 and HIV incidence in high income countries. Lancet 385: 1177.

44. Fazito E, Cuchi P, Fat DM, Ghys PD, Pereira MG, et al. (2012) Identifying and quantifying misclassified and under-reported AIDS deaths in Brazil: A retrospective analysis from 1985 to 2009. Sex Transm Infect 88: i86-94.

45. Szwarcwald CL, Barbosa Júnior A, Souza-Júnior PR, Lemos KR, Frias PG, et al. (2008) HIV testing during pregnancy: Use of secondary data to estimate 2006 test coverage and prevalence in Brazil. Braz J Infect Dis 12: 167-172.

46. Piñeirúa A, Sierra-Madero J, Cahn P, Guevara Palmero RN, Martínez Buitrago E, et al. (2015) The HIV care continuum in Latin America: Challenges and opportunities. Lancet Infect Dis 15: 833-839.

47. Grinsztejn B, Hosseinipour MC, Ribaudo HJ, Swindells S, Eron J, et al. (2014) Effects of early versus delayed initiation of antiretroviral treatment on clinical outcomes of HIV-1 infection: Results from the phase 3 HPTN 052 randomised controlled trial. Lancet Infect Dis 14: 281-290.

48. Barbosa Júnior A, Szwarcwald CL, Pascom AR, Souza Júnior PB (2009) Trends in the AIDS epidemic in groups at highest risk in Brazil, 1980-2004. Cad Saude Publica 25: 727-737.

49. Beyrer C, Sullivan P, Sanchez J, Baral SD, Collins C, et al. (2013) The increase in global HIV epidemics in MSM. AIDS 27: 2665-2678.

50. Snowden JM, Wei C, McFarland W, Raymond HF (2014) Prevalence, correlates and trends in seroadaptive behaviours among men who have sex with men from serial cross-sectional surveillance in San Francisco, 2004-2011. Sex Transm Infect 90: 498-504.

51. Lippman SA, Périssé AR, Veloso VG, Sullivan PS, Buchbinder S, et al. (2014) Acceptability of self-conducted home-based HIV testing among men who have sex with men in Brazil: Data from an online survey. Cad Saude Publica 30: 724-734.