

## RESEARCH ARTICLE

# Estimation of the Undiagnosed Intervals of HIV-Infected Individuals by a Modified Back-Calculation Method for Reconstructing the Epidemic Curves

Ngai Sze Wong<sup>1,2,3</sup>, Ka Hing Wong<sup>4</sup>, Man Po Lee<sup>5</sup>, Owen T. Y. Tsang<sup>6</sup>, Denise P. C. Chan<sup>1</sup>, Shui Shan Lee<sup>1\*</sup>

**1** Stanley Ho Centre for Emerging Infectious Diseases, The Chinese University of Hong Kong, Hong Kong, China, **2** Institute for Global Health & Infectious Diseases, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America, **3** University of North Carolina Project-China, Guangzhou, Guangdong, China, **4** Special Preventive Programme, Department of Health, Hong Kong Special Administrative Region Government, Hong Kong, China, **5** Department of Medicine, Queen Elizabeth Hospital, Hong Kong, China, **6** Department of Medicine and Geriatrics, Princess Margaret Hospital, Hong Kong, China

\* [sslee@cuhk.edu.hk](mailto:sslee@cuhk.edu.hk)



CrossMark  
click for updates

## OPEN ACCESS

**Citation:** Wong NS, Wong KH, Lee MP, Tsang OTY, Chan DPC, Lee SS (2016) Estimation of the Undiagnosed Intervals of HIV-Infected Individuals by a Modified Back-Calculation Method for Reconstructing the Epidemic Curves. PLoS ONE 11 (7): e0159021. doi:10.1371/journal.pone.0159021

**Editor:** Kok Keng Tee, University of Malaya, MALAYSIA

**Received:** March 6, 2016

**Accepted:** June 24, 2016

**Published:** July 12, 2016

**Copyright:** © 2016 Wong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data used in the study have been included in the paper and its Supporting Information files.

**Funding:** This study was supported by a grant approved by the Council for the AIDS Trust Fund, Hong Kong Special Administrative Region Government (project code: MSS229R). First author NSW is supported by Guangdong Provincial Centers for Skin Diseases and STI Control, the UNC-South China STD Research Training Center (FIC1D43TW009532-01) and the Fogarty International Center.

## Abstract

### Background

Undiagnosed infections accounted for the hidden proportion of HIV cases that have escaped from public health surveillance. To assess the population risk of HIV transmission, we estimated the undiagnosed interval of each known infection for constructing the HIV incidence curves.

### Methods

We used modified back-calculation methods to estimate the seroconversion year for each diagnosed patient attending any one of the 3 HIV specialist clinics in Hong Kong. Three approaches were used, depending on the adequacy of CD4 data: (A) estimating one's pre-treatment CD4 depletion rate in multilevel model; (B) projecting one's seroconversion year by referencing seroconverters' CD4 depletion rate; or (C) projecting from the distribution of estimated undiagnosed intervals in (B). Factors associated with long undiagnosed interval (>2 years) were examined in univariate analyses. Epidemic curves constructed from estimated seroconversion data were evaluated by modes of transmission.

### Results

Between 1991 and 2010, a total of 3695 adult HIV patients were diagnosed. The undiagnosed intervals were derived from method (A) (28%), (B) (61%) and (C) (11%) respectively. The intervals ranged from 0 to 10 years, and were shortened from 2001. Heterosexual infection, female, Chinese and age >64 at diagnosis were associated with long undiagnosed interval. Overall, the peaks of the new incidence curves were reached 4–6 years ahead of

**Competing Interests:** The authors have declared that no competing interests exist.

reported diagnoses, while their contours varied by mode of transmission. Characteristically, the epidemic growth of heterosexual male and female declined after 1998 with slight rebound in 2004–2006, but that of MSM continued to rise after 1998.

## Conclusions

By determining the time of seroconversion, HIV epidemic curves could be reconstructed from clinical data to better illustrate the trends of new infections. With the increasing coverage of antiretroviral therapy, the undiagnosed interval can add to the measures for assessing HIV transmission risk in the population.

## Introduction

Before progression to AIDS, HIV infection is largely asymptomatic in the period since seroconversion, the duration of which can be as long as 7 years or more in the absence of treatment.[1] An HIV-infected individual remains undiagnosed, unless he/she receives an HIV test for different reasons. Within this undiagnosed period, infected individuals are not aware of their HIV status. Their transmission risk can be substantial in the presence of a high partner exchange rate and the practice of unprotected sex. After HIV diagnosis, transmission risk may fall as a result of self-initiated reduction of risk behaviours and/or interventions.[2] Moreover, good coverage of highly active antiretroviral treatment (HAART) could reduce the population viral burden, thereby minimizing the transmission risk, as concluded in the HPTN052 study.[3, 4] Therefore, the status of being undiagnosed, the first stage of the care continuum cascade, constitutes a major gap for achieving effective interventions through HAART. Epidemiologically, the lag time between infection and diagnosis is an obstacle for proper interpretation of epidemic curves plotted by annual numbers of new HIV diagnoses, as recent and past infections could not be differentiated. Quantification of the undiagnosed intervals is, therefore instrumental for reconstructing epidemic curves for supporting the effective monitoring of the epidemic and evaluating interventions introduced.

In the past, HIV incidence back-calculated by computing the number of diagnosed AIDS cases and distribution of incubation period between HIV infection and AIDS diagnosis[5] was a reasonable approach. The widespread use of HAART since the mid-1990s has however distorted the natural history of HIV/AIDS. While a few studies have introduced modified back-calculation method that incorporated diagnosed HIV cases,[6] the estimation of new infections was often made at aggregate level. Other studies have used biological approaches such as tests for recent infection (TRIs), recent infection testing algorithm (RITA) and BED HIV-1 Capture Enzyme Immunoassay to determine whether a diagnosed individual was recently infected.[7, 8] However, such method was limited by the availability of samples, technologies and resources, and could only broadly distinguish between recent and non-recent infections. To date, some studies have estimated the prevalence of undiagnosed HIV-infected individuals and investigated their epidemiological characteristics, as reported in China,[9] India,[10] Spain, Italy, Slovakia, Romania, Slovenia and Czech Republic.[11] While these studies have provided insights into the size of the hidden infective populations, temporal patterns were not systematically evaluated.

At an individual level, the undiagnosed interval has largely been ignored so far, except for a limited number of studies which computed seroconverters' pre-treatment CD4 depletion rate to back-calculate the year of seroconversion,[12, 13] or used diagnosed individuals' specific

test-seeking behaviours and clinical status at diagnosis to estimate the time interval between infection and diagnosis.[14, 15] Against these backgrounds, we propose to expand the modified back-calculation method to make full use of currently available clinical data to determine the undiagnosed interval for each HIV positive individual, using seroconverters' CD4 depletion as the reference group. With this approach, we estimated the seroconversion year and undiagnosed interval at individual level and reconstructed the HIV incidence curves to help explain the temporal trend of virus transmission in the city of Hong Kong.

## Methods

### Data source

In Hong Kong, 3 HIV specialist clinics in the public service are providing care to almost all reported HIV/AIDS cases in the territory, the data from which constituted the cohort database described in this study.[16] HAART is initiated in accordance with local guidelines adapted from the recommendations of Department of Health and Human Services in USA.[17] Demographics, baseline clinical conditions and regularly collected laboratory measures (CD4 and viral load) in 1985–2012 were accessed. Approvals from the Research Ethics Committee of the Joint Chinese University of Hong Kong—New Territories East Cluster, and Kowloon Central/Kowloon East Cluster of the Hospital Authority were obtained. In compliance with the Personal Data (Privacy) Ordinance, data access was approved by the Department of Health, Hong Kong Special Administrative Region Government. No informed consent was obtained as the collected data were anonymized and collected retrospectively. This study was conducted in compliance with Declaration of Helsinki.

### Modified back-calculation methods for estimating seroconversion year

In our study, the annual number of new infections, total number of undiagnosed infections, and every HIV infected person's undiagnosed interval were derived after estimating the seroconversion year individually. Estimation was made by the modified back-calculation approach, taking reference from the pre-treatment CD4 trajectory of known HIV-infected patients. As a set-point viral load is reached about 6 months after infection,[1] this interval was deducted from the time-point when CD4 count reaches a plateau following acute infection. According to the number of pre-treatment CD4 measurements in the clinical dataset, patients were classified into Group A (>3 CD4 counts), Group B (1–3 CD4 counts) and Group C (no CD4 count) for the estimation.

**Group A (individuals with >3 pre-treatment CD4 counts).** Methodologically, if over 3 pre-treatment CD4 counts have been recorded, the levels were input in multilevel model (lme4 R package) to derive the pre-treatment CD4 depletion rate for each patient. Linear multilevel model was performed because of the repeated CD4 measurements available per patient.[18] In the model, CD4 measurements (outcome variable) were nested by patients, with time (defined as months from the first CD4 measurement after diagnosis) as random intercept and random slope, and ethnicity and gender as random intercepts. The seroconversion year of HIV patients who were infected via injection drug use or sexual transmission were estimated separately in multilevel model. Different combination of variables in multilevel model were explored and the best multilevel model was selected with reference to intraclass correlation coefficient (ICC). [19] By randomly selecting a CD4 count within the normal reference range of a healthy adult for 1000 times,[20–22] and using these as the starting points for CD4 depletion, the seroconversion year of a patient was calculated in Eq 1 [23, 24] for 1000 simulations:

$$\text{seroconversion year} = \text{diagnosis year} + \frac{\text{random normal CD4 reference range}_{i,j} - a}{\text{monthly CD4 slope} * 12} \quad (1)$$

Where  $i$  is gender (male, female),  $j$  is ethnicity (1. Asian, 2. White, 3. African & others), [20–22]  $a$  is intercept of an individual’s regression line in multilevel model, **CD4 slope** is the adjusted coefficient of the individual’s regression line in multilevel model

The simulation results falling outside the lower boundary (a. the last negative HIV testing year; or b. 1980, which might be the year with the first possible infection case in Hong Kong; or c. year of attaining 12 year-old, possible minimum age of being sexually active) and the upper boundary (year of HIV diagnosis) were discarded.

We have selected the pre-treatment CD4 counts of seroconverters (patients with  $\leq 2$  years’ time interval from last negative HIV test to first positive HIV test) to form a reference dataset to validate the simulation results. Their CD4 measurements were input to the same multilevel model as patients in Group A, with the removal of the year of last negative HIV testing date as the lower boundary of filter. The simulation results were compared with the mid-point of the interval between the last negative and the first positive HIV tests, assuming the latter method gave the most precise estimation. [25] The seroconversion time of patients in Group A was derived from the third quartile of the simulation results as these were closest to the mid-point estimation by comparison. (S1 Table) The discrepancy would be the smallest if we take the 90th percentile, but this would not be meaningful unless everyone had seroconverted a year before diagnosis.

**Group B (individuals with 1–3 pre-treatment CD4).** Multilevel models were performed in a different way for patients with 1–3 CD4 measurements before treatment initiation, or if Group A estimation failed due to minimal changes in CD4 count pre-treatment or rising pre-treatment CD4 count with slope of depletion above  $-1/\mu\text{L}/\text{month}$ . Reference range of coefficients showing the relationship between seroconversion time (mid-point of interval between last negative and first positive HIV testing  $\leq 2$  years) and CD4 count was first established from seroconverters in the same cohort. The seroconversion year of injection drug users (IDU) and sexually acquired HIV patients were estimated separately by linear multilevel model. Time (months from seroconversion time) and ethnicity (being White or not) were random intercepts in the model. Gender was not included in the selected multilevel model as it was not significantly associated with CD4 count. With the coefficients determined in multilevel model, the year of seroconversion was estimated by:

$$\text{CD4 measurement year} = \frac{\text{CD4} - \text{random intercept} - \text{coefficient} * \text{White}}{12 * \text{random CD4 slope}} \quad (2)$$

The randomly generated intercept and coefficient were derived from the mean and standard deviation of coefficients calculated in multilevel model. A total of 2500 simulations were performed for a clinical measurement available. As the seroconversion time of patients in Group A were estimated by the third quartile of simulation result, the same logic applied here.

**Group C (individuals without pre-treatment CD4 count).** For individuals without pre-treatment CD4 count or with flat or rising pre-treatment CD4 slope (above  $-1/\mu\text{L}/\text{month}$ ), estimation of seroconversion year was made from the estimation results of Group B. This was based on the distribution of estimated undiagnosed interval (from year of seroconversion to diagnosis) in patients in Group B, as the in care condition between the two Groups were similar, both with less frequent or irregular visit to the clinics before treatment. Due to their similarity, the seroconversion year was calculated by subtracting the median undiagnosed interval from the year of diagnosis, stratified by mode of transmission and status of late HIV diagnosis. (Table 1) Late HIV diagnosis here refers to the diagnosis of AIDS within 3 months of HIV diagnosis. Sensitivity analyses were performed to examine the difference of seroconversion estimation results between selection of reference groups 1) Group A, 2) Group B and 3) Group A

**Table 1. Summary for median of intervals between HIV diagnosis year and 3<sup>rd</sup> quartile of the simulation results of seroconversion year in Group B patients.**

Subgroups by transmission route	Median of interval (year)	n	Median of interval (year)	n
	Non-late diagnosis		Late diagnosis*	
Blood recipient	1	39	1	8
Heterosexual female	2	333	3	166
Heterosexual male	2	518	3	499
Injection drug use	2	148	2	35
Men who have sex with men	2	739	3	295
Undetermined	3	37	4	31

\*Late diagnosis—patients diagnosed with AIDS within 3 months of HIV diagnosis

doi:10.1371/journal.pone.0159021.t001

and B that were used to derive undiagnosed intervals, and selection of central tendency (a. median, b. first quartile and c. third quartile) of undiagnosed intervals.

**Analyses.** With the estimation of seroconversion year at individual level, we plotted the HIV incidence curves as the yearly estimated count of individuals who had seroconverted. The annual number of undiagnosed individuals were plotted as the total number of HIV-infected remaining undiagnosed in the respective year. Undiagnosed interval was defined as the difference between year of seroconversion and year of diagnosis. The temporal trend of diagnosed, newly diagnosed, newly infected and undiagnosed were smoothed in 2-year windows by Seasonal-Trend Decomposition Procedure based on Loess in R3.2.2. [26] Associations between characteristics of patients and undiagnosed interval >2 years were examined in univariate analyses. We examined the confounding effects of ethnicity (Chinese vs non-Chinese), gender (female vs male), age at diagnosis (continuous variable) and mode of transmission (MSM vs non-MSM) in multivariable logistic regression models in SPSS. Confounding variables affecting >10% change of crude odds ratio (OR) of other independent variables were kept in the model for calculating adjusted odds ratio (aOR) of other independent variables. Patients were selected for analysis if they were diagnosed in 1991–2010, during which a CD4-guided approach was the norm for treatment initiation. We also used an alternative approach to estimate seroconversion time in sensitivity analysis. First, we estimated the seroconversion time with the original method proposed in this study. Second, we used Group B method to estimate the seroconversion time for patients with any pre-treatment CD4 counts (i.e. patients in Group A and Group B), while Group C method for Group C patients. Third, we used Group C method to estimate the seroconversion time for all patients. Lastly, we took the average undiagnosed interval among the 3 estimations for each patient, and performed univariate analysis to examine the association between long undiagnosed interval and independent variables.

## Results

As of 2012, 74,541 clinical measurements of 4551 HIV patients have been collected, of which data from 3695 adult (aged 18 or older) HIV patients diagnosed in 1991–2010 were selected for inclusion in the study. The sample size accounted for about 80% of diagnosed HIV cases recorded in the surveillance system during the 20-year period. Among them, 27% (999/3695) were in late HIV diagnosis, and 14% (528/3695) were seroconverters with an interval ≤2 years between last HIV negative testing date and first HIV positive testing date. The seroconversion year of each case was estimated, of which 1050 (28%) were in Group A, 2241 (61%) in Group B and 404 (11%) in Group C.

**Table 2. Pre-treatment CD4 depletion rate in cells/month estimated in the linear multilevel models for patients in Group A and Group B.**

	sexually acquired infections	infections in IDU
<b>Group A</b>		
<i>Dataset for multilevel model: CD4 data of patients with &gt;3 pre-treatment CD4 count</i>		
<b>Sample size</b>	8398 CD4 counts in 831 patients	728 CD4 counts in 95 patients
<b>ICC</b>	86%	89%
<b>Random effect</b>		
Months from baseline CD4	variance = 13.4, SD = 3.7	variance = 4.5, SD = 2.1
<b>Fixed effect</b>		
Months from baseline CD4	coe: -5.0 (s.e. = 0.16, t-value = -0.7)	coe: -4.2 (s.e. = 0.4, t-value = -11.9)
Being White	coe: 82.7 (s.e. = 28.6, t-value = 2.9)	/
<b>Group B</b>		
<i>Dataset for multilevel model: CD4 data of seroconverters</i>		
<b>Sample size</b>	2971 CD4 counts in 347 patients	72 CD4 counts in 8 patients
<b>ICC</b>	84%	51%
<b>Random effect</b>		
Months from seroconversion	variance = 17.8, SD = 4.2	variance = 0.3, SD = 0.5
<b>Fixed effect</b>		
Months from seroconversion	coe: -4.7 (s.e. = 0.3, t-value = -15.2)	coe: -2.2 (s.e. = 0.5, t-value = -4.6)
Being White	coe: 73.5 (s.e. = 30.1, t-value = 2.4)	/

ICC (intraclass correlation coefficient) = intercept variance / (intercept variance + residual variance) [19] coe-coefficient, s.e.-standard error

doi:10.1371/journal.pone.0159021.t002

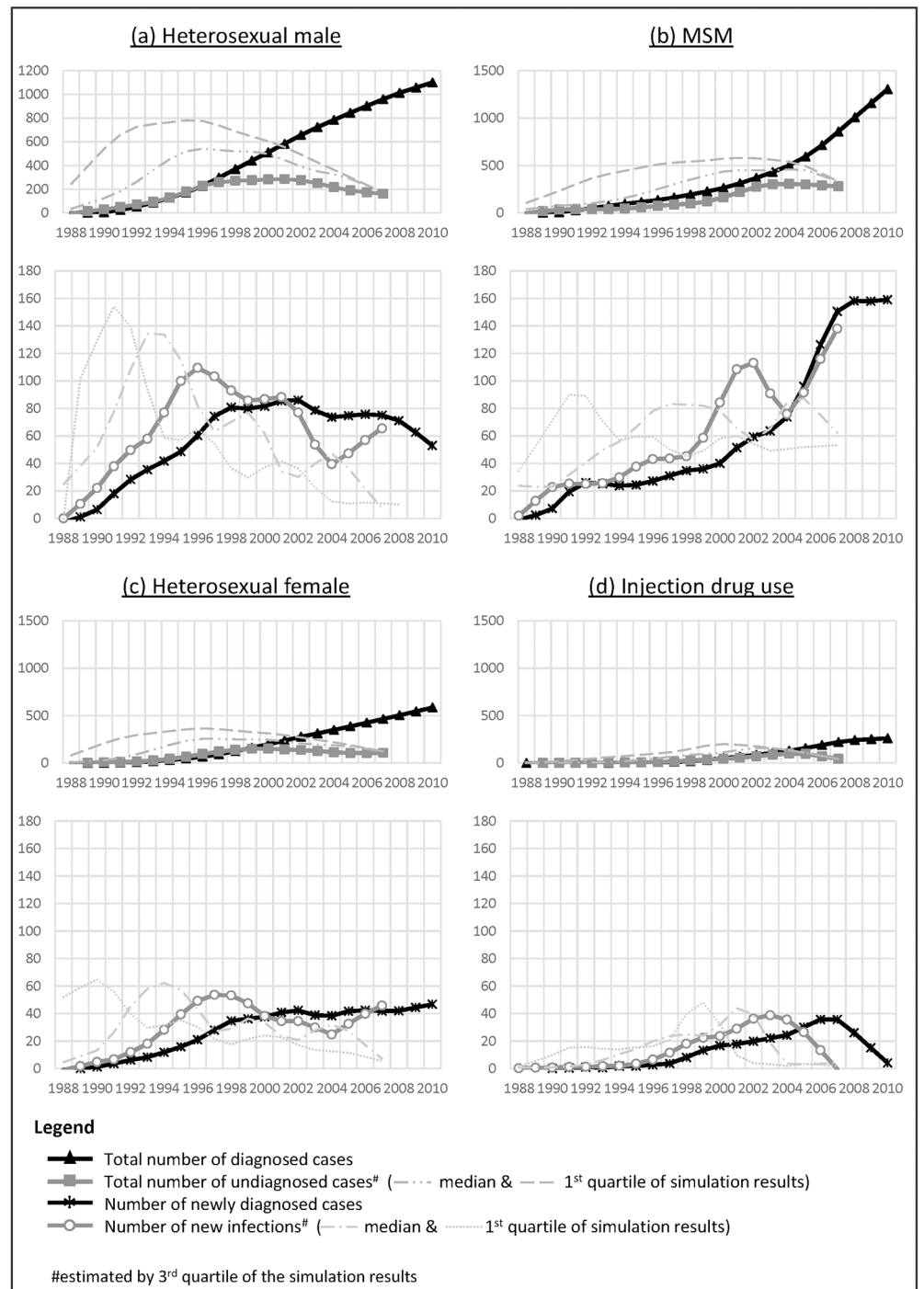
### Multilevel model results

Linear multilevel models were performed to estimate pre-treatment CD4 depletion rate in both Group A and Group B. There was high heterogeneity of CD4 depletion rate among patients, with ICC ranging between 51% and 89% (Table 2). In Group A, the CD4 depletion rate of patients who contracted HIV through sexual contact (n = 831 patients) and contaminated needle sharing (n = 95 patients) were similar at around -5 cells/μL per month from baseline. Among individuals infected through sexual contact, the White were adjusted to have higher CD4 (82.7/μL) than their counterparts. In Group B, pre-treatment CD4 depletion rate among seroconverters infected through sexual contact (n = 347 patients) was similar to those in Group A. However, the CD4 depletion rate of IDU seroconverters (n = 8 patients) was lower (-2/μL) than those in Group A. The third quartile of simulation results in Group A and Group B are shown in S1 Fig.

### Trends of annual number of new infections and undiagnosed infections

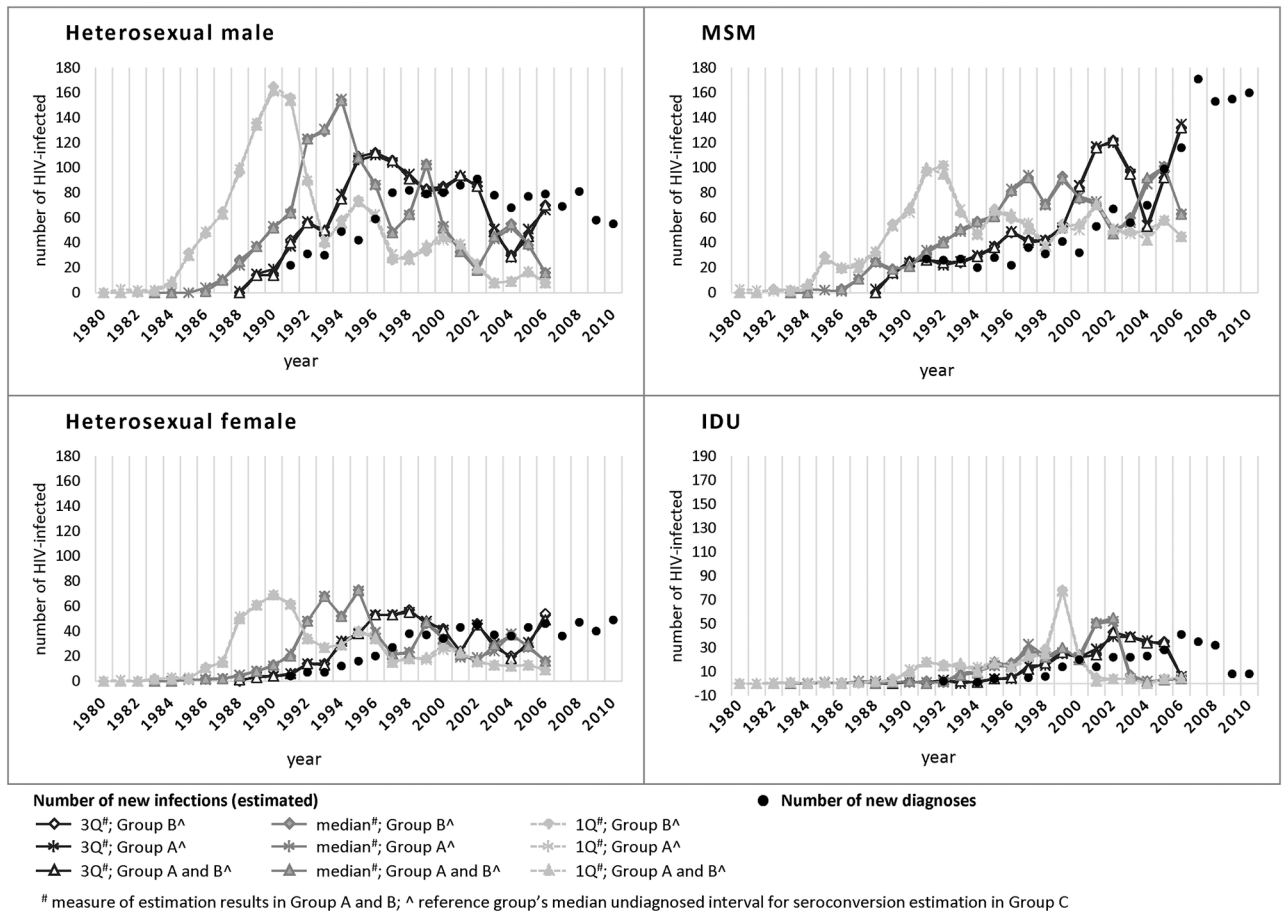
Compared with the annual new diagnoses, the peak of new infections occurred 4–6 years in advance (Fig 1). The peak for new heterosexual male and female infections was reached in 1996–1998, followed by a decline in 1998–2004. The rebound of new heterosexual male infections after 2004 on the newly constructed curve contrasted significantly with the plateauing of the reported incidence curve from 1998 onward. The annual total number of undiagnosed heterosexual male and female was at least two-folds that of new diagnoses in 1995–2007. The estimated proportion of undiagnosed individuals dropped linearly from 52% in 1996 to 16% in 2006 among heterosexual male, and from 60% to 20% in among heterosexual female.

For men-who-have-sex-with-men (MSM), the epidemic curves were distinctly different between new infections and new diagnoses. (Fig 1) Following a steady increase in the 1980s and 1990s, rapid upsurge of new infections could be seen in 1998–2001, six years in advance of



**Fig 1. Annual number of reported new diagnoses and prevalent cases (number of diagnosed HIV-infected cases who were alive), estimated new infections and undiagnosed infections (total number of infections remaining undiagnosed) with uncertainty intervals smoothed by Seasonal-Trend Decomposition Procedure based on Loess for (a) heterosexual male, (b) MSM, (c) heterosexual female and (d) IDU.**

doi:10.1371/journal.pone.0159021.g001



**Fig 2. Sensitivity analysis of new infection curves constructed by computing the first quartile, third quartile and median of the simulation results in Group A and B, and the median of undiagnosed interval in reference group (Group A, Group B, Group A and B) for Group C, by mode of transmission.**

doi:10.1371/journal.pone.0159021.g002

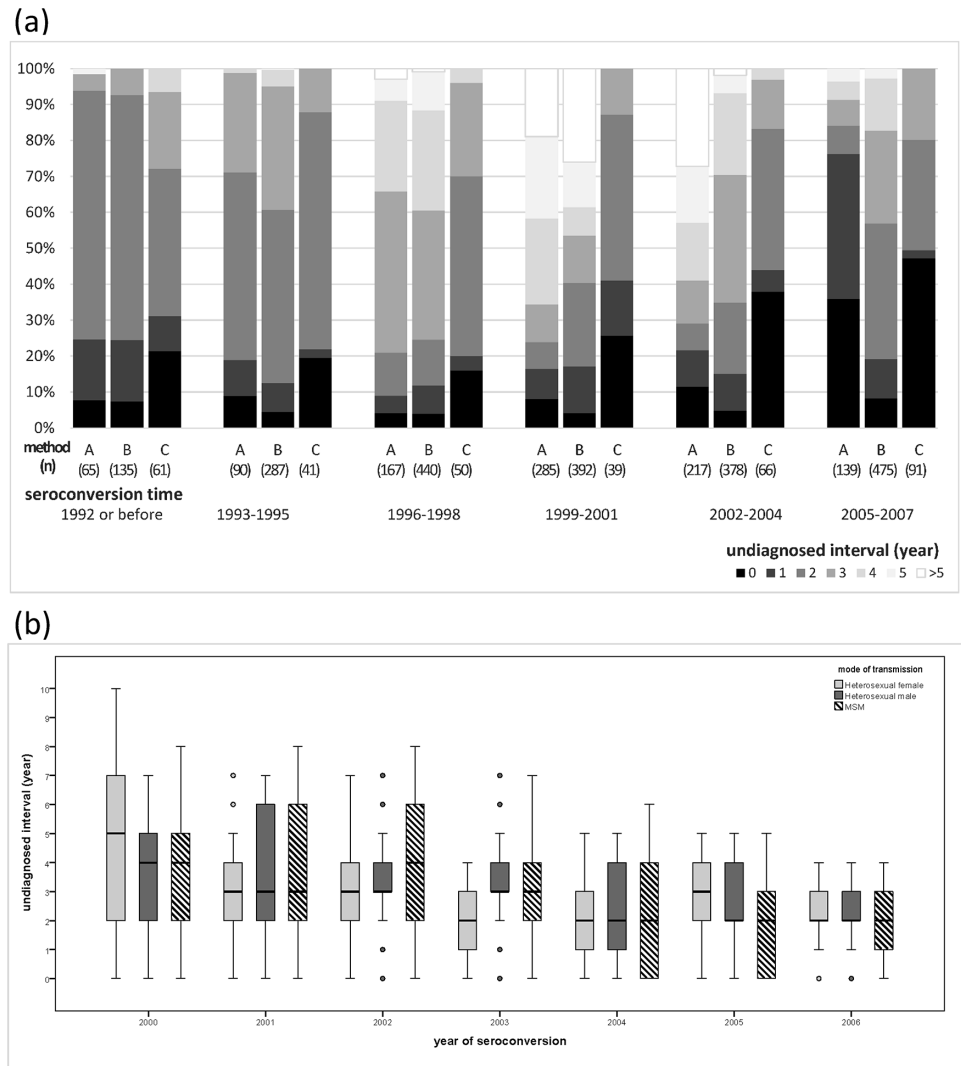
the new diagnoses curve. After a drop in 2001–2004, the MSM epidemic rose again, as shown by the increasing number of new infections. The annual number of undiagnosed MSM was at least twice that of the new diagnoses in 1995–2007, with a widening difference after 1999. The estimated proportion of undiagnosed MSM dropped and became stable at 32%–43% in 1996–2006. The contours of the injection drug users' (IDU) curves for new infections, new diagnoses and undiagnosed were similar, with a rise followed by a decline. There was however an earlier peak for the former (2003 vs 2007). Similar to the heterosexuals, the proportion of undiagnosed IDU fell from 56% in 1996 to 27% in 2006.

In the sensitivity analyses, the difference between annual number of new infections using different reference groups (Group A, Group B, or Group A and B) for estimating seroconversion year in Group C were minimal (Fig 2). The contours of new infection curves using different measures (first quartile, median and third quartile) from the 3 groups of patient data were largely similar, though the specific time of rise and fall varied.

### Undiagnosed interval

The interquartile range of estimated undiagnosed intervals was 1–4 years (range: 0–10 years). The proportion of short undiagnosed intervals ( $\leq 2$  years) increased across time. Also, the





**Fig 3.** A) Temporal variation of the proportion of undiagnosed intervals (years) and the seroconversion estimation methods applied. B) Yearly variation of undiagnosed intervals (years) by mode of transmission.

doi:10.1371/journal.pone.0159021.g003

distribution of estimated undiagnosed intervals varied both by the estimation methods and mode of transmission. Higher variation of undiagnosed interval was observed in Group A than the other two groups. (Fig 3a) The proportion of longer undiagnosed interval was higher in Group B than those in Group A, while the proportion in Group C was the lowest. By mode of transmission, the first quartile of undiagnosed interval in MSM was much shorter than that of heterosexual male and female. (Fig 3b) The interquartile range of undiagnosed intervals was at least 3 years in MSM after 1999, while the high variation of intervals among heterosexual male and female patients was seen in 2000–2001 only.

Dichotomizing the undiagnosed intervals by a threshold of 2 years, two groups (>2 years vs ≤2 years) were compared (Table 3). Patients with longer undiagnosed interval were more likely to have contracted HIV through heterosexual contact (OR = 1.61, 95% C.I. = 1.42–1.84), be female (OR = 1.20, 95% C.I. = 1.02–1.43), of Chinese ethnicity (OR = 1.29, 95% C.I. = 1.12–1.49), and age >64 at diagnosis (OR = 2.96, 95% C.I. = 2.05–4.26, aged ≤35 as reference group). At diagnosis, they were more likely to have higher baseline viral load (>log<sub>10</sub> 5 copies/mL)

(aOR = 1.25, 95% C.I. = 1.07–1.47) and to be in late HIV diagnosis (aOR = 2.87, 95% C.I. = 2.45–3.37), after adjusting for age at diagnosis, being MSM and Chinese. Adjusted by the same confounders, clinically, they were more likely to have initiated HAART (aOR = 1.82, 95% C.I. = 1.54–2.14), and be diagnosed with AIDS (aOR = 1.95, 95% C.I. = 1.69–2.25). In the sensitivity analysis, factors associated with undiagnosed interval >2 years as shown in [S2 Table](#) were similar to results in [Table 3](#), but gender was no longer significantly associated.

## Discussion

In this study, we estimated the seroconversion time of individual HIV-infected adults to reconstruct the epidemic curves for Hong Kong, an approach that managed to remove the influence of the highly varied patterns of HIV diagnoses. Compared to a previous local study which used a modified back-calculation method at population level for estimating HIV incidence,[\[6\]](#) our incidence curve showed a similar epidemiologic trend with their estimated curve, but our curve ran 4–5 years ahead of theirs. Apart from the different sources and characteristics of data for estimation, the variation of parameters used could be another reason of discrepancy. In our study, we used pre-treatment CD4 depletion rates to back-calculate the year of seroconversion, while the previous local study was parametrized by the lengths of incubation period, settings of diagnosis (routine testing and symptom-related testing) and the annual number of HIV diagnoses and AIDS diagnoses.[\[6\]](#) Elsewhere, back-calculation of the seroconversion year with CD4 depletion rate has been applied,[\[12, 13, 15\]](#) though most had used the depletion rate of seroconverters as reference group. The latter was adopted as one of the three approaches for estimation in this study. Another approach was applied for patients with very comprehensive CD4 reference data, while a third approach was adopted for patients without any pre-treatment CD4 count. Our mixed-methods approach for estimation illustrated the complexity of data availability in reality, reflecting that a single method might not be able to solve all problems. Our estimated pre-treatment CD4 depletion rate was 56/ $\mu$ L–60/ $\mu$ L per year for sexually acquired HIV infections, which was consistent with the estimated range for patients from Europe and Australia.[\[27\]](#)

Apart from showing an earlier peak in the incidence curves compared to new diagnoses curves, it is noted that the contour of the epidemic curves varied remarkably by the route of HIV transmission. Instead of a continuous rise in the number of newly diagnosed MSM through 2010, our seroconversion curve peaked in 2000–2001, followed by a temporal decline in 2001–2004, and then rose back to the 2001 level. With the increasing number of undiagnosed MSM in the community, the number of newly diagnosed MSM was expected to rise significantly and continuously. However, the observed increase was relatively modest. This could be partly explained by the relatively short undiagnosed interval in HIV-infected MSM. The transmission risk tended to fall after diagnosis as a result of behavioural changes [\[2\]](#) and viral load suppression following HAART.[\[4\]](#) The trend was also consistent with the relatively stable HIV prevalence of MSM reported at different time-points in Hong Kong in 2006, 2008 and 2011.[\[28\]](#)

The epidemic growth for heterosexually acquired infection was very different. The total number of undiagnosed HIV infected heterosexuals increased in 1996–2002, followed by a slower growth afterwards. Unlike MSM, HIV-infected heterosexuals were more likely to have longer undiagnosed interval. This might be partly due to the high proportion of non-locally acquired heterosexual infections,[\[29\]](#) and could also be a result of their different partnership pattern. A Taiwan study on Chinese heterosexual partnership showed that many (around 56%) were in serial monogamy [\[30\]](#), implying on-going transmission could be limited. When a heterosexual person is infected with HIV, virus transmission might occur to affect 1 or 2 sex

**Table 3. Characteristics of HIV-infected patients with long (>2 years) undiagnosed interval from estimated seroconversion to HIV diagnosis, compared to patients with a shorter interval.**

	Undiagnosed interval ≤2 years (n = 1929)		Undiagnosed interval >2 years (n = 1766)		Crude Odds Ratio (OR)		Adjusted Odds Ratio (aOR)‡	
	no.	%	no.	%	OR	95% CI	aOR	95% CI
<b>Demographics</b>								
Gender								
Male	1612	53%	1428	47%		<i>ref</i>		
Female	317	48%	338	52%	1.20	1.02–1.43*		
Ethnicity								
Non-Chinese (Asian, White, others, NA)	593	57%	452	43%		<i>ref</i>		
Chinese	1336	50%	1314	50%	1.29	1.12–1.49*		
Mode of transmission								
Non-heterosexuals (MSM, IDU, BL and UN)	1036	58%	739	42%		<i>ref</i>		
Heterosexuals	893	47%	1027	53%	1.61	1.42–1.84*		
<i>Heterosexual female</i>	298	48%	326	52%		<i>ref</i>		
<i>Heterosexual male</i>	595	46%	701	54%	1.08	0.89–1.3		
Non-MSM (heterosexuals, IDU, BL and UN)	1082	47%	1223	53%		<i>ref</i>		
MSM	847	61%	543	39%	0.57	0.5–0.65*		
<b>Conditions at diagnosis</b>								
Age at diagnosis								
Aged ≤35	1075	61%	696	39%		<i>ref</i>		
Aged 36–64	807	45%	980	55%	1.88	1.64–2.14*		
Aged >64	47	34%	90	66%	2.96	2.05–4.26*		
Baseline VL (copies/mL)								
≤ log <sub>10</sub> 5	823	47%	924	53%		<i>ref</i>		<i>ref</i>
>log <sub>10</sub> 5	432	40%	660	60%	1.36	1.17–1.59*	1.25	1.07–1.47*
Late HIV diagnosis <sup>#</sup>								
No	1622	60%	1074	40%		<i>ref</i>		<i>ref</i>
Yes	307	31%	692	69%	3.4	2.92–3.97*	2.87	2.45–3.37*
<b>Treatment</b>								
Initiated HAART by 2012								
No	552	64%	313	36%		<i>ref</i>		<i>ref</i>
Yes	1377	49%	1453	51%	1.86	1.59–2.18*	1.82	1.54–2.14*
Time from HAART to SVL								
>3 months	783	48%	845	52%		<i>ref</i>		<i>ref</i>
≤3 months	606	50%	618	50%	0.94	0.81–1.1	0.97	0.83–1.13
<b>Outcomes</b>								
Years with NSVL after diagnosis								
≤ 3 years	1030	47%	1173	53%		<i>ref</i>		<i>ref</i>
>3 years	899	60%	593	40%	0.58	0.51–0.66*	0.66	0.57–0.76*
Ever diagnosed with AIDS								
No	1387	60%	932	40%		<i>ref</i>		<i>ref</i>
Yes	542	39%	834	61%	2.29	2–2.62*	1.95	1.69–2.25*
Deceased by 2012								
No	1725	54%	1498	46%		<i>ref</i>		<i>ref</i>
Yes	204	43%	268	57%	1.51	1.25–1.84*	1.03	0.84–1.27

\*p-value <0.05

MSM-men who have sex with men, IDU-injection drug use, BL-blood transfusion, UN-undetermined, VL-viral load, HAART- highly active antiretroviral therapy, SVL-suppressed viral load (≤500copies/mL), NSVL-non-suppressed viral load (>500copies/mL)

‡adjusted by mode of transmission (MSM vs non-MSM), age at diagnosis (continuous) and ethnicity (Chinese vs non-Chinese) in multivariable logistic regression model

<sup>#</sup>Late HIV diagnosis refers to patients diagnosed with AIDS within 3 months of HIV diagnosis

doi:10.1371/journal.pone.0159021.t003

partners regardless of the length of the undiagnosed interval. This pattern differs considerably from the dense networking of some MSM, whose HIV infection may be characterised by rapid growth.[31] While female sex workers (FSW) could form a bridge in intensifying virus spread, their low HIV prevalence of below 0.2%,[32, 33] means that rapid dissemination from HIV infected FSW to their clients was unlikely.

For IDU, the flat contour of newly infected IDU lent support to the uncommon occurrence of local transmission, which could be attributed to the longstanding methadone treatment programme introduced since the 1970s.[34] For both heterosexuals and IDU, the epidemic curves redrawn with estimated seroconversion years ran 4 years in advance of their diagnoses. Evidently, the estimation of new infections has contributed to the assessment of HIV transmission risk in the population and subpopulations.

The length of undiagnosed intervals carries significant public health implications. Our study has identified the following associating factors with long undiagnosed interval in Hong Kong: heterosexual, female, Chinese ethnicity, and age >64 at diagnosis. The longer undiagnosed interval in heterosexual male/female and elderly was probably related to their lower perceived risks of infection, compared to younger MSM in the local community, an observation that has also been made elsewhere.[35] Because of the low perceived risk of infection, HIV testing experiences of heterosexuals, especially the elderly, are predictably low. With 27% of patients in late HIV diagnosis, expanded HIV testing is therefore one most imminent strategy for shortening the undiagnosed interval, which can serve not just to improve clinical outcomes but also reduce transmission risk in the community. An earlier diagnosis of HIV infection is desirable regardless of the route of transmission. Our results showed that early diagnosis was negatively associated with high baseline viral load and AIDS diagnosis, adjusted by ethnicity, mode of transmission and age at diagnosis. Paradoxically, they had a longer duration of non-suppressed viral load, which was probably due to higher baseline CD4 and therefore longer interval between diagnosis and treatment initiation, in accordance with the CD4-guided approach to HAART. Expanded treatment initiation has recently been recommended by World Health Organization to cover all diagnosed patients with any CD4 count,[36] a strategy supported by increasing number of clinical studies which would predictably shorten the duration of non-suppressed viral load if early diagnosis could be achieved. Separately, a meta-analysis study concluded that there was a lower prevalence of risk behaviour after one became aware of the HIV status.[2] A combination of early diagnosis and universal treatment without regard to the prevailing CD4 count would predictably reduce the undiagnosed interval and facilitate the plateauing of the HIV epidemic curves.

We acknowledge that the study carries some limitations. As it takes time to capture a majority of the infected patients in a year, estimation by back-calculation approach was limited to the use of data before 2008, five years before the end of our data collection period. The estimation for patients in Group C was based on the estimation result in Group B, which might not be the best reference group to represent patients in Group C. However, from the sensitively analyses for reference group selections, Group B might be a better option under existing data availability. Also, in the course of conducting CD4 depletion rate estimation, the small proportion of rapid progressors and long-term non-progressors had been ignored. Their contribution to the epidemic growth is however predicted to be small. We assumed linear trajectory of pre-treatment CD4 decline and ignored the association between CD4 and viral load levels. We acknowledge there exists a wide range of uncertainties in applying our proposed estimation methods, and there are rooms for improvement in future epidemiologic estimation. Finally, it may not be easy to have Group A method regularly applied as the immediate treatment strategy is becoming a routine in clinical services. In practice, the future approach might rely more on Group B and C with additional information such as patients' perception of their infection date,

viral load measurement, seroconversion illnesses and molecular study. Nonetheless, we believe our mixed-methods approach has demonstrated the feasibility of using readily available longitudinal clinical data for epidemiologic estimation.

In conclusion, we have endeavoured to estimate the undiagnosed interval at individual level with currently available clinical data, without resorting to sophisticated laboratory technologies, to reconstruct HIV epidemic curves founded on seroconversion time rather than clinical diagnoses. The study was conducted in a small city where a majority of patients were managed according to similar clinical guidelines. In principle, a plausible estimation approach can be adopted, which can be turned into a public health tool for describing and projecting HIV epidemiology in places where pre-treatment CD4 data are available from the HIV services.

## Supporting Information

**S1 Table. The central tendency of discrepancy between simulation results in Group A and mid-point interval of seroconverters.**

(PDF)

**S1 Fig. Distribution of 3rd quartile of simulation results for seroconversion year estimation in Group A and Group B.**

(TIF)

**S2 Table. Characteristics of HIV-infected patients with long (>2 years) undiagnosed interval from estimated seroconversion to HIV diagnosis in sensitivity analysis, compared to patients with a shorter interval.**

(PDF)

## Acknowledgments

The authors thank all patients and staff of the Integrated Treatment Centre, Queen Elizabeth Hospital and Princess Margaret Hospital for their assistance in the maintenance of clinical data used in this study. Li Ka Shing Institute of Health Sciences and Stanley Ho Centre for Emerging Infectious Diseases of The Chinese University of Hong Kong are acknowledged for providing technical support in conducting the research. NSW is supported by Guangdong Provincial Center for Skin Diseases and STI Control and the UNC-South China STD Research Training Center (FIC1D43TW009532) and NIH Fogarty International Center.

## Author Contributions

Conceived and designed the experiments: SSL NSW. Performed the experiments: NSW DPCC. Analyzed the data: NSW. Contributed reagents/materials/analysis tools: KHW MPL OTYT. Wrote the paper: NSW SSL.

## References

1. Maartens G, Celum C, Lewin SR. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet* (London, England). 2014; 384:258–71. doi: [10.1016/S0140-6736\(14\)60164-1](https://doi.org/10.1016/S0140-6736(14)60164-1)
2. Marks G, Crepaz N, Senterfitt JW, Janssen RS. Meta-analysis of high-risk sexual behavior in persons aware and unaware they are infected with HIV in the United States: implications for HIV prevention programs. *Journal of Acquired Immune Deficiency Syndromes* (1999). 2005; 39:446–53.
3. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *The New England Journal of Medicine*. 2011; 365:493–505. doi: [10.1056/NEJMoa1105243](https://doi.org/10.1056/NEJMoa1105243) PMID: [21767103](https://pubmed.ncbi.nlm.nih.gov/21767103/)
4. Cohen MS, Chen Y, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Final results of the HPTN 052 randomized controlled trial: antiretroviral therapy prevents HIV transmission. IAS 2015

- 8th Conference on HIV Pathogenesis, Treatment and Prevention; July 19–22; Vancouver, Canada2015. p. 9.
5. Brookmeyer R, Gail MH. Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States. *Lancet*. 1986; 2(8519):1320–2. PMID: [2878184](#).
  6. Chau PH, Yip PSF, Cui JS. Reconstructing the incidence of human immunodeficiency virus (HIV) in Hong Kong by using data from HIV positive tests and diagnoses of acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2003; 52:237–48. doi: [10.1111/1467-9876.00401](#)
  7. Mastro TD, Kim AA, Hallett T, Rehle T, Welte A, Laeyendecker O, et al. Estimating HIV Incidence in Populations Using Tests for Recent Infection: Issues, Challenges and the Way Forward. *J HIV AIDS Surveill Epidemiol*. 2010; 2(1):1–14. PMID: [21743821](#); PubMed Central PMCID: PMCPMC3130510.
  8. Hall HI, Song R, Rhodes P, Prejean J, An Q, Lee LM, et al. Estimation of HIV incidence in the United States. *JAMA*. 2008; 300(5):520–9. doi: [10.1001/jama.300.5.520](#) PMID: [18677024](#); PubMed Central PMCID: PMCPMC2919237.
  9. Jia Z, Huang X, Wu H, Zhang T, Li N, Ding P, et al. HIV burden in men who have sex with men: a prospective cohort study 2007–2012. *Sci Rep*. 2015; 5:11205. doi: [10.1038/srep11205](#) PMID: [26135810](#).
  10. Armstrong G, Medhi GK, Mahanta J, Paranjape RS, Kermode M. Undiagnosed HIV among people who inject drugs in Manipur, India. *AIDS care*. 2015; 27:288–92. doi: [10.1080/09540121.2014.972322](#) PMID: [25345544](#)
  11. Ferrer L, Furegato M, Foschia J-P, Folch C, González V, Ramarli D, et al. Undiagnosed HIV infection in a population of MSM from six European cities: results from the Sialon project. *European Journal of Public Health*. 2015; 25:494–500. doi: [10.1093/eurpub/cku139](#) PMID: [25161202](#)
  12. Rice BD, Elford J, Yin Z, Delpech VC. A new method to assign country of HIV infection among heterosexuals born abroad and diagnosed with HIV. *AIDS*. 2012; 26(15):1961–6. doi: [10.1097/QAD.0b013e3283578b80](#) PMID: [22781226](#).
  13. Infections HPACf. Longitudinal analysis of the trajectories of CD4 cell counts.
  14. Ndawinz JD, Costagliola D, Supervie V. New method for estimating HIV incidence and time from infection to diagnosis using HIV surveillance data: results for France. *AIDS*. 2011; 25(15):1905–13. doi: [10.1097/QAD.0b013e32834af619](#) PMID: [21811147](#).
  15. Lodwick RK, Nakagawa F, van Sighem A, Sabin CA, Phillips AN. Use of surveillance data on HIV diagnoses with HIV-related symptoms to estimate the number of people living with undiagnosed HIV in need of antiretroviral therapy. *PLoS One*. 2015; 10(3):e0121992. doi: [10.1371/journal.pone.0121992](#) PMID: [25768925](#); PubMed Central PMCID: PMCPMC4358920.
  16. Wong NS, Wong KH, Wong PKH, Lee SS. Incorporation of estimated community viral load before HIV diagnosis for enhancing epidemiologic investigations: a comparison between men who have sex with men and heterosexual men in Hong Kong. *Asia Pac J Public Health*. 2015; 27:756–64. doi: [10.1177/1010539515589340](#) PMID: [26041836](#)
  17. Adolescents. PoAGfAa. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents.
  18. Wong NS, Reidpath DD, Wong KH, Lee SS. A multilevel approach to assessing temporal change of CD4 recovery following HAART initiation in a cohort of Chinese HIV positive patients. *J Infect*. 2015; 70(6):676–8. doi: [10.1016/j.jinf.2014.10.012](#) PMID: [25452038](#).
  19. Merlo J, Yang M, Chaix B, Lynch J, Råstam L. A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *Journal of Epidemiology and Community Health*. 2005; 59:729–36. doi: [10.1136/jech.2004.023929](#) PMID: [16100308](#)
  20. Jentsch-Ullrich K, Koenigsmann M, Mohren M, Franke A. Lymphocyte subsets' reference ranges in an age- and gender-balanced population of 100 healthy adults—a monocentric German study. *Clinical Immunology (Orlando, Fla)*. 2005; 116:192–7. doi: [10.1016/j.clim.2005.03.020](#)
  21. Wong WS, Lo AWI, Siu LP, Leung JNS, Tu SP, Tai SW, et al. Reference ranges for lymphocyte subsets among healthy Hong Kong Chinese adults by single-platform flow cytometry. *Clinical and vaccine immunology: CVI*. 2013; 20:602–6. doi: [10.1128/CVI.00476-12](#) PMID: [23408529](#)
  22. Kassu A, Tsegaye A, Petros B, Wolday D, Hailu E, Tilahun T, et al. Distribution of lymphocyte subsets in healthy human immunodeficiency virus-negative adult Ethiopians from two geographic locales. *Clinical and Diagnostic Laboratory Immunology*. 2001; 8:1171–6. doi: [10.1128/CDLI.8.6.1171-1176.2001](#) PMID: [11687459](#)
  23. Forbi JC, Forbi TD, Agwale SM. Estimating the time period between infection and diagnosis based on CD4+ counts at first diagnosis among HIV-1 antiretroviral naïve patients in Nigeria. *Journal of Infection in Developing Countries*. 2010; 4:662–7. PMID: [21045361](#)

24. Taffé P, May M, Study SHC. A joint back calculation model for the imputation of the date of HIV infection in a prevalent cohort. *Statistics in Medicine*. 2008; 27:4835–53. doi: [10.1002/sim.3294](https://doi.org/10.1002/sim.3294) PMID: [18444229](https://pubmed.ncbi.nlm.nih.gov/18444229/)
25. Vanhems P, Lambert J, Guerra M, Hirschel B, Allard R. Association between the rate of CD4+ T cell decrease and the year of human immunodeficiency virus (HIV) type 1 seroconversion among persons enrolled in the Swiss HIV cohort study. *The Journal of Infectious Diseases*. 1999; 180:1803–8. doi: [10.1086/315110](https://doi.org/10.1086/315110) PMID: [10558934](https://pubmed.ncbi.nlm.nih.gov/10558934/)
26. Cleveland R, Cleveland W, Mcrae J, Terpenning I. STL: a seasonal-trend decomposition procedure based on Loess. *J Official Statistics*. 6:3–73.
27. Wolbers M, Babiker A, Sabin C, Young J, Dorrucchi M, Chene G, et al. Pretreatment CD4 cell slope and progression to AIDS or death in HIV-infected patients initiating antiretroviral therapy—the CASCADE collaboration: a collaboration of 23 cohort studies. *PLoS Med*. 2010; 7(2):e1000239. doi: [10.1371/journal.pmed.1000239](https://doi.org/10.1371/journal.pmed.1000239) PMID: [20186270](https://pubmed.ncbi.nlm.nih.gov/20186270/); PubMed Central PMCID: PMCPCMC2826377.
28. Wong HTH, Wong KH, Lee SS, Leung RWM, Lee KCK. Community-based surveys for determining the prevalence of HIV, chlamydia, and gonorrhoea in men having sex with men in Hong Kong. *Journal of Sexually Transmitted Diseases*. 2013; 2013:958967. doi: [10.1155/2013/958967](https://doi.org/10.1155/2013/958967) PMID: [26316969](https://pubmed.ncbi.nlm.nih.gov/26316969/)
29. Department of Health HKSAR. HIV surveillance report– 2014 update. Hong Kong: December 2015. Report No.
30. Chang H, Ruan F, Chien S. The study in the survey of the marriage and the affairs cognition: for married and single men and women. *Shu-Te Online Studies of Humanities and Social Sciences*. 2013; 9(1).
31. Chen JH, Wong KH, Li P, Chan KC, Lee MP, Lam HY, et al. Molecular epidemiological study of HIV-1 CRF01\_AE transmission in Hong Kong. *J Acquir Immune Defic Syndr*. 2009; 51(5):530–5. doi: [10.1097/QAI.0b013e3181aac516](https://doi.org/10.1097/QAI.0b013e3181aac516) PMID: [19521252](https://pubmed.ncbi.nlm.nih.gov/19521252/).
32. Department of Health HKSAR. CRiSP—community based risk behavioural and seroprevalence survey for female sex workers in Hong Kong 2006. October 2007. Report No.
33. Department of Health HKSAR. CRiSP—community based risk behavioural and seroprevalence survey for female sex workers in Hong Kong 2009. May 2010. Report No.
34. Lee SS. A humble service that has delivered public health good. *Public Health*. 2007; 121:884–6. doi: [10.1016/j.puhe.2007.02.017](https://doi.org/10.1016/j.puhe.2007.02.017) PMID: [17570450](https://pubmed.ncbi.nlm.nih.gov/17570450/)
35. Ellman TM, Sexton ME, Warshafsky D, Sobieszczyk ME, Morrison EAB. A forgotten population: older adults with newly diagnosed HIV. *AIDS patient care and STDs*. 2014; 28:530–6. doi: [10.1089/apc.2014.0152](https://doi.org/10.1089/apc.2014.0152) PMID: [25211596](https://pubmed.ncbi.nlm.nih.gov/25211596/)
36. World Health Organization. WHO | Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV 2015 [updated 2015-10-09 06:58:51].