

## Current Status of the Estimation on the Number of People Who Living with HIV and the Rate of Undiagnosed Cases

Myeongsu Yoo, Cheol-Hee Yoon, Byeong-Sun Choi\*

Division of Viral Disease Research, Center for Infectious Diseases Research, Korea National Institute of Health, Cheongju 28159, Republic of Korea

### Corresponding

Byeong-Sun Choi, Ph.D.  
Director, Division of Viral Disease  
Research, Center for Infectious Disease  
Research, Korea National Institute of  
Health, 187 Osongsaengmyeong2-ro,  
Osong-yeup, Cheongju 28159, Republic  
of Korea

Phone : +82-43-719-8410

Fax : +82-43-719-8459

E-mail : [byeongsun@korea.kr](mailto:byeongsun@korea.kr)

Received : September 2, 2020

Revised : September 23, 2020

Accepted : September 24, 2020

Human immunodeficiency virus (HIV) is a major public health issue worldwide. As of 2018, 37.9 million people worldwide live with HIV, 1.7 million of which are new HIV infections, and 770,000 are surmised to have died from Acquired immune deficiency syndrome (AIDS) related illnesses. However, the exact number of HIV infections cannot be confirmed; The Joint United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO) have computed and disclosed the number of HIV infections, new HIV infections, and AIDS mortality in participating countries for several years to track and estimate the current HIV prevalence, and the organizations are striving to enhance the accuracy of estimation of current HIV infections by ameliorating various techniques. In South Korea, the government only discloses the number of new HIV infections as an official statistic, and there is no agreed method of estimating current HIV infections. Thus, in this article, we introduce various methods for estimating HIV infections and methods to reflect the number of undiagnosed HIV infections in Korea to the estimate.

**Key Words:** Undiagnosed HIV/AIDS, Estimation, Back-calculation method

## INTRODUCTION

No potential conflict of interest relevant to this article was reported.

Copyright © 2020 Journal of Bacteriology and Virology

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>).

Human immunodeficiency virus (HIV) remains a major public health issue in the world (1). In 2018, the number of people infected HIV around the world was 37.9 million, of which 1.7 million were newly diagnosed, while 0.77 million people died of Acquired immune deficiency syndrome (AIDS)-related diseases (2). Worldwide, new cases of HIV infection are gradually decreasing. By continent, the trend has decreased in southeast Africa, west and central Africa, and the Asia Pacific region, and increase in Eastern Europe, Central Asia, the Middle East and North Africa (3). In the European Union, it has been reported that about one-third of HIV infected people were unaware that they were infected (4), and in some eastern European countries up to 60% of persons infected with AIDS/HIV have not been diagnosed (5). In Korea, the first cases of HIV infection were found and reported in 1985, and the number of cases has increased continuously since 1988. In 2018, 1,206 new cases (989 Koreans, 217 foreign nationals) were diagnosed and reported, bringing the current number of people living with HIV infection to 12,991 (6).

However, HIV does not immediately lead to the onset of AIDS symptoms upon infection, so the number of HIV diagnoses and number of undiagnosed HIV infections are both estimates. Hence, there is a method that estimates the number of HIV infections as measures to respond to the risk. This plays a crucial role in reducing the burden of HIV infection and strengthening the rationale for appropriate treatment and HIV testing.

The method of estimating the new HIV infected persons are an important evidence for the true economic burden of HIV infected, the corresponding need for treatment and the strengthening of HIV testing. Two organizations, The Joint United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO), have calculated the number of HIV-infected people, the number of newly infected people, and the AIDS mortality rate in participating countries over a period of several years in a bid to follow and estimate the trend of HIV infection. As there is no agreed method of estimating the number of HIV-infected people in Korea at present, the purpose of this paper is to review the integrated and comprehensive methodology, and the approaches to calculating the number of people infected with HIV that are currently available.

## THE ESTIMATION OF THE NUMBER OF HIV-INFECTED PEOPLE

### Selection of major variables for estimating the number of HIV-infected people

It is most important to select the major factors to be considered in estimating the number of HIV-infected people, and the major variables with priority selected through the methods of calculation presented in several research papers are as follows.

#### Prevalence calculation

Prevalence calculation is a method of obtaining an estimate of the rate of HIV-infected people from a sample group of persons in a specific risk category. As prevalence is a quantitative method, it can produce sampling errors, and the estimate can be biased if the sample does not fully represent the risk group; however, the bias can be reduced by increasing the size of the sample or dividing the risk group (7).

#### Collection of clinical information at diagnosis

Using the HIV surveillance system, divided stratification of the estimation of HIV-infected people makes it possible to estimate the number of people infected with HIV through such data as basic demographic data, date of diagnosis, HIV infection pathway, CD4+ T cell count at diagnosis, symptoms at diagnosis (AIDS-related or non-related symptoms), antibody testing or recent infections. Clinicians in some European countries including the United Kingdom (UK) must periodically report treatment outcomes (such as the most recent CD4+ T cell count, the patient's treatment status). Ideally, as in the case of the Netherlands, the clinical information of all diagnosed HIV-infected people is collected on a continuous basis for use in public health.

#### AIDS cases

Although the effectiveness of using reported cases of AIDS as basic data has declined since the use of antiretroviral treatment, it can provide information on the size of the undiagnosed population through late HIV finding (diagnosis after AIDS development), and thus the related data should be collected continuously as a matter of priority (8).

#### Deaths among people with HIV/AIDS

Because the number of mortalities should be excluded when calculating the number of HIV-infected people, information

on HIV/AIDS mortality is a useful variable for estimating the number of infected people. But, for a more accurate calculation of the number of infections, the bias in mortality should also be considered, as in the case for undiagnosed HIV/AIDS-infected mortality.

Most methods of calculating the number of HIV-infected people recommend collecting as many of those variables as possible so as to be able to estimate the number of cases more accurately.

## Estimation of undiagnosed HIV-infected people

### Prevalence estimation

HIV prevalence within a population can be estimated by summing the HIV prevalence estimate within an at-risk group and sample size estimates for each at-risk group. In other words, the HIV prevalence estimate can be calculated by estimating HIV prevalence within each group and multiplying it by the population size, and if this is calculated for all at-risk groups, the total HIV infections can be estimated (9).

The number of undiagnosed HIV infections can be estimated by subtracting the number of HIV diagnoses for each at-risk group from the total HIV infection estimate. One major limitation of prevalence estimation is the need to match the size of at-risk population to HIV prevalence in those populations. For example, the male HIV prevalence in the MSM (Men who have Sex with Men) group estimated by estimating the MSM population does not apply to the stratification of the entire male population. This method is utilized by UNAIDS (UNAIDS, <http://www.unaids.org/en/KnowledgeCentre/HIVDATA/Epidemiology/E>), and the shortcoming of infection estimation can be overcome with enhanced accuracy of the method if the high-risk and low-risk groups are appropriately distributed (10, 11).

### Estimation of high-risk population size

While multiple methods have been employed to estimate the size of high-risk populations, it has been confirmed that these populations are underestimated due to reasons such as stigma. Thus, more accurate methods of estimation, such as network scale-up method (weighting responses from acquaintances) and developing indices that minimize potential bias, have been proposed (7).

### The multi-parameter evidence synthesis method (MPES) in AIDS patients

The MPES method is based on prevalence survey and sample size of at-risk groups, but it was designed to include the whole body of information pertaining to other approaches, prevalence estimation, and sample size of at-risk groups in its data source (12-14). The MPES technique integrates various bodies of evidence through mathematical triangulation and establishes a framework for detecting the link between evidences that show an equal quantity or mediating variables. Some information may be biased, but the MPES technique can model parameters without such biases. Estimation is performed using the Bayesian inference, and this technique attempts to produce an accurate and consistent estimate of HIV infections without uncertainty of the estimation process.

### HIV cumulative incidence method

HIV cumulative incidence method enables estimation of both current HIV incidence and past HIV incidence. Cumulative HIV deaths is needed to estimate the current HIV prevalence, and this can be calculated by subtracting from HIV cumulative incidence. One major limitation of the HIV cumulative incidence method is that incidence at the time of infection cannot be estimated solely based on diagnosis results. Therefore, adding data at the time of diagnosis, such as AIDS onset and CD4+ T cell count could enhance the estimation accuracy. The HIV cumulative incidence method assumes that all HIV infections are diagnosed during the patient's lifetime or upon death, and this varies according to the HIV diagnostic testing situations in each country. Furthermore, data on total cumulative HIV deaths is also needed, and the possibility of underestimation due to a gap from the number of registered deaths must be taken into consideration (7).

### Back calculation method

The back-calculation method, which was the most widely used approach in early HIV infection estimation, estimates HIV infections by estimating the distribution of annual AIDS infections and time from HIV infection to AIDS onset. If the data on cumulative HIV-related deaths is accurate and available, total HIV infections can be estimated, and the number of undiagnosed HIV infections is estimated based on the data on the diagnosed HIV infections (5, 7). Due to the lack of an effective treatment at the time, most HIV infections progressed to AIDS. However, the introduction of an effective antiretroviral therapy (ART) stalled the progression of most HIV infections to AIDS, and for this reason, estimating the distribution of time from HIV infection to AIDS onset has become obsolete. Hence, the basic back-calculation method is no longer used as a technique for estimating undiagnosed HIV infections. In addition, even if the distribution of time from infection to AIDS onset can be estimated, it is possible that the patients diagnosed with AIDS had not been diagnosed with HIV infection; thus, it cannot provide reliable information on the trend of recent HIV incidence among AIDS patients. However, the technique is used after some modification and expansion, such as replacing number of AIDS patients with number of HIV diagnoses.

### Expanded back-calculation method

(Bayesian back-calculation method, UK) The Biostatistics Unit at Cambridge, UK, calculates the number of undiagnosed HIV infections with the Bayesian back-calculation using a multi-state model using the MSM group data from the Health Protection Agency (HPA) data (15). This approach requires data on new HIV diagnoses by quarter and AIDS diagnosis in the same quarter (late diagnosis), and CD4+ T cell count is strongly recommended but not essential for some data. Data is computed by strata of risk groups.

(The serological method, US CDC) The US CDC in Atlanta, Georgia computes HIV incidence in the US using an expanded back-calculation method and compares this with the results of classifying recent infections and long-term infections based on a serological method (16). The expanded back-calculation method was used to estimate the number of undiagnosed HIV infections and HIV incidence in the US. Data required for such estimation include demographic data, whether AIDS and HIV were diagnosed on the same year, and number of annual HIV diagnoses (disease progression).

### (Used MSM, Persons who Inject Drugs (PWID), and heterosexual risk group)

National Centre in HIV Epidemiology and Clinical Research in Sidney, Australia developed another expanded back-calculation method in collaboration with the research team at Public Health Agency of Canada. The team restructured the MSM, PWID, and heterosexual risk group data (17, 18), and this approach has been used to estimate HIV incidence and prevalence in Canada as well. Data required for this approach include HIV diagnosis results and distinguishing between recent HIV infection and long-term HIV infection (previous negative test results, seroconversion, western blot indeterminate result within 1 year). This methodology does not require an experiment on biomarkers, such as CD4+ T cell count.

### Simultaneous HIV/AIDS diagnosis based on the CD4+ T cell count

Every year, some undiagnosed patients are diagnosed with HIV as they are diagnosed with AIDS (defined as concurrent HIV/AIDS diagnosis). This method estimates the size of undiagnosed population (number of undiagnosed HIV/AIDS patients) in a specific risk group based on the data on concurrent HIV/AIDS diagnoses in a particular period using AIDS incidence in CD4+ T cell count (19). However, estimation of the number of undiagnosed people with a high CD4+ T cell count can be more unstable than the number of concurrent HIV/AIDS diagnoses with a specific CD4+ T cell count. This methodology is conducted under the assumption that both the number of concurrent HIV/AIDS diagnoses over a year and the uncertainty of concurrent HIV/AIDS diagnosis can be factored into the estimate. The uncertainty of this estimate is evaluated under the assumption that AIDS incidence follows a normal distribution, and the number of concurrent HIV/AIDS diagnoses is assumed to follow a Poisson distribution. Both techniques can be improved if it is assumed that a continuous function describes the relationship between CD4+ T cell count and AIDS rate.

#### **(Use the CD4+ T cell count method, UK)**

With method 1, the CD4+ T cell count at time of HIV/AIDS diagnosis is required. The number of undiagnosed HIV infections can be estimated by dividing the number of concurrent HIV/AIDS diagnoses for a group by the CD4-specific AIDS rate. The results for all groups can be added to determine the total number of undiagnosed HIV infections. AIDS rate estimation can be inaccurate for high CD4 cell counts (4, 19).

With method 2, it is assumed that the CD4+ T cell count in the undiagnosed population can be approximated based on the CD4+ T cell count at time of diagnosis of asymptomatic HIV patients under treatment. This requires cohort data or data on whether cases were asymptomatic at time of diagnosis (4, 19).

#### **(Collection of major variables at the KCDC (Korea National Institute of Health))**

The KCDC collects major HIV parameters for the management of HIV patients. In this section, we review the major parameters collected by the KCDC (KNIH) and needs for improvement.

### **Prevalence calculation**

Currently, the KCDC (KNIH) calculates HIV prevalence, an estimated percentage of HIV-infected individuals, by HIV testing facility. In other words, the sample population for prevalence calculation is set to each testing facility. While this differs from the routine prevalence calculations based on high-risk populations such as the MSM population or PWID population, prevalence by facility would be a meaningful estimate of HIV infections in Korea, owing to the fact that HIV prevalence significantly varies across facilities in Korea (1.68 per 10,000 population in three facilities—public health centers, blood services, and military manpower administration in 2015; prevalence of 0.43 per 10,000 population in blood services, 8.46 per 10,000 population in public health centers, and 0.86 per 10,000 population in military manpower administration) (20). However, computing prevalence in more specific testing facilities or prevalence by high-risk groups would further reduce bias.

### **Collection of clinical information at diagnosis**

In Korea, information about HIV-positive individuals is collected via HIV/AIDS Supporting Network System (HASNet), and the major parameters, namely basic demographics, date of diagnosis, route of HIV infection (if available), antibody testing, CD4+ T cell count at diagnosis (if available), and symptoms at diagnosis (AIDS-related or unrelated symptoms), are collected through an epidemiologic study of infected individuals. However, route of infection differed from that found by the Korean HIV/AIDS cohort study. Further, immunoassays for infected individuals have several missing values and thus cannot be used as a major parameter.

Aside from this, HIV infections are distinguished between recent infections and long-term infections using a serologic kit. More accurate data could be collected if immunoassay results and results of counseling for HIV-infected individuals can be accessed.

### **AIDS patients**

Korean data does not classify patients with reference to the onset of AIDS. Epidemiologic data at diagnosis are collected, so it would be possible to classify patients according to the definition of AIDS. However, with the increased use of ART over the years, whether AIDS diagnosis can be utilized in the estimation of population size calls for thorough review.

### **Death of HIV patients**

When an HIV patient dies, the patient's data is uploaded to HASNet, and data on HIV deaths can be accessed through the Statistics Korea death data. However, more data are needed to determine the major cause of death of HIV patients.

Moreover, because HIV testing is not performed on expired patients, bias pertaining to undiagnosed HIV deaths must be taken into consideration.

## Cohort study

Through the Korean HIV/AIDS cohort study, treatment outcomes of approximately 1,000 patients yearly—a total of 1,500 patients thus far—are available for use in public health. To overcome the limitations with representativeness of the data due to inadequate sample sizes, efforts are being made to enhance the external validity of the data through measures such as analyzing the National Health Insurance Service (NHIS) data.

Status of KCDC (KNIH) estimates (possibility)

The KCDC computes epidemiologic data for HIV, such as prevalence by HIV testing facility (focused on public health centers), recent HIV infection rate, basic demographics of HIV patients, CD4+ T cell count for some HIV patients, route of infection of some patients, information about HIV deaths, antibody testing, and indeterminate testing results. Prevalence and recent infection rate are computed based on this information and are currently being prepared as a paper, and in the future, incidence estimates computed based on the prevalence and recent infection rates could be analyzed with an expanded back-calculation method (modified Atlanta method including HIV cumulative incidence method). Significant results could be obtained by further stratifying groups for parameters with inadequate data to reduce bias.

## CONCLUSION

We reviewed several methodologies for estimating the number of undiagnosed HIV infections. As each approach has both strengths and weaknesses, it is desirable to establish the most reliable estimation technique by utilizing as many approaches as possible to address the limitations of uncertainty with the number of HIV infections. The prevalence estimation method can potentially provide a comprehensive epidemiologic evaluation, and collection of clinical data would play a crucial role in estimating the number of undiagnosed HIV infections. In Korea, it is difficult to study estimation of undiagnosed HIV infections for reasons such as stigma. Thus, to estimate undiagnosed HIV infections, a broader scope of surveillance data must be collected, including data to be collected by the government. Collecting a wealth of data and utilizing multiple approaches as done so in European countries (21-23) would enable researchers to determine the presence of systematic differences in the results of various estimation approaches. Determining the exact number of HIV infections is practically impossible. However, with effort to gain an understanding of various problems with HIV in Korea, that is, the cause of mismatch, the methods of data collection could be continuously improved in the coming years, and it would gravely contribute to estimating the number of undiagnosed HIV infections.

## ACKNOWLEDGMENTS

This study was supported by a fund from the Chronic Infectious Disease Cohort Study (4800-4859-304) from the Korea Center for Disease Control and Prevention (KCDC).

## REFERENCES

- 1) WHO. Fact sheet; HIV/AIDS. 2019. <https://www.who.int/news-room/fact-sheets/detail/hiv-aids>
- 2) UNAIDS. Fact sheet; 2018 Global HIV statistics. 2019. <https://www.unaids.org/en/resources/fact-sheet>
- 3) UNAIDS. UNAIDS DATA 2019. 2019. [https://www.unaids.org/sites/default/files/media\\_asset/2019-UNAIDS-data\\_en.pdf](https://www.unaids.org/sites/default/files/media_asset/2019-UNAIDS-data_en.pdf)

- 4) European Centre for Disease Prevention and Control. HIV prevention in Europe: action, needs and challenges; 2-3 October 2006; Stockholm, Sweden. [https://www.ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/0610\\_MER\\_HIV\\_prevention\\_in\\_Europe.pdf](https://www.ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/0610_MER_HIV_prevention_in_Europe.pdf)
- 5) Hamers FF, Phillips AN. Diagnosed and undiagnosed HIV-infected populations in Europe. *HIV Med* 2008; 9:6-12. <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1468-1293.2008.00584.x>
- 6) Korea Center for Disease Control Prevention. 2018 Annual Report on the Notified HIV/AIDS in Korea. Jul. 2020. <http://www.cdc.go.kr/contents.es?mid=a20602070000>
- 7) UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance. Guidelines on Estimating the Size of Populations Most at Risk to HIV. Co-published by World Health Organization and UNAIDS; 2010. [https://www.unaids.org/en/resources/documents/2011/2011\\_Estimating\\_Populations](https://www.unaids.org/en/resources/documents/2011/2011_Estimating_Populations)
- 8) Kaldor JM, Delpech V, Guy RJ. AIDS case reporting: do we still need it?. *Lancet* 2009;373:181-3.
- 9) Giesecke J, Johnson A, Hawkins A, Noone A, Nicoll A, Wadsworth J, et al. An Estimate of the Prevalence of Human-Immunodeficiency-Virus Infection in England and Wales by Using A Direct Method. *J Roy Stat Soc Ser A Stat Soc* 1994;157:89-103.
- 10) Lyerla R, Gouws E, Garcia-Calleja JM, Zaniewski E. The 2005 Workbook: an improved tool for estimating HIV prevalence in countries with low level and concentrated epidemics. *Sex Transm Infect* 2006;82 Suppl 3:iii41-44.
- 11) Brown T, Bao L, Raftery AE, Salomon JA, Baggaley RF, Stover J, et al. Modelling HIV epidemics in the antiretroviral era: the UNAIDS estimation and projection package 2009. *Sex Transm Infect* 2010;86 Suppl 2:ii3-10.
- 12) Goubar A, Ades AE, De Angelis D, McGarrigle CA, Mercer CH, Tookey PA, et al. Estimates of human immunodeficiency virus prevalence and proportion diagnosed based on Bayesian multiparameter synthesis of surveillance data. *J Roy Stat Soc Ser A Stat Soc* 2008;171:541-80.
- 13) Presanis AM, De Angelis D, Spiegelhalter DJ, Seaman S, Goubar A, Ades AE. Conflicting evidence in a Bayesian synthesis of surveillance data to estimate human immunodeficiency virus prevalence. *J Roy Stat Soc Ser A Stat Soc* 2008;171:915-37.
- 14) Presanis AM, Gill ON, Chadborn TR, Hill C, Hope V, Logan L, et al. Insights into the rise in HIV infections, 2001 to 2008: a Bayesian synthesis of prevalence evidence. *AIDS* 2010;24:2849-58.
- 15) Sweeting MJ, De Angelis D, Aalen OO. Bayesian back-calculation using a multistate model with application to HIV. *Stat Med* 2005;24:3991-4007.
- 16) Hall HI, Song RG, Rhodes P, Prejean J, An Q, Lee LM, et al. Estimation of HIV incidence in the United States. *JAMA* 2008;300:520-9.
- 17) Wand H, Yan P, Wilson D, McDonald A, Middleton M, Kaldor J, et al. Increasing HIV transmission through male homosexual and heterosexual contact in Australia: results from an extended back-projection approach. *HIV Med* 2010;11:395-403.
- 18) Wilson D, Hoare A, Regan D, Wand H, Law M. Mathematical models to investigate recent trends in HIV notifications among men who have sex with men in Australia; 2008. <http://www.nchecr.unsw.edu.au>.
- 19) Lodwick RK, Sabin CA, Phillips AN. A method to estimate the number of people in a country or region with HIV who are undiagnosed and in need of ART. *Journal of the international AIDS society*. 2010, (suppl 4):P165.

- 20) Yoo M, Seong J, Yoon JG, Cha Jo, Chung YS, KIM K, et al. Characteristics of Adolescents and Young adults with HIV in the Republic of Korea from 2010 through 2015. *Sci Rep* 2020;10:9384.
- 21) Bezemer D, de Wolf F, Boerlijst MC, van Sighem A, Hollingsworth TD, Prins M, et al. A resurgent HIV-1 epidemic among men who have sex with men in the era of potent antiretroviral therapy. *AIDS* 2008;22:1071-7.
- 22) Van Veen MG, Presanis AM, Conti S, Xiridou M, Stengaard AR, Donoghoe MC, et al. National estimate of HIV prevalence in the Netherlands: comparison and applicability of different estimation tools. *AIDS* 2011;25:229-37.
- 23) Working Group on Estimation of HIV Prevalence in Europe. HIV in hiding: methods and data requirements for the estimation of the number of people living with undiagnosed HIV. *AIDS* 2011;25:1017-23.