**Principles and uses of HIV incidence estimation from recent infection testing - a review**

S Le Vu, J Pillonel, Caroline Semaille, P Bernillon, Y Le Strat, L Meyer, J C Desenclos

1. Department of Infectious Diseases, HIV/AIDS-STI-HCV Unit, Institut de veille sanitaire (French Institute for Public Health Surveillance, InVS), Saint-Maurice, France
2. Department of Epidemiology, Institut national de la santé et de la recherche médicale/Institut national d'études démographiques (National Institute of Health and Medical Research/National Institute for Demographic Studies, INSERM/INED/Paris XI U569), Le Kremlin-Bicêtre, France

Since the 1990s, the development of laboratory-based methods has allowed to estimate incidence of human immunodeficiency virus (HIV) infections on single samples. The tests aim to differentiate recent from established HIV infection. Incidence estimates are obtained by using the relationship between prevalence, incidence and duration of recent infection. We describe the principle of the methods and typical uses of these tests to characterise recent infection and derive incidence. We discuss the challenges in interpreting estimates and we consider the implications for surveillance systems.

Overall, these methods can add remarkable value to surveillance systems based on prevalence surveys as well as HIV case reporting. The assumptions that must be fulfilled to correctly interpret the estimates are mostly similar to those required in prevalence measurement. However, further research on the specific aspect of window period estimation is needed in order to generalise these methods in various population settings.

**Introduction**

Estimating HIV incidence, the number of new infections during a time period, is critically important for assessing the dynamics of human immunodeficiency virus (HIV) transmission and evaluating the impact of prevention policies. A conceptual improvement in surveillance methods has been made in the past ten years to make incidence estimation more feasible. By using a biomarker measurement to identify seropositive individuals who have recently been infected, incidence estimates can be obtained from a single specimen. This laboratory-based method can take advantage of the collection of specimen intended to assess prevalence (the proportion or number of persons cumulatively infected at a given time) and to obtain valid incidence data without the expensive and logistically complex requirement of following a cohort of uninfected individuals over time. However, as for other methods based on repeated prevalence data and mathematical modelling, the use of biomarkers to estimate incidence requires a substantial number of assumptions, some being difficult to assess, and an appropriate definition of the population the incidence is estimated for.

In this article based on the literature, we attempt to give an overview of the methods that allow estimating HIV incidence based on biomarker detection at the early stage of infection. After defining the principles, we review some typical uses of serological incidence assays and the challenges for each type of application.

**Principles**

**Incidence based on detection of virological markers before seroconversion**

In 1995, Brookmeyer and Quinn introduced a simple approach for estimating HIV incidence from a cross-sectional survey [1]. They used a two-step algorithm combining diagnostic tests for the p24 antigen and HIV-1 antibodies to determine the prevalence of p24 antigenemia among antibody-negative individuals (Figure 1). The HIV incidence rate was then calculated by using the classical epidemiologic relation between prevalence, incidence, and duration of the period between the onset of detectability of p24 and the first HIV antibodies.

The disadvantage of this approach was that the time during which p24 antigen is detectable prior to seroconversion is short (the mean duration of this period was 22.5 days in 1995 and has become shorter since then due to the development of new diagnostic assays that allow to detect antibodies earlier [2]). The first consequence of this is that the estimation of this period comes with a considerable uncertainty which can have a large impact on the incidence estimate. The second consequence is that large samples and/or high HIV incidence are required to identify a sufficient number of individuals with detectable p24 antigen who have not seroconverted. Nevertheless, Brookmeyer and Quinn provided the conceptual framework for subsequent laboratory-based methods to estimate incidence from single cross-sectional surveys.

Within the range of methods to identify early infection through virological markers before seroconversion, testing of pooled HIV RNA now seems to be the most appropriate approach because RNA can be detected earlier than p24 antigen, which allows characterisation of a longer time period (Figure 1). Moreover, pooling of specimens improves the predictive value of the amplification assays and substantially lowers the costs. However, in order to obtain accurate incidence estimates, this method requires the inclusion of very large sample populations, such as those provided by blood donations [2] or by the large testing programme in the United States (US) described by Pilcher et al. [3].

**Serologic incidence assays**

Janssen et al. were the first to describe in 1998 an approach based on a test specifically developed for the purpose of estimating incidence [4]. This approach named “Serologic testing algorithm
for recent HIV seroconversion (STARHS)” aimed at detecting a transient state reached after the antibody conversion. It thus offered the advantage of testing only positive individuals and defining a period sufficiently short to fulfill the requirements of stationarity of the incidence over the study period, while sufficiently long to minimise the inaccuracy in its estimation. The work of Janssen et al. can be considered as a milestone for the concept of serological methods for the estimation of HIV incidence.

Following the same principle, various applications of laboratory-based incidence estimation from cross-sectional population surveys have been described and a growing number of assays have been developed (see the article of Parry et al. in this issue). These assays measure the immunological response against the virus, based on specific HIV antibody concentration [4-6], proportion [7], isotype [8] or avidity [9]. This measure should define a transient state from the onset of detectability by a standard HIV screening test to the cut-off value defining the “established” infection status of the test for recent infection (Figure 1). This period is called the window period. Because of the individual variability in antibody response, window periods may differ widely from person to person. Their mean duration is measured in advance by testing serial specimens from infected individuals with known dates of seroconversion [10]. The STARHS methods have been compared to classical incidence measurements obtained in cohorts to assess their validity [4,11,12]. Provided that the compared estimates are not affected by population sampling bias, the estimates are reported to be similar [10,12].

**Incidence estimation**

The incidence estimation is calculated as the frequency of the transient state (i.e. the prevalence of recent infection) divided by its duration (the mean window period). As stated above, this calculation is based on the relation “prevalence = incidence * mean duration”. This relation assumes that the condition, in our context “recent HIV infection”, is a rare event so that the prevalence odds can be approximated by the prevalence [13]. And the relation is valid for a stationary population with a constant level of incidence during the study period [1]. In Figure 2, we present an example of an incidence calculation using the formula developed by Janssen et al. with a window period of 180 days [4].

Various adjustments have been made to Janssen’s formula in order to correctly express the number of people at risk and to account for misclassification of long-term infections. The first adjustment consisted in varying the assumed number of people at risk of having had a recent HIV infection during one year. As in the estimation of incidence in a cohort, HIV-negative individuals are considered at risk during the whole period, while infected individuals can be considered at risk during half a year on average [14].

In addition, concerns have been expressed that the mean window period for the BED capture enzyme immunoassay (BED-CEIA) does not properly take into account people who have a very long individual window period and can be falsely classified as

**Figure 1**

Kinetics of virological markers and host immune response used to define transient states in the early phase of HIV infection

A constant incidence rate of 0.4% persons/year is observed in a population of 1,000 individuals seronegative from the beginning of year 2005. Prevalence, incidence and rate of recent infection are estimated cross-sectionally at the end of 2005, 2006 and 2007. The number of HIV-positive individuals includes those with recent infection, tested within window period (NR), and those with established infection, tested after the window period (NE), represented respectively in light blue and dark blue in the figure. HIV-negative individuals (Nneg) are represented in grey. While incidence estimates are nearly constant over the years, the recent infection rate, being influenced by the prevalence of established infection, is decreasing.

Estimates are calculated as follows [4]:

\[
\text{Prevalence} = \frac{N_R + N_E}{N_{\text{neg}} + N_R + N_E} \\
\text{Incidence} = \frac{N_{\text{neg}} + N_R}{N_{\text{neg}} + N_R + N_E} \times \frac{365}{\text{mean WP}} \\
\text{Recent infection rate} = \frac{N_R}{N_{\text{neg}} + N_E}
\]

This illustration was inspired by the presentation of Ruigang Song “Modeling HIV Testing Behavior and Its Impact on Incidence Estimation” at the 15th International AIDS Conference, July 15, 2004, Bangkok, Thailand.

**Figure 2**

Relation between HIV prevalence, recent infection rate, and the incidence estimation in a cross-sectional survey

1: RNA-to-seroconversion transient state as defined by Busch et al., 2005 [2]
2: p24-to-seroconversion transient state as defined by Brookmeyer et al., 1995 [1]
3: Antibody-based mean window period as defined by Janssen et al., 1998 [4]

recent. This issue is probably a general one, affecting all the tests that have been calibrated using a disproportionate number of short term infections (for less than one year). It should have an impact on incidence estimation since the cross-sectional populations on which the method is to be applied are expected to contain a larger number of long-term infections. Two adjustments have been proposed to correct this issue about the specificity [15]. They share the principle of applying a corrective factor in the incidence formula to compensate for the false recent cases due to very long window period. Other algorithms have been proposed that, rather than correcting the formula, combine two incidence assays in order to avoid misclassification [12,16].

Applications

While a comprehensive review of applications for serological incidence assays is beyond the scope of this paper, the purpose of this chapter is to point out typical settings in which they may be used.

Typical applications

The most common context in which incidence assays are used are prevalence sero-surveys. Some were dedicated to incidence estimation, but the majority were set up to observe the recent infection status of stored HIV-positive serum specimens.

Numerous serial cross-sectional surveys have been applied in the setting of testing for HIV or other sexually transmitted diseases in countries such as the US [17-19], some European countries [20;21] or Brazil [22]. In these studies, temporal trends in incidence rate could be derived and helped to assess retrospectively epidemic phenomena among high-risk subgroups. But concerns about representativeness and selection bias can be raised about such voluntary testing sites (as reviewed below in the section “Issues”).

Similarly, already existing sentinel surveillance systems have provided insight into underlying trends in transmission in particular risk groups. Specimens gathered at enrolment in syringe exchange programmes or serial street surveys allowed the estimation of trends in HIV incidence among intravenous drug users in New York City, US [23] and San Francisco, US [24] over a long period.

For purposes of precision and as done for prevalence estimation, targeting a more general population than particular high-risk groups requires testing a very large number of people or setting the study in a country with a high incidence level. At least one of these conditions was met in studies that estimated the HIV incidence by means of recent infection testing in antenatal screening programmes in Cambodia [25], South Africa [26], the US [27] and Brazil [28], in screening programmes for blood donation in the US [2;4], France [29] and the Ivory Coast [30], and a national household survey in South Africa [31].

In all these settings, specimens are collected routinely and can be tested for recent infection retrospectively or prospectively. Some demographic and behavioural data on the targeted population are usually collected along with the specimens, both for positive and negative individuals. Taking advantage of specimens from prevalence serosurveys allows to derive incidence data for these populations with only minor expenses in terms of cost and logistics.

In certain contexts, the most obvious added value of the incidence assays approach is that the incidence could not have been estimated by any other means. This is what happens when no accurate data on prior testing or exposure period can be obtained such as for the population of blood donors screened during their first donation [29].

Identifying recent infection

A particular use of incidence assays is identifying recent infection status per se, for individual patient management such as contact tracing or assessment of primary resistance. It is helpful to bear in mind that characterisation of recent infection was initially a by-product in the method described by Janssen et al. which considered incidence derivation as the main outcome. In particular, the use of the mean value of an incidence assay window period assumes that individual window periods are variable and that a certain number of individuals in a given population will have a window period shorter or longer than the mean. Consequently, some misclassifications of established infection (false positives) and of recent infection (false negatives) are to be expected. For the purpose of incidence estimation, the respective misclassifications are supposed to cancel each other out, so that the number of recent infection at a population level is correctly estimated. At the level of individual patients, however, this could lead to serious misinterpretation.

On the other hand, some assays have been developed for the specific purpose of classifying infections in individual patients as recent or established with given predictive criteria. This is the case for the enzyme immunoassay for recent HIV-1 infections (EIA-RI) developed by Barin et al. [6]. This assay uses a logistic regression classification algorithm in which the cut-off was chosen to detect individuals infected for less than 180 days with a enhanced focus on the level of specificity of detection. It is to be noted that a lack of specificity, because it affects the population of established infections that is generally larger, should have a wider impact on misclassification than a lack of sensitivity, considering the low prevalence of recent infection status [30]. On-going development of the EIA-RI test aims to re-calibrate it for the purpose of incidence derivation.

Expressing the proportion of recent infection

Some applications define the proportion of recent infection in a population of positive individuals as an outcome. This is the way Puchhammer et al. analysed the results of the avidity assay among new diagnoses from case-reporting in Austria [32]. This is also the way that correlates of recent infection among new diagnoses are interpreted in France [33] (see also the article by Semaille et al. in this issue). However, this quantity that is somehow related to incidence depends also on the prevalence of non-recent infection and thus can not be considered as a good proxy for incidence. In fact, in the context of diagnostic testing, the proportion of recent infection has a lot to do with the testing framework capacity as well as the incidence rate in the population. Since the prevalence of undiagnosed infection affects the proportion of recent infection independently of any change in incidence (Figure 2), such results are difficult to interpret.

Incidence estimation from HIV case-reporting data

While it seems especially promising to take advantage of recent infection testing among reported HIV diagnoses at province or country level, there are several specific difficulties with regards to deriving a valid incidence measurement. Unlike cross-sectional surveys, a case-reporting system collects information only for individuals with positive test results and generally can not provide information on those who were negative. Therefore, the denominator of the formula, i.e. the number of people at risk, is not available. Another approach is needed to derive an incidence that can be generalised for the population targeted by the surveillance, and to take account of the fact that negative test results are not reported.
Such an approach has been described by Lee et al. for the estimation of the national HIV incidence in the US [34]. The statistical framework considers the reported cases identified as recently infected as a sample selected from all annual new cases, with a probability of inclusion related to their testing pattern. According to this probability, each case identified as recently infected is assigned a weight, and the sum of weights provides the incidence count. This approach represents a good opportunity to improve large scale surveillance of HIV dynamics, especially where a framework of HIV case reporting already exists and can provide data on testing patterns.

Finally, another approach has been described to bypass the issue that only positive individuals are reported to the surveillance system. In Ontario, Canada, an enhanced surveillance system has been established that requires diagnostic laboratories to collect information (number and risk factor) on a random subset of individuals with a negative test result in parallel to the information on those that were positive [35]. This system then allows the use of the Janssen’s formula to derive the incidence in different risk groups.

Issues
There are issues that pertain to the estimation HIV incidence by characterising recent infections. We can distinguish issues that are related to the determination of recent HIV infection from those that affect the validity of incidence estimation.

Limitations in determining recent infection
The first issues are due to the limitations of the assays in detecting recent HIV infection. As the majority of assays are based on quantitative measurement of the antibody response, factors that affect the patient’s immune response lead to some misclassification. Qualitative assays such as the avidity assay may be affected to a lesser extent [36].

Firstly, people with acquired immunodeficiency syndrome (AIDS) may falsely be identified as recently infected due to declining antibody levels. The same appears to be true in some individuals in the late stage of non-AIDS HIV infection. As for the AIDS stage, clinical data or CD4+ T-cell counts would need to be collected in order to exclude these patients from the calculation and avoid overestimation. A correction for misclassification due to late-stage non-responsive patients, has been proposed by Mc Dougal et al. and Hargrove et al. [15].

Secondly, antiretroviral drugs affect the antibody level by decreasing the viral load [37]. Again, to correctly assess recent infection, patients with ongoing treatment need to be identified and excluded by gathering declarative information (from clinician or patient) or alternatively by detecting drugs in serum specimens by, for example, mass-spectrometry.

Thirdly, test results are affected by the virus subtype and/or the patient’s genetic background. It has been shown that all tests that have been developed mainly on specimens from patients infected with subtype B viruses give inconsistent results when used for infections with non-B subtypes. Therefore, an assessment of the test properties (cut-off and window period) in different population settings is needed before applying any method [30].

We have seen how the correct interpretation of test results relies on the availability of clinical data that characterise the population [38]. In order to further interpret incidence estimates, data on sex, mode of contamination, testing patterns, and possibly virus subtypes must be gathered along with tests results.

Representativeness and selection bias
A general issue of incidence estimation arises from the fact that the populations tested are not randomly selected and may not be representative of the populations at risk of infection. This is particularly the case in the context of HIV testing or sexually transmitted diseases clinics. The bias may go in either direction. People at high risk may seek testing more frequently with the consequence of raising the incidence estimation. On the other hand, people attending HIV testing settings as part of a prevention strategy might be at lower risk than people who do not do a test because they do not recognise the risk or are afraid of a positive result.

Schoenbach et al. raised this issue in 2001 and questioned the rationale of inferring HIV incidence in testing settings and in particular, whether it is possible to extrapolate these incidence estimates to a larger population [39]. With regard to generalising incidence, it may be preferable to collect specimens from surveillance settings such as blood donation facilities or antenatal clinics where people are not self-selected but tested in a systematic manner, and where large sample size can be obtained.

Nevertheless, it can be argued that every design of an incidence study suffers from some kind of selection bias, even longitudinal studies [11]. Moreover, studying the level of the infection among the attendees of testing sites can still provide insights over time, especially in conjunction with behavioural data.

Even more problematic seems to be the issue of a selection bias occurring if recently infected people tended to seek testing sooner than expected because of seroconversion illness or identified recent exposure. This leads to an increase in the number of detected recent infections and an overestimation of the incidence. Remis et al. refer to this bias as the “seroconversion effect” and proposed a way to measuring it by making different incidence estimates based on varying window periods [40]. Song et al. formulated the hypothesis of independence between testing and the occurrence of infection and proposed a procedure to test this hypothesis [41]. All these biases can be found when inferring HIV incidence from case-reporting of new diagnoses which also include individuals seeking testing or health care.

Finally, as it is not always possible to test the whole positive study population for recent infection, the proportion of recent infection obtained among those tested is classically assigned to those for whom a test result is not available. This extrapolation assumes that the availability of specimens for recent infection testing is randomly determined in the population.

Conclusion
Overall, the use of laboratory-based methods to estimate HIV incidence can add remarkable value to surveillance systems based on prevalence surveys or on HIV case reporting. The estimation of HIV incidence provides a clear public health benefit in that it allows better monitoring of HIV transmission and targeting of preventive initiatives. We have seen that the application of those methods in cross-sectional settings have been well described in terms of incidence estimation and limitations, one of the most important limitations being the lack of representativeness. The assumptions that must be fulfilled to correctly interpret the estimates are to a
large extent similar to those required in prevalence measurement. However, further research on the more specific aspect of window period estimation may be needed in order to generalise these methods. In particular, efforts are needed to correctly define the mean window periods for different virus subtypes and stages of infection so that the essential relation between prevalence and incidence holds true in various population settings.

Acknowledgements
We thank Dr Michael P. Busch for his helpful suggestions on the literature review, Dr Robert S. Remis for his valuable comments on the draft version.

References


This article was published on 4 September 2008.