Published: April 2010

HIV incidence using cross-sectional surveys: recent theoretical developments

Thomas McWalter - Adjunct lecturer,
School of Computational and Applied Mathematics, University of the Witwatersrand

Prevalence and incidence are the two most important indicators of the state of an epidemic. Prevalence is the proportion of a population that has contracted an infection, while incidence is a measure of the risk of uninfected individuals contracting the disease and is usually expressed as a rate (i.e. the proportion of the at-risk (uninfected) population that become infected per unit time). Prevalence is the easier of the two indicators to measure, requiring only that the proportion of infected individuals be sampled directly in the population of interest. By contrast, considerable effort must be expended to estimate incidence.

In South Africa it is particularly important to have information on the state of the HIV epidemic since it has the largest population of HIV infected individuals of any country in the world (one in six South Africans aged 15-49 is infected). Large-scale intervention is required to reduce the rate of new infections. The South African National Strategic Plan for HIV AIDS (1) has stated that one of its primary aims is to "Reduce the rate of new HIV infections by 50% by 2011." It is therefore necessary to have good estimates of incidence to ensure effective targeting and evaluation of interventions.

Measuring incidence in a cohort study

The most common way in which incidence is measured is by follow-up of an initially uninfected cohort. Over the duration of surveillance, individuals in the cohort are regularly tested, and incidence is estimated as the number of new infection events observed divided by the number of person-years of observation. The incidence estimated in this way is effectively an average incidence over the duration of the survey. Unfortunately, such longitudinal surveillance is expensive, logistically complex and prone to biases. These biases include the fact that certain individuals may become unavailable for follow-up, which may be correlated with risky behaviour, and that risk-reduction counselling, which must be extended to participants on ethical grounds, may affect behaviour during participation.

Measuring incidence in a cross-sectional survey

For infections with a relatively short duration (e.g. less than a year), another method for estimating incidence is available. By performing a cross-sectional survey of the population of interest, one can identify the number of individuals infected. Incidence may then be calculated by inverting the well-known epidemiological relationship that prevalence is equal to incidence multiplied by the duration of infection. So, incidence can be calculated by dividing prevalence by the average duration of infection (given that an accurate estimate of the mean duration is available). Simplistically, this means that the difficult incidence measurement problem has been replaced by an easier problem of measuring infection prevalence. The incidence measured in this manner is effectively an average of the incidence over an historical period with length approximately equal to the duration of the infection.

Using biomarkers of recent HIV infection

Unfortunately, HIV has a long asymptomatic phase before the onset of immune failure and AIDS. This means that HIV infections last for many years and may not be diagnosed until long after the infection event. Furthermore, with the advent of antiretroviral therapy (ART), individuals who are enrolled on treatment programs may now survive for many decades. As a result, if one implemented the cross-sectional approach described above for HIV, the incidence estimate would be an average of incidence over decades of epidemic history. Such an incidence estimate would not be very useful. In the mid 1990s, however, a novel way of using cross-sectional surveys to estimate HIV incidence was proposed. The idea is to observe a biological marker (also known as a biomarker) indicating an immune system response to early infection and classify individuals as either recently infected or non-recently infected. Since individuals remain classified by the biomarker as recent infections for a much shorter period than they remain infected with HIV, the biomarker results can be used to estimate incidence in the same way that a short duration infection facilitates incidence estimation. An incidence estimate is computed in the same way as before, with one slight difference – the prevalence of recently infected individuals must be determined in the sub-population of the cross-section that excludes those that are non-recently infected.

Prior to conducting cross-sectional surveys to estimate incidence using a biomarker, it is necessary to estimate the mean duration that individuals spend in

the recently-infected state. Since being classified as recent by the biomarker is sometimes referred to as "being in the window period", this mean duration is usually called the mean window period. Estimating (or calibrating) the mean window period requires longitudinal follow-up of individuals with approximately known infection dates. This requires considerable effort and cost, but need only be conducted once. Unless there is good reason to suspect that the mean window period is different in different contexts (e.g. as a result of sub-type diversity), all subsequent incidence surveys use the same value of the mean window period.

Brookmeyer and Quinn were the first to propose the biomarker-based approach in HIV monitoring (2). They used the presence of p24 antigen (present in the HIV protein shell) in a blood sample prior to HIV antibody production by the immune system (seroconversion) as an indication of recent infection. Unfortunately, the mean window period for this biomarker is very short (about three weeks), which means that to get good statistics, unreasonably large sample sizes for the incidence cross-section surveys are needed. Later, Janssen et al. (3) proposed a method based on the increase of a serological response (in particular they used 'detuned' assays to detect recently infected individuals). Depending on how the assays are applied, the mean window period for this approach is longer (between 100 and 200 days), facilitating better precision in the incidence estimates. This approach later became known as the Serological Testing Algorithm for Recent HIV Seroconversion (STARHS). Unfortunately, the use of detuned assays did not prove reliable due to the variability in immune response due to subtype diversity. In order to improve biomarker characteristics, a number of other assays that test for recent HIV infection have been developed, including the much used BED assay, which is a capture enzyme immunoassay (CEIA) based on protein sequences from the B, E and D HIV subtypes (4).

Handling anomalous results

As the method of using biomarkers of recent infection in cross-sectional surveys was more widely applied, it became apparent that the results obtained from surveys invariably overestimated incidence. It was then realized that for the most useful assays (i.e. those with the mean window periods long enough for good statistics) a proportion of individuals remained perpetually classified as "recent" as a result of insufficient immunological response to HIV infection. The reasons for these anomalous responses are not completely understood, but it is well known that elite controllers (those individuals that have an innate ability to control the progress of HIV thus avoiding AIDS) fall into this category. It was also found that certain individuals on ART or with end stage disease experience immune system changes that may result in incorrect classification. These complications mean that the simple relationship between incidence and the prevalence of recently infected individuals is no longer obeyed. As a result alternative techniques for analysing data from such imperfect biomarker based surveys were needed.

To account for these anomalous results, McDougal and colleagues (5) at the Centers for Disease Control and Prevention (CDC), using statistical concepts from diagnostic testing, introduced an 'adjusted' incidence estimator which required not only the mean window period to be calibrated, but also a sensitivity and two specificity parameters (short- and long-term). This meant that, although imperfect tests like the BED assay could not be used for individual diagnosis, incidence could be estimated in surveys, provided that the calibration parameters were accurate for the population being surveyed. While this approach provided a way forward in analysing data, it did so at the cost of requiring a complicated calibration exercise. To date, this calibration has only been carried out once in the context of North America.

A new estimator

Since this innovation, it was realized that the most important parameter for characterizing anomalous results is the long-term specificity parameter of the McDougal estimator. John Hargrove and colleagues (6) from SACEMA were the first to realize that the McDougal estimator is over-parameterised and provided a new estimator that depends only on the mean window period and a false-recent rate. The false-recent rate can be expressed as one minus the long-term specificity, and is essentially the proportion of non-recently infected individuals, infected for more than a certain time, that are incorrectly classified as recently infected. Later, Alex Welte and I (7) showed that an incidence estimator (different from that of Hargrove et al.), which also depends only on the mean window period and the false-recent rate, could be directly derived using a survival analysis of the problem. This estimator has been shown to be the least biased of the estimators available (8). We were also able to derive a theoretically consistent way of reducing the number of calibration parameters in the McDougal estimator, in effect conclusively showing why their approach is over-parameterised (8).

An important consequence of the reduction of the number of calibration parameters in the McDougal estimator is that the complexity of the techniques required to calibrate the parameters that remain is less than the complexity required for the parameters that are eliminated. Another important fact that emerges from our analysis is that, under realistic conditions, the false-recent rate will vary with location and time. Our new estimator has been validated in an incidence study conducted in rural KwaZulu-Natal by the Africa Centre for Health & Population Studies (9). This study

showed that the local estimate for the false-recent rate was significantly lower than the equivalent false-recent recent rate found by McDougal et al. in the North American calibration exercise.

More precise definition of the calibration experiments needed

Recently, there has been some debate as to whether incidence estimates should be adjusted for false-recent results or not. Brookmeyer (10-13) has suggested that rather than using an estimator that accounts for false-recent results, one should consider all recent classifications as valid (i.e. within the window period) and use the simple estimator with a "better" estimate of the mean window period. While this approach is theoretically correct, it has some serious practical problems, including the fact that a calibration of such a mean window period must naturally happen over a very long period, since it must include a small proportion of individuals who may remain classified as recently infected for decades. Since such individuals are only removed from the population as a result of death, the calibration must be locally relevant. For example, it would be inappropriate to use a mean window period of this type calibrated in North America for a study conducted in Africa, since access to basic healthcare, ART, nutrition and the mix of opportunistic infections in these two locations will lead to very different survival profiles for HIV infected individuals. Clearly, as these factors may change with time, the mean window period thus defined would also vary as a function of time.

While we disagree with the approach of Brookmeyer, the debate has, however, highlighted that there is a need for a more precise definition of the calibration experiments that are needed to characterise the performance of biomarkers. Further work in this direction is currently being undertaken at SACEMA (14).

While there has been some controversy on how best to interpret data derived from cross-sectional surveys using biomarkers that test for recent HIV infection, the work undertaken at SACEMA has shown that there is a consistent, tractable and reliable method for producing incidence estimates using biomarker-based cross-sectional surveys. SACEMA continues to contribute to ongoing research in this area, and has also been active in ensuring that these results are publicised to a wider audience of researchers and public health officials through various journal publications and contributions to the World Health Organization Working Group on HIV Incidence Assays.

Thomas McWalter, Adjunct lecturer, School of Computational and Applied Mathematics, University of the Witwatersrand and research associate, SACEMA. Areas of interest: Application of stochastic processes and optimal control in epidemiology, finance and physics. *mcwalter@cam.wits.ac.za*

Reference list

1. Government of South Africa. HIV and AIDS and STI strategic plan for South Africa, 2007-2011. Pretoria: Government of South Africa; 2007.
2. Brookmeyer R, Quinn TC. Estimation of current Human Immunodeficiency Virus incidence rates from a cross-sectional survey using early diagnostic tests. Am J Epidemiol. 1995;141(2):166-172 .
3. Janssen JS, Satten GA, Stramer SL, et al. New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes. JAMA. 1998;280(1):42-48.
4. Parekh BS, Kennedy MS, Dobbs T, et al. Quantitative detection of increasing HIV type 1 antibodies after seroconversion: a simple assay for detecting recent HIV infection and estimating incidence. AIDS Res Hum Retroviruses. 2002;18(4):295-307.
5. McDougal JS, Parekh BS, Peterson ML, et al. Comparison of HIV type 1 incidence observed during longitudinal follow-up with incidence estimated by cross-sectional analysis using the BED capture enzyme immunoassay. AIDS Res Hum Retroviruses. 2006;22(10):945-952.
6. Hargrove JW, Humphrey JH, Mutasa K, et al. Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay. AIDS. 2008;22(4):511-518.
7. McWalter TA, Welte A. Relating recent infection prevalence to incidence with a sub-population of assay non-progressors. J Math Biol. 2009 [epub ahead of print].
8. McWalter TA, Welte A. Comparison of biomarker based incidence estimators. PLoS ONE. 2009;4(10):e7368.
9. Bärnighausen T, Wallrauch C, Welte A, et al. HIV incidence in rural South Africa: comparison of estimates from longitudinal surveillance and cross-sectional cBED assay testing. PLoS ONE. 2008;3(11):e3640.
10. Brookmeyer R. Should biomarker estimates of HIV incidence be adjusted? AIDS. 2009;23(4):485-491.
11. Hargrove JW. BED estimates of HIV incidence must be adjusted. AIDS. 2009;23(15):2061-2062.
12. Welte A, McWalter TA, Bärnighausen T. Reply to 'Should biomarker estimates of HIV incidence be adjusted?' AIDS. 2009;23(15):2062-2063.
13. McDougal JS. BED estimates of HIV incidence must be adjusted. AIDS. 2009;23(15):2064-2065.
14. McWalter TA, Kassanjee R, Welte A. Biomarker Based Incidence Estimation: Accounting for False Recent Results (working paper). Stellenbosch: SACEMA; 2009.