Update on HIV incidence estimation from prevalence data

Guy Mahiane - Biostatistician, post-doc at SACEMA
Hilmarie Brand - Student at Stellenbosch University/SACEMA

Accurate HIV incidence estimates are essential for monitoring the HIV epidemic, determining public health priorities and assessing the impact of interventions. There are many approaches to estimating incidence. The “gold standard” approach is through observational studies, in which subjects are periodically monitored for HIV infection. However, such studies are time consuming, expensive and may provide biased estimates. An alternative, which has attracted considerable attention, is the use of HIV tests that can discriminate recent infections, based on a single specimen. The theoretical underpinnings of the latter approach have recently been put on a sound footing under general assumptions (1). Nevertheless, the tests themselves have not yet evolved to the point where they yield consistently informative estimates (2-4). Another alternative is estimating incidence from (possibly multiple) cross-sectional measures of prevalence. Previous efforts to relate incidence to (possibly multiple) cross sectional measures of prevalence (5-7) have been based on the idea of a formal solution to a complex dynamical problem, in which a later population state is expressed explicitly in terms of an earlier population state and dynamical rules presumed to have been in effect over the intervening period. This can however not be done in closed form without the use of simplifying assumptions. A recent work which presents a new method, avoids these difficulties by extracting the estimator directly from the dynamical rules/equations, rather than from a ‘solution’ (8).

Estimating incidence from individual level HIV serostatus data

Brunet and Struchiner (6) derived the fundamental formula expressing the age and time specific incidence rate as a function of the excess mortality rate, the prevalence and the rates of change in prevalence, in an age and time structured susceptible-infected population. The new proposed approach uses the Maximum Likelihood Estimation method to estimate each of the elements in this formula, except for the differential mortality which was assumed to be known, using individual level HIV serostatus data (8). This is in contrast with earlier methods for estimating incidence from prevalence data, which work with aggregated data and the aggregated effect of demographic and epidemiological rates over the time interval between prevalence surveys. Estimates of incidence can therefore in principle be obtained for any chosen age and time, and no particular assumptions are made about the epidemiological or demographic context.

The performance of the new method was assessed using a simulated epidemic with four phases: the age specific incidence was constant in time in the first phase, increasing in the second, constant in the third and decreasing in the last phase. Two cross-sectional surveys were simulated in each phase and the generated age and time specific prevalence was used to simulate the infection statuses of individuals in these surveys. The new method appeared to be the most successful at reproducing the input incidence rates, compare to the methods proposed by Brunet and Struchiner (6) and Hallett et al (7).

The extent to which the approach can be used to estimate incidence in birth cohorts was also investigated. The performance of the method was also assessed by applying it to simulated data. The closed form of the prevalence at the time of the second survey was used to simulate the prevalence of an infection with differential mortality in a population given: 1) the prevalence at the beginning of the observation period; 2) the incidence rate between the two surveys; and 3) a constant excess mortality rate.

The performance of the new method, for the case of the birth cohort, was compared the performance of the methods proposed by Hallett et al. (7) and Brookmeyer and Konikoff (5). A wide range of scenarios were investigated and results indicated that: 1) the estimators proposed by Hallett et al. and Brookmeyer and Konikoff tend to underestimate the incidence rate; 2) the bias of the estimator of Brookmeyer and Konikoff is comparable to the bias of the estimator of Hallett et al.; 3) the bias of the new estimator was much smaller than that of the other two estimators across all of the investigated scenarios; 4) the asymptotic standard error of the estimator proposed by Brookmeyer and Konikoff was lower than the asymptotic standard error of the new estimator, which in turn was lower than the asymptotic standard error of the estimator proposed by Hallett et al., except when the excess mortality rate was very small. The new estimation method appeared to not be very sensitive to non-constant incidence. It also appeared to be the most robust of the three methods with respect to uncertainty regarding differential mortality.
In conclusion, these results motivate for: 1) efforts to be done to obtain accurate estimates of age and time specific excess mortality rates among HIV infected individuals, and 2) the use of individual level rather than aggregated data in order to estimate HIV incidence rates at times between (possibly multiple) prevalence surveys.

**Guy Mahiane**, Biostatistician, post-doc at SACEMA. Areas of interest: mathematical and statistical modelling the effects of interventions on HIV incidence, HIV incidence estimation methods. mahanies@yahoo.fr

**Hilmarie Brand**, Student at Stellenbosch University/SACEMA. hbrand@sun.ac.za

**References:**


