Combining demographic and biomarker-based approaches to estimate age-specific incidence

Eduard Grebe - staff researcher at SACEMA.

Incidence is the most sensitive indicator of epidemiological trends, but it is hard to estimate for non-remissible and long-lasting conditions like HIV. Epidemic surveillance therefore largely relies on population-level prevalence surveys, antenatal clinic surveys and various forms of routine data.

A range of methods are used to estimate HIV incidence. Widely-used methods include the 'gold standard' of observing infection events in an HIV-negative cohort, which is logistically challenging and hard to generalise to the population as a whole, and using biomarkers of 'recent infection' in a cross-sectional survey. This method is very expensive, given the large sample sizes required to attain sufficient precision. 'Back calculation' from new diagnoses with disease staging information like CD4 count is also used, but is also problematic unless a very robust disease progression model is available. Dynamic mathematical models fit primarily to prevalence data rely heavily on mechanistic assumptions and can be hard to calibrate when complex.

A purely data-driven and general approach is a demographic or 'synthetic cohort' approach, in which the age and time structure of HIV prevalence, together with estimates of mortality, can be used to infer incidence. A number of specific methods for doing so have been proposed, notable examples including Brunet and Struchiner's method (1), and one proposed by Brian Williams and colleagues (2). Most of these methods exhibited significant bias or relied on strong assumptions unlikely to be obeyed in real surveillance contexts. In 2012, Guy Mahiane and others at SACEMA reviewed existing methods and derived a simple incidence estimator that relies purely on estimates of (age- and time-specific) HIV prevalence, the derivatives of (i.e., rates of change in) prevalence in the age and time directions and estimates of (age- and time-specific) HIV-related excess mortality (3). It was shown to exhibit less bias than previously-proposed estimators.

Unfortunately, the Mahiane estimator has found limited uptake, representing a missed opportunity to optimally analyse rich data sources like the HSRC's survey in South Africa. Further, no single approach attains the levels of precision one might hope for with the data available, even in a well-studied epidemic like the one in South Africa.

Combining estimates of HIV prevalence and biomarkers of recent infection

We, therefore, developed an approach that implements an optimal weighting of the Mahiane and Kassanjee (4) incidence estimators (the latter being the standard method for cross-sectional incidence estimation using biomarkers of recent infection) and applied it to data from a survey conducted by Médecins Sans Frontières (MSF) in KwaZulu-Natal (5). The survey ascertained HIV status as well as recency using two serological markers and viral load.

In this study we adapted the Mahiane incidence estimator for use with cross-sectional data, which is valid in the case of a stable epidemic – i.e. the age structure of prevalence is not changing rapidly in the time direction. Estimates of age-specific prevalence in KwaZulu-Natal obtained from the THEMBISA model showed very small prevalence gradients in time at the time of the survey, indicating that this assumption is likely to be innocuous. A sensitivity analysis reported in an appendix to the paper (5) showed that our estimates were not very sensitive to plausible rates of change. We further adapted the Kassanjee estimator to produce age-specific incidence estimates using the recency biomarker data.

We obtained age-specific incidence estimates by fitting regression models for HIV-positivity and for recent infection (amongst HIV-positives), both as functions of age. We were then able to obtain 'predicted' HIV prevalence and prevalence of recency for any age and utilise the former in both estimators and the latter in the Kassanjee estimator. We evaluated uncertainty, as well as the covariance between the two estimators, through a bootstrapping approach that replicated the complex sampling frame used in the survey. Excess mortality estimates came from the THEMBISA model for KwaZulu-Natal since direct estimates were not available. Perhaps the greatest challenge in utilising the Mahiane estimator is the need for excess mortality estimates, which are generally very hard to obtain. This represents much less of a problem at younger ages, where excess mortality is low, and prevalence is also usually low. This is important because excess mortality enters the estimator as a factor of the (age and time-specific) prevalence.
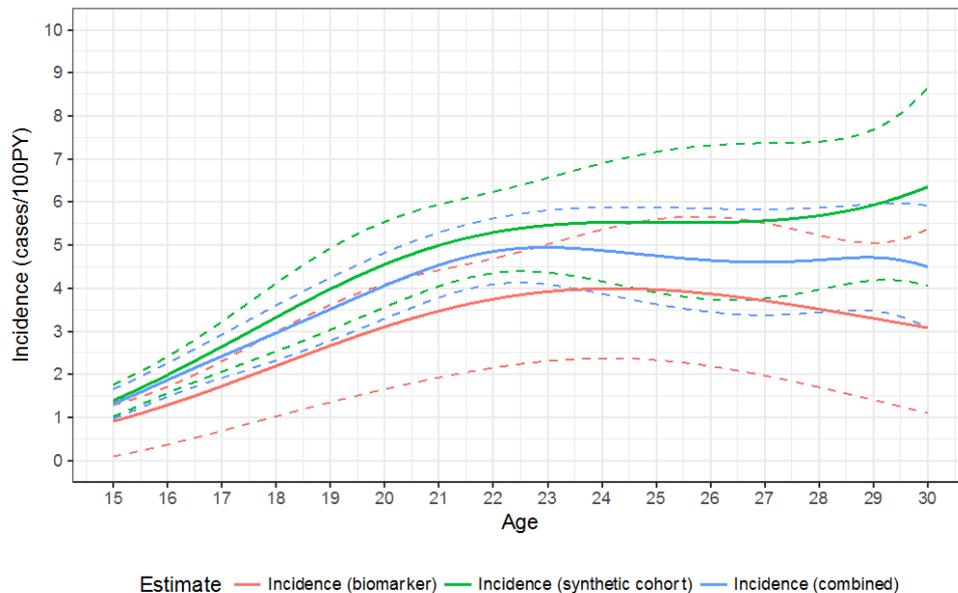
Figure 1. HIV incidence by age in females aged 15-30, using the synthetic cohort, recency biomarker and combined methods

We restricted our analysis in this study to under-35s since even the combined method became very uncertain at ages above 35. This limitation may to some extent reflect our parametric choices, and further work is needed to optimise the method for use in older population groups. It is important to note, however, that in many settings – including this one, young girls experience much higher incidence than any other group and therefore constitute a natural sentinel population. Even restricted to this group the method can be a very valuable tool.

The most significant limitation of the study is the absence of explicit information on the change in the age structure of prevalence over time. MSF is currently conducting a second survey of the same population, and we expect to extend the analysis by incorporating the data from the new survey. With data from two surveys, we expect to estimate age-specific incidence at the time of each survey, but the most informative estimates will be for the midpoint between the two surveys, based on an implementation of the full Mahiane estimator that relies on prevalence gradients in the age and time directions.

This method has great potential for obtaining improved incidence estimates – and perhaps most importantly, incidence trends – from large population-level prevalence surveys. These include the HSRC survey in South Africa and the Population HIV Impact Assessments (PHIAs) conducted in many African countries in recent years – most of which also included recency biomarker measurements. We are actively pursuing collaborations that would allow us to utilise these rich data sources for HIV incidence surveillance optimally.

Eduard Grebe - *staff researcher at SACEMA. Areas of interest: disease surveillance – focusing on incidence estimation, diagnostic and staging biomarkers (particularly of HIV and Tuberculosis) and blood product safety.* eduardgrebe@sun.ac.za

References:
1. Brunet RC, Struchiner CJ. Rate Estimation from Prevalence Information on a Simple Epidemiologic Model for Health Interventions. Theoretical Population Biology. 1996; 50(3):209–226. https://doi.org/10. 1006/tpbi.1996.0029 PMID: 9000488.
2. Williams B, Gouws E, Wilkinson D, Karim SA. Estimating HIV incidence rates from age prevalence data in epidemic situations. Statistics in Medicine. 2001; 20(13):2003–2016. https://doi.org/10.1002/sim.840 PMID: 11427956.
3. Mahiane GS, Ouifki R, Brand H, Delva W, Welte A. A General HIV Incidence Inference Scheme Based on Likelihood of Individual Level Data and a Population Renewal Equation. PLoS ONE. 2012; 7(9): e44377. https://doi.org/10.1371/journal.pone.0044377 PMID: 22984497.
4. Kassanjee R, McWalter TA, Bärnighausen T, Welte A. A New General Biomarker-based Incidence Estimator. Epidemiology. 2012; 23(5):721–728. https://doi.org/10.1097/EDE.0b013e3182576c07 PMID: 22627902.
5. Grebe E, Welte A, Johnson LF, van Cutsem G, Puren A, Ellman T, et al. (2018) Population-level HIV incidence estimates using a combination of synthetic cohort and recency biomarker approaches in KwaZulu-Natal, South Africa. PLoS ONE 13(9): e0203638. https://doi.org/ 10.1371/journal.pone.0203638.