Letter to the Editor

A Simplified Formula for Inferring HIV Incidence from Cross-Sectional Surveys Using a Test for Recent Infection

EDITOR: The paper of McDougal *et al.*, previously published in this journal,[1] is becoming a standard reference used for the estimation of HIV incidence from applications of the BED IgG-capture enzyme immunoassay (BED assay) to cross-sectional blood samples.[2,3] Their approach provides an estimate for an annual risk of infection in a hypothetical cohort, using an estimate for the true proportion, P_t, of "recent infections" among HIV-seropositive individuals. The estimate P_t is in turn derived from the proportion, P_o, of seropositive individuals in a survey who test below a threshold value for normalized BED optical density (OD-n).[4] The condition of being below the OD-n threshold is declared to be an imperfect test for recent infection.

True "recent infection" is defined as having been infected for less than a period ω, where ω is the mean time individuals spend below the OD-n threshold. Since it is well known that not all individuals progress to a given threshold, even after arbitrarily long times, ω needs to be carefully defined as the mean threshold crossing time among those who do progress. It is also known that during late stage illness, or under the influence of antiretroviral therapy, individuals may regress to OD-n values below the recency threshold. It is further plausible, and indeed appears to be the case,[5,6] that the parameters characterizing progression through the BED-defined states of infection vary regionally. These complications have caused doubt about the prospects for using the BED assay as a robustly characterizable test for recent infection for the purposes of estimating HIV incidence, as reflected in a UNAIDS statement in 2006[7] recommending it not be used for this purpose.

Hence, new assays, or combinations of assays (such as a BED and an antibody avidity test), are being developed to provide more robust tests for recent infection. The fraction of individuals that progresses atypically through an assay-defined class of "recently infected" may thus be reduced, but is unlikely to be zero. Therefore, the methodology developed to deal with this problem for the BED assay appears, at face value, to be immediately transferable, requiring only minor modification (namely in the values of its parameters) to be applicable to other imperfect tests for recent infection. We argue that several subtle points need to be addressed to ensure that incidence inferences based on imperfect tests for recent infection are not unnecessarily limited, or even in error, and we do this by a critique of the original application.

The interindividual variability of BED OD-n progression is captured in the McDougal model by three parameters:

- The sensitivity (σ) of the BED assay as a test for the condition of being "recently infected," as defined above.
- The short-term specificity (ρ_1) of the BED assay as a test for the condition of being "recently infected," when restricted to persons who have been infected for a time between ω and 2ω.
- The long-term specificity (ρ_2) of the BED assay as a test for the condition of being "recently infected," when restricted to persons who have been infected for a time longer than 2ω.

Using data from a major epidemiological and demographic surveillance study in South Africa,[8,9] we and our collaborators are currently comparing various approaches to HIV incidence estimation using the BED assay.[5,10] Given the long intervals between follow-up visits in this study (about a year), it was not possible to calibrate the McDougal formula in its published form. Calibration of σ and ρ_1 requires a follow-up interval of at most ω (which is of the order of half a year[1]).

While trying to address this issue, we discovered that a simplification of the McDougal formula is possible. In their paper, the key result relating P_t to the calibration parameters is given by

$$P_t = \frac{P_o + \rho_2 - 1}{\sigma - \rho_1 + 2\rho_2 - 1} \quad (1)$$

As is shown by McWalter and Welte in a separate short note,[11] the above equation can be simplified using the following identity:

$$\sigma - \rho_1 + \rho_2 = 1 \quad (2)$$

This identity is derived using no more assumptions than are used by McDougal *et al.* to derive their formula; these assumptions are, however, stated with greater precision.[11] The idea that these parameters might be related was inspired by the analysis of the incidence estimation problem previously undertaken.[12] Inserting the identity into (1) gives

$$P_t = \frac{P_o + \rho_2 - 1}{\rho_2} \quad (3)$$

This means that in order to estimate incidence, it is only necessary to calibrate the long-term specificity ρ_2 (to estimate P_t) and the window period ω (to convert P_t to an annual risk of infection). Unlike σ and ρ_1, these can both be inferred from

infrequent follow-up. Incidentally, using the values of σ, ρ_1, and ρ_2 previously reported,[1] we find that

$$\sigma - \rho_1 + \rho_2 = 0.989 \approx 1 \qquad (4)$$

which manifests the combined fluctuations in the estimates of σ, ρ_1, ρ_2, and ω. Although ω is superficially absent in the identity, it enters as the period over which the other parameters are defined.

The appropriately simplified form[10] is amenable to calibration using data obtained with long intervals of follow-up.[10] This seems to us to be an important point, as many demographic and epidemiological surveillance studies we are aware of, or expect to see implemented, are characterized by follow-up intervals of the order of a year, almost ideal for calibrating the reduced formula and clearly inadequate for calibrating the previously published form. There is likely to be substantial data of this sort available. On the other hand, the cost of obtaining short interval follow-up data is high and the opportunities for doing so are rare.

Note that even given an appropriate data set for estimating σ, ρ_1, and ρ_2, the use of the naive formula, for the purpose of systematically quantifying uncertainty due to imperfect calibration, would require additional specification of nontrivial covariances implied by the identity.[5]

The attraction of using a test for recent infection for HIV surveillance, program evaluation, and policy making lies in the fact that it allows HIV incidence estimation from cross-sectional blood samples. Cross-sectional HIV status information alone, however, does not allow estimation of the calibration parameters. These must be estimated in separate studies, involving follow-up of an intensity comparable to a prospective observation of incidence. Only after this has been done can the more efficient cross-sectional survey be employed on a suitably similar population. The more robust and locally validated the calibration parameters are, the more informative cross-sectional surveys can be. Therefore it is important that the necessary parameters be calibrated as widely and thoroughly as possible, using such data as is available. The parameters of the simplified formula are independent and can be estimated from comparatively long interval follow-up data, while the parameters used by McDougal *et al.*[1] have nontrivial correlation and require short intervals of follow-up.

Disclosure Statement

No competing financial interests exist.

References

1. McDougal JS, Parekh BS, Peterson ML, Branson BM, Dobbs T, Ackers M, and Gurwith M: 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.
2. Jiang Y, Wang M, Ni M, Duan S, Wang Y, Feng J, Xiao Y, Dong Y, Wang D, Han M, Xiang L, Ma L, and Zhou Q: HIV-1 incidence estimates using IgG-capture BED-enzyme immunoassay from surveillance sites of injection drug users in three cities of China. AIDS 2007;21(Suppl. 8):S47–S51.
3. Rehle T, Shisana O, Pillay V, Zuma K, Puren A, and Parker W: National HIV incidence measures—new insights into the South African epidemic. S Afr Med J, 2007;97(3):194–199.
4. Parekh BS, Kennedy MS, Dobbs T, Pau C-P, Byers R, Green T, Hu DJ, Vanichseni S, Young NL, Choopanya K, Mastro TD, and McDougal JS: 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. Bärnighausen T, Wallrauch C, Welte A, McWalter TA, Mbizana N, Viljoen J, Graham N, Tanser T, Puren A, and Newell M-L: HIV incidence in rural South Africa: comparison of estimates from longitudinal surveillance and cross-sectional cBED assay testing. PLoS ONE 2008;3:e3640.
6. Karita E, Price M, Hunter E, Chomba E, Allen S, Fei L, Kamali A, Sanders EJ, Anzala O, Katende M, and Ketter N: Investigating the utility of the HIV-1 BED capture enzyme immunoassay using cross-sectional and longitudinal seroconverter specimens from africa. AIDS 2007;21:403–408.
7. Modelling UNAIDS Reference Group on Estimates and Projections: Statement on the use of the BED assay for the estimation of HIV-1 incidence for surveillance or epidemic monitoring. Wkly Epidemiol Rec 2006;81(4):403–408.
8. Tanser F, Hosegood V, Bärnighausen T, Herbst K, Nyirenda M, Muhwava W, Newell C, Viljoen J, Mutevedzi T, and Newell M-L: Cohort profile: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey. Int J Epidemiol 2008;37(5):956–962.
9. Bärnighausen T, Tanser F, Gqwede Z, Mbizana C, Herbst K, and Newell M-L: High HIV incidence in a community with high HIV prevalence in rural South Africa: Findings from a prospective population-based study. AIDS 2008;22(1):139–144.
10. Bärnighausen T, Welte A, McWalter TA, and Newell M-L: Calibrating the BED capture immunoassay using data from a large population-based HIV surveillance in rural South Africa. In preparation.
11. McWalter TA and Welte A: On the estimation of the proportion of true recent infections using the BED capture enzyme immunoassay. Available at http://arxiv.org/abs/0806.1297, June 2008.
12. McWalter TA and Welte A: Relating recent infection prevalence to incidence with a sub-population of nonprogressors. Available at http://arxiv.org/abs/0801.3380, June 2008.

Alex Welte[1]
Thomas A. McWalter[2]
Till Bärnighausen[3]

Address reprint requests to:
Alex Welte
Department of Computational and Applied Mathematics
University of the Witwatersrand
Private Bag 3
Johannesburg, South Africa 2050

E-mail: alex.welte@wits.ac.za

[1]DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, and School of Computational and Applied Mathematics, University of the Witwatersrand, Johannesberg, South Africa.
[2]School of Computational and Applied Mathematics, University of the Witwatersrand, Johannesberg, South Africa.
[3]Africa Centre for Health and Population Studies, University of KwaZulu-Natal, Mtubatuba, Kwazulu-Natal, South Africa.