Estimating HIV Incidence: A Mathematical Modelling Approach

ALI NADAF[1,4], ALEXANDER R RUTHERFORD[1,4], BOJAN RAMADANOVIC[1], KRISZTINA VASARHELYI[2,3], BENITA YIP [2], ART POON[2], RICHARD LIANG[2], RICHARD HARRIGAN[2,5], RALF W WITTENBERG[4], JULIO SG MONTANER[2,5]

1 The IRMACS Centre, Simon Fraser University, Burnaby, BC, Canada; 2 British Columbia Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada; 3 Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada; 4 Department of Mathematics, Simon Fraser University, Burnaby, BC, Canada; 5 Division of AIDS, Department of Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Introduction

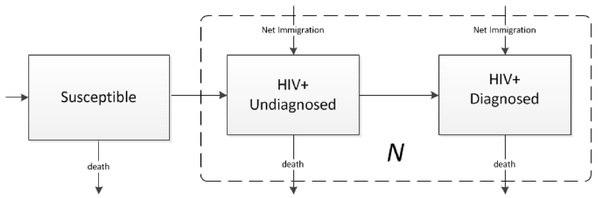
HIV incidence, the rate of new infections in the population, is an important measure of the success of public health strategies for the HIV/AIDS epidemic. HIV incidence is difficult to estimate because people infected with HIV may remain asymptomatic and undetected for as long as eight years. The rate of new HIV positive tests is not necessarily a good measure of incidence.

We present a mathematical modelling approach to estimate incidence using existing public health data and data from genotypic drug resistance tests, that are a recommended component of the patient treatment protocol [2]. We develop two separate and independent models for the diagnosed fraction of the HIV positive subpopulation and use numerical optimisation methods to find the values of the parameters for which the two models most closely agree. Subsequently, we calculate HIV incidence from our estimate of the time series for the fraction of the HIV positive diagnosed population.

Models

a) Transmission Model

The transmission model is based on the principle that new HIV infections in a population occur either through transmission from another HIV positive individual in the population or through immigration of HIV-positive individuals into the population. Diagnosed and undiagnosed segments of the infected population are treated separately. Surveillance data is used to calibrate model by matching them to the known properties of the diagnosed population. The remaining model parameters, linking the diagnosed and undiagnosed people in the infected population are estimated from the other sources of data, either from surveillance studies or from literature. The time rate of change of the number in HIV positive individuals N is given by the differential equation



$$\frac{dN}{dt} = e(1-\alpha_T)N + p_b p_h e\alpha N - d(1-\alpha_T)N - D + F_d + F_u \tag{1}$$

The differential equation (1) can be rewritten to give the Bernoulli differential equation

$$\frac{d\alpha_T}{dt} = \alpha_T^2((1-p_b(1-h_{eff}\frac{H}{M}))e - d - \frac{F_d - f_u - D}{M}) + \alpha_T(d - e + \frac{F_d + R - D}{M}) \tag{2}$$

Parameters used in the model

- α_T : Proportion of the population that is diagnosed
- e : Number of new HIV infections generated by each undiagnosed HIV-positive individual per unit time
- p_b : Factor by which the undiagnosed transmission rate is reduced due to behaviour change after diagnosis
- F_u : Net immigration for undiagnosed individuals
- $1-p_h$: Proportion of diagnosed individuals on HAART
- d : Deaths due to all causes for individuals with undiagnosed HIV infection
- D : Deaths due to all causes for individuals with diagnosed HIV infection
- F_d : Net immigration for diagnosed individual
- H : Number of diagnosed on HAART
- H_{eff} : Proportion on HAART that are virally suppressed
- M : Number of people known to be living with HIV
- R : Rate of new positive HIV diagnoses

All quantities except α_T, e, p_b and F_u are obtained from the surveillance data. The form of equation (2) means that value of α_T as a parameter is only needed for one anchor year as long as the other parameters are known. Nonetheless, this leaves us three unknown parameters (α_T, e and p_b) to be calibrated from the sources other than the surveillance data - under the assumption that the baseline infectivity of undiagnosed individuals within the modelled population (e) is constant and net immigration for undiagnosed individuals (F_u) is zero over the period under consideration. We assume that individuals who die of AIDS-related causes are diagnosed before death. Therefore, the death rate d of individuals with undiagnosed HIV infection is equal to the all-cause death rate in the general population.

For any given choice of e, p_b, and initial value $\alpha_0 = \alpha_T(t_0)$, we can solve the differential equation (2) numerically using Eulers method to obtain a monthly time series estimate $\alpha_1(t; e, p_b, \alpha_0)$. The initial time t_0 is an arbitrary time in the interval covered by the data.

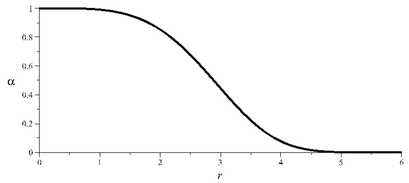
Models

b) Genetic Model

Alternative approach relies on the HIV virus genetic data collected by BCCfE. The basic idea underlying this approach is to compare the difference of the genetic sequence of the HIV virus in every newly diagnosed case with the database of sequences of the viruses in already diagnosed patients. We define the population genetic distance for each individual receiving a genotypic drug resistance test by first calculating, at time t, the minimum viral genetic distance to all individuals tested prior to time t. The population-level genetic distance time series $r(t)$ is calculated by averaging this result over all individuals tested at time t. Viral genetic distance is computed using the Tamura-Nei model of pairwise genetic distance [4].

We view the diagnosed fraction α_G as a function of the population genetic distance $r(t)$. The population genetic distance $r(t)$ being close to 0 means that for any tested individual, the viral sequence from the individual who infected them is highly likely to be in the tested database. This implies that α_G is close to 1. Furthermore, the derivative of α_G with respect to $r(t)$ should be zero at $r = 0$. If $r(t)$ is large, it is unlikely that the tested database contains many viral sequences from individuals that infected other indiviuduals in the database. In this case, α_G approaches 0 as r becomes large. The general shape of the dependence of α_G on r shown below.

Functional form of the relationship between fraction diagnosed and population genetic distance



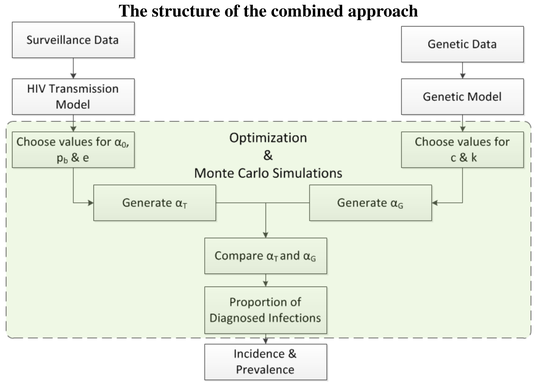
If we make the assumption that the function relating $r(t)$ and α_G is uniform and smooth, we are left with a two-parameter family of functions satisfying the limiting conditions above. This two-parameter family of functions is described by the equation:

$$\alpha_G(r) = e^{-cr^k} \tag{3}$$

where $c > 0$ and $k > 1$ are constants.

Combined Approach

The structure of the combined approach



Results

The Continuous Tabu Search method [1] is used to find the values of e, p_b, α_0, c and k which minimise the sum of the weighted squared differences

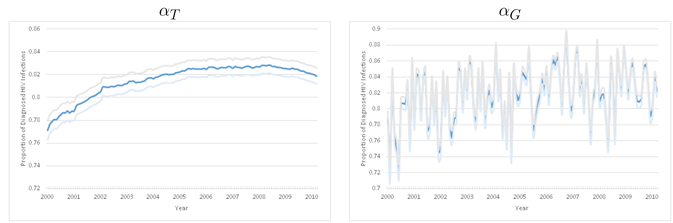
$$\Sigma_t \frac{dr}{d\alpha_G}|_{\alpha=\alpha_G(t;c,k)}(\alpha_T(t; e, p_b, \alpha_0) - \alpha_G(t; c, k))^2 \tag{4}$$

The values for e, p_b and α_0 are substituted back into α_T to obtain an estimate of the time series α, the fraction of the HIV positive population that is diagnosed. This estimate is biased towards predicting a small error in the estimate of $\alpha(t)$ for values of t near t_0. However this error is inconsequential to the model because t_0 is an arbitrary time. Therefore, we conduct a Monte Carlo simulation by repeating the optimisation procedure for randomly chosen t_0 in the time interval 2000 to 2009. The results of this Monte Carlo simulation are used to determine confidence intervals for $\alpha_T(t)$ and $\alpha_G(t)$.

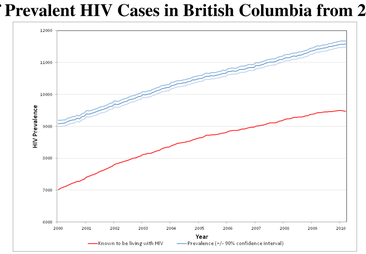
Results

A Monte Carlo simulation with 200 iterations is used to calculate the 90% confidence interval.

Proportion of Diagnosed HIV infections in British Columbia from 2000 to 2010



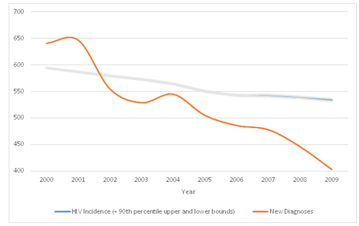
Number of Prevalent HIV Cases in British Columbia from 2000 to 2010



An estimate of the HIV incidence time series is obtained from the times series for α using the equation

$$I = Me(\frac{1}{\alpha} - 1 + p_b(1 - h_{eff}\frac{H}{M}) \tag{5}$$

Number of New HIV Infections and New Diagnoses in British Columbia from 2000 to 2009



Conclusions

We have developed a new method for estimating HIV incidence from routinely collected public health data combined with genotypic resistance test data. We produce estimates for (i) the proprotion of diagnosed HIV infections, (ii) HIV incidence and (iii) HIV prevalence by developing two independent models based on two independent data sets from the same population, . The model also generates estimates for number of new HIV infections generated by each undiagnosed HIV positive individual per unit time (e) and factor by which the undiagnosed transmission rate is reduced due to behavioural changes after diagnosis (p_b).

HIV incidence estimates provide public health officials, HIV clinicians, and healthcare policy makers with the ability to monitor and evaluate the effectiveness of programmes to control the HIV epidemic. A particular advantage of this method is that it utilises existing data and avoids the need for costly cohort studies.

References

[1] Cvijovi , D. and J. Klinowski. (1995) Taboo Search: An Approach to the Multiple Minima Problem. Science, 267(5198): 664666.

[2] Hirsch, M. S., H. F. Günthard, J. M. Schapiro, F. Brun-Vézinet, B. Clotet, S. M. Hammer, V. A. Johnson, D. R. Kuritzkes, J. W. Mellors, D. Pillay, P. G. Yeni, D. M. Jacobsen, and D. D. Richman. (2008) Antiretroviral Drug Resistance Testing in Adult HIV-1 Infection: 2008 Recommendations of an International AIDS SocietyUSA Panel. Clinical Infectious Diseases, 47: 266285.

[3] Prejean, J., R. Song, A. Hernandez, R. Ziebell, T. Green, F. Walker, L. S. Lin, Q. An, J. Mermin, A. Lansky, H. I. Hall, and the HIV Incidence Surveillance Group. (2011) Estimated HIV Incidence in the United States, 20062009. PloS ONE, 6: e17502.

[4] Tamura, K. and M. Nei. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution, 10: 512526.

[5] UNAIDS (2010) Methods for estimating HIV incidence. Epi Alert: UNAIDS Quarterly Update on HIV Epidemiology.

[6] Montaner JSG, Lima VD, Barrios R, Yip B, Wood E, et al. (2010) Association of highly active antiretroviral therapy coverage, population viral load, and yearly new HIV diagnoses in British Columbia, Canada: a population-based study. Lancet 376: 532–539.

We are grateful to the IRMACS Centre for providing a stimulating research environment.