Estimating HIV Incidence: A Mathematical Modelling Approach


1 The IMBACS 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 undiagnosed for as long as eight years. The rate of new HIV positives 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 in to 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

$$\frac{dN}{dt} = (1-\alpha)N + \gamma pN - dN - \alpha N - D + F_D = F_D$$

(1)

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

$$\frac{dN}{dt} = \alpha(1-(1-\alpha)N) + \gamma pN - dN - \alpha N - D + F_D = F_D$$

(2)

Parameters used in the model

- $\alpha$: Proportion of the population that is diagnosed
- $\gamma$: Number of new HIV infections generated by each undiagnosed HIV-positive individual per unit time
- $p$: Factor by which the undiagnosed transmission rate is reduced due to behaviour change after diagnosis
- $d$: Deaths due to causes other than HIV infected individuals with undiagnosed HIV infection
- $D$: Deaths due to causes for individuals with HIV diagnosis
- $F_D$: Net immigration for diagnosed individual
- $N$: Number of diagnosed individual
- $H$: Number of diagnosed individual on HAART
- $D$: Number of diagnosed individual

b) Genetic Model

Alternative approach relies on the HIV virus genetic data collected by RCOE. 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 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 diagnostic fraction $\frac{d}{N}$ as a function of the population genetic distance $r(t)$. The population genetic distance $r(t)$ being close to 0 means that the viral test individual from the individual who infected them is highly likely to be in the tested database. This implies that $\frac{d}{N}$ is close to 1. Furthermore, the derivative of $\frac{d}{N}$ with respect to $r(t)$ should be zero at $r(t)$. If $r(t)$ is large, it is unlikely that the tested database contains many viral sequences from individuals that infected other individuals in the database. In this case, $\frac{d}{N}$ approaches 0 as $r(t)$ becomes large. The general shape of the dependence of $\frac{d}{N}$ on $r(t)$ is shown below.

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

$$\frac{d}{N}(r(t)) = e^{-ar(t)}$$

(3)

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

Combined Approach

The structure of the combined approach

$$\frac{dN}{dt} = \alpha(1-(1-\alpha)N) + \gamma pN - dN - \alpha N - D + F_D = F_D$$

(2)

Results

The Continuous Tabu Search method [1] is used to find the values of $\alpha$, $\gamma$, $p$, $N$, $d$, and $k$ which minimise the sum of the weighted squared differences

$$\sum_{i=0}^{n}(\hat{x}_i - x_i)^2$$

(4)

The values for $\alpha$, $\gamma$, and $p$ are substituted back in $\frac{d}{N}$ to obtain an estimate of the time series $\frac{d}{N}$, the fraction of the HIV positive population that is diagnosed. This estimate is biased towards predicting a small error in the estimate of $\frac{d}{N}$ for values of $t$ near 0. However this error is inconsequential to the model because $t$ is an arbitrary time. Therefore, we conduct a Monte Carlo simulation by repeating the optimisation procedure for randomly chosen $t$ in the time interval 2000 to 2009. The results of this Monte Carlo simulation are used to determine confidence intervals for $\alpha$, $\gamma$, and $p$.

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 proportion 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 and factor by which the undiagnosed transmission rate is reduced due to behavioral changes after diagnosis (\(w(t)\)).

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

References