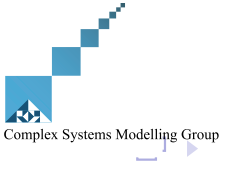
# A New Method for Estimating HIV Incidence Using Public Health Surveillance and Viral Genetic Data

A. R. Rutherford[1,2] Ali Nadaf[1,2] Bojan Ramadanovic[2]
Krisztina Vasarhelyi[2,3,4] Benita Yip[4] Art Poon[4,5] Richard Liang[4]
Richard Harrigan[4,5] Ralf W. Wittenberg[1] J. S. G. Montaner[4,5]

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

IDM Symposium
April 20, 2016



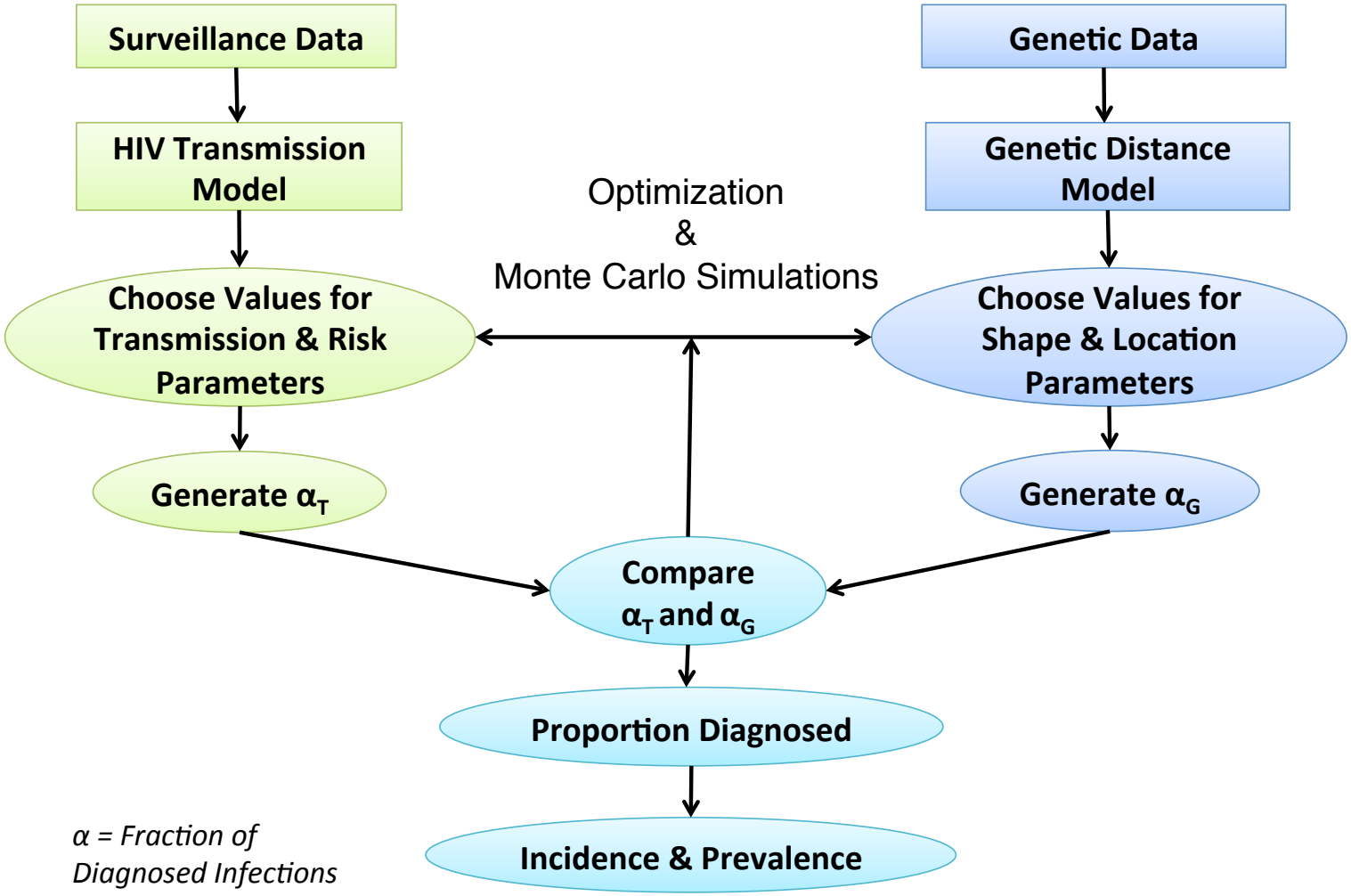

# HIV Incidence

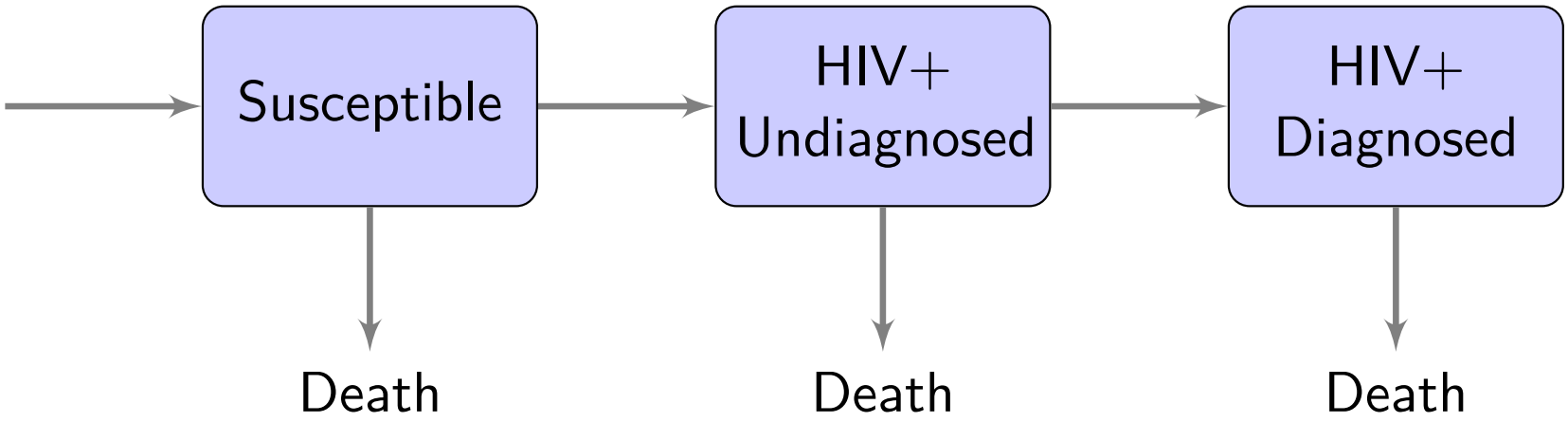
- Important indicator for the success of public health programs.
- Difficult to measure, because people infected with HIV may be asymptomatic for as long as 8 years.
- Currently estimated using back-calculation methods or longitudinal cohort studies.
- Our goal is to develop a method for estimating HIV incidence from *existing* public health and medical data.

# Data Sources
## British Columbia

- Reconciled Ministry of Health, BC Centre for Disease Control, and BC-CfE Drug Treatment Program data on new HIV diagnoses, mortality, immigration, and emigration.
- Drug treatment program data on number of patients receiving HAART and number virally suppressed.
- BC-CfE viral genetic data from genotypic drug resistance tests.

# Algorithm


Surveillance Data
HIV Transmission Model
Choose Values for Transmission & Risk Parameters
Generate $\alpha_T$
Optimization & Monte Carlo Simulations
Genetic Data
Genetic Distance Model
Choose Values for Shape & Location Parameters
Generate $\alpha_G$
Compare $\alpha_T$ and $\alpha_G$
Proportion Diagnosed
Incidence & Prevalence
$\alpha$ = Fraction of Diagnosed Infections




- The change in the number of prevalent HIV cases is the result of HIV+ individuals being added and removed from the population.
- New HIV infections are generated by transmission from both diagnosed and undiagnosed HIV+ individuals.
- HIV+ individuals are removed from the population through death — both AIDS related and other causes.

The number of HIV+ individuals $N$ is given by

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

$\alpha_T$ = fraction of the HIV+ subpopulation that is diagnosed

$e$ = number of new HIV infections generated by each undiagnosed HIV+ individual per unit time

$p_b$ = factor by which the undiagnosed transmission rate is reduced after diagnosis

$1 - p_h$ = diagnosed fraction of HIV+ subpopulation who are on HAART and virally suppressed

$d$ = per person all cause death rate for individuals with undiagnosed HIV infection

$D$ = all cause death rate for individuals with diagnosed HIV infection

$F_u$ = net immigration for undiagnosed HIV+ individuals

$F_d$ = net immigration for diagnosed HIV+ individuals

The fraction diagnosed of the HIV+ subpopulation is given by

$$\frac{d\alpha_T}{dt} = \alpha_T^2 \left( \left( 1 - p_b \left( 1 - h_{\text{eff}} \frac{H}{M} \right) \right) e - d - \frac{F_d + F_u - D}{M} \right) + \alpha_T \left( d - e + \frac{F_d + T - D}{M} \right)$$

$H$ = number of diagnosed on HAART

$h_{\text{eff}}$ = fraction virally suppressed on HAART

$M$ = number of people known to be living with HIV/AIDS

$T$ = rate of new HIV diagnosis

The equation for $\alpha_T$ has the form of the Bernoulli differential equation:

$$\frac{d\alpha_T}{dt} + p(t)\,\alpha_T(t) = q(t)\,\alpha_T^2(t)$$

Can be solved by integrating factors.

However, $p(t)$ and $q(t)$ are only known as data time series.

- The differential equation for $\alpha_T$ is solved numerically using Euler's method.
- It has three free parameters:

$e$ = number of new HIV infections generated by each HIV+ individual per unit time

$p_b$ = factor by which the undiagnosed transmission rate is reduced after diagnosis

$\alpha_0$ = value of $\alpha_T$ at an arbitrary "initial" or reference time $t_0$

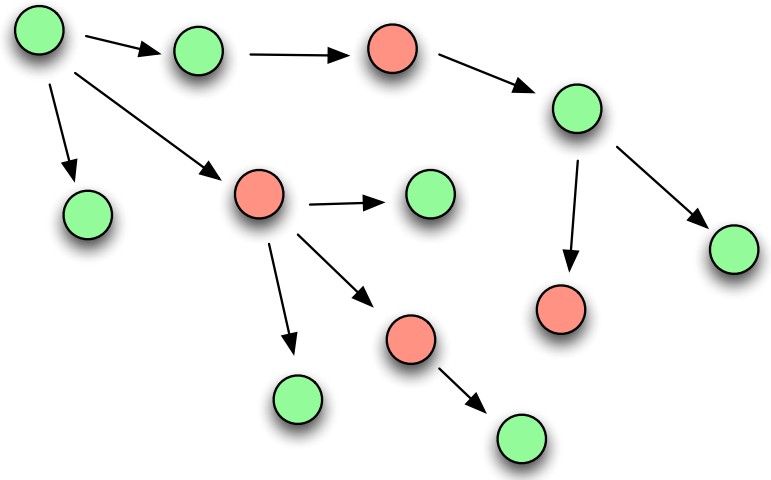
# Genetic Distance Model

## Overview

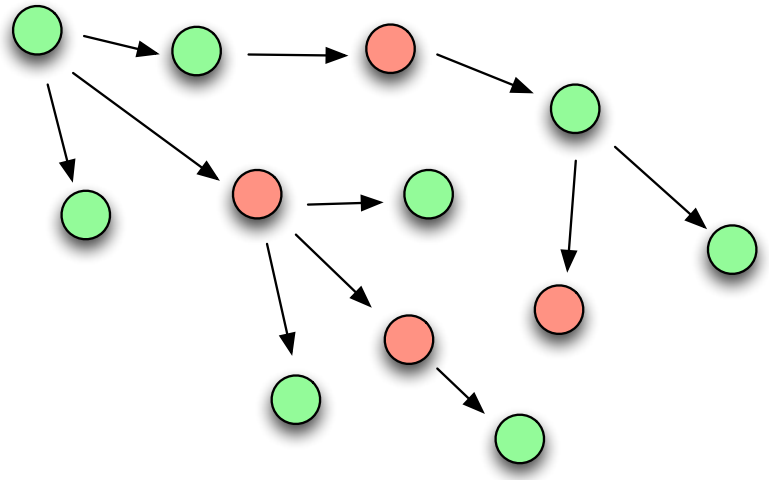
- Utilises viral RNA sequence data from genotypic drug resistance tests, which are done for every new diagnosis.
- HIV evolves continually and its genetic sequence diverges from a given ancestor with each transmission.
- As the fraction diagnosed increases it becomes more likely that a sequence and its closes ancestor are both in the genetic database.

# Genetic Distance Model
## Transmission Network

Genetic distance computations using genotypic drug resistance data can be used to construct the HIV transmission network.

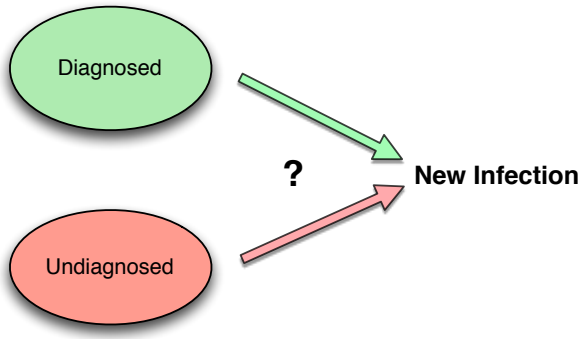


Genetic distance computations using genotypic drug resistance data can be used to construct the HIV transmission network.



Constructing the transmission network for a large population is very computationally expensive.

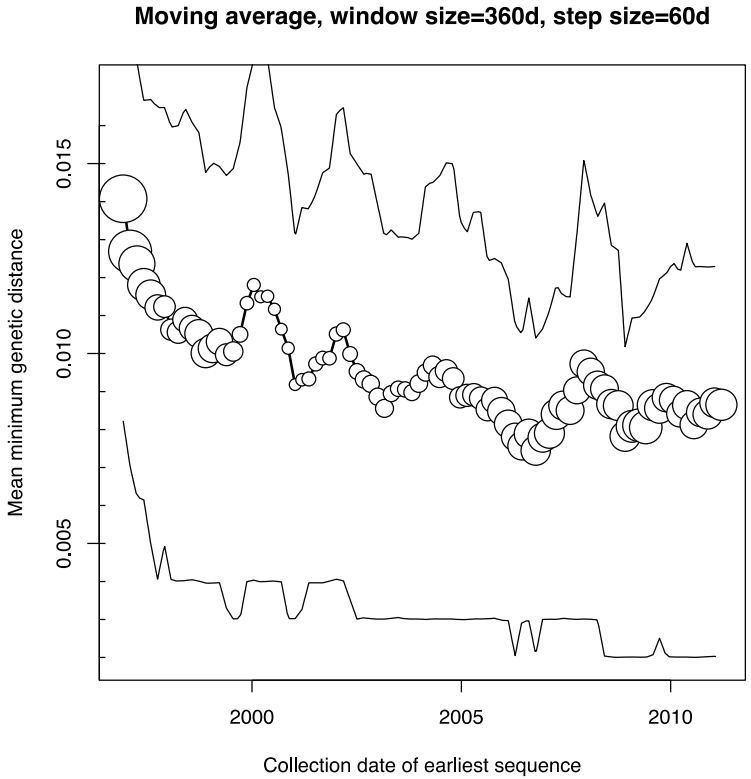
**Do we really need to construct the network?**



# Genetic Distance Model
## Population Genetic Distance

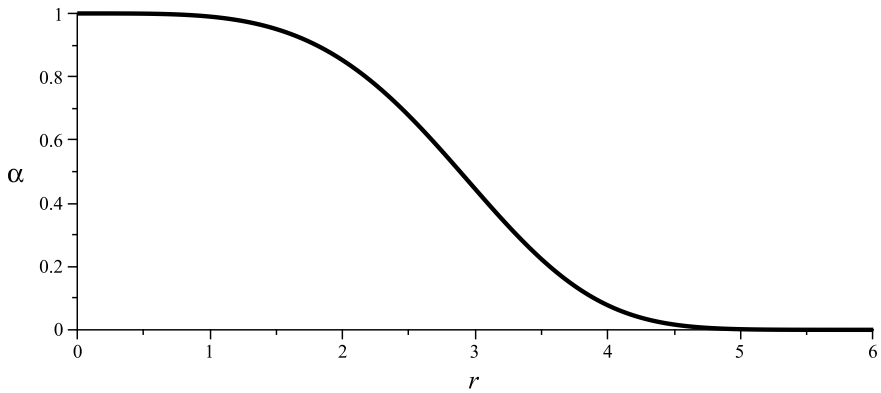
$$r(t) = \underset{\substack{i \text{ at} \\ \text{time } t}}{\text{ave}} \left[ \min_{\substack{j \text{ at time} \\ \tau < t}} g(i,j) \right]$$

where $g(i,j)$ is the Tamura-Nei genetic distance between virus $i$ and virus $j$.




K. Tamura and M. Nei, *Mol Biol Evol*, **10** (1993).

The fraction diagnosed as a function of the population genetic distance has the form of a decreasing function from 1 to 0.




This family of functions is modelled by

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

where $c > 0$ and $k > 0$ are free parameters.

# Optimisation and Monte Carlo Simulations

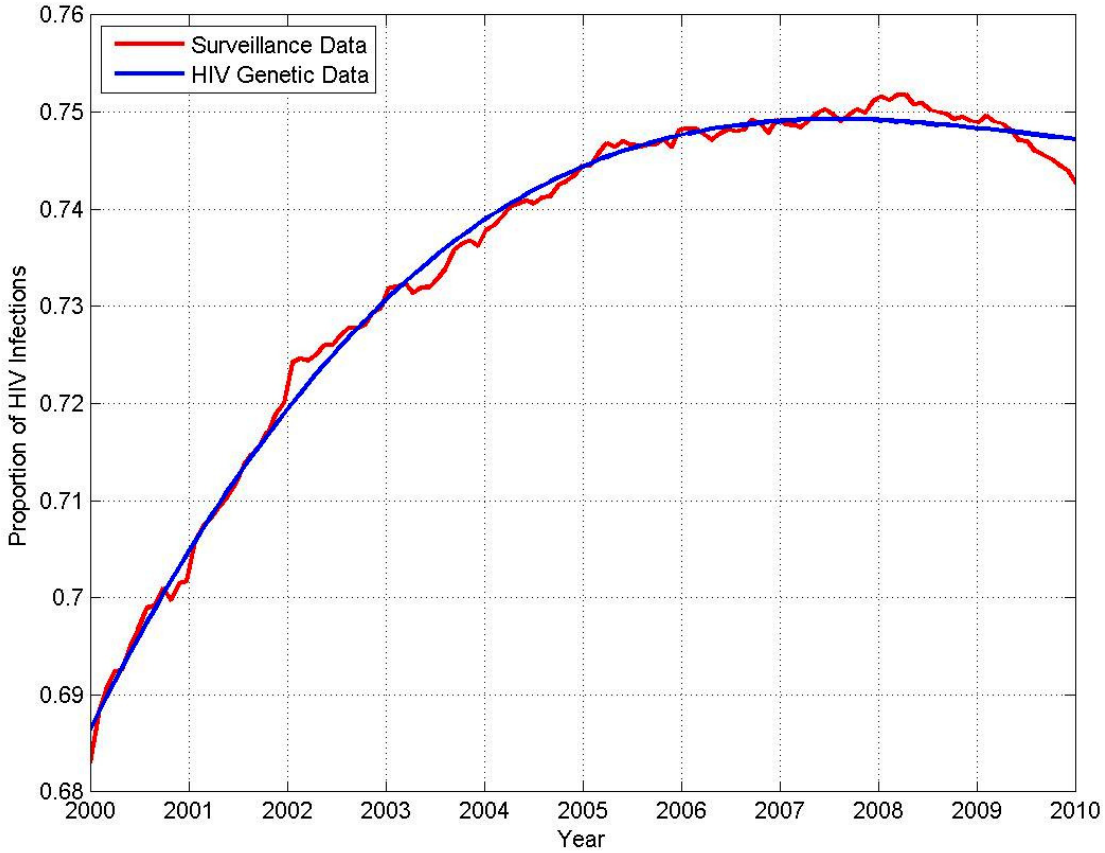
Tabu search is used to find the values of the 5 parameters $e$, $p_b$, $\alpha_0$, $c$, and $k$ that minimise the objective function

$$F(\alpha_0, e, p_b, c, k) = \sum_i \frac{\left(\alpha_T(\alpha_0, e, p_b; t_i) - \alpha_G(c, k; t_i)\right)^2}{\alpha'_G(t_i)} .$$

A bias is introduced into the solution for $\alpha_T$ by the choice of reference time $t_0$. A Monte Carlo simulation is done in which the optimisation is repeated for randomly chosen $t_0$.

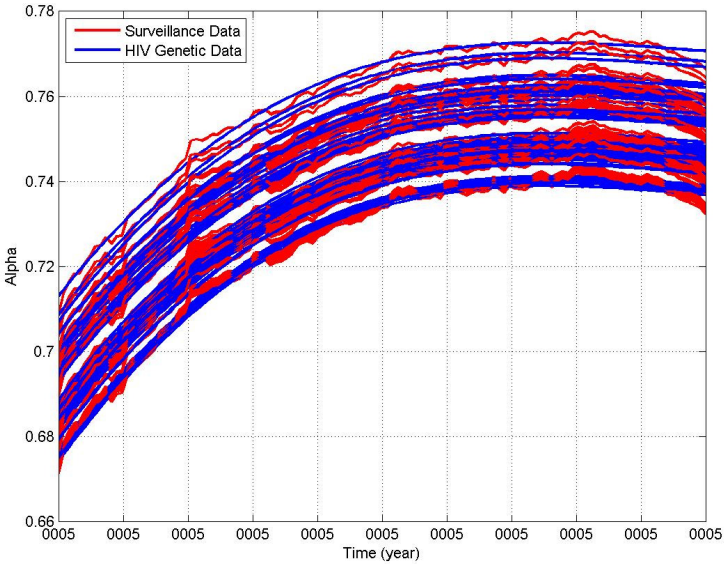
The results of the Monte Carlo are used to calculate $\alpha$, HIV prevalence, and HIV incidence. The empirical distribution of Monte Carlo results is used to calculate confidence intervals.

A single tabu search result for the fraction diagnosed from the transmission model $\alpha_T$ to a spline-smoothing of the fraction diagnosed from the genetic distance model $\alpha_G$.

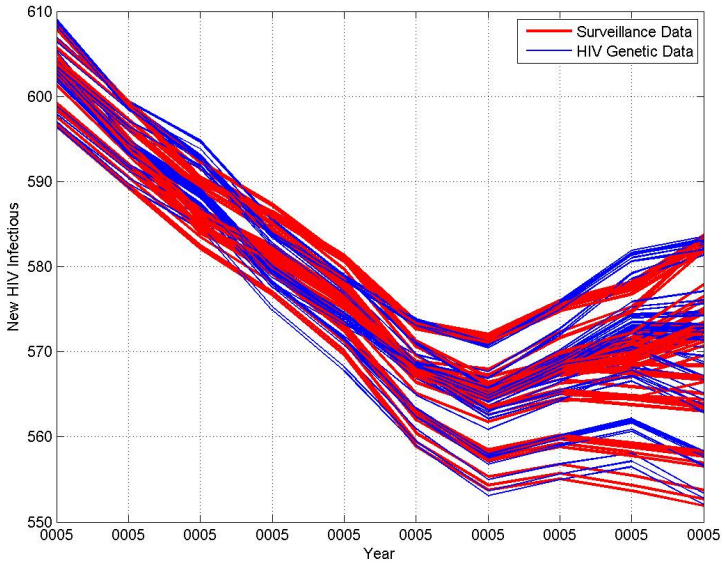


# Monte Carlo Simulations

Results of the Monte Carlo simulation for 71 tabu search optimisations with randomly chosen reference times $t_0$ for the transmission model.



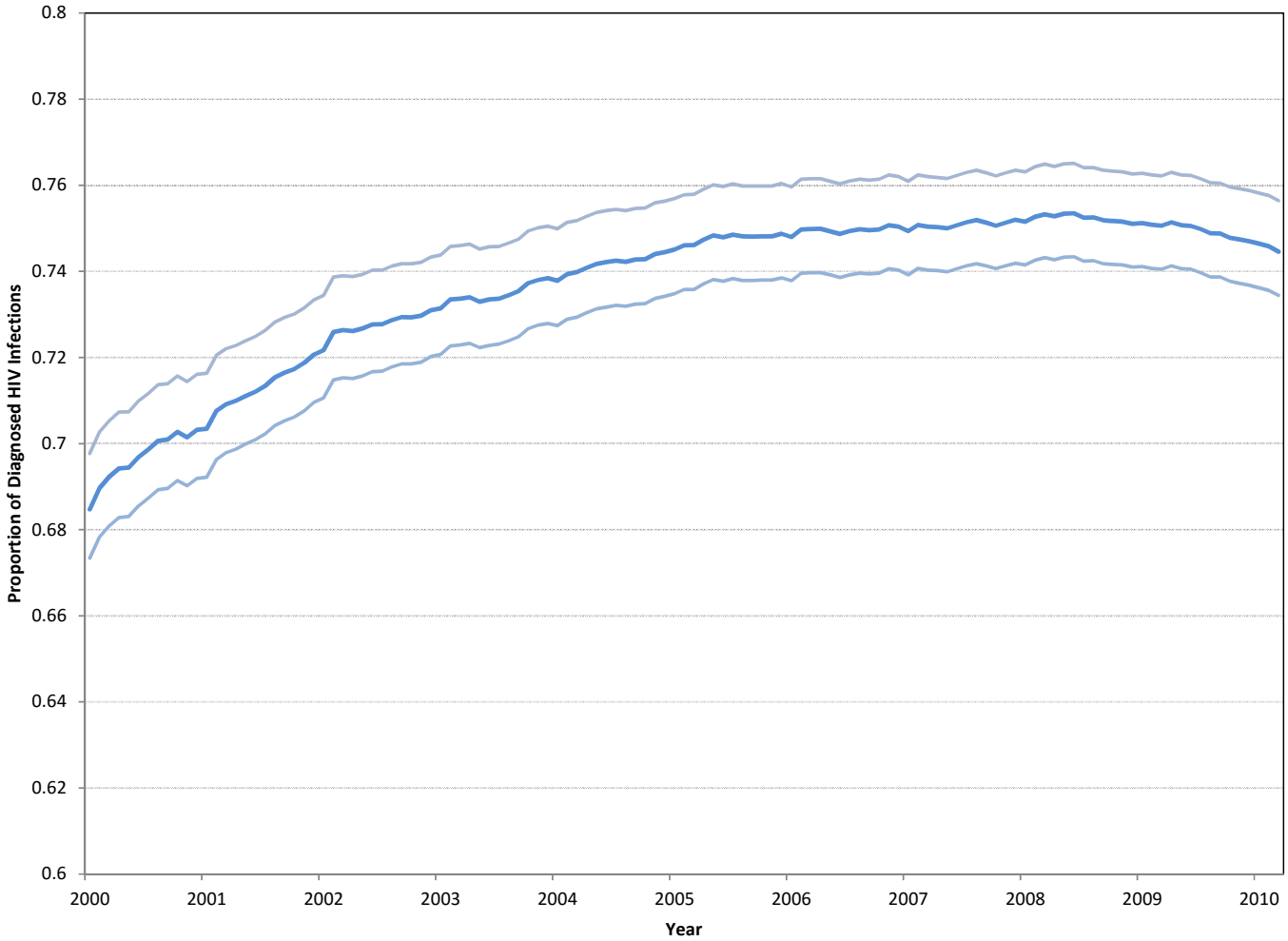

Fraction Diagnosed




HIV Incidence

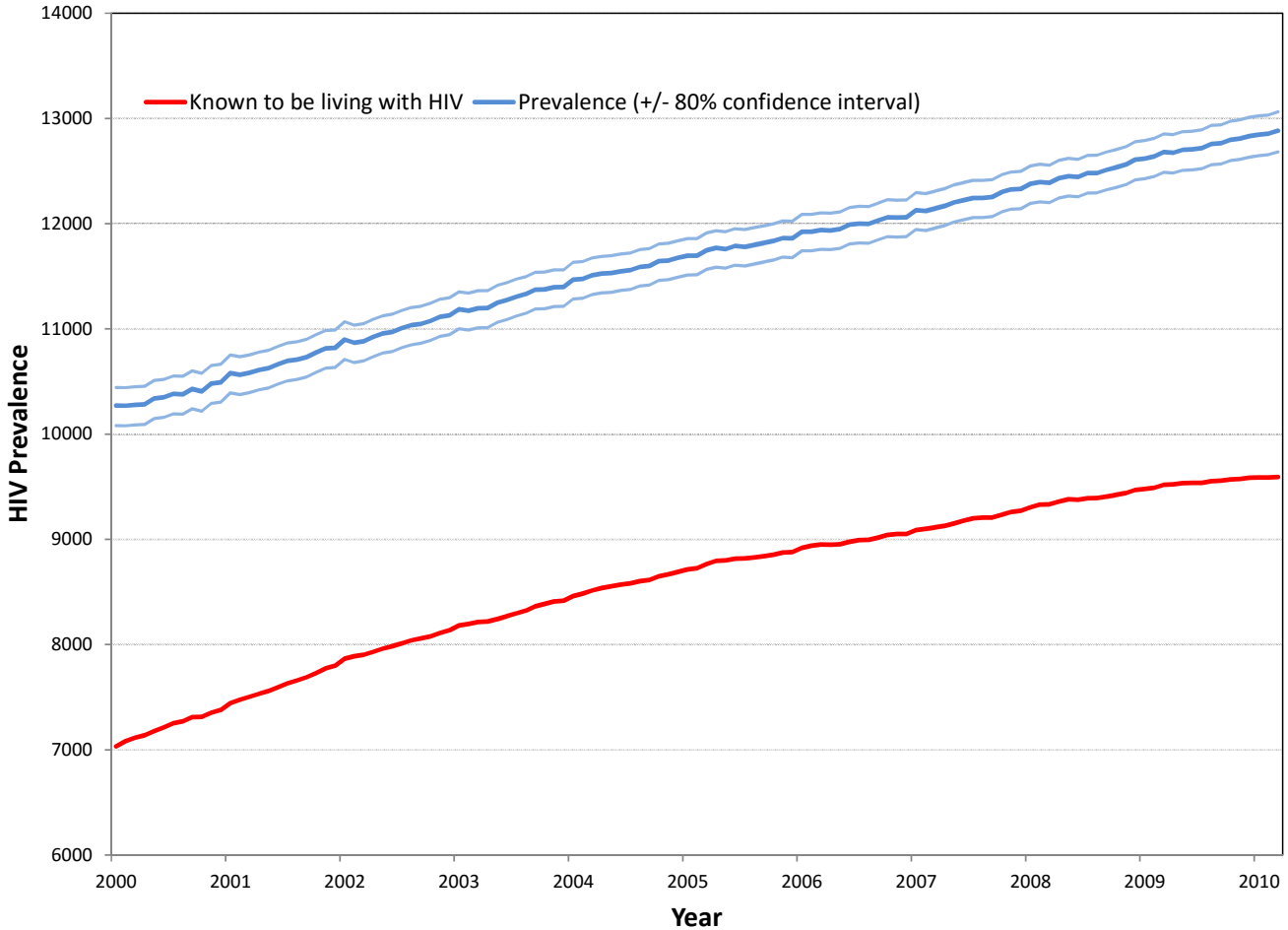
# Diagnosed Fraction of HIV+ Subpopulation
## British Columbia

Model results for the fraction of the HIV+ subpopulation in British Columbia that was diagnosed from 2000 to 2010.



# HIV Prevalence
## British Columbia

Comparison of model results for HIV prevalence in British Columbia to the number of people known to be living with HIV/AIDS.



# HIV Incidence

HIV incidence may be obtained from the model results for $\alpha$, the fraction diagnosed, using the equation

$$I = Me\left(\frac{1}{\alpha} - 1 + p_b\left(1 - h_{\text{eff}}\frac{H}{M}\right)\right)$$

where

$M$ = number of people known to be living with HIV/AIDS

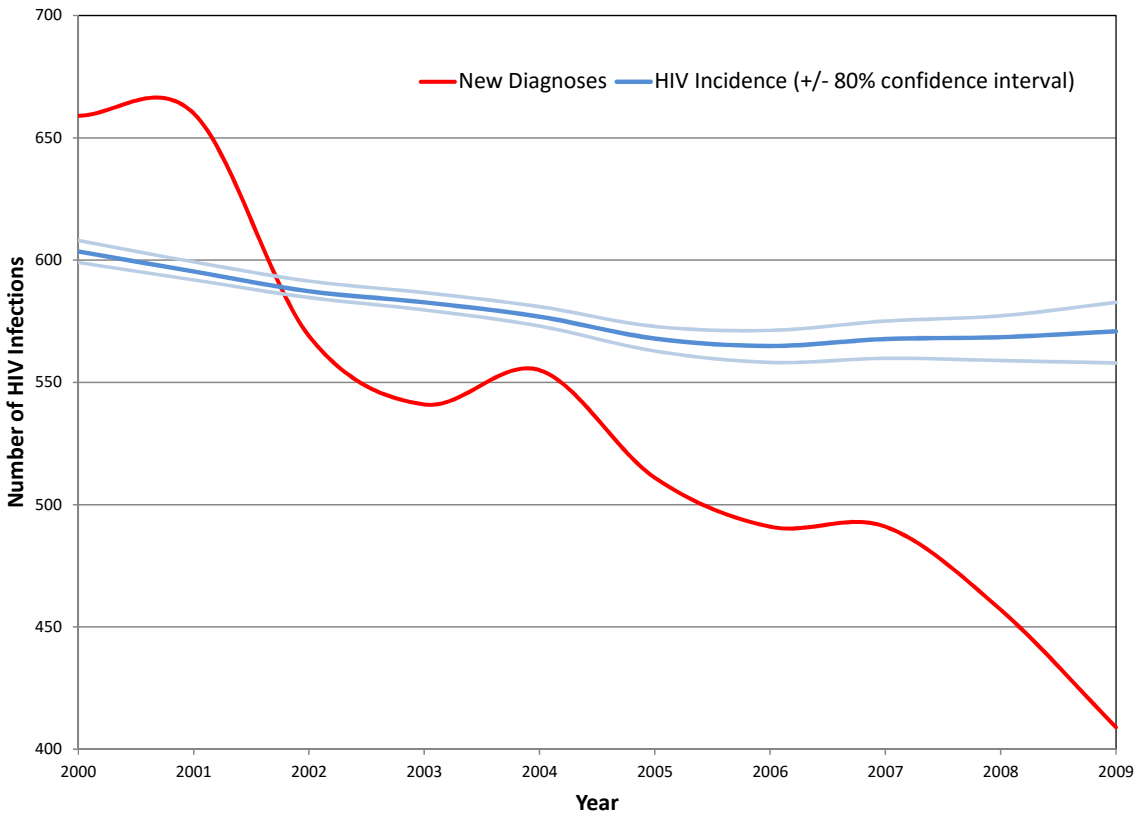
$p_b$ = factor by which the undiagnosed transmission rate is reduced after diagnosis

$h_{\text{eff}}$ = fraction virally suppressed on HAART

$H$ = number of diagnosed on HAART

# HIV Incidence
## British Columbia

Comparison of model results for HIV incidence in British Columbia to the number of new diagnoses.



# Validation

| Parameter | Estimate from Model | Estimate from Literature | Reference |
|---|---|---|---|
| Proportion diagnosed in 2002 | 0.80 | 0.76 | PHAC 2006 |
| Proportion diagnosed in 2005 | 0.82 | 0.79 | PHAC 2006 |
| Proportion diagnosed in 2008 | 0.83 | 0.81 | PHAC 2008 |
| Reduction in transmission due to behaviour change after diagnosis ($p_b$) | 0.45 | 0.29 | Marks, et al., AIDS, 2006 |
| Transmission rate for undiagnosed ($e$) | 0.11 | 0.069 | Marks, et al., AIDS, 2006 |
| HIV incidence in 2005 | 548 | 320 - 620 | BCCDC 2010 |
| HIV incidence in 2008 | 538 | 280 - 540 | BCCDC 2010 |
| HIV prevalence in 2005 | 10,560 | 10,350 (8,300 – 12,400) | BCCDC 2010 |
| HIV prevalence in 2008 | 11,108 | 11,400 (9,300 – 13,500) | BCCDC 2010 |

# Future Work

- Analyse more recent data from British Columbia
  - STOP HIV/AIDS programme data
- Time-dependent model parameters
  - requires sufficient data
- Apply to other jurisdictions
- Compare to other methods for estimating HIV incidence
  - retrospective cohort studies
    - often focused on specific risk groups
  - incidence assays

# Thank You

**Acknowledgements:**

Canadian Institutes of Health Research (CIHR)
MITACS
Providence Health Care
BC Centre for Disease Control
Vancouver Coastal Health