# Run simulated epidemic history

*Stephane*

*3/27/2017*

## Epidemic simulation

Epidemic simulations are produced by code in `model0.r` were carried out with the aim of comparing different methods to infer epidemiological quantities from labelled viral sequence data. Our motivation was to obtain a simple though realistic transmission history in a population that is comparable to MSM in London. Epidemic simulation was based on a system of ordinary differential equations which describe the dynamics of the number of infected hosts in different categories. Additionally, genealogical trees were simulated conditioning on the epidemic history, and trees were matched to the real UK metadata data pertaining to times of sampling and clinical stage of infection. For some analyses, sequence data were also simulated conditioning on simulated genealogies. Clustering and source attribution approaches were applied to simulated data with known transmission risk profiles to evaluate the ability of different approaches to identify transmission risk factors and to correctly estimate transmission risk ratios. %2.3 Compartmental epidemic model

## Compartmental epidemic model

The epidemic history in London MSM was modeled with a system of ordinary differential equations determining transmission and transition through 5 CD4 stages, 4 age groups and 3 diagnosis states (undiagnosed, diagnosed untreated and diagnosed under treatment). Individuals were further stratified by an arbitrary binary risk characteristic influencing transmission. The population was thus structured in 120 states. Furthermore, we modeled importation of infections into London, which can have a dramatic effect on HIV genetic diversity. All code used to simulate epidemic histories and genealogical trees is available online [**?**]. %2.3.1 Model parameters

### Model parameters

Mean time of progressions to CD4 stages and proportion in each CD4 category after seroconversion were obtained from Cori et al. [**?**]. Transmission was allowed to vary according to weights provided by risk category, treatment status and according to age assortativity. A proportion of 20% of the population were deemed to be at high risk with a ten-fold increase in transmission than low risk counterparts. Relative to undiagnosed individuals, diagnosed and treated patients had a reduction in transmission by respectively a factor 2 and 20. An age assortativity parameter was introduced in the transmission matrix which caused transmission rates to decrease as a power law function of the difference in age. Age groups were based on quantiles of observed age distribution of MSM diagnosed with HIV in London [**?**] and transmission rates were independent of age. Two variations in this simulation were explored in terms of how transmission rate varies with age of infection in order to evaluate rates of false-positive identification of transmission risk factors. We let infection stage influence transmission rates with a ten-fold increase in probability of transmission in early HIV infection and a three-fold increase in AIDS stage (baseline scenario), or transmission was independent of infection stage (equal rates scenario). Expressed mathematically, the total transmission rate of a patient with CD4 stage $i$, continuum of care status $j$, and generic risk factor $k$ is $\lambda_{ijk}(t) \propto r_i r_j r_k$, where $r_.$ are risk ratios for each category. Individual transmission rates are normalised so that total incidence is given by $\iota(t)$ based on a previous study [**?**] assuming that dynamics of new infections in MSM was the same at the country level and in London.

Incidence and diagnosis rates were modeled as logistic functions of time and jointly calibrated to match the number of MSM living with diagnosed HIV in London in 2012 [**?**]. Rates of treatment were modelled as zero before 1995 and then increase according to a logistic function with maximum 1 and steepness 0.5.

```
library(knitr)
read_chunk('model0.R')
```

Load packages

```
require(phydynR) # replaces rcolgem
require(deSolve)
require(Rcpp)
```

Source C functions

```
sourceCpp( 'model0.cpp' ) # F_matrix and G_matrix fns
```

Define input parameters

```
##- Age progression
##(by quantiles: 18.0, 27.0, 33.0, 40.0, 80.5)
age_rates <- c(agerate1 = 1/9/365
    , agerate2 = 1/6/365
    , agerate3 = 1/7/365
    , agerate4 = 1/40.5/365
)

##- Natural history (from Cori et al. AIDS 2015)
stage_prog_yrs <- c( .5, 3.32, 2.7, 5.50, 5.06 )
stageprog_rates <- setNames( 1 / (stage_prog_yrs  * 365 )
 , c('gamma1', 'gamma2', 'gamma3', 'gamma4', 'gamma5')  )

pstarts <- c( pstartstage1 = 0 #NA
 , pstartstage2 = 0.76
 , pstartstage3 = 0.19
 , pstartstage4 = 0.05
 , pstartstage5 = 0
)

theta <- c( age_assort_factor = .5 # power of age difference
  , pRiskLevel1 = .8 # proportion in low risk group
  , srcMigrationRate = 1/50/365 # per lineage rate of migration to source
  , srcGrowthRate = 1 / 3 / 365 #
  , src0 = 1e3  # initial source size
  , inc_scale = 0.09401734 # based on docking (see below) # initial = .03
  , max_diag_rate = 0.66227809 # based on docking (see below) # initial = 1/3 (time 2 diag of 3 yrs)
  , diag_rate_85 = 1/10
  , accel_diag_rate = 0.03196171 # based on docking (see below ) # initial = 1/7 # accel of logistic fu
  , treatmentEffectiveness = .95 # slows stage progression
  , pstarts
    , age_rates
    , stageprog_rates
)

theta_default <- theta
```

| Notation | Parameter | Value |
|---|---|---|
| | *Age progression rate* [a] | |
| $\alpha_1$ | Group 1 [18-27) | $1/9/365\ day^{-1}$ |
| $\alpha_2$ | Group 2 [27-33) | $1/6/365\ day^{-1}$ |
| $\alpha_3$ | Group 3 [33-40) | $1/7/365\ day^{-1}$ |
| $\alpha_4$ | Group 4 [40-80.5) | $1/40.5/365\ day^{-1}$ |
| | *Stage progression rate*[b] | |
| $\gamma_1$ | Stage 1 (early HIV infection) | $1/0.5/365\ day^{-1}$ |
| $\gamma_2$ | Stage 2 (CD4 $>$ 500 cells/mm3) | $1/3.32/365\ day^{-1}$ |
| $\gamma_3$ | Stage 3 (350 $<$ CD4 $\leq$ 500 cells/mm3) | $1/2.7/365\ day^{-1}$ |
| $\gamma_4$ | Stage 4 (200 $<$ CD4 $\leq$ 350 cells/mm3) | $1/5.5/365\ day^{-1}$ |
| $\gamma_5$ | Stage 5 (CD4 $\leq$ 200 cells/mm3) | $1/5.06/365\ day^{-1}$ |
| | *Fraction of individuals transitioning from* [b] | |
| $\pi_1$ | Stage 1 to stage 2 | 0.76 |
| $\pi_2$ | Stage 1 to stage 3 | 0.19 |
| $\pi_3$ | Stage 1 to stage 4 | 0.05 |
| $\pi_4$ | Stage 1 to stage 5 | 0 |
| $a$ | Age assortativity factor | 0.5 |
| $p$ | Proportion of individuals in low-risk group | 0.8 |
| $m$ | Per lineage rate of migration to source compartment | $1/50/365\ day^{-1}$ |
| $g$ | Rate of growth of source compartment | $1/3/365\ day^{-1}$ |
| $s$ | Initial size of source compartment | 1000 |
| $i$ | Incidence scaling factor for London MSM [c] | 0.03 |
| | *Diagnosis rate* | |
| $d_{85}$ | Fixed rate prior to 1985 | $1/10\ year^{-1}$ |
| $\mu_d$ | Maximum value of logistic function after 1985 [c] | $1/3\ year^{-1}$ |
| $k_d$ | Steepness of logistic function after 1985 [c] | $1/7\ year^{-1}$ |
| | *Treatment rate* | |
| $t_{95}$ | Fixed rate prior to 1995 | 0 |
| $\mu_t$ | Maximum value of logistic function after 1995 | 1 |
| $k_t$ | Steepness of logistic function after 1995 | 0.5 |
| $e$ | Treatment effectiveness | 0.95 |
| | *Transmission weight conferred to individuals in* | |
| $w_{s1}$ | Stage 1 | 1 |
| $w_{s2}$ to $w_{s4}$ | Stages 2 to 4 | 0.1 |
| $w_{s5}$ | Stage 5 | 0.3 |
| $w_{a1}$ to $w_{a4}$ | Age groups 1 to 4 | 1 |
| $w_{c1}$ | Care status 1 (undiagnosed) | 1 |
| $w_{c2}$ | Care status 2 (diagnosed and untreated) | 0.5 |
| $w_{c3}$ | Care status 3 (diagnosed and treated) | 0.05 |
| $w_{r1}$ | Risk status 1 (low risk) | 1 |
| $w_{r2}$ | Risk status 2 (high risk) | 10 |

[a] From quartiles of age of MSM diagnosed in London available in UKDRDB

[b] From Cori et al. AIDS 2015

[c] Initial value later calibrated to retrieve observed number of diagnosed cases from surveillance data

Note: incidence and diagnosis rate scaling factors are a priori. Now, there are fitted doi:10.1371/journal.pone.0055312.g002 (Fig 2.A)

every individuals start infection at EHI stage

prRecipMat: - prob that recipient get infection, conditionning on - prob of being risk level 1 (80%) vs risk level 2 (20%) - EHI stage (only those recipient get infection) - care status (only undiagnosed get infection) - age assortativity (power of age class difference) - intervenes in F matrix [ F(i,j) = incidence * w * prRecipMat(i,j),

with w = beta_NH * beta_age * beta_care * beta_risk ]

prStageRecipMat: - Prob for EHI recipient to jump to next other CD4 stage