

# Modelling the growth of an HIV incidence assay

Michael Sweeting   Daniela De Angelis

MRC Biostatistics Unit, Cambridge

Acknowledgments: Barbara Suligoj, National AIDS & STI Unit,  
Rome

# Outline

- ▶ Background
  - ▶ Estimating incidence using HIV incidence assays
  - ▶ The *window period*
- ▶ The Rome AI Cohort
- ▶ Longitudinal modelling of assay growth
- ▶ Inverse prediction
- ▶ Summary and conclusions

## HIV incidence assays

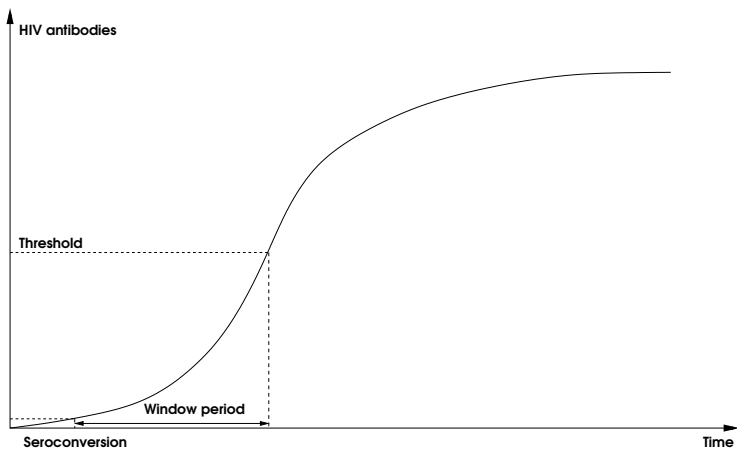
- ▶ HIV incidence assays (biomarkers) have been developed to allow discrimination of ...
  - ▶ antibody titre: measures the amount of antibodies the body has produced.
  - ▶ antibody avidity: measures the amount of binding between antigens and antibodies.
- ▶ Commonly a specific threshold is chosen
- ▶ Individuals whose biomarker level less than the threshold are classified as *recently infected*

## Using a biomarker to estimate HIV incidence from a cross-sectional sample

- ▶ Suppose a (random) sample of infected individuals identifies  $r$  out of  $n$  as *recently infected* using an incidence assay technique with estimated prevalence  $\hat{P} = r/n$

$$I \approx \frac{\hat{P}}{\mu}$$

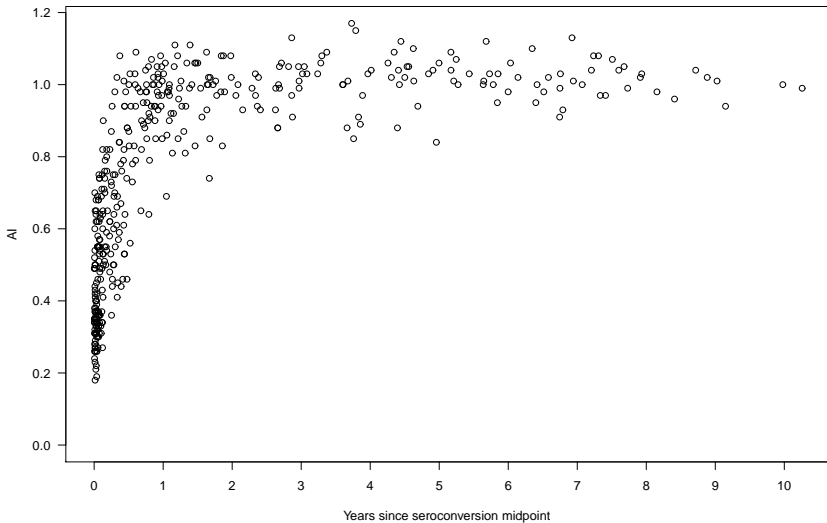
- ▶  $\mu$  is the mean time spent in the *recent infection* state - the mean *window period*
- ▶ This equation relies on two assumptions...
  - ▶ There exists a maximum window period  $w_m$
  - ▶ Incidence remains constant over the duration of window period
- ▶ Aim: to estimate  $\mu$  and the *distribution* of the window period associated with a new avidity assay



# The Rome AI Cohort

- ▶ Data available from 103 seroconverters with known last negative ( $d_i^{-ve}$ ) and first positive ( $d_i^{+ve}$ ) test dates
- ▶ Each individual has measurements of the Avidity Index (AI) assay  $y_{ij}$  taken at times  $t_{ij}$  after first positive,  $j = 1, \dots, m_i$
- ▶ The AI is a ratio of two assays (a treated and control)
- ▶ Hence:  $AI < 1$  in the initial period after infection and should approach 1 for established infection
- ▶ At least two AI measurements per person
- ▶ Mean seroconversion interval is 3.6 months





## Longitudinal modelling of assay growth

- ▶ A non-linear random-effects model is used
- ▶ Individual-level parameters allow person-specific growth curve estimation
- ▶ The unknown seroconversion date gives the model an errors-in-variables component
- ▶ Let  $T_{ij}^* = t_{ij} + \tau_i$  be the unknown time from seroconversion to the  $j$ th sampling time for patient  $i$

$$y_{ij} = \phi_{0i} + (\phi_{1i} - \phi_{0i})\exp(-\exp(\phi_{2i})T_{ij}^*) + \epsilon_{ij}$$

- ▶  $\epsilon_{ij} \sim_{\text{iid}} N(0, \sigma_w^2)$  are the residual error terms

$$(\phi_{0i}, \phi_{1i}, \phi_{2i})^T \sim N_3 \left( (\mu_0, \mu_1, \mu_2)^T, \Sigma_b \right)$$

- ▶  $\phi_{0i}$  is the asymptote of the non-linear function
- ▶  $\phi_{1i}$  is the intercept (*i.e.* AI at seroconversion)
- ▶  $\phi_{2i}$  is the logarithm of the rate constant

- ▶ *Naïve model*

$$\tau_i = \frac{d_i^{+ve} - d_i^{-ve}}{2}$$

- ▶ *Uniform prior model*

$$\tau_i \sim \text{Uniform}(0, d_i^{+ve} - d_i^{-ve})$$

- ▶ Analysis performed within an MCMC Bayesian framework

## Results

- ▶ No evidence of between-individual variation in the asymptote, so this is modelled as a fixed effect:

$$\phi_{0i} = \mu_0$$

Parameter	<i>Naïve</i> model Posterior median (SD)	<i>Uniform prior</i> model Posterior median (SD)
<b>Population means</b>		
Asymptote $\mu_0$	1.017 (0.007)	1.016 (0.006)
Intercept $\mu_1$	0.346 (0.023)	0.349 (0.021)
Log-rate $\mu_2$	0.934 (0.119)	0.964 (0.122)
<b>Between-individual SD</b>		
Intercept	0.145 (0.021)	0.125 (0.019)
Log-rate	0.835 (0.109)	0.860 (0.110)
Correlation	-0.53 (0.14)	-0.59 (0.14)
<b>Within-individual SD</b>		
	0.076 (0.003)	0.074 (0.003)
DIC	-870.6	-892.0

## Inverse prediction

- ▶ We wish to use the model to estimate the window period distribution in the study population for a given threshold  $\alpha$
- ▶ For individual  $i$  in the sample this is just

$$T_i^*(\alpha) = \log \left( \frac{\phi_{1i} - \phi_{0i}}{\alpha - \phi_{0i}} \right) \exp(-\phi_{2i})$$

where we use the posterior distributions of  $\phi_{0i}$ ,  $\phi_{1i}$ , and  $\phi_{2i}$ .

- ▶ Mean in-sample window period is

$$E [T^*(\alpha)] = \frac{\sum_{i=1}^n T_i^*(\alpha)}{n}$$

- ▶ Out-of-sample prediction

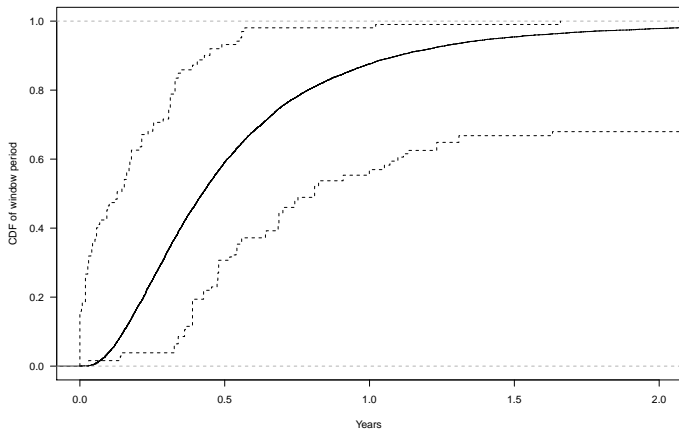
$$T_{new}^*(\alpha) = \log \left( \frac{\phi_{1new} - \phi_{0new}}{\alpha - \phi_{0new}} \right) \exp(-\phi_{2new})$$

where  $\phi_{0new}$ ,  $\phi_{1new}$ , and  $\phi_{2new}$  are first drawn from the posterior distribution of the random-effects.

## Estimating the window period for the *Uniform prior* model

	Threshold		
	0.6	0.7	0.8
<hr/>			
<b>Mean in-sample window period, days</b>			
	71 (61, 85)	125 (108, 149)	202 (174, 245)
<hr/>			
<b>Predicted out-of-sample probability of reaching threshold</b>			
0-3 months	0.75	0.46	0.25
0-6 months	0.95	0.81	0.59
0-9 months	0.98	0.92	0.78
0-12 months	0.99	0.96	0.88

## CDF of window period for 0.8 threshold



## Summary

- ▶ The advantages of modelling AI growth are
  - ▶ The functional form of the non-linear model makes biological sense
  - ▶ The window period can be estimated for any given threshold using inverse prediction
  - ▶ In-sample and out-of-sample predictions can be made
  - ▶ Method utilizes all available longitudinal measurements
- ▶ Models that assume a known seroconversion time (e.g. the midpoint) do not properly account for the variability in the data
- ▶ A Bayesian framework easily allows incorporation of unknown seroconversion times