Defining and Measuring 'Recent infection'

Application to Incidence Estimation

Alex Welte

SACEMA

DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis

What is 'Recent Infection'?

- Introduction of a time scale into a biological state definition poses some fundamental quandaries:
  - Is recency of fixed duration?
  - Is it linked to the presence of a biomarker?
  - Is there a 'gold standard' for evaluating a test?
- Why would we bother with knowing about recent infection?
  - **Incidence is related to 'prevalence' of recent infection**
  - Clinical/sociological response to diagnosis

Crucial Epidemiological Indicators

- Reliable estimation of prevalence and incidence (even more so) are central to the determination of epidemiological trends.
  - Prevalence is the number of infected individuals.
  - Incidence is the 'rate' of occurrence of new infections.
- These estimates are important for:
  - Assessing outbreaks
  - Allocating resources
  - Evaluating interventions
  - Planning cohort and community based studies

The Problem of Estimating 'Rates'

- The "gold standard" for incidence measurement was long said to be a prospective cohort follow-up (a.k.a. "direct observation").
- It's not all gold:
  - Expensive and logistically complex to run
  - Time consuming
  - Biased due to selection, intervention and follow-up
- Various approaches:
  - Direct observation such as in ongoing cohort studies
  - Mathematical modeling to interpret trends in prevalence data
  - Cross-sectional surveys using 'recent infection tests'

Outline of the Talk

(Almost Consistent) Colour Coding

- **Terms on first use / at time of definition**
- **Jargon which is hopefully clear**
- **General Emphasis**

Outline

Longitudinal Study Design

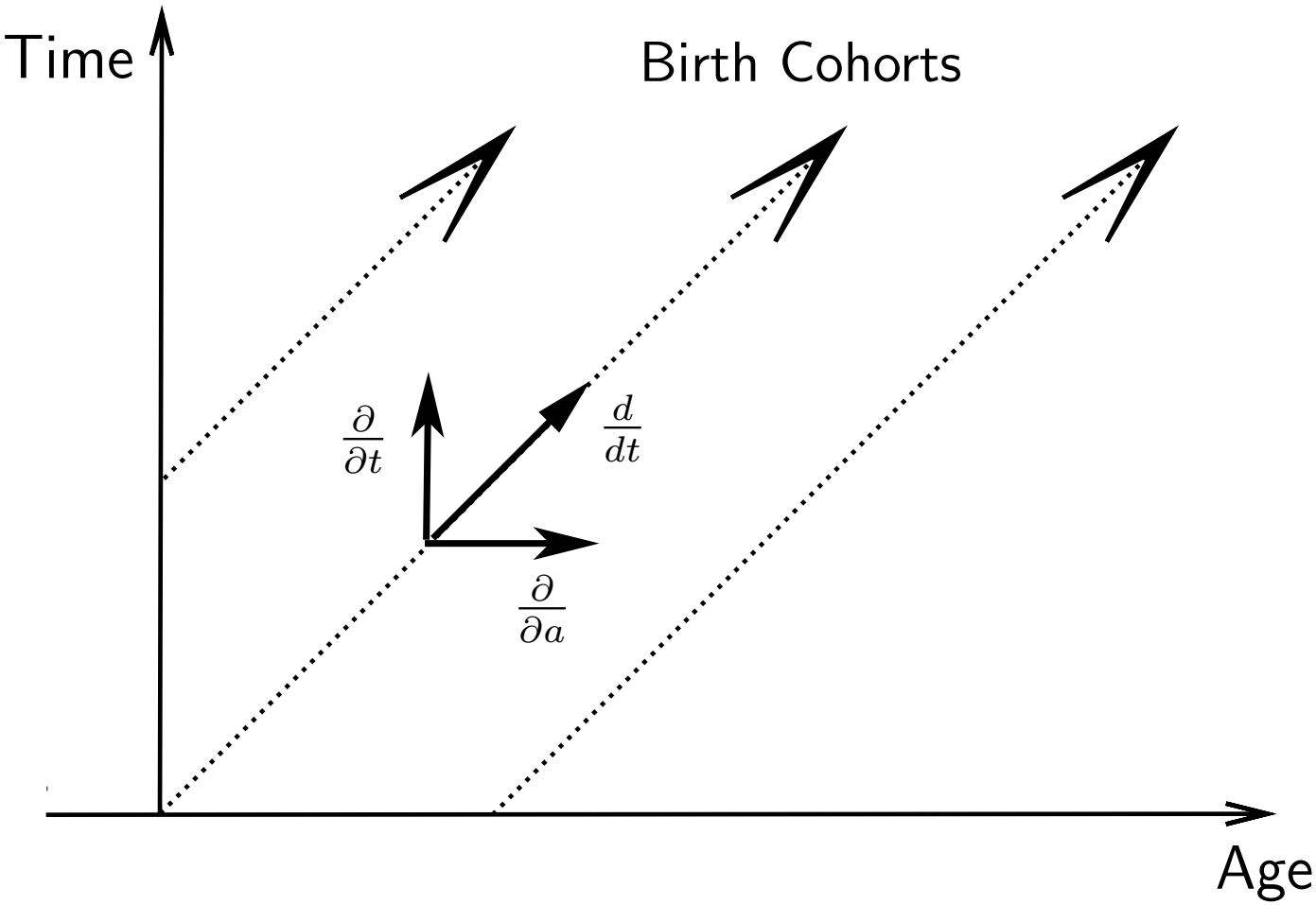
- Follow up (for 8 months) a cohort of E (1500) exposed individuals.
- Record N_{IE} (25) infection events.
- Estimate an exposure time T - a little (about 8) less than 1000 person years

$$I_{\mathrm{est}} = \frac{N_{\mathrm{IE}}}{T} = \frac{25}{992} \approx \frac{25}{1000} = 2.5\%\ p.a.$$

'Synthetic' Cohort

- Given age (and preferably also *time*) structured prevalence data
  - Need to get external *excess mortality* estimate
  - Combine with prevalence *trend* estimate
  - Obtain Incidence estimate
  - (Brunet and Struchiner, Williams et al, Hallett et al, Brookmeyer and Konikoff)
  - **Mahiane et al Plos ONE 2012**
- Note: Considerable Complexity in optimization, power/uncertainty analysis

Birth Cohorts



$$\frac{d}{dt}\,(\text{Birth Cohort}) = \left(\frac{\partial}{\partial a} + \frac{\partial}{\partial t}\right)(\text{Population Density})$$

'Fitting' Models

- Given a general mix of country/region level data
  - Antenatal Clinic Data
  - Household Survey
  - AIDS Indicator Survey
  - Demographic and Health Survey
  - Burden of disease estimates
  - Life Tables
- Fit your best/favourite model and read off the fitted incidence
- Note:
  - Parametric assumptions unavoidable
  - Difficulty in weighting diverse sources of data
  - No formal uncertainty estimation
  - No validation

Instantaneous Population State

- Let persistence in the infected state have a mean duration D.
- For example

$$\dot{N}_{\mathrm{I}} = IE - \frac{1}{D} N_{\mathrm{I}}$$

- At equilibrium, the number of currently infected individuals is

$$N_{\mathrm{I}} = EID$$

- Rearrange:

$$I_{\mathrm{est}} \approx \frac{N_{\mathrm{I}}}{ED}$$

- Note: Same basic structure as longitudinal estimate!

Cross Sectional Study Design

- Let mean persistence in the infected state $D = 1$ week (≈ 0.02 years).
- 'Survey' 100 000 civil servants
- 4000 have flu on a given date

$$I_{\text{est}} \approx \frac{4000}{96,000 \times 0.02} = 208\% \ p.a.$$

- (which may mean a 10% Risk of infection in a month long flu outbreak)

When Survival in the Infected State is Long.

- If D is long $\Longrightarrow$ an uninformative, widely smoothed, 'weighting'.
- Survival post infection is not well known, and it is changing (ARV's).
- So - can we use a transient Recent Infection state?

Cross Sectional HIV estimate

- Survey 3,000 Individuals
- Prevalence 10%
- Mean duration of recent infection of test is half a year
- 15 'recent infections' detected

$$I_{\text{est}} = 1.1\%\ p.a. \quad (0.6 - 1.7)$$

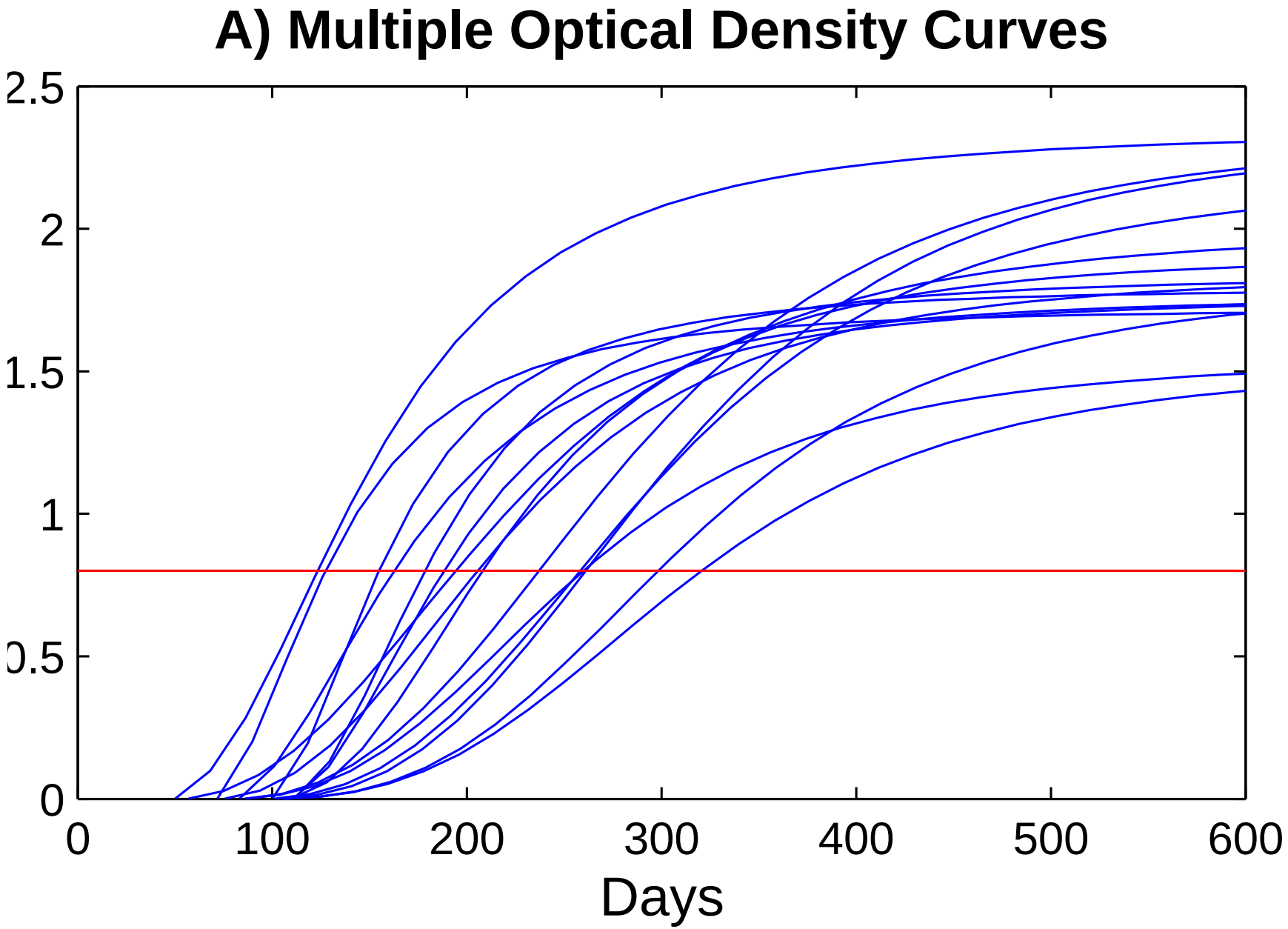
Outline

Can Recent Infection be Defined by a Fixed Time?

- In principle, Yes - **BUT**
- As a purely time based definition, it requires an omniscient investigator
- No conceivable study design uses this level of information
- If you had a perfectly discriminating biomarker - how would you know?
  - (Think seroconverter cohort sampling intervals)

Can Recent Infection be Usefully Defined?

- Need a biomarker which correlates well to a meaningful time scale
- Many biomarkers evolve post infection to provide heuristic correlates of 'recent infection'
- Need to understand what makes a biomarker useful for surveillance


A) Multiple Optical Density Curves
2.5
2
1.5
1
0.5
0
0
100
200
300
400
500
600
Days

Talking Points

- There is always inter-subject variability
- There is always a dynamic range of some kind - and hence the need for a threshold
- 'Raising the threshold' eventually, always breaks the concept of 'recent':
- Just measurement noise alone means there is always some FRR
- Choosing assay conditions and threshold **is everything**

What Matters for Incidence Surveillance?

- Only two things really matter:
  - Minimization of bias (maximization of accuracy)
  - Minimisation of variance / Coefficient of Variation // (maximization of precision)

- In cross sectional surveillance of biomarkers, accuracy is linked to:
  - Good knowledge of the test properties
  - Unbiased sample frame
  - Use of correct analytical/statistical methodology

  i.e. **NOT** the actual test properties themselves

- Precision, in turn is crucially linked to:
  - Long Mean Duration of Recent Infection
  - Low False Recent Rate

i.e. the actual test properties themselves!

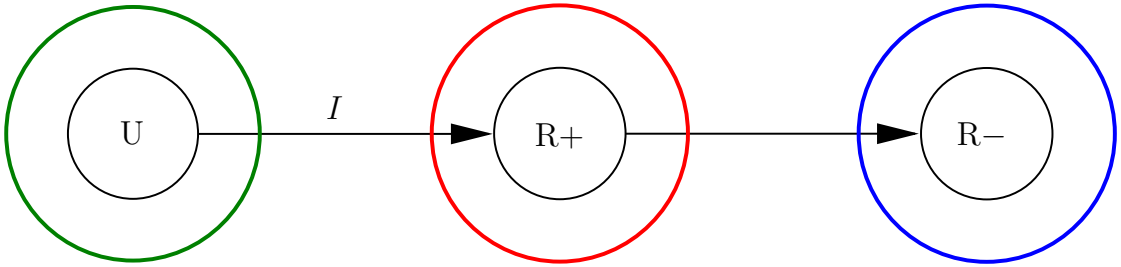
Recent Infection Defined **Directly** by Biomarkers

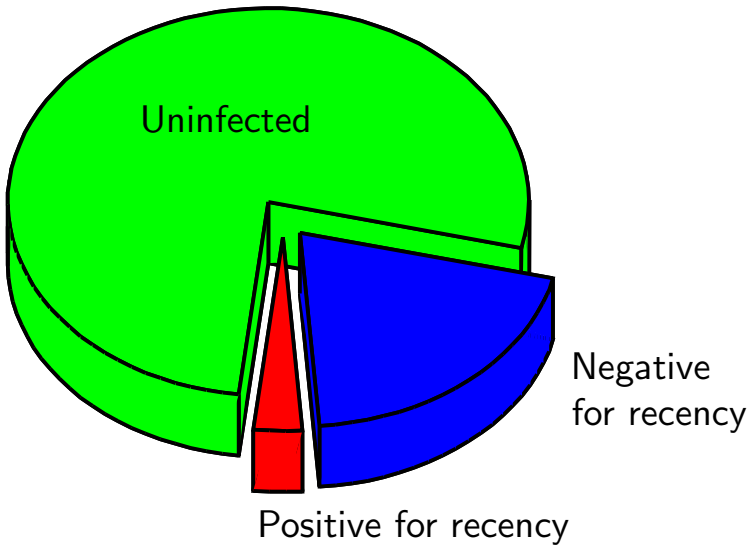
- Now there is, *a priori*, no fixed time in the definition
- Lack of a diagnostic gold standard - the biomarker is everything!
- *Something like* Survival Analysis gives mean recency duration
- Recover the naive estimator
- Explicit formula for implicit temporal averaging/weighting

$$\hat{I} = \frac{\int IW}{\int W}$$

- (No need for equilibrium assumptions!)

Simple Estimator based on knowledge of Idealised Recency



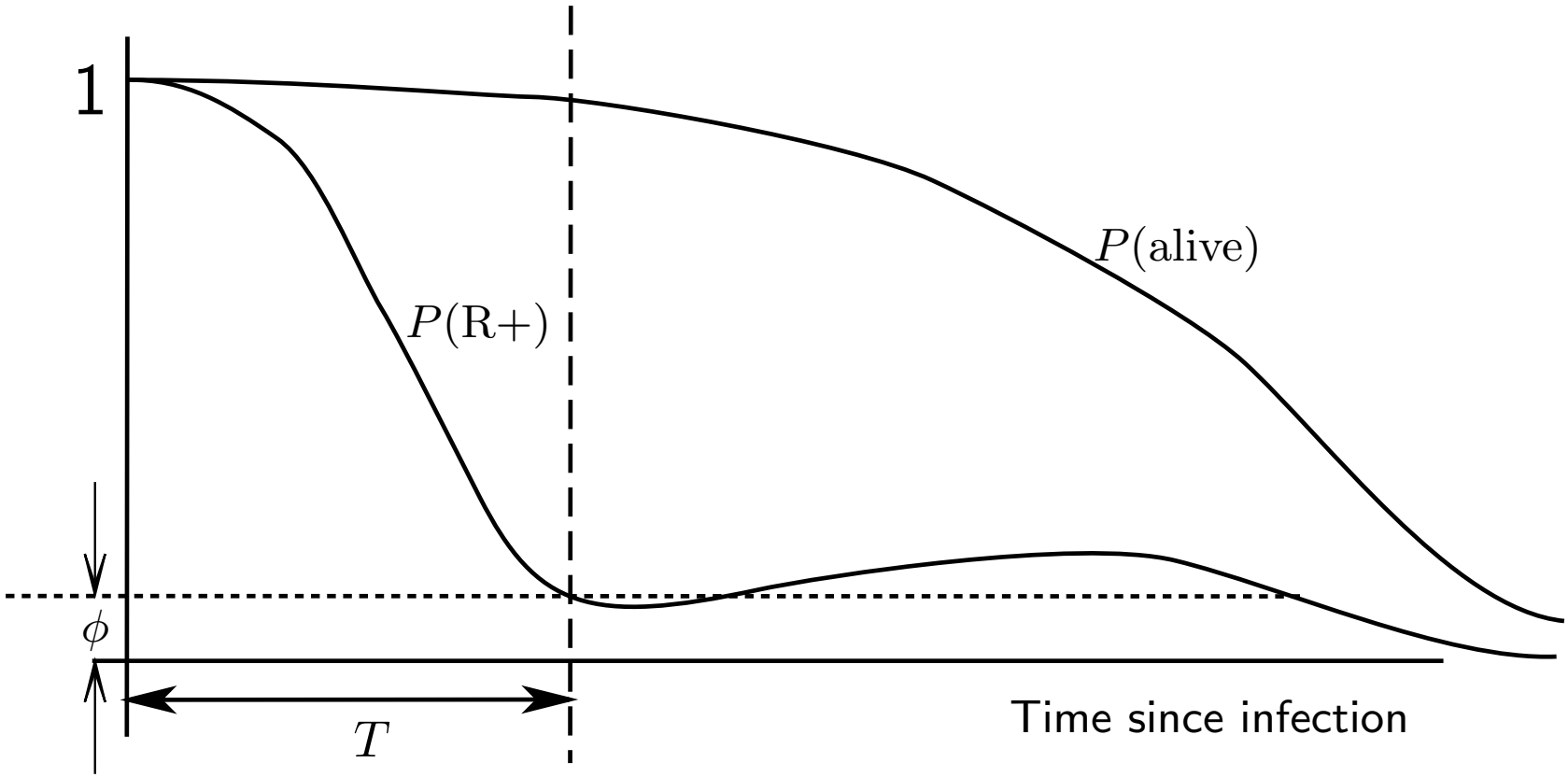


Simple incidence estimator:

$$I_{\text{est}} = \frac{N_{\text{R+}}}{N_{\text{U}}\tau_{\text{R}}}$$

Summarising Recency Test Performance into **Two** parameters

(MDRI Ω and False Recent Rate ϵ)



Talking Points

- There is some persistent disagreement about use of T
- (but actually a storm in a T cup)
- Choosing T requires some care but is not a conventional 'optimization'
- (but choosing a threshold is a conventional, context specific, optimization)
- MDRI, importantly, contains (or should contain) mainly biology
- FRR will vary, and will inevitably be imperfectly known
- FRR should be small enough to:
  - Not imply a huge effect
  - Be acceptably approximated with low precision

Estimating a Mean Duration of Recent Infection

- Recruit newly infected individuals
- Follow them up frequently
- Apply full recent infection test repeatedly
- Regression of various kinds can estimate mean recency
- Once data is reasonable - rough estimates are easy
- Uncertain infection/seroconversion time is a computational bottleneck

Estimating a False Recent Rate

- Recruit non-newly infected individuals
- If possible, stage by time of infection
- Apply full recent infection test (including VL, etc)
- Estimate FRR 'globally' and by 'risk factors'
- Provide context for 'risk factor' *weighting* (tricky)
- Is there prior information/FRR estimate or this is all we have?

Evaluating Recent Infection Biomarkers - I

- Context matters (stage of epidemic)
- Is new candidate biomarker supplemental or stand-alone?
- How ambitious is the funding programme?
- Can devise a variety of estimates and hypothesis tests on performance

Evaluating Recent Infection Biomarkers - II

- Does MDRI Confidence Interval include/exclude critical values?
- Does FRR Confidence Interval include/exclude critical values?
- Is MDRI/FRR superior/non-inferior to other products (accounting for uncertainty in those)
- Given context, is CoV of Incidence Estimate, superior, non inferior to other products, or targets
- How orthogonal are biomarkers