Modern Concepts in Incidence Estimation

Alex Welte

SACEMA

DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis

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

Outline

Longitudinal Study Design ('Real' Cohorts)

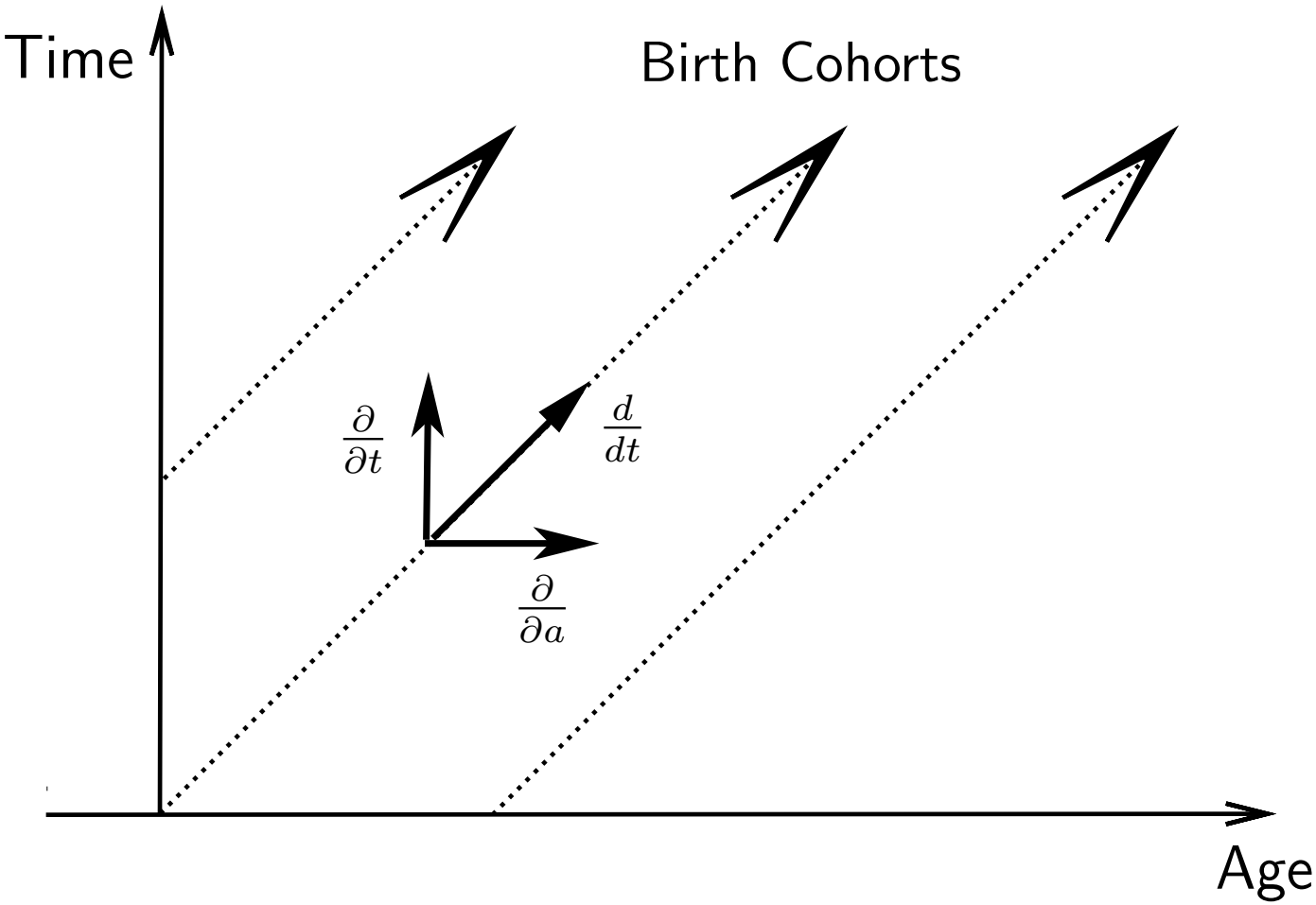
- Follow up (for 8 months) a cohort of E (1500) exposed individuals.
- Record N_{IE} (25) infection events.
- Estimate an exposure time T - a little (about 8) less than 1000 person years

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

'Synthetic' Cohort

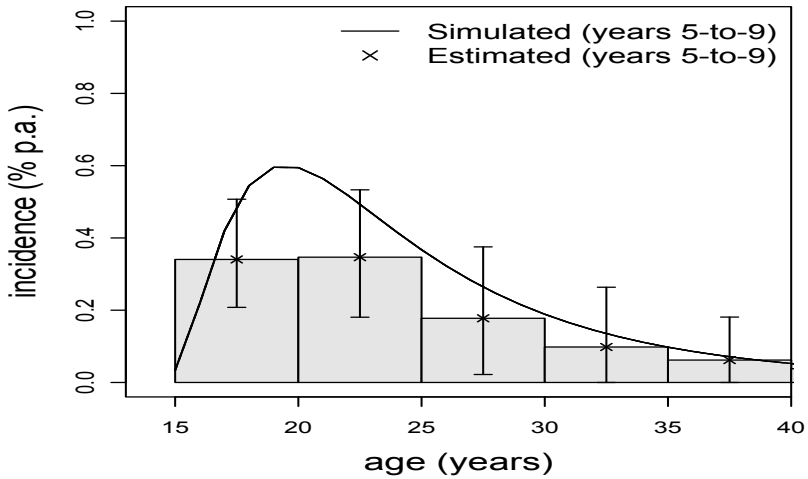
- Given age (and preferably also *time*) structured prevalence data
  - Need to get external *excess mortality* estimate
  - Combine with prevalence *trend* estimate
  - Obtain Incidence estimate
  - (Brunet and Struchiner, Williams et al, Hallett et al, Brookmeyer and Konikoff)
  - **Mahiane et al Plos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

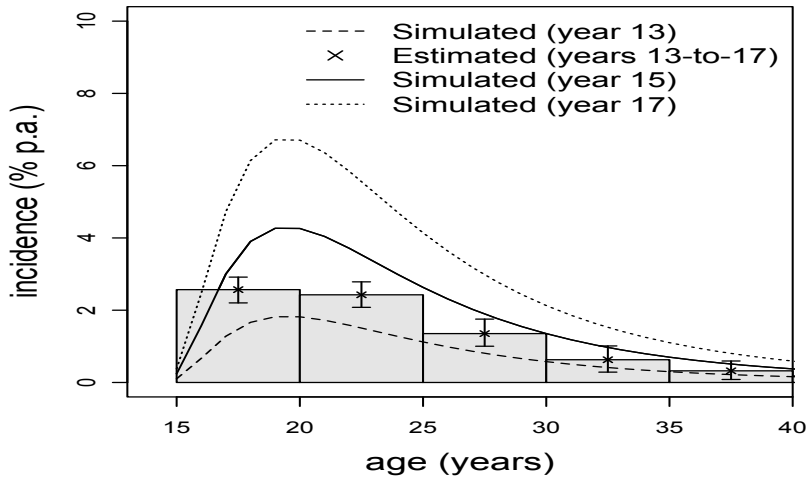
Birth Cohorts

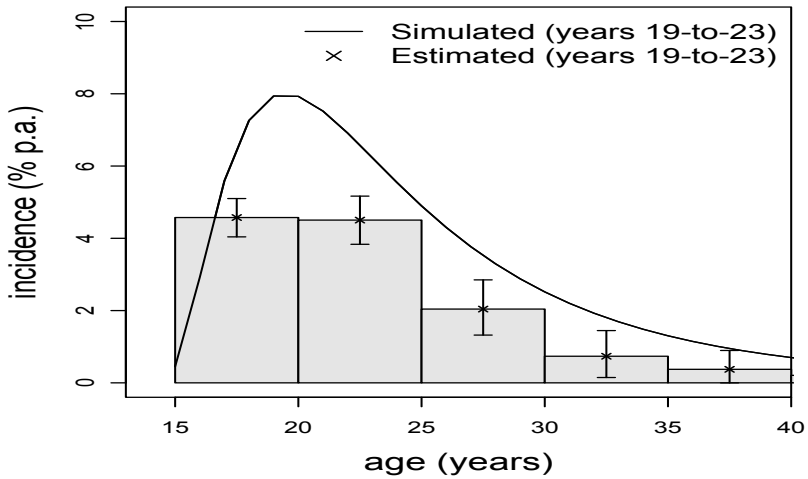


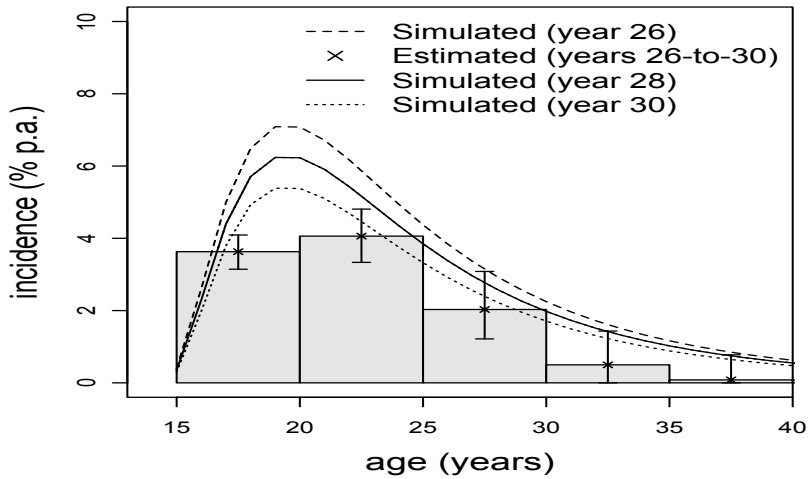
$$\frac{d}{dt}\,(\text{Birth Cohort}) = \left(\frac{\partial}{\partial a} + \frac{\partial}{\partial t}\right)(\text{Population Density})$$

5 year Bins

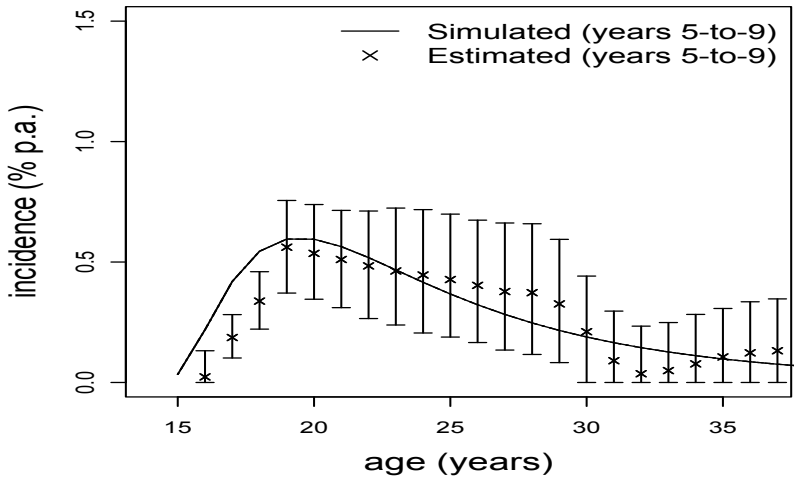

Simulated (years 5-to-9)
Estimated (years 5-to-9)
incidence (% p.a.)
age (years)

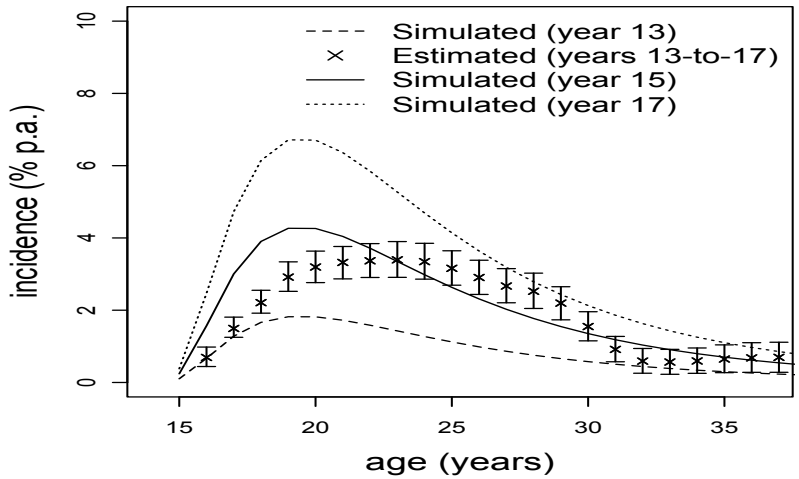

Simulated (year 13)
Estimated (years 13-to-17)
Simulated (year 15)
Simulated (year 17)
incidence (% p.a.)
age (years)

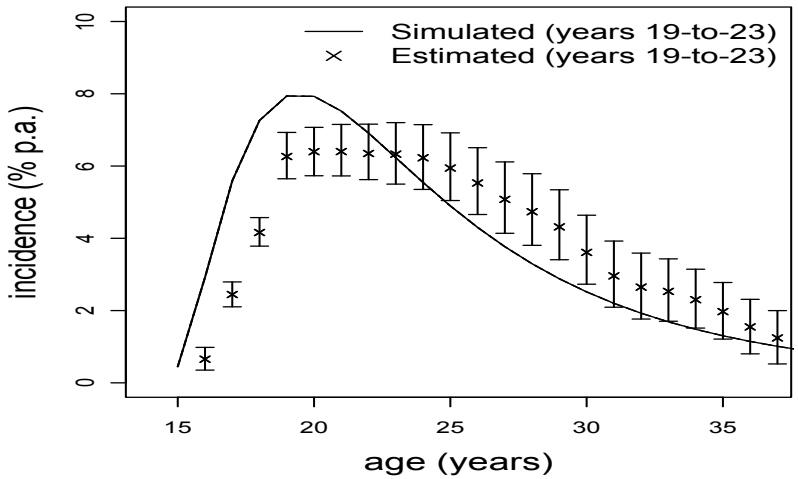

Simulated (years 19-to-23)
Estimated (years 19-to-23)
incidence (% p.a.)
age (years)

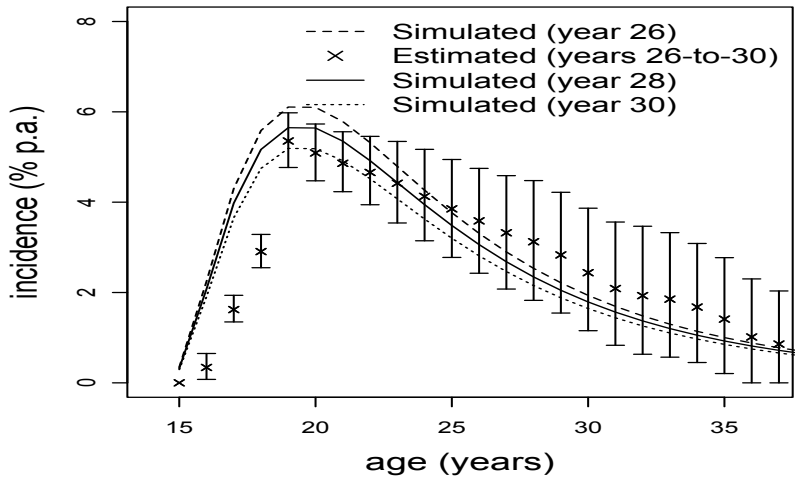

Simulated (year 26)
Estimated (years 26-to-30)
Simulated (year 28)
Simulated (year 30)
incidence (% p.a.)
age (years)

Age-Smoothing Prevalence, but averaging over time

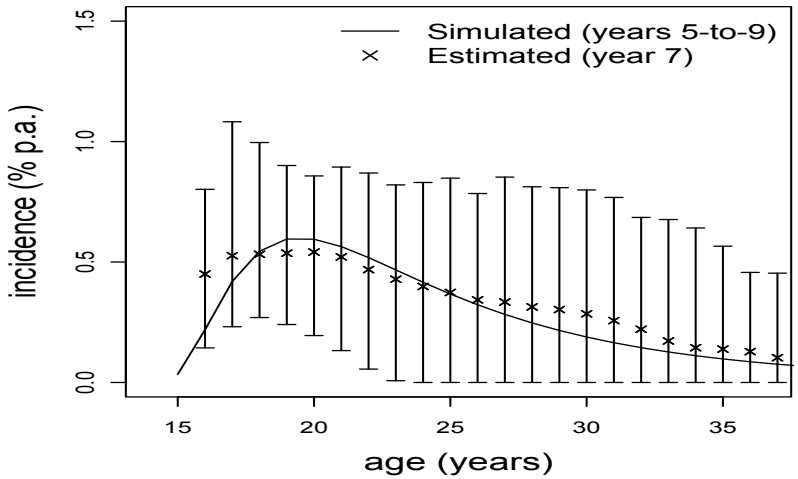


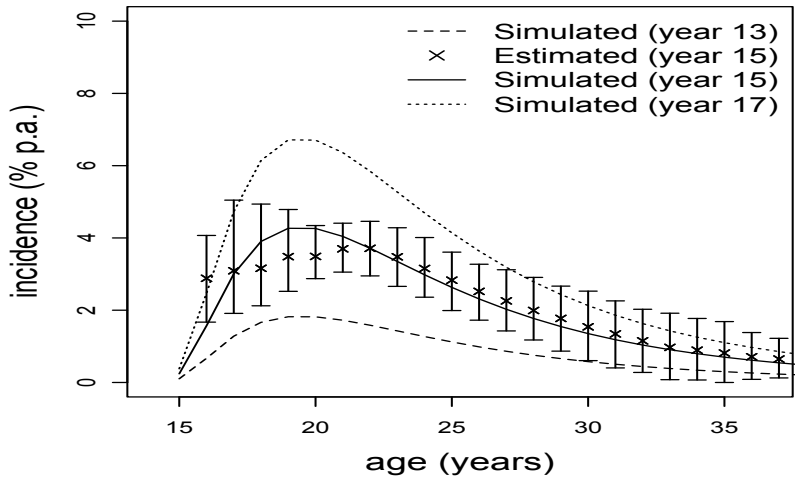


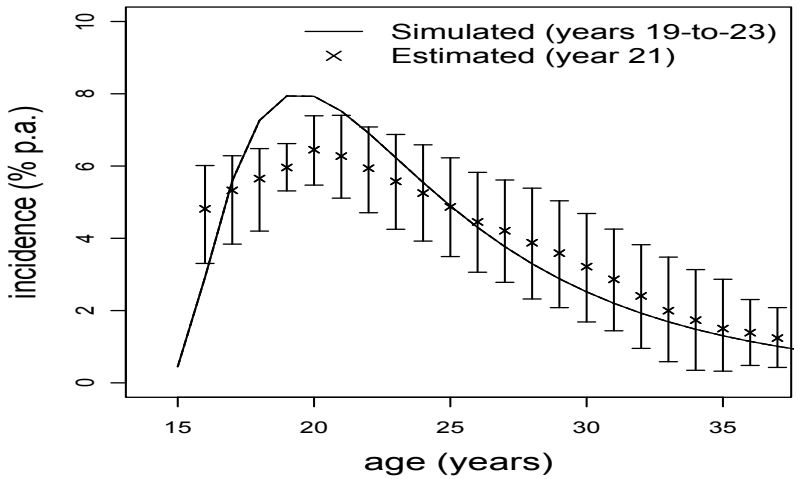


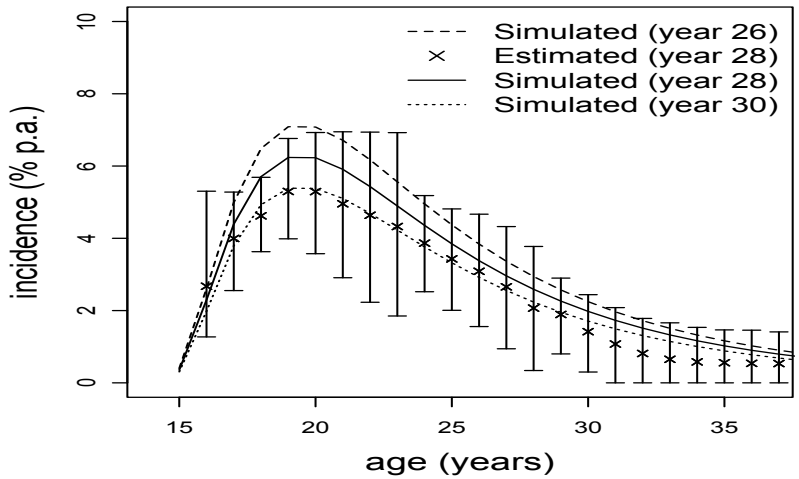


Fully Instantaneous Formulation









'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

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

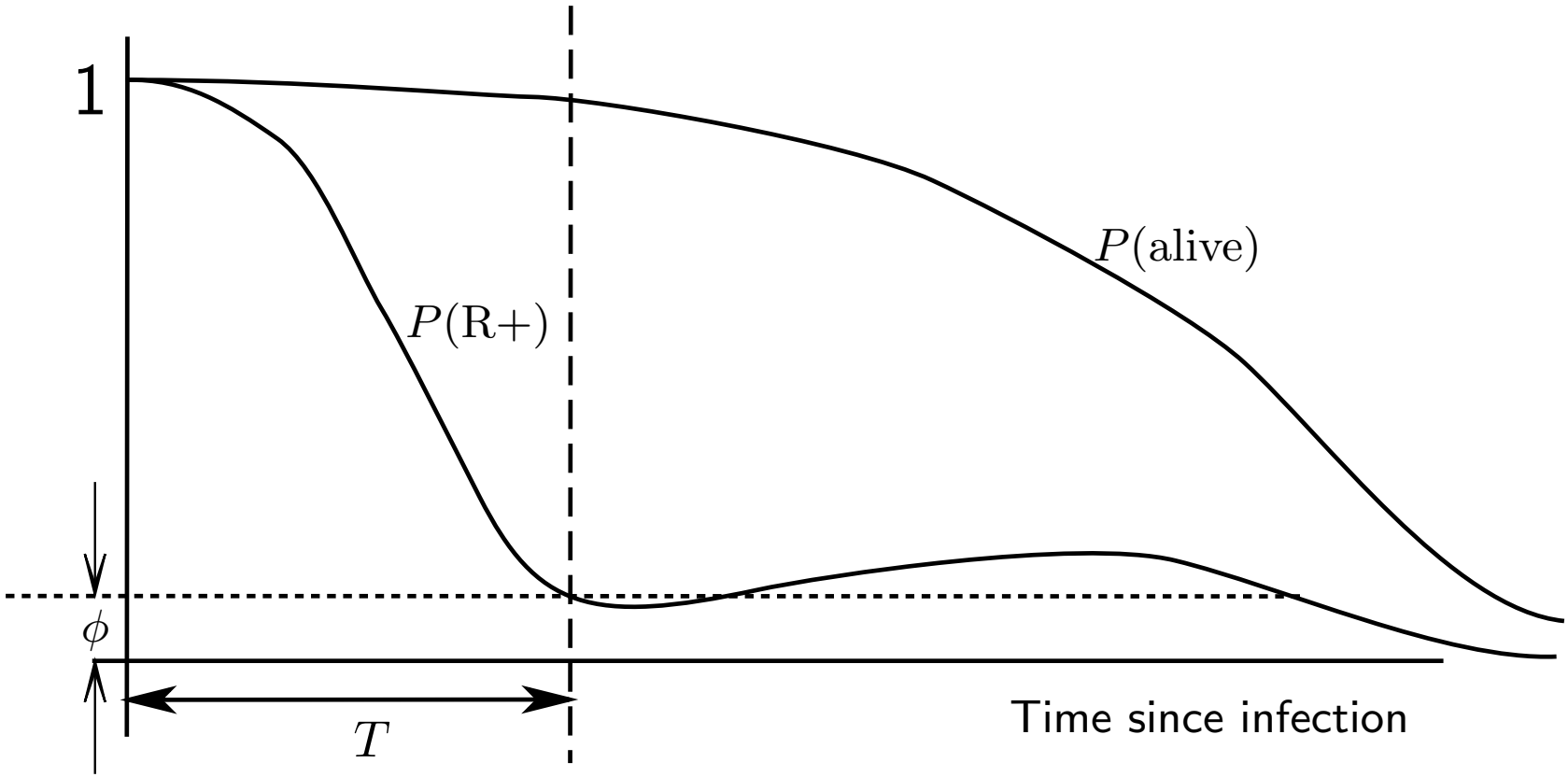
Outline

Can Recent Infection be 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

Summarising Recency Test Performance into **Two** parameters

(MDRI Ω and False Recent Rate ϵ)



Talking Points

- 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