

Inferring Rates From Instantaneous State Variables

Application To HIV Incidence Estimation

Alex Welte, Tom McWalter



South African Centre for Epidemiological Modelling and Analysis (SACEMA),
Stellenbosch University



Computational and Applied Mathematics, University of the Witwatersrand

Basic Epidemiological Indicators

- Reliable estimation of **prevalence** and **incidence** (even more so) are central to the determination of epidemiological trends.
- 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 is supposedly a cohort follow-up survey (“direct observation”).
- It's not all gold:
 - Expensive and logistically complex to run
 - Time consuming
 - Biased due to intervention and loss to follow-up
 - Not free from intrinsic limitations under transient dynamics
- 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

- 1 Introduction
- 2 Measuring Incidence
- 3 Survival Analysis with Immortality and Resurrection
- 4 Statistical Issues
- 5 Conclusion

(Almost Consistent) Colour Coding

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

Outline

- 1 Introduction
- 2 Measuring Incidence
 - Ongoing Observation
 - Instantaneous Observation
- 3 Survival Analysis with Immortality and Resurrection
- 4 Statistical Issues
- 5 Conclusion

What is a rate?

- In the real world, **there are no rates** - there are only **events**.
- Any measurement of a rate has at least an implicit **smoothing/weighting** scheme.
- In a **continuum model world**, we can define $I(t)$:

$$dN_{\text{EI}} = I(t)E(t)dt$$

where

- $N_{\text{EI}}(t)$ is the **N**umber of **E**ver **I**nfected individuals
- $E(t)$ is the **E**xposed population.

Even in this case, inference from incomplete information is subtle.

Observing a Continuous Population over Time

- Observing a population over an **infinitesimal** time slice dt :

$$I(t) = \frac{1}{E} \frac{dN_{EI}}{dt}$$

- In a cohort of exposed individuals, which varies as $E(t)$,

$$N_{EI} = \int IE dt \quad \text{and} \quad T = \int E dt$$

give the total number of **infection events** and **exposure time**, and we can determine the weighted incidence:

$$I_w = \frac{\int IE dt}{\int E dt} = \frac{N_{EI}}{T}$$

which is the naive result, reinterpreted **away from equilibrium**.

The Ideal Cohort

- Consider the case of instantaneous recruitment of an exposed cohort, followed up, without loss for a time Δt .

$$I_W = \frac{N_{EI}}{T}$$

where the **weighting function**

$$W(t) = E(t) = S_U(t) = \exp\left(-\int_0^t I(s)ds\right)$$

is the survival in the uninfected state and the **exposure time** is

$$T = E(0) \int_0^{\Delta t} S_U(s)ds = E(0)\Delta t \left[1 + O\left(\frac{\Delta E}{E}\right) \right]$$

Observing a Continuous Population at One Time Point

- If post infection state is described by $S_I(t)$, with a mean duration D , the number of **currently** infected individuals is

$$N_I(t) = \int_{-\infty}^t I(s)E(s)S_I(t-s)ds$$

- If at least E is constant, we can obtain a **weighted average**

$$I_W(0) = \frac{\int_{-\infty}^0 I(t)W(t) dt}{\int_{-\infty}^0 W(t) dt} = \frac{N_I(0)}{E D}$$

where $W(t) = E(t)S_I(-t)$ which is **proportional to the probability**, that individuals are:

- 1 available for being infected at time $t < 0$, **AND**
- 2 still living and infected at time 0 if they are infected around time t

When the Exposed Population is Non-Constant

- If $E(t)$ is not constant, we can still write

$$I_w = \frac{\int_{-\infty}^0 I(t)W(t) dt}{\int_{-\infty}^0 W(t) dt}$$

but **we cannot simplify the denominator.**

- Let $E(t) = E_0 + E_1t + E_2t^2 + \dots$

$$I_w = \frac{N_I(0)}{E_0\mathbb{E}[\tau_I] - \frac{E_1}{2}\mathbb{E}[\tau_I^2] + \dots} = \frac{N_I(0)}{E_0D} \left[1 + O\left(\frac{\Delta E}{E}\right) \right]$$

which has the **same structure as in the case of the ideal cohort!**

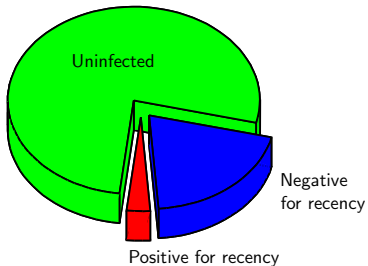
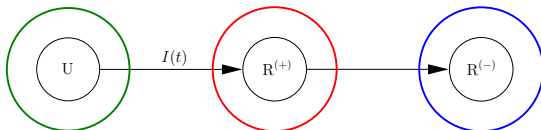
When Survival in the Infected State is Very Long.

- If $D = \mathbb{E}[\tau_1]$ is long (as for HIV), what use the weighted incidence?
- Survival post infection is **not well known**, and is **changing**.
- Considerable interest in development of **recent infection tests**.
- All known tests appear to have **anomalous subpopulations** - i.e. where the 'recent infection' biomarker is either **permanent**, or **recurring**.
- Little consensus even on the use of an '**ideal**' test!

Outline

- 1 Introduction
- 2 Measuring Incidence
- 3 Survival Analysis with Immortality and Resurrection**
 - Immortality
 - Resurrection
- 4 Statistical Issues
- 5 Conclusion

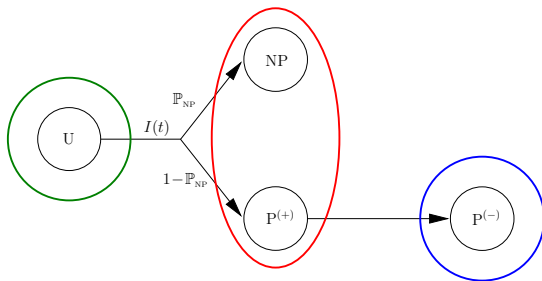
Simple Estimator With Generally Transient Recency



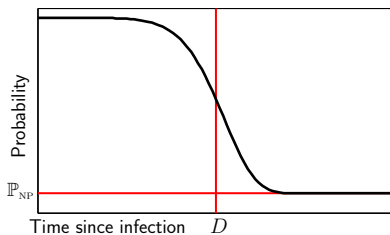
Simple **incidence estimator**:

$$I_{\text{est}} = \frac{N_R}{N_U D}$$

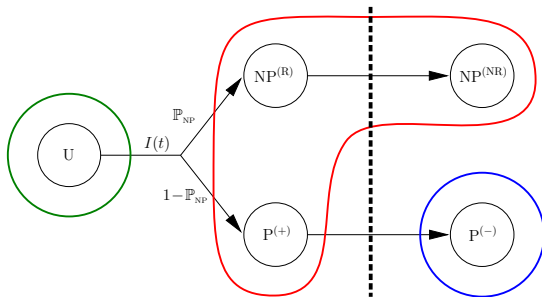
When Recent Infection Lasts Forever



Probability of Testing Positive for Recency



Solution: Endowing Immortals with Unobservable Deaths



- Declare (unobservable) R/NR transition in the **test-non-progressors**
- Use the **same distribution** of waiting times.
- Now a transient **recency** applies to **everyone!**

Reinterpreted Estimator

- Distinguish recent/non-recent from Recency-test-positive/Recency-test-negative.
- Same simple estimator, **if** survival is the same for test-progressors and test-non-progressors

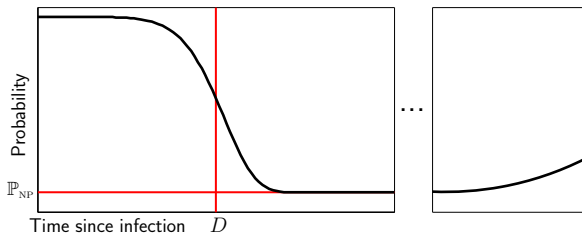
$$I_{\text{est}} = \frac{N_{\text{R}}}{N_{\text{U}}D}, \quad N_{\text{R}} = N_{\text{R}+} - \left(\frac{\mathbb{P}_{\text{NP}}}{1 - \mathbb{P}_{\text{NP}}} \right) N_{\text{R}-}$$

- Recency-test-sensitivity is perfect by definition.
- Need to calibrate recency-test-specificity ($\rho_{\text{R}} = 1 - \mathbb{P}_{\text{NP}}$).

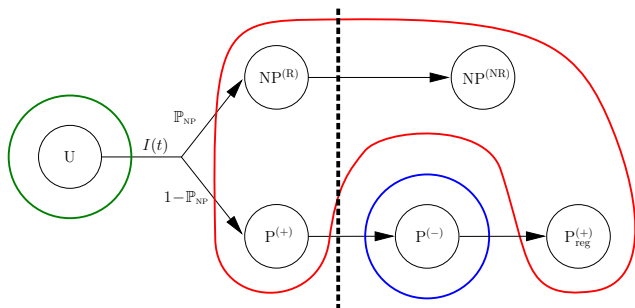
Reverting to Recency-Test-Recent

- Some recency-test-progressors may regress to being Recency-test-positive.

Probability of Testing Positive for Recency



A Further Chance to be 'Falsely' Recency-Test-Positive



- **Recency-test-sensitivity** is still perfect.
- Can still define **specificity**. (Calibration is a bit trickier.)
- Can still use **the same** simple estimator!
- (Also addresses unequal survival for **Recency-Test-(non-)progressors**.)

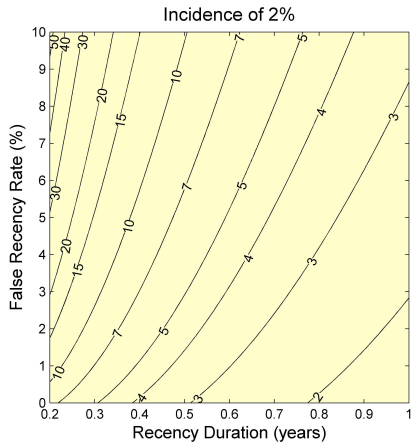
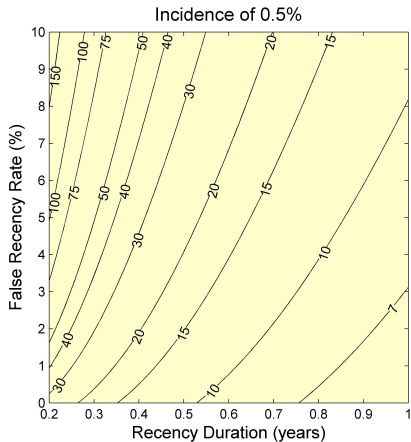
Outline

- 1 Introduction
- 2 Measuring Incidence
- 3 Survival Analysis with Immortality and Resurrection
- 4 Statistical Issues**
 - Power
 - Calibration
- 5 Conclusion

Finite sample sizes

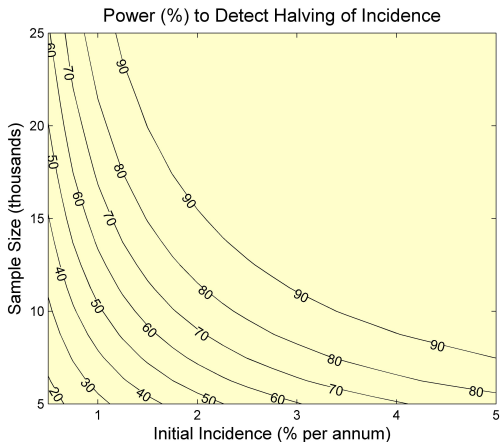
- In practice, we have only **estimates** for population proportions and test characteristic parameters.
- Given point estimates of the calibration, our estimator is **likelihood maximising**.
- Full forward model has survey count uncertainty and calibration uncertainty.

Reproducibility



- Sample sizes (**thousands**) required to obtain a coefficient of variation of 20%, for various **recency test** characteristics.

Trend detection



- Probability of correctly detecting a reduction in incidence between two surveys, using a 'BED-assay-like' recent infection test, **if the incidence actually halved!**

Calibration of Recency Test Characteristics

- Estimation of **recency duration**:
 - **seroconverter** cohorts
 - blood donors
 - survey populations
- Estimation of **false recency rate**:
 - cohorts of **non-recently** infected individuals
 - survey populations with **repeat positive** HIV results

Outline

- 1 Introduction
- 2 Measuring Incidence
- 3 Survival Analysis with Immortality and Resurrection
- 4 Statistical Issues
- 5 Conclusion**

Summary

- No **fundamental** obstacle to estimating rates from imperfect knowledge of instantaneous population states.
- Considerable depletion of statistical power with non-ideal characteristics of a test for recent infection.
- The methodology outlined mainly serves to:
 - power surveys
 - produce consistent estimates
 - characterise potential recent infection tests

Challenges

- Statistical power poses a **trade-off** for recent infection tests:
 - Large counts - long recency durations
 - High specificity - near universal progression out of test-recency
- Excellent tests are not currently available for HIV.
- Better analysis is no substitute for better data.

References

- T. McWalter and A. Welte, *Relating recent infection prevalence to incidence with a sub-population of assay non-progressors*, J. Math Biol, DOI 10.1007/s00285-009-0282-7, (2009)
- T. Bärnighausen et. al., *HIV Incidence in Rural South Africa: Comparison of Estimates from Longitudinal Surveillance and Cross-Sectional cBED Assay Testing*, Plos-ONE, 3(11): e3640, (2008)
- T. McWalter and A. Welte, *A Comparison of Biomarker Based Incidence Estimators*, Plos-ONE, in press

Acknowledgements

- John Hargrove, Reshma Kassanjee ([SACEMA](#))
- Till Bärnighausen ([Harvard](#)), Mike Busch ([Blood Systems Research](#)), Tim Hallet ([Imperial](#)), Oliver Laeyendecker ([NIH](#))
- [WHO](#) Incidence Assays Working Group
- Canadian International Development Agency