Incidence from 'Recent Infection' Prevalence

Alex Welte
(with Norman Ives and Tom McWalter)

School of Computational and Applied Mathematics,
University of the Witwatersrand

and

DST/NRF Centre of Excellence in Epidemiological
Modelling and Analysis (SACEMA)

DIMACS/SACEMA Workshop
25 July 2007

Outline of Talk

Motivation

- Determination of epidemiological trends requires knowledge of both *incidence* and *prevalence*.
- This is required for:
  - Allocating resources
  - Evaluating interventions
  - Planning cohort based studies

Measuring Incidence is Problematic

- 'Gold Standard' of cohort follow up is:
  - Expensive
  - Slow
  - Not free from bias
    - Intervention
    - Loss to follow up

Alternatives to 'Gold Standard'

- Incidence from snapshots of 'recent infection' prevalence
  - Many pasts can produce one present
  - Problem of calibrating 'recent'
  - Forever 'recent'?
  - Sample sizes

Stages of Infection

- As observed through *two* assays of different sensitivity, infected individuals experience the following events:

0. The moment of infection, at t
1. Infection becomes detectable by assay one, at $t + t_1$
2. Infection becomes detectable by assay two, at $t + t_2$
3. Death, from whatever cause, at $t + t_3$

Counting Infectees

- All historically infected individuals

$$N_{\mathrm{i}}(0) = \int_{-\infty}^{0} \lambda(t)\, dt = \int_{-\infty}^{0} I(t) N_{\mathrm{s}}(t)\, dt$$

- After infection at time $t < 0$ the probability of a being alive and classified as recently infected at time zero is given by

$$\mathbb{P}\big\{(t_1 < -t) \text{ AND } (t_2 > -t) \text{ AND } (t_3 > -t)\big\} = \int_{0}^{-t} \int_{-t}^{\infty} \int_{-t}^{\infty} \rho(t_1, t_2, t_3)\, dt_3\, dt_2\, dt_1$$

- The probability of a being alive and classified as **not** recently infected at time zero is given by

$$\mathbb{P}\big\{(t_2 < -t) \text{ AND } (t_3 > -t)\big\}$$

Recently and Long Infected Individuals

- Counts of recently and long infected individuals observed in a cross-sectional survey

$$R = \int_{-\infty}^{0} I(t) N_{\mathrm{s}}(t) \mathbb{P}\left\{ (t_1 < -t) \text{ AND } (t_2 > -t) \text{ AND } (t_3 > -t) \right\} dt$$

$$L = \int_{-\infty}^{0} I(t) N_{\mathrm{s}}(t) \mathbb{P}\left\{ (t_2 < -t) \text{ AND } (t_3 > -t) \right\} dt$$

Simplifications

- In order to infer incidence from observables, we need to make numerous simplifications of the general model.
  - Redefine incidence to start at t_1.
  - Factorise $\rho(t_2, t_3)$ into $\rho_{\mathrm{r}}(t_2)\rho_{\mathrm{a}}(t_3)$ (independent events).
  - Use only a few 'moments' of ρ's (preferably just the means).
- Now define the survival times

$$S_{\mathrm{r}}(t) = 1 - \int_0^t \rho_{\mathrm{r}}(s)\, ds$$

$$S_{\mathrm{a}}(t) = 1 - \int_0^t \rho_{\mathrm{a}}(s)\, ds$$

Which Incidence Can Possibly Be Measured?

If $I(t)$ is not constant then define a *weighted* average

$$I_{\mathrm{w}} = \frac{\int_{-\infty}^{0} I(t)W(t)\,dt}{\int_{-\infty}^{0} W(t)\,dt}$$

where $W(t)$ is a statistical weight implied by the sampling methodology and formal assumptions. It will be useful to consider

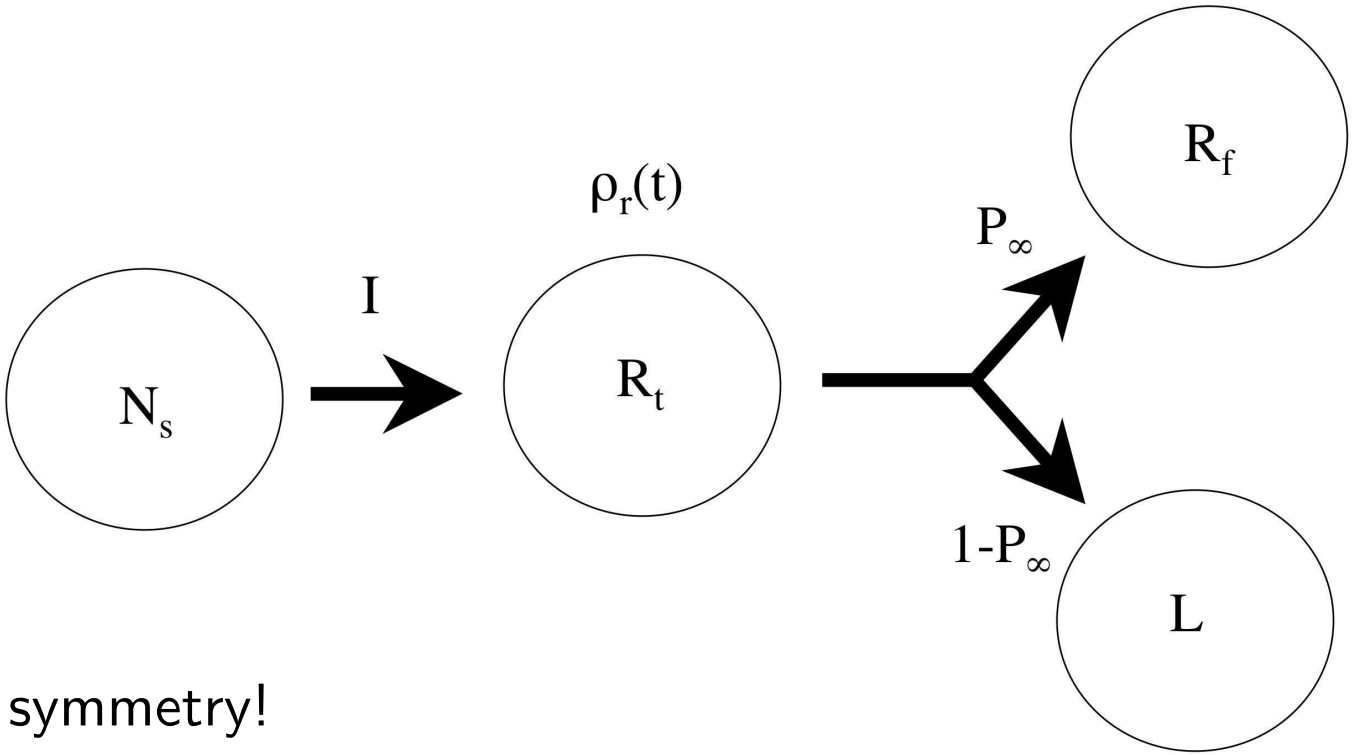
$$W(t) = S_{\mathrm{a}}(-t)S_{\mathrm{r}}(-t) = S_{\mathrm{ar}}(-t)$$

in which case

$$\int_{-\infty}^{0} W(t)\,dt = \int_{-\infty}^{0} S_{\mathrm{ar}}(-t)\,dt = \mathbb{E}\left[\tau_{\mathrm{r}}\right]$$

When Recent is Forever

- BED assay may be best bet for a practical definition of 'recent'.
- **But**: not everyone leaves this category.



- Note symmetry!

'General Practical' Model

- Counts of infection categories

$$R_{\mathrm{t}} = \int_{-\infty}^{0} I(t) N_{\mathrm{s}}(t) S_{\mathrm{r}}(-t) S_{\mathrm{a}}(-t)\, dt$$

$$L = (1 - \mathbb{P}_{\infty}) \int_{-\infty}^{0} I(t) N_{\mathrm{s}}(t) (1 - S_{\mathrm{r}}(-t)) S_{\mathrm{a}}(-t)\, dt$$

$$R_{\mathrm{f}} = \frac{\mathbb{P}_{\infty}}{(1 - \mathbb{P}_{\infty})} L$$

- Unbiased estimation of R_{t} is possible!

$$R_{\mathrm{t}} = R_{\mathrm{o}} - \frac{\mathbb{P}_{\infty}}{(1 - \mathbb{P}_{\infty})} L$$

Constant Number of Susceptibles

If $N_{\mathrm{s}}(t) = n_0$ then

$$\begin{aligned} R_{\mathrm{t}} &= \frac{n_0 \int_{-\infty}^{0} I(t) S_{\mathrm{ar}}(-t)\, dt}{\mathbb{E}\left[\tau_{\mathrm{r}}\right]} \mathbb{E}\left[\tau_{\mathrm{r}}\right] \\ &= n_0 I_{\mathrm{ar}} \mathbb{E}\left[\tau_{\mathrm{r}}\right] \end{aligned}$$

Hence

$$I_{\mathrm{ar}} = \frac{R_{\mathrm{t}}}{n_0 \mathbb{E}\left[\tau_{\mathrm{r}}\right]}$$

is an unbiased estimator.

When $N_{\mathrm{s}}(t)$ is Non-Constant

- It is not possible to obtain unbiased estimator for I_{w} with non-constant $I(t)$ and non-constant $N_{\mathrm{s}}(t)$.
- Need to assume specific functional forms for $N_{\mathrm{s}}(t)$ and $I(t)$.
- Can infer up to two parameters in $I(t)$, $N_{\mathrm{s}}(t)$.
- In the case where $I(t) = i_0$ and $N_{\mathrm{s}}(t) = n_0 + n_1 t$ we recover

$$I_{\mathrm{w}} = i_0 = \frac{R_{\mathrm{t}}}{n_0 \mathbb{E}\left[\tau_{\mathrm{r}}\right] - \frac{n_1}{2}\mathbb{E}\left[\tau_{\mathrm{r}}^2\right]}$$

When $N_{\mathrm{s}}(t)$ and $I(t)$ are Both Non-constant

- Even the simple case of

$$N_{\mathrm{s}}(t) = n_0 + n_1 t \qquad \text{and} \qquad I(t) = i_0 + i_1 t$$

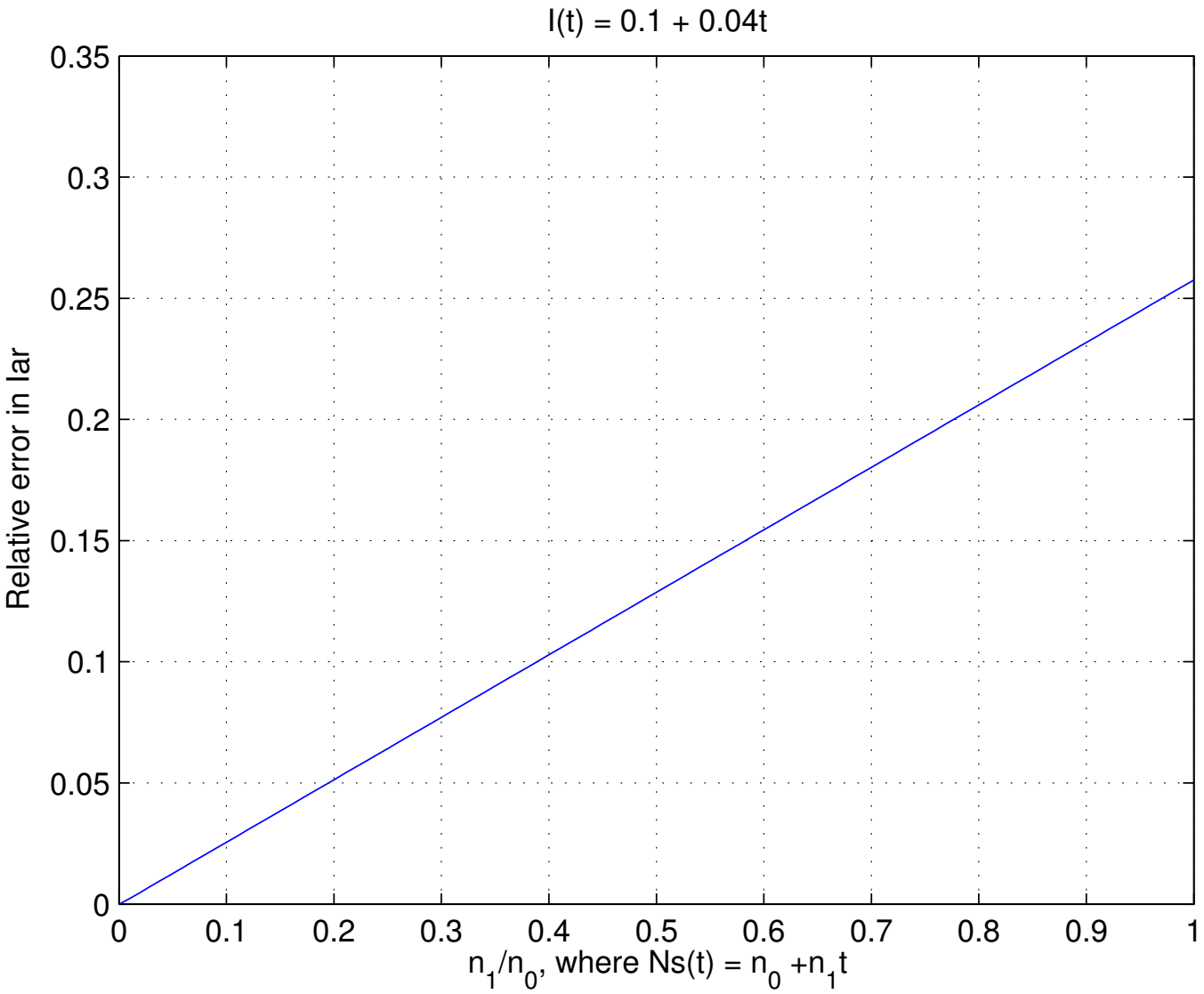
  requires calibration for $\mathbb{E}\left[\tau_{\mathrm{r}}^2\right]$, $\mathbb{E}\left[\tau_{\mathrm{r}}^3\right]$ and $\mathbb{E}\left[\tau_a\right]$, which seems impossible.
- **But**: we can use a simple formula and then see how wrong it is when we violate the assumptions under which it would be correct.

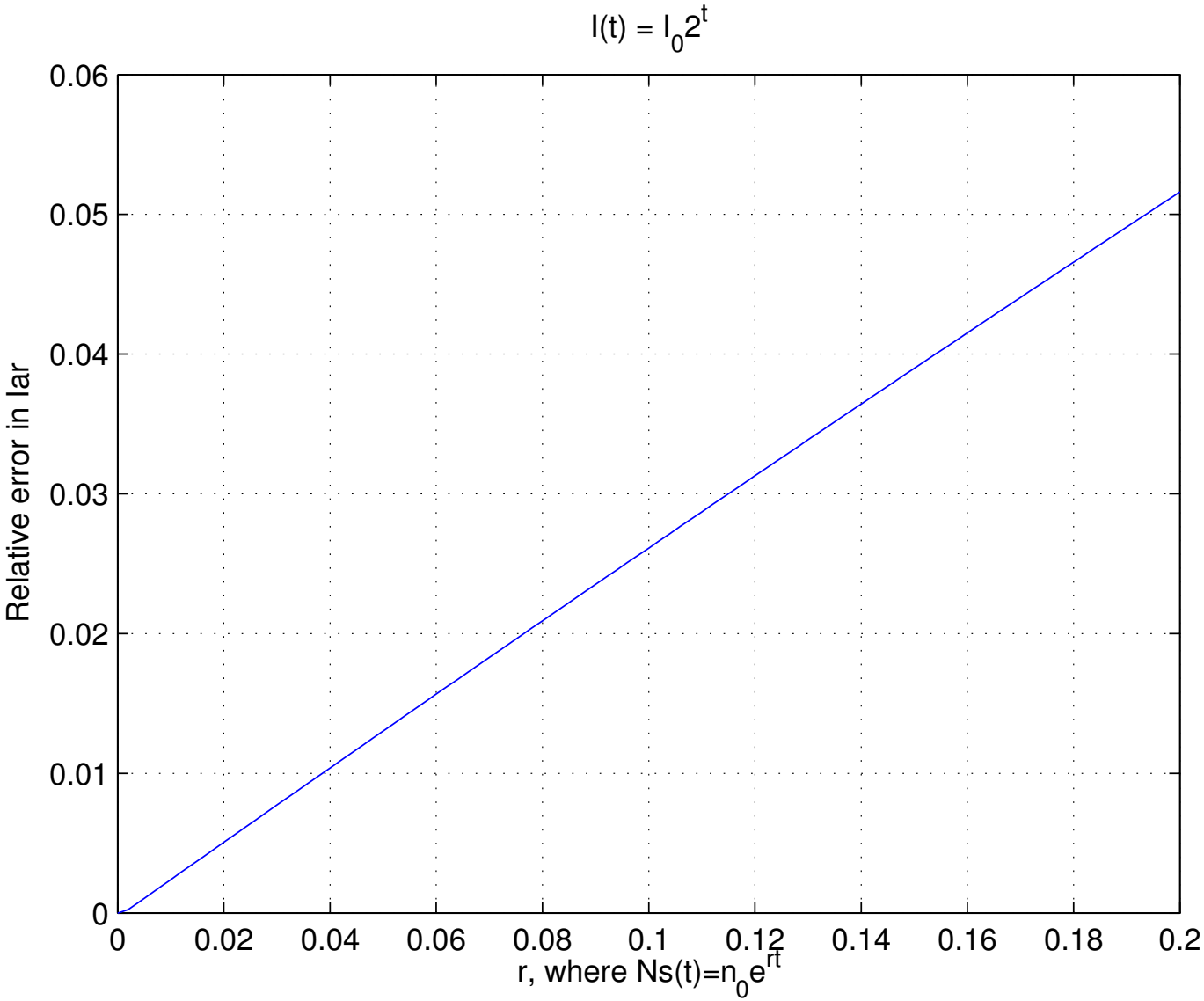
Closed-Form Calculation of Bias

- Declare $I(t)$, $N_{\mathrm{s}}(t)$, $\rho_{\mathrm{a}}(t)$ and $\rho_{\mathrm{r}}(t)$.
- Evaluate I_{ar} from knowledge of declared inputs.
- Evaluate naive formula to estimate I:

$$I_{\mathrm{est}} = \frac{R_{\mathrm{t}}}{\mathbb{E}\left[\tau_{\mathrm{r}}\right] N_{\mathrm{s}}(0)}$$

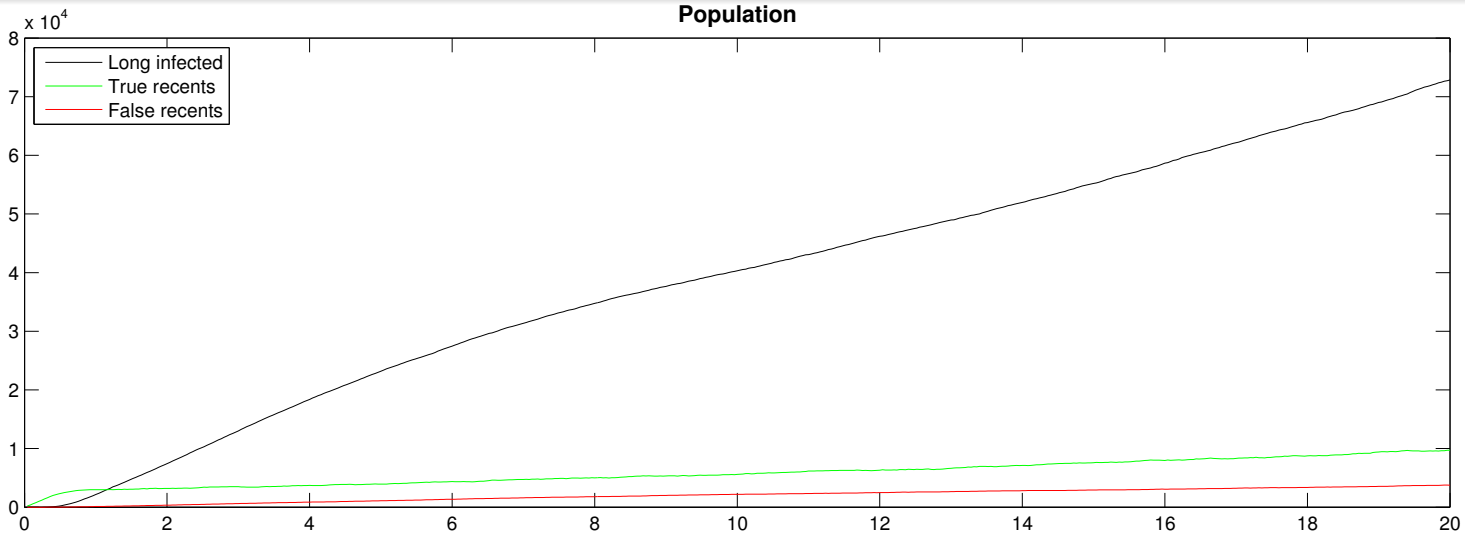
- Compare I_{ar} and I_{est}.
- Note bias as a function of parameters which break assumption in naive formula.

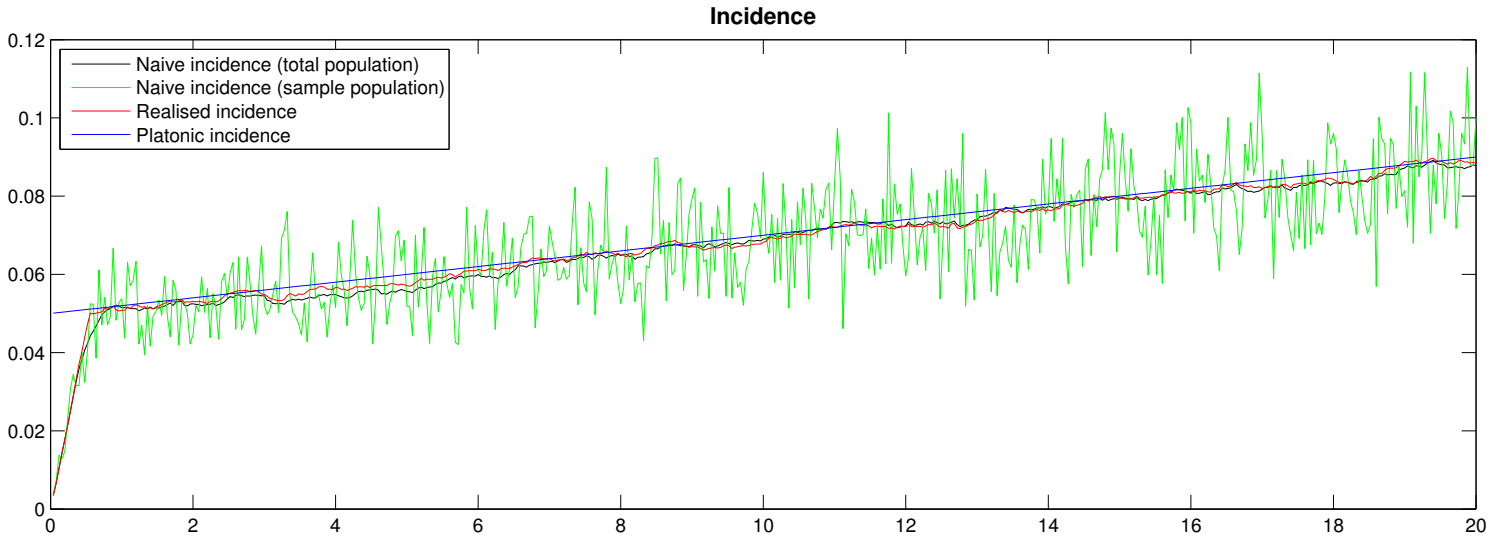

I(t) = 0.1 + 0.04t
Relative error in Iar
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
n1/n0, where Ns(t) = n0 +n1t


I(t) = I0 2^t
Relative error in Iar
0.06
0.05
0.04
0.03
0.02
0.01
0
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2
r, where Ns(t)=n0 e^rt

Matlab Simulations

- Produce infection events according to $\lambda(t) = N_{\mathrm{s}}(t)I(t)$.
- Obtain samples after 'burn-in' time.
- Confirms analytical results.
- Can explore sample-path specific I_{w}.
- Can combine sample size effects with sample path effects.
- Future implementation of hierarchical Bayesian inference.


Population
x 10^4
Long infected
True recents
False recents
0
1
2
3
4
5
6
7
8
0
2
4
6
8
10
12
14
16
18
20


Incidence
Naive incidence (total population)
Naive incidence (sample population)
Realised incidence
Platonic incidence
0
0.02
0.04
0.06
0.08
0.1
0.12
0
2
4
6
8
10
12
14
16
18
20

How Big Must Surveys Be?

- Precision limiting step in a survey is the count of recent infections.
- Need to find tens of recent infections for reasonable precision.
- Example: suppose we wish to distinguish $I = 0.05$ confidently from $I = 0.04$, then:
  - BED definition of recent leads to $N \sim 5000$.
  - RNA+/AB- definition of recent leads to $N \sim 100000$.

Determining $S_{\mathrm{ar}}(t)$ and $\mathbb{P}_\infty$

- Most important parameter is $\mathbb{P}_\infty$.
- Need to observe this 'failure to progress' several tens of times in a calibration study.
- Need to estimate the mean of $S_{\mathrm{ar}}(t)$ *in a cohort censored for progression*.
- Non equilibrium formulas require estimation of long term survival times or short term standard deviations.

Long Term Incidence Surveillance

- Can't run from the need for incidence surveillance.
- There is no magic bullet.
- BED type definition is perhaps the best short term option.
- Just discussed 'one-shot' methods.
- Need to set up open ended calibration.
- Need to manage multi-survey meta analysis to:
  - construct informative 'priors';
  - calibrate demographic/epidemiological parameters.