Serocalculator: An open-source R package for estimating seroincidence from cross-sectional serosurveys

2024 IDM Annual Symposium

October 01, 2024

Kristen Aiemjoy, PhD & Ezra Morrison, PhD Assistant Professors of Epidemiology & Biostatistics

Department of Public Health Sciences

UC Davis School of Medicine

UC**DAVIS** HEALTH

GOAL: Easily and reproducibly translate quantitative antibody responses at the population level into meaningful and accurate epidemiological measures of infection burden



Seroepidemiology & Environmental Surveillance for Enteric Fever (SEES)



Blood culture confirmed

Blood culture performed

Sought care at a surveillance site

Symptomatic infections

All infections (including subclinical)

BILL& MELINDA GATES foundation









THE AGA KHAN UNIVERSITY



Aiemjoy et al, Lancet Infectious Diseases, 2022



Typhoid conjugate vaccines are effective but have yet to be widely adopted

Together We Can Take on Typhoid

Within-host model of antibody dynamics

Received: 18 December 2019 Revised: 9 April 2020 Accepted: 28 April 2020

DOI: 10.1002/sim.8578

RESEARCH ARTICLE

Estimation of seroconversion rates for infectious diseases: Effects of age and noise

seroclassification ai

This article expands models of the service

P. F. M. Teunis¹ | J. C. H. van Eijkeren²

1Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA ²Consultant Mathematical Modelling, Houten, Netherlands

Correspondence



Contents lists available at ScienceDirect Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Linking the seroresponse to infection to within-host heterogeneity in antibody production



EPIDEMIC

P.F.M. Teunis^{a,b,*}, J.C.H. van Eijkeren^c, W.F. de Graaf^d, A. Bonačić Marinović^a, M.E.E. Kretzschmar^{a,e}

* Centre for Infectious Disease Control, RIVM, Bilthoven, Netherlands

^b Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA

In Medicine WILEY

*Statistics and Methodology Group, Information Technology Division, RIVM, Bilthoven, Netherlands

^d Department of Mathematics, Utrecht University, Utrecht, Netherlands

* Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands



0 0

00

• •

(• •)

 $\begin{pmatrix} \circ & \circ \\ & & \end{pmatrix}$

00

 $(\circ \circ)$

00

00

0 0

(• •)

00

...

••

00

00

00

00

(• •)

Within-host model of antibody dynamics



0 0

00

0 0

0 0

 $(\circ \circ)$

စစ

 $(\circ \circ)$

(0 0

00

••

 $\left(\circ \circ \right)$

00

 $\begin{pmatrix} \circ & \circ \\ & & \end{pmatrix}$

00

00

်၀၀ဲ

(• • •

Within-host model of antibody dynamics



0 0

00

0 0

0 0

 $(\circ \circ)$

စစ

 $\left(\circ\circ\right)$

00

် ၀ ၀

 $\left(\circ \circ \right)$

00

 $\begin{pmatrix} \circ & \circ \\ & & \end{pmatrix}$

(• • •

00

(0 0

00

ဴ၀၀ဲ

(• • •

Within-host model of antibody dynamics



်၀၀

00

0 0

ဴ၀၀

(° °)

00

 $(\circ \circ)$

(0 0

် ၀ ၀

• • `

 $(\circ \circ)$

(0 0)

 $\begin{pmatrix} \circ & \circ \\ & & \end{pmatrix}$

0 0

00

(0 0

00

(• •)

Within-host model of antibody dynamics



်၀၀

00

0 0

ဴ၀၀

(° °)

00

 $(\circ \circ)$

(0 0

် ၀ ၀

 $(\circ \circ)$

(0 0)

 $\begin{pmatrix} \circ & \circ \\ & & \end{pmatrix}$

0 0

00

(0 0

00

(• •)

Within-host model of antibody dynamics



Within-host model of antibody dynamics







00

 $\left(\begin{array}{c} \circ \circ \\ \end{array} \right)$

 $\begin{pmatrix} \circ & \circ \\ & & \end{pmatrix}$

00

00

••

 $\left(\begin{array}{c} \circ \circ \end{array} \right)$

(• •)

1 ...

+

 $(\circ \circ)$

(0 0

 $\left(\begin{array}{c} \circ \circ \\ \circ \end{array} \right)$

(• • `

 $\begin{pmatrix} \circ \circ \\ \hline \end{pmatrix}$

00

Within-host model of antibody dynamics





Defining incidence

The **incidence rate** of a disease over a **specific time period** is the rate at which individuals in a population are **acquiring** the **disease** per **person-time at risk**.

Example: if there are 10 new cases of typhoid in a population of 1000 over a one month time period, then the incidence rate for that time period is *"10 new cases per 1000 persons per month"*.

Incidence from an individual's perspective

From the perspective of an individual in the population:

+the incidence rate at a given time point (t) is the instantaneous probability (density) of becoming infected at that time point, given that they are at risk at that time point.
+That is, the incidence rate is a hazard rate.

+Notation: let's use λ_t to denote the incidence rate at time t_s

Cross-sectional antibody surveys

+We recruit participants from the population of interest.

- +For each survey participant, we measure antibody levels (Y) for the disease of interest
- +Each participant was **most recently infected** at some time (*T*) **prior** to when we measured their antibodies.

+*T* is a **latent, unobserved variable**.

Modeling assumptions

We **assume** that:

+The incidence rate is approximately **constant over time** and **across the population** ("**constant and homogenous incidence**")

+Participants are always at risk of a new infection, regardless of how recently they have been infected ("**no lasting immunity**").

Time since infection and incidence

Under those assumptions:

+T has an **exponential distribution**:

 $+\mathbf{p}(T = t) = \lambda \exp\{-\lambda t\}$

+the rate parameter is the incidence rate

Likelihood of latent infection times

$$\begin{aligned} +\mathcal{L}^{*}(\lambda) &= \prod_{i=1}^{n} p \left(T = t_{i}\right) = \prod_{i=1}^{n} \lambda \exp(-\lambda t_{i}) \\ +\ell^{*}(\lambda) &= \log\{\mathcal{L}^{*}(\lambda)\} = \sum_{i=1}^{n} \log\{\lambda\} - \lambda t_{i} \\ +\ell^{*'}(\lambda) &= \sum_{i=1}^{n} \lambda^{-1} - t_{i} \\ +\hat{\lambda}^{*}_{\mathsf{ML}} &= \frac{n}{\sum_{i=1}^{n} t_{i}} = \frac{1}{\overline{t}} \end{aligned}$$

Likelihood of observed data

$$+p(Y = y) = \int_{t} p(Y = y, T = t) dt$$

+p(Y = y, T = t) = p(Y = y|T = t)p(T = t)



Model for active infection period

Notation:

+x(t): Pathogen concentration at time t+y(t): Antibody concentration at time tModel:

 $+x'(t) = \alpha x(t) - \beta y(t)$ $+y'(t) = \delta y(t)$

Within-host model for post-infection

antibody decay

b(t) = 0+y'(t) = -\alpha y(t)^r

Interactive Shiny app:





Open source analytical package for R available on GitHub https://github.com/UCD -SERG/serocalculator



→ C 🏠 🔤 ucd-serg.github.io/serocalculator/index.html

serocalculator 1.2.0.9006 Get started Reference Articles - Changelog

serocalculator

Antibody levels measured in a cross-sectional population sample can be translated into an estimate of the frequency with which seroconversions (infections) occur in the sampled population. In other words, the presence of many high antibody titers indicates that many individuals likely experienced infection recently and the burden of disease is high in the population, while low titers indicate a low frequency of infections in the sampled population and therefore a lower burden of disease.

The **serocalculator** package was designed to use the longitudinal response characteristics using a set of modeled parameters characterizing the longitudinal response of the selected serum antibodies. More details on the underlying methods can be found in <u>Getting Started</u>.

<u>View on CRAN</u> <u>Browse source code</u>

License

<u>GPL-3</u>

Links

Community Contributing_guide Code of conduct

Citation <u>Citing serocalculator</u>









National Institute of Allergy and Infectious Diseases

९ ☆

R CODE

Import longitudinal antibody parameters from OSF
curves <- "https://osf.io/download/rtw5k/" %>%
 load curve params()

Visualize curve parameters
curves %>% autoplot()

Import sample population data from OSF
xs_data <- "https://osf.io/download//n6cp3/" %>%
 load pop data()

Visualize antibody data
xs_data %>%
 autoplot(strata = "Country", type='density')

OUTPUT



R code using serocalculator

Estimate seroincidence stratified by country
and age

```
est_Country_ageCat = est.incidence.by(
    strata = c("Country", "ageCat"),
    pop_data = xs_data,
    curve_params = curves,
    noise_params = noise,
    antigen_isos = c("HlyE_IgG", "HlyE_IgA")
```

summary(est_Country_ageCat)

ResultsCountryageCatnest.startincidence.rateSECl.lwrCl.uprBangladesh<5</td>1010.10.399982930.03950.32970.4855

otratam	Country	ageeat		cotistart	meracitecitate			- anapi
Stratum 1	Bangladesh	<5	101	0.1	0.39998293	0.0395	0.3297	0.4853
Stratum 2	Bangladesh	5-15	256	0.1	0.47701125	0.032	0.4183	0.544
Stratum 3	Bangladesh	16+	44	0.1	0.44929893	0.0763	0.3221	0.6267
Stratum 4	Nepal	<5	171	0.1	0.02026628	0.0044	0.0132	0.0311
Stratum 5	Nepal	5-15	378	0.1	0.0354936	0.0031	0.0299	0.0421
Stratum 6	Nepal	16+	211	0.1	0.0935101	0.0078	0.0795	0.11
Stratum 7	Pakistan	<5	126	0.1	0.10592089	0.0136	0.0823	0.1363
Stratum 8	Pakistan	5-15	261	0.1	0.1145304	0.0084	0.0991	0.1323
Stratum 9	Pakistan	16+	107	0.1	0.19011951	0.0204	0.1541	0.2346

Stratum



Serocalculator

The serocalculator R package provides a rapid and computationally simple method for calculating seroconversion rates, as originally published in Simonsen (2009) and Teunis (2012), and further developed in subsequent publications by de Graaf (2014), Teunis (2016), and Teunis (2020).

In short, longitudinal seroresponses from confirmed cases with a known symptom onset date are assumed to represent the time course of human serum antibodies against a specific pathogen. Therefore, by using these longitudinal antibody dynamics with any cross-sectional sample of the same antibodies in a human population, an incidence estimate can be calculated.

Further details on the methodology can be found on the main package website.

This app provides a user-friendly interface to use the serocalculator methodology without the need for specialized coding knowledge. Users should follow the steps to:

- Import the required datasets
- Inspect their data
- Estimate seroincidence
- Prepare a report (optional)

Required datasets:

- · Cross-sectional population-based dataset with age and quantitative antibody results
- Noise parameters
- Longitudinal curve parameters

If you need assistance or encounter a clear bug, please file an issue with a minimal reproducible example on GitHub

https://ucdserg.shinyapps.io/shiny_serocalculator/

Filling the gaps with Serocalculator

Stanford University

💥 SEACTN

South and Southeast Asian Community-based-Trials Network

International Vaccine Institute



World Health Organization

*+ :

South Sudan



Typkoid Vaccine Acceleration Consortium

CENTER FOR VACCINE DEVELOPMENT . OXFORD VACCINE GROUP . PATH

ACKNOWLEDGEMENTS



BILL& MELINDA GATES foundation





THE AGA KHAN UNIVERSITY



International

Vaccine Institute







Mahidol University Wisdom of the Lond

CHRF









Extra slides

Biological noise

When we measure antibody concentrations in a blood sample, we are essentially counting molecules (using biochemistry).

We might miss some of the antibodies (undercount, false negatives) and we also might incorrectly count some other molecules that aren't actually the ones we are looking for (overcount, false positives, cross-reactivity).

We are more concerned about overcount (cross-reactivity) than undercount. For a given antibody, we can do some analytical work beforehand to estimate the distribution of overcounts, and add that to our model p(Y = y | T = t).

Measurement noise

There are also some other sources of noise in our bioassays; user differences in pipetting technique, random ELISA plate effects, etc. This noise can cause both overcount and undercount. We can also estimate the magnitude of this noise source, and include it in p(Y = y|T = t).

Variation in antibody kinetics by:

