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

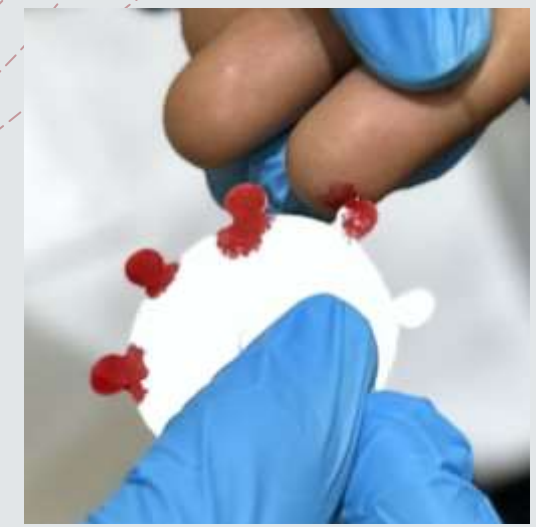
Assistant Professors of Epidemiology & Biostatistics

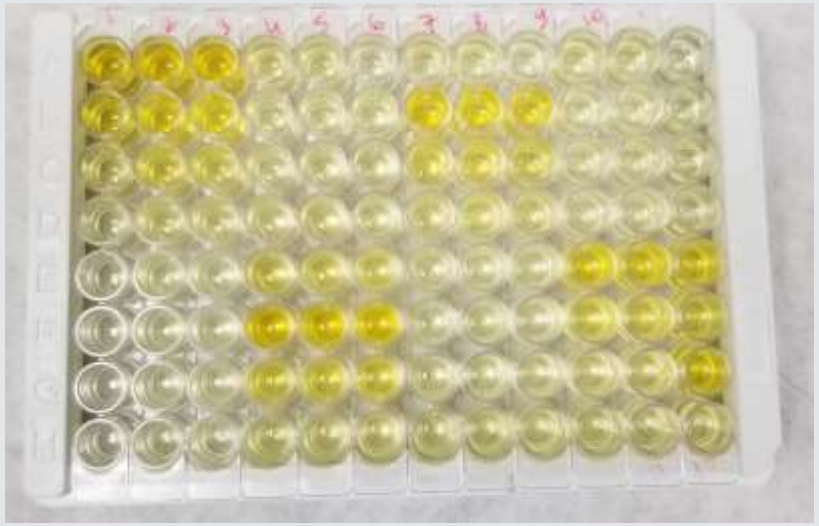
Department of Public Health Sciences

UC Davis School of Medicine

UCDAVIS HEALTH

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



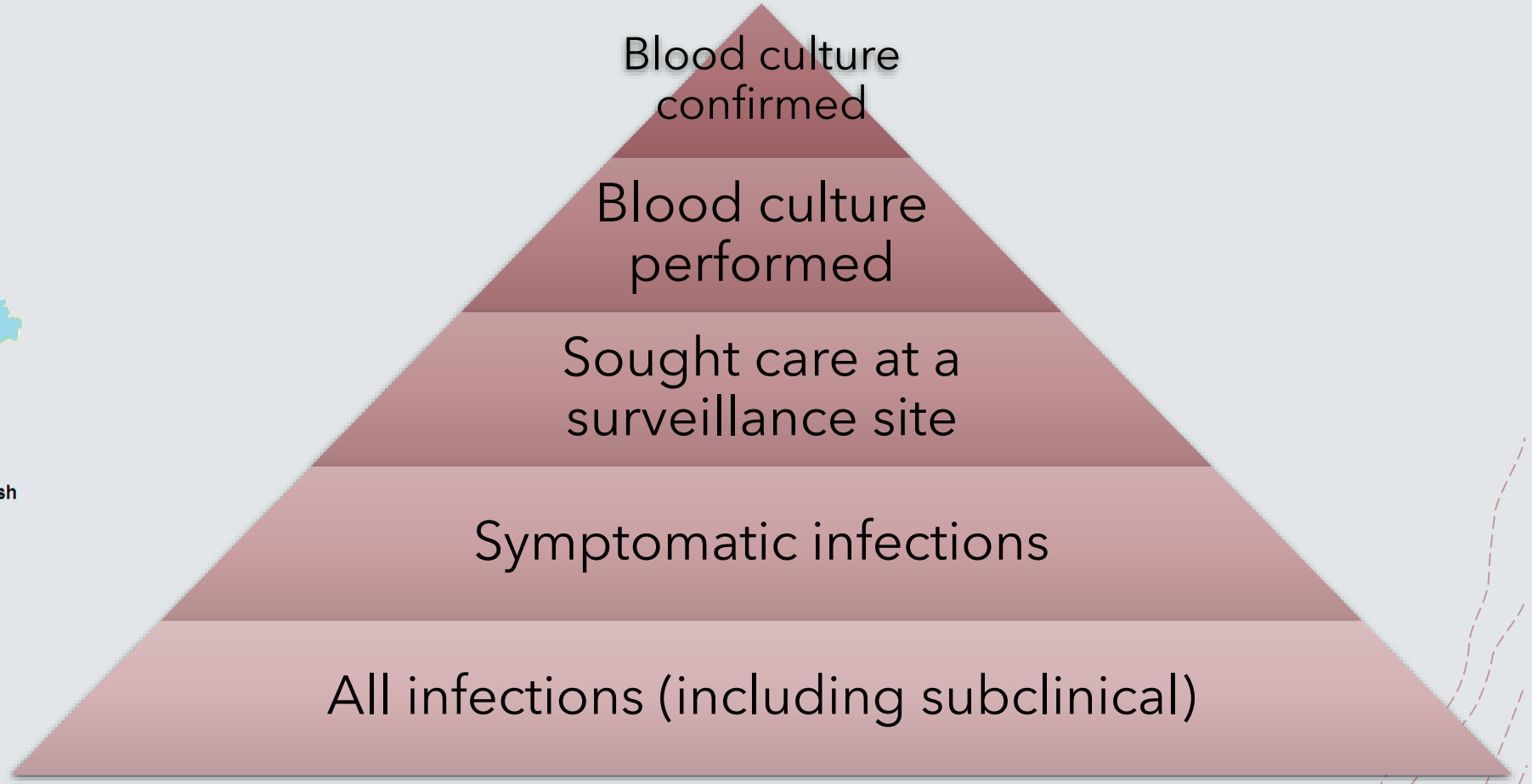


serocalculator 1.2.0.9006 Get started

serocalculator

Incidence (λ)

Seroepidemiology & Environmental Surveillance for Enteric Fever (SEES)



BILL & MELINDA GATES foundation

SABIN VACCINE INSTITUTE

Estd. 1996 DHULIKHEL HOSPITAL

CHRF

THE AGA KHAN UNIVERSITY

Stanford University

Aiemjoy et al, Lancet Infectious Diseases, 2022

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

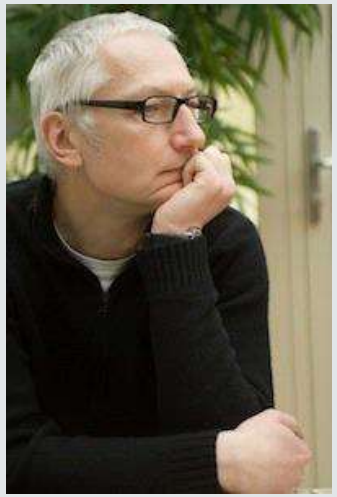
Together We Can

Take on Typhoid

Cross-sectional population-based serosurveys

+

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

Statistics in Medicine WILEY

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

P. F. M. Teunis[1] | J. C. H. van Eijkeren[2]

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

Correspondence

The presence of ser
seroclassification ai
serum antibody leve
This article expands
models of the serore

ELSEVIER

Contents lists available at ScienceDirect

Epidemics

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

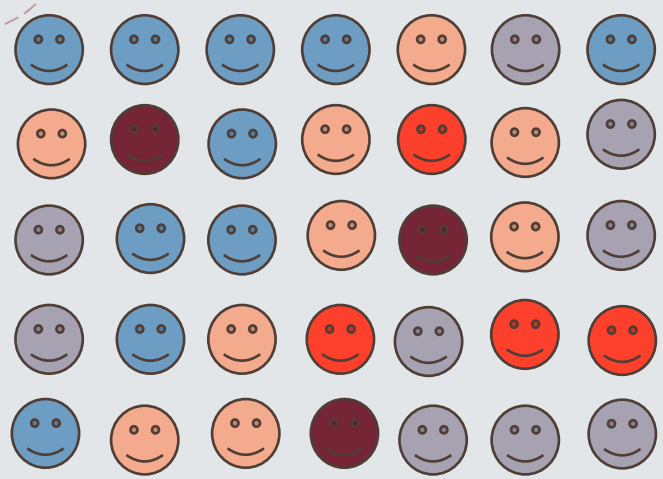
Linking the seroresponse to infection to within-host heterogeneity in antibody production

CrossMark

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]

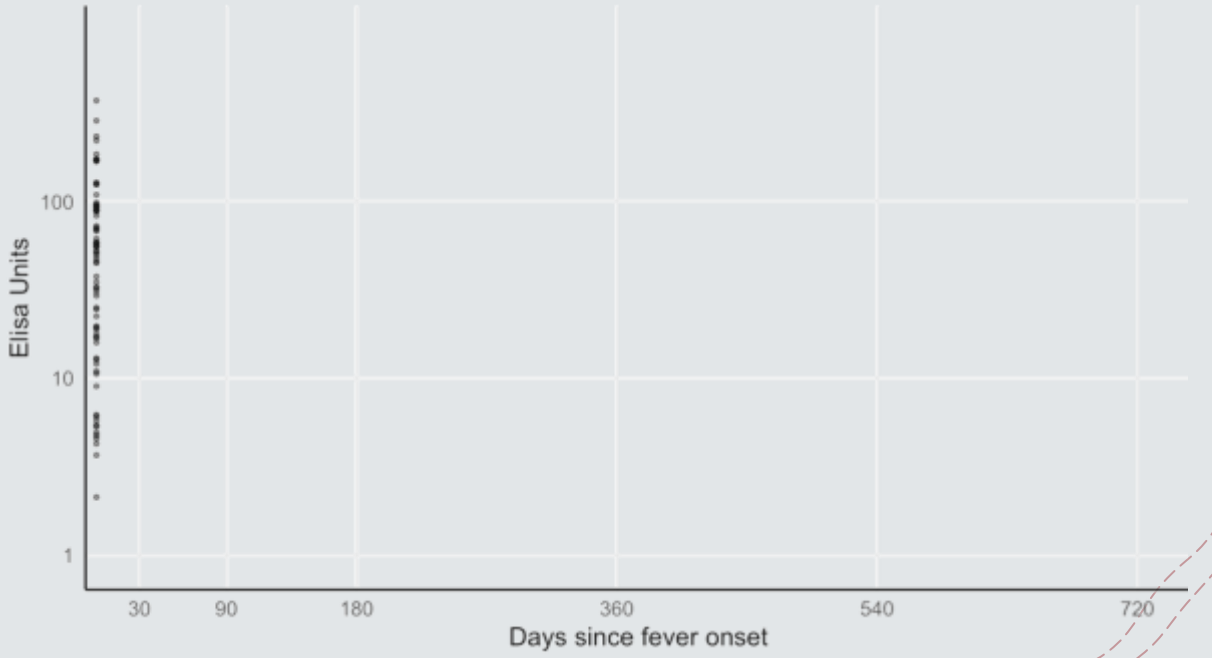
[a] Centre for Infectious Disease Control, RIVM, Bilthoven, Netherlands
[b] Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA
[c] Statistics and Methodology Group, Information Technology Division, RIVM, Bilthoven, Netherlands
[d] Department of Mathematics, Utrecht University, Utrecht, Netherlands
[e] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands

Cross-sectional population-based serosurveys



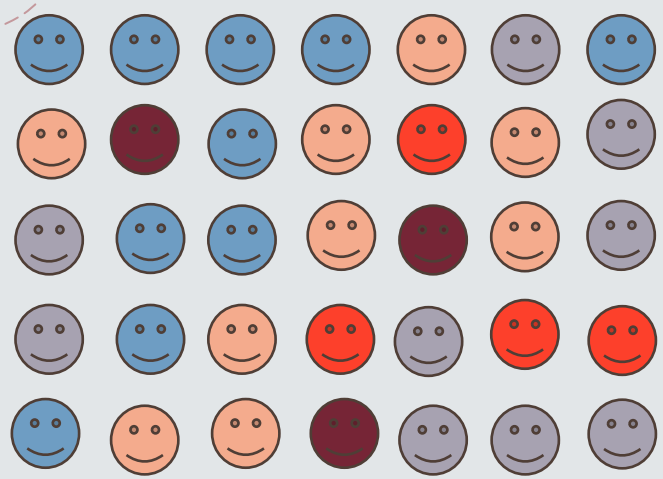
+

Within-host model of antibody dynamics



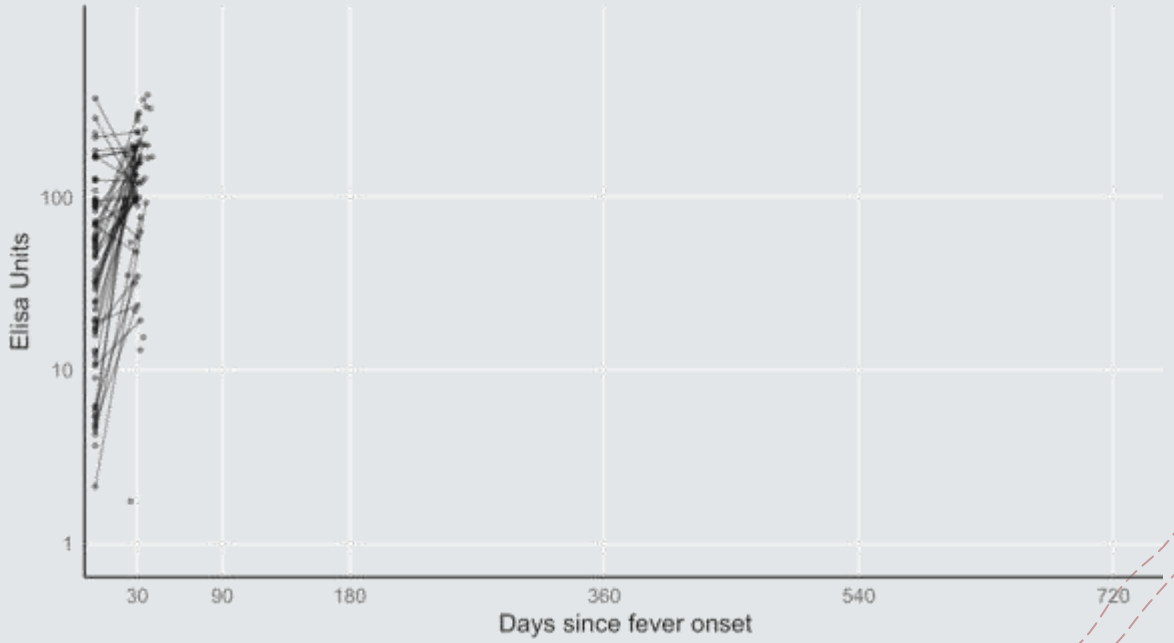
Teunis et al, Stats in Medicine, 2020
Teunis et al, Epidemics, 2016

Cross-sectional population-based serosurveys



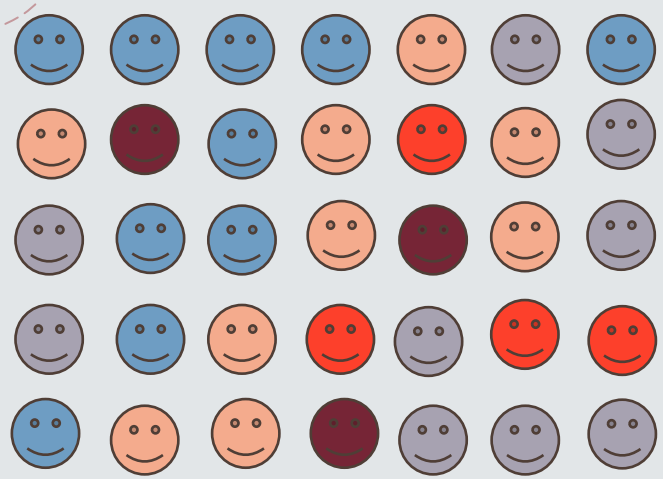
+

Within-host model of antibody dynamics



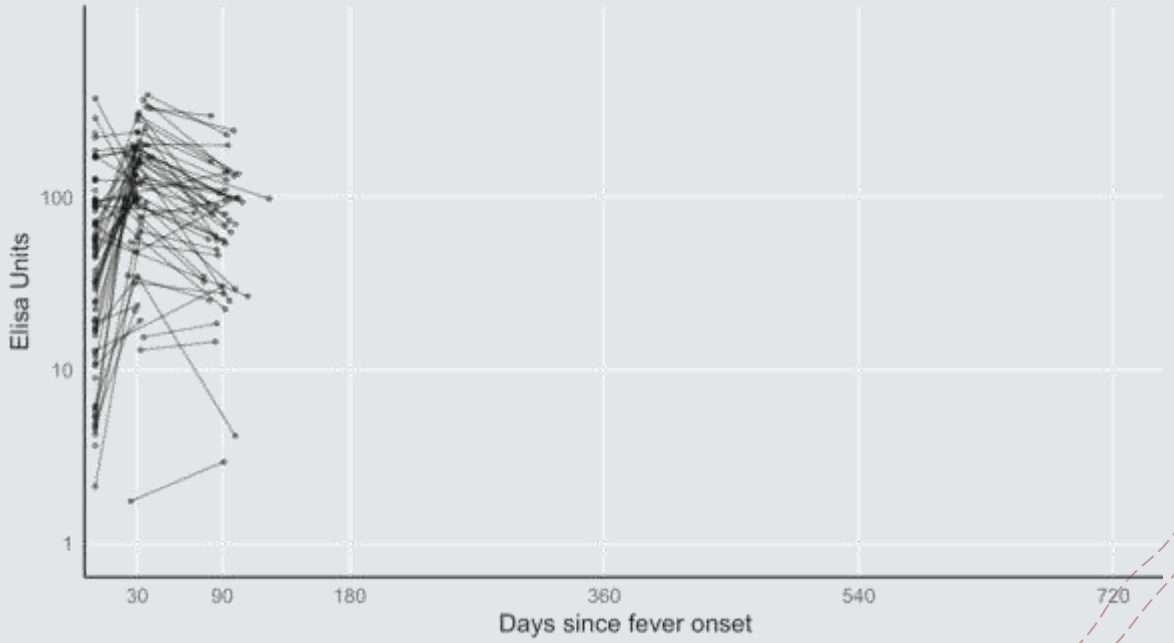
Teunis et al, Stats in Medicine, 2020
Teunis et al, Epidemics, 2016

Cross-sectional population-based serosurveys



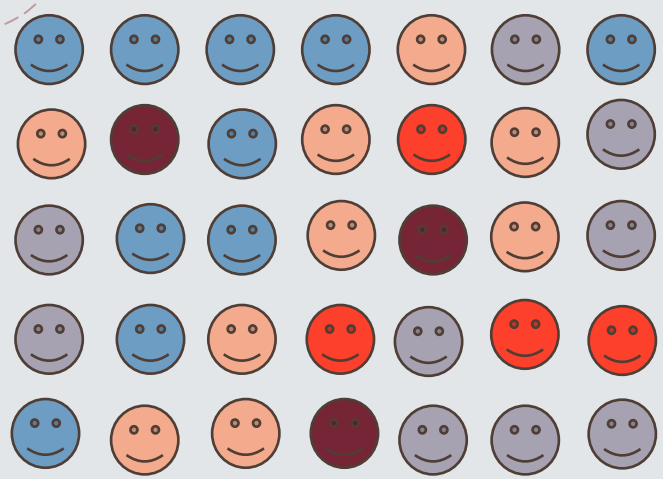
+

Within-host model of antibody dynamics



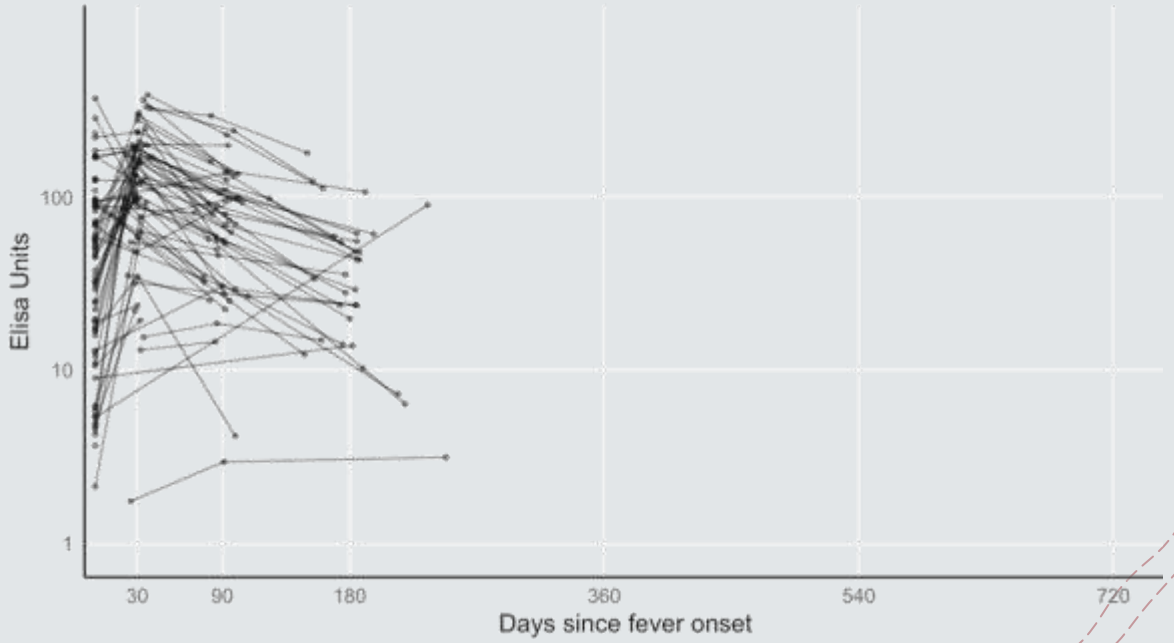
Teunis et al, Stats in Medicine, 2020
Teunis et al, Epidemics, 2016

Cross-sectional population-based serosurveys



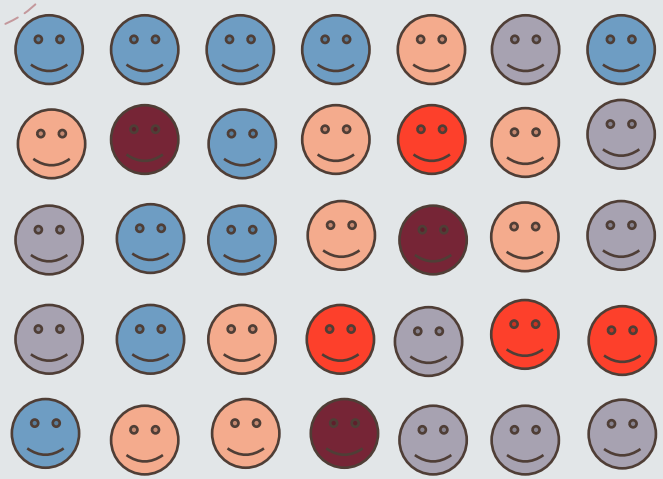
+

Within-host model of antibody dynamics



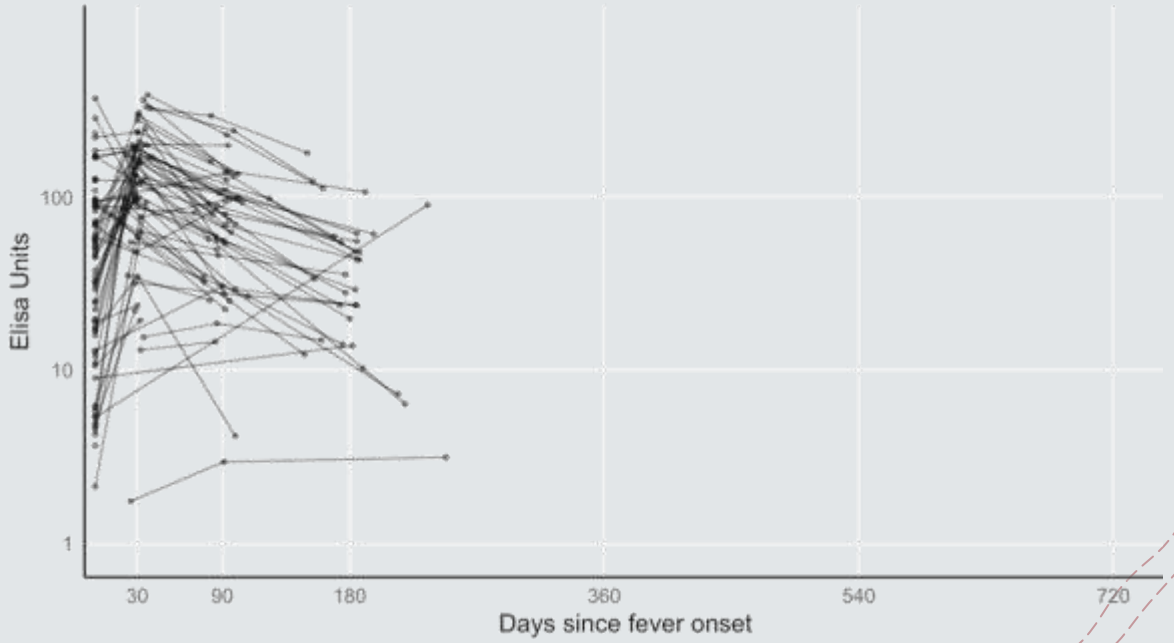
Teunis et al, Stats in Medicine, 2020
Teunis et al, Epidemics, 2016

Cross-sectional population-based serosurveys



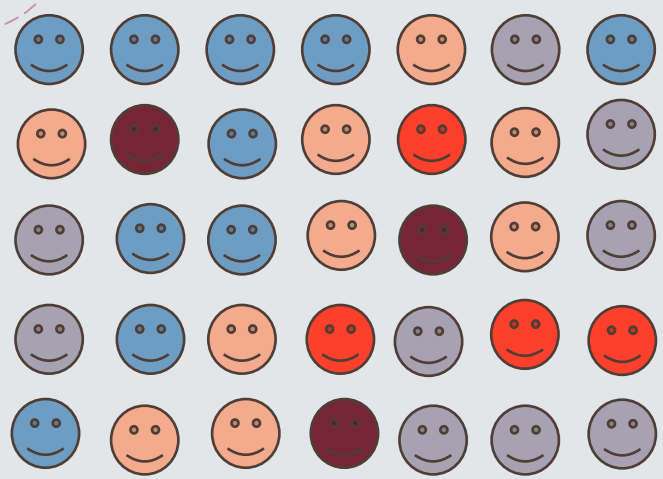
+

Within-host model of antibody dynamics



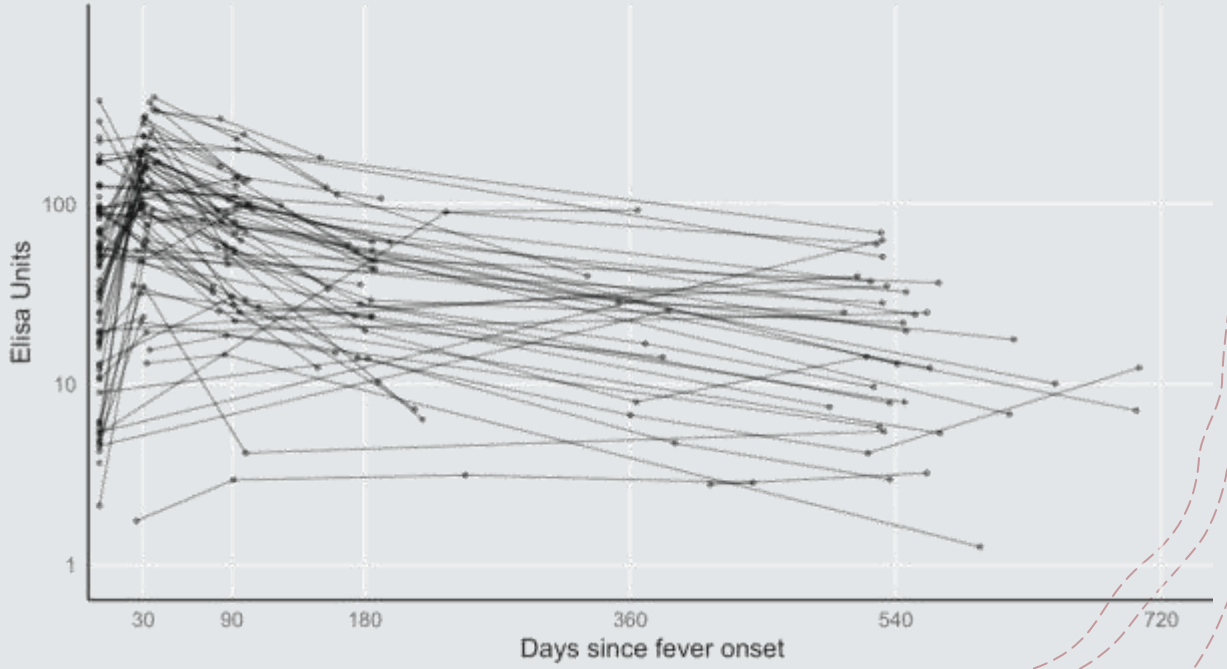
Teunis et al, Stats in Medicine, 2020
Teunis et al, Epidemics, 2016

Cross-sectional population-based serosurveys



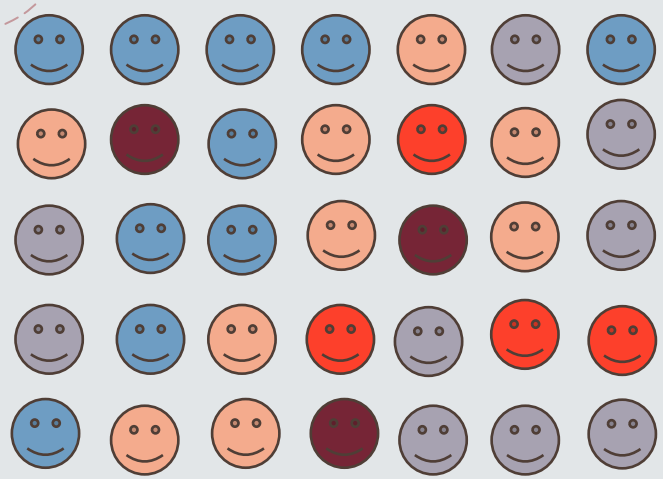
+

Within-host model of antibody dynamics



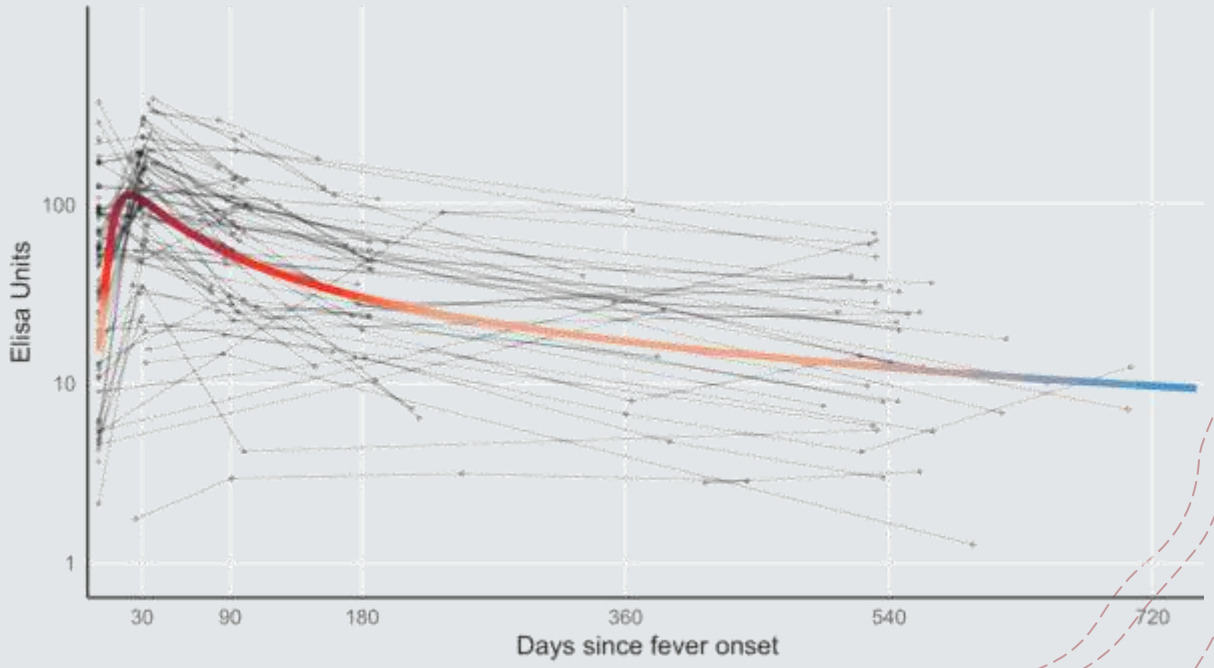
Teunis et al, Stats in Medicine, 2020
Teunis et al, Epidemics, 2016

Cross-sectional population-based serosurveys

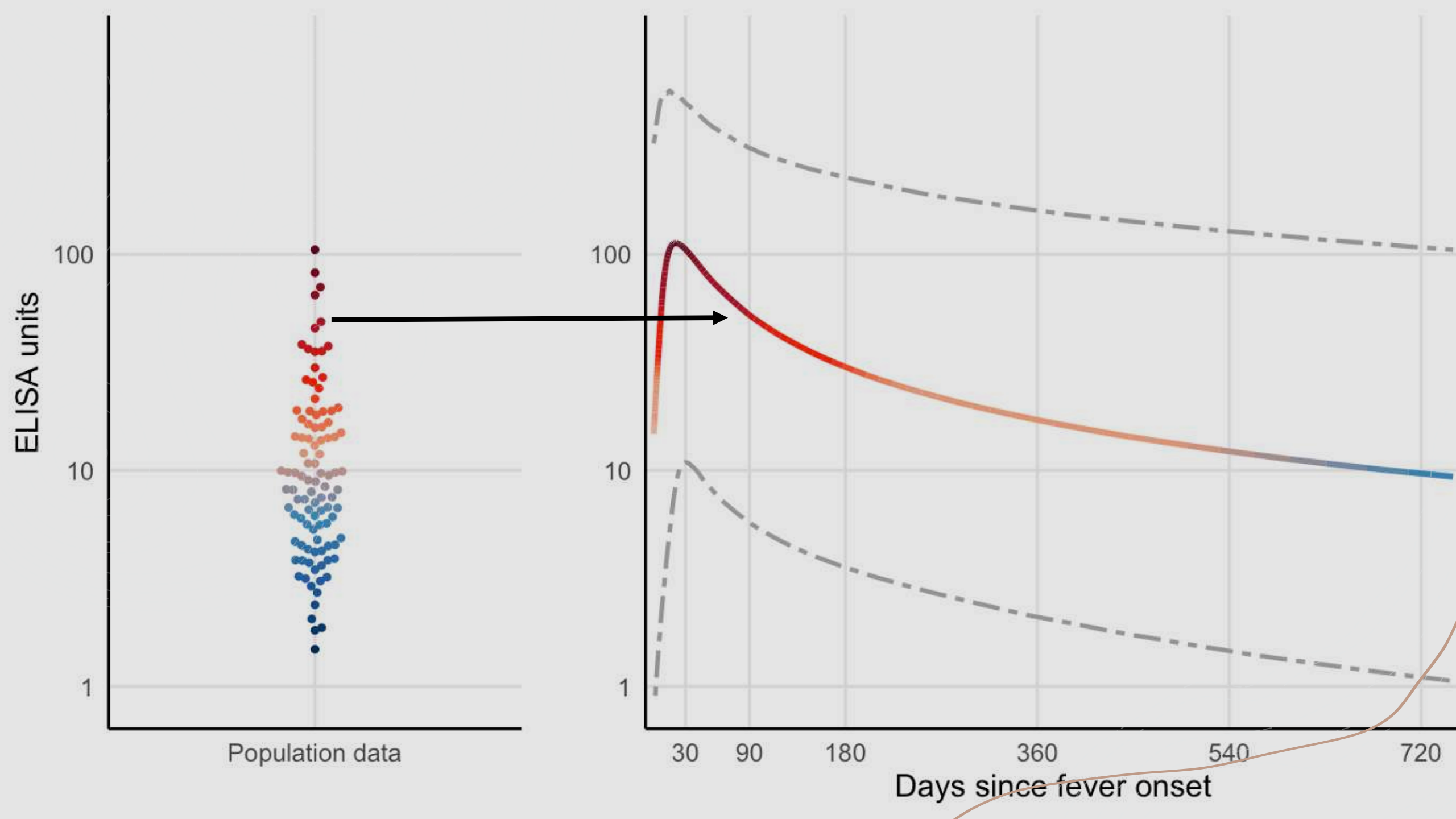


+

Within-host model of antibody dynamics



Teunis et al, Stats in Medicine, 2020
Teunis et al, Epidemics, 2016


ELISA units
100
10
1
Population data
100
10
1
30
90
180
360
540
720
Days since fever onset

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.

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**:
+ $\mathrm{p}(T = t) = \lambda \exp\{-\lambda t\}$
+ the rate parameter is the incidence rate

Likelihood of latent infection times

+ $\mathcal{L}^*(\lambda) = \prod_{i=1}^{n} p\,(T = t_i) = \prod_{i=1}^{n} \lambda \exp(-\lambda t_i)$
+ $\ell^*(\lambda) = \log\{\mathcal{L}^*(\lambda)\} = \sum_{i=1}^{n} \log\{\lambda\} - \lambda t_i$
+ $\ell^{*\prime}(\lambda) = \sum_{i=1}^{n} \lambda^{-1} - t_i$
+ $\hat{\lambda}^*_{\mathrm{ML}} = \frac{n}{\sum_{i=1}^{n} t_i} = \frac{1}{\bar{t}}$

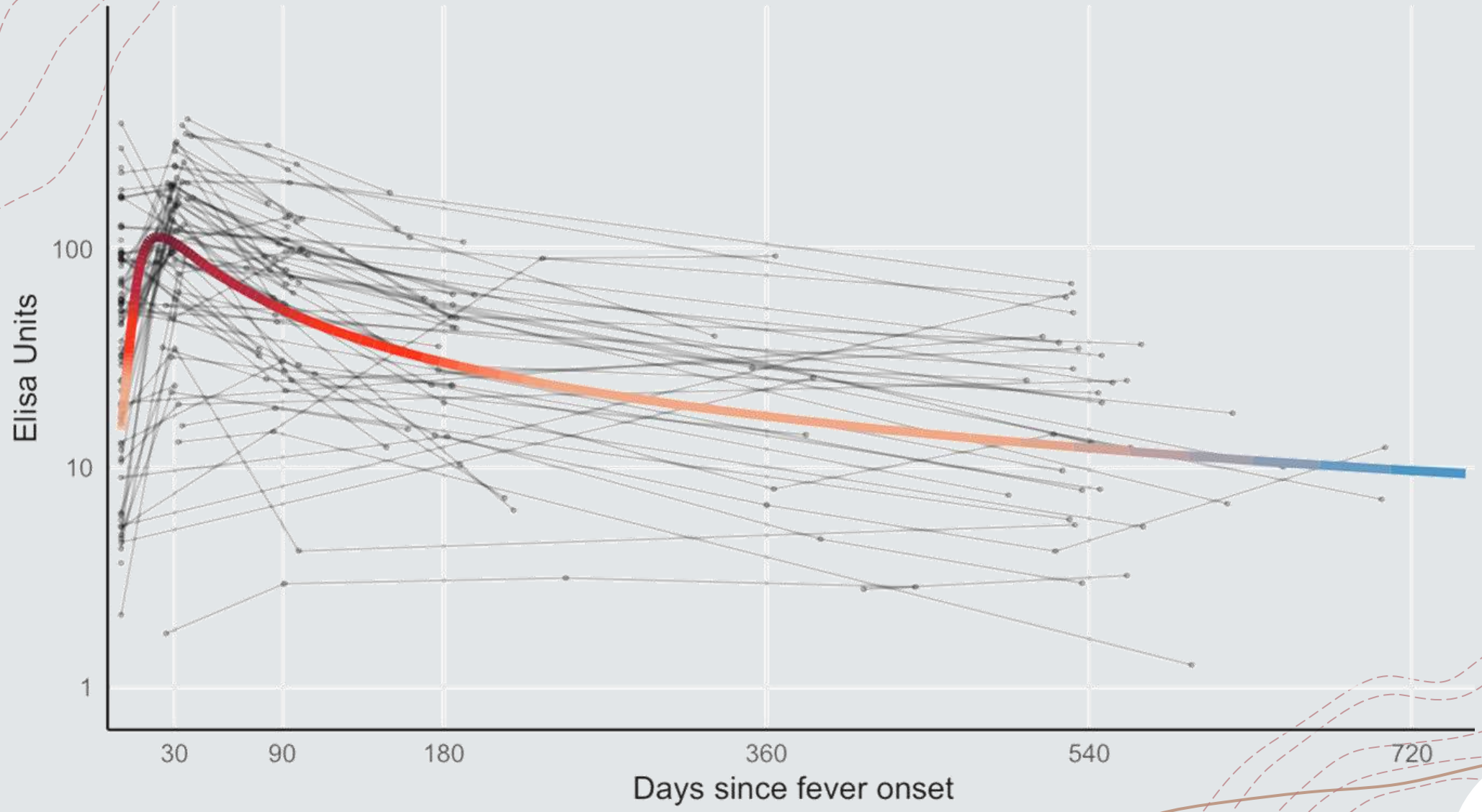
Likelihood of observed data

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

+ $\mathrm{p}(Y = y, T = t) = \mathrm{p}(Y = y|T = t)\mathrm{p}(T = t)$

Antibody response curves

$$p(Y = y|T = t)$$



Model for active infection period

Notation:

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

Model:

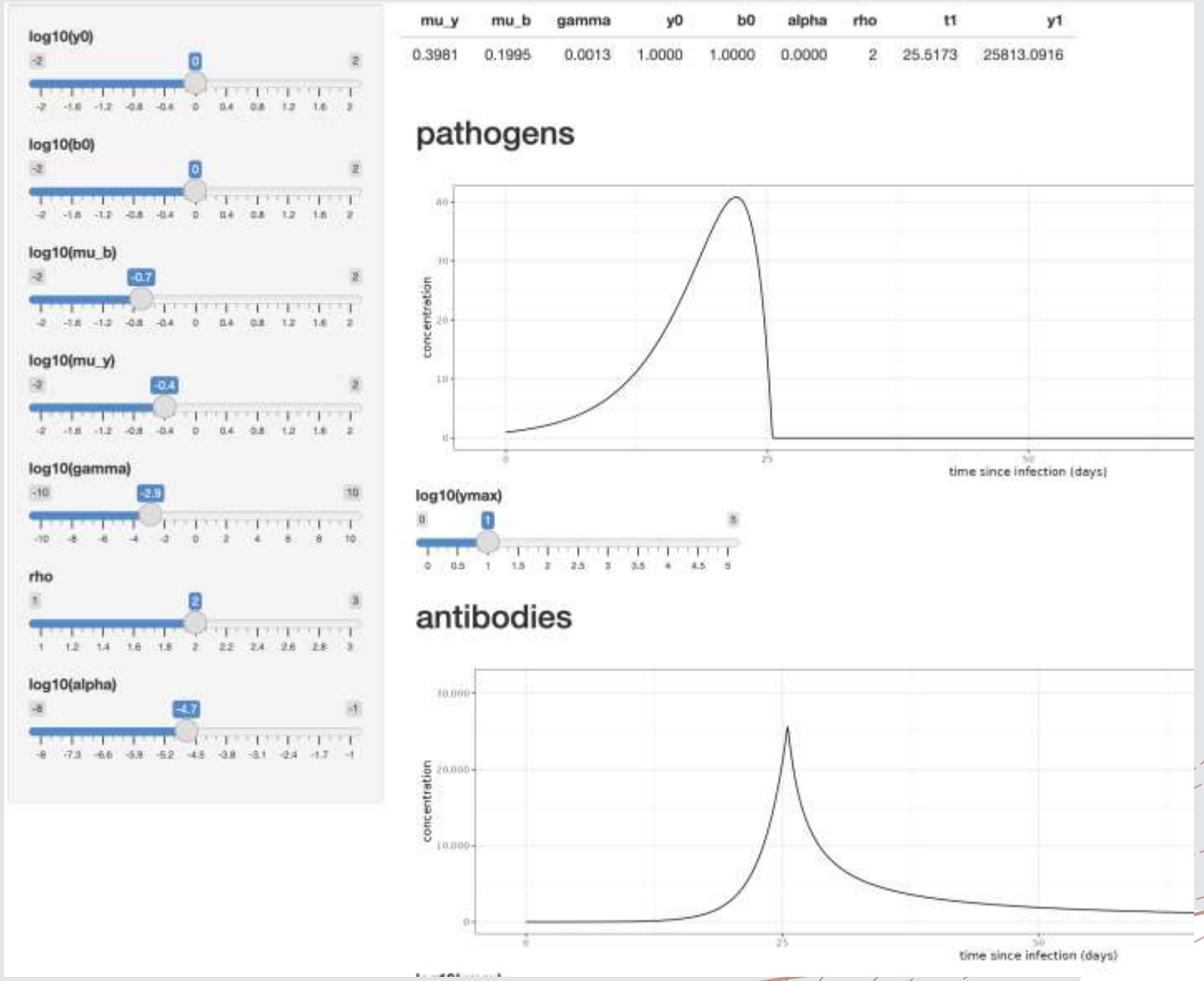
+ $x'(t) = \alpha x(t) - \beta y(t)$
+ $y'(t) = \delta y(t)$

Within-host model for post-infection antibody decay

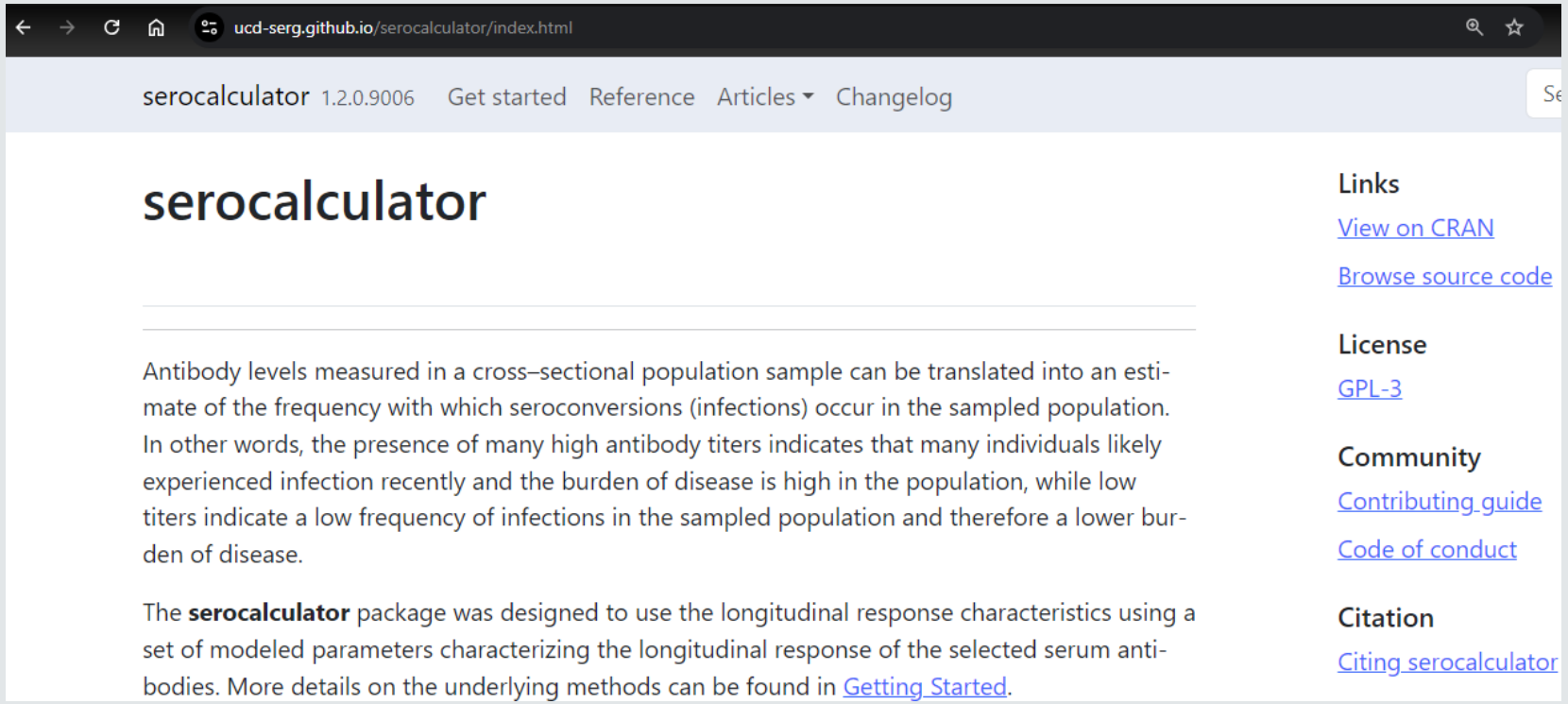
$$+b(t) = 0$$

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

Interactive Shiny app:



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



SkyeHI Technologies

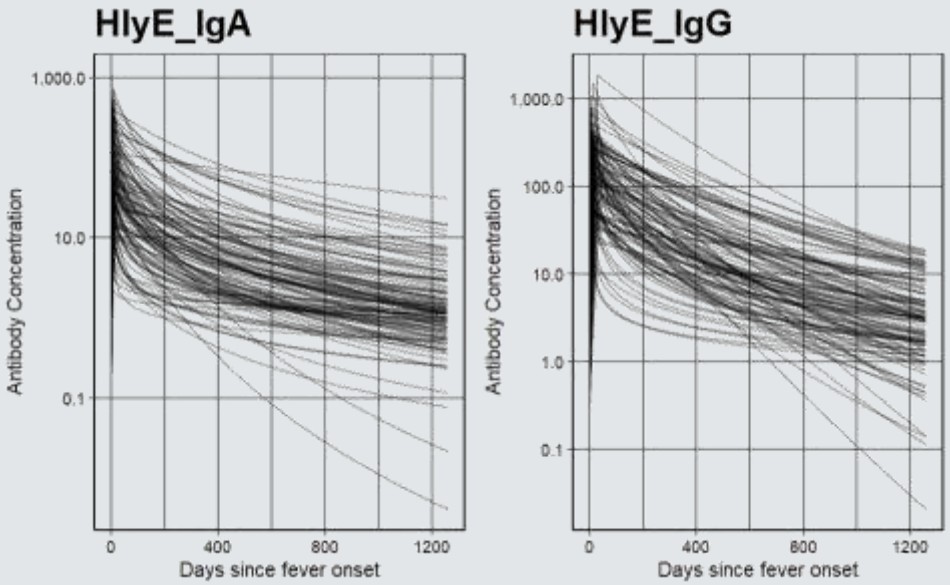
NIH National Institute of Allergy and Infectious Diseases

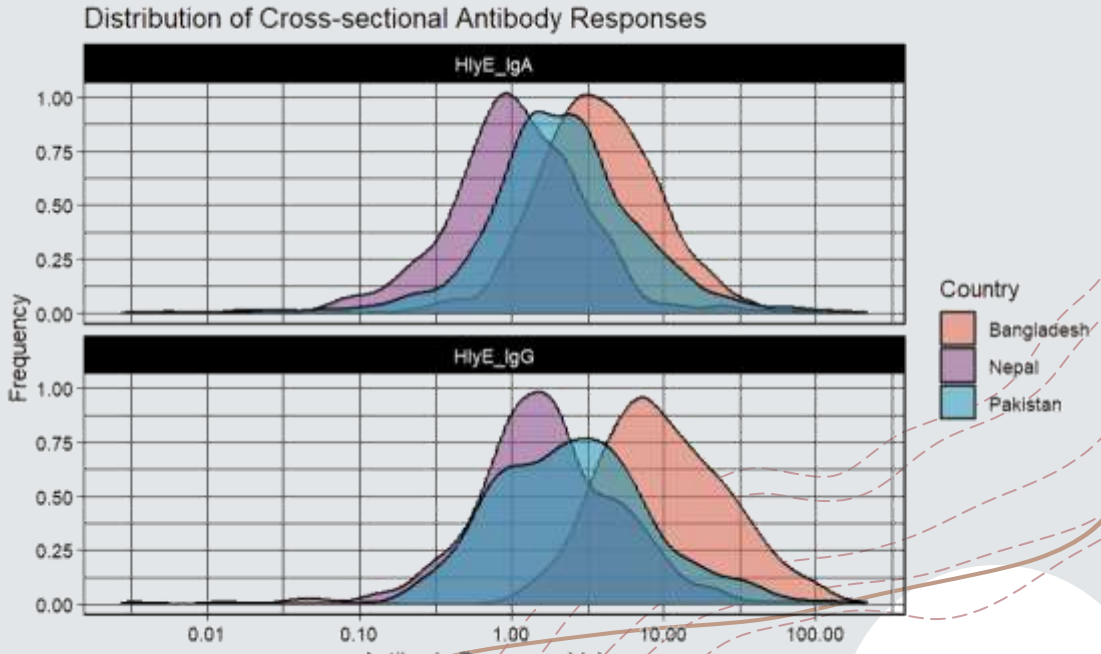
R CODE

```
# 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



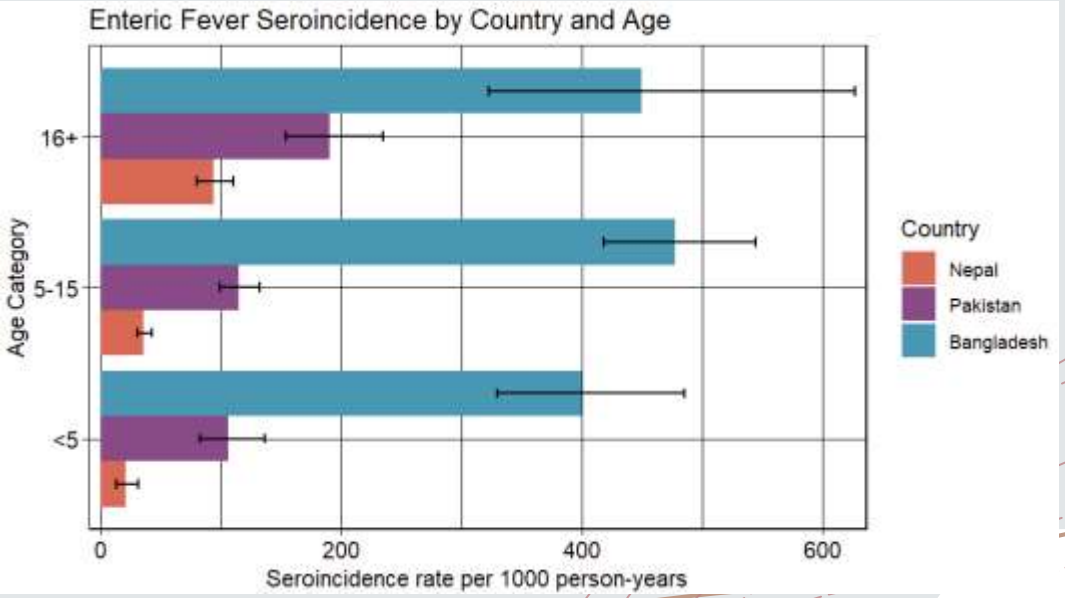


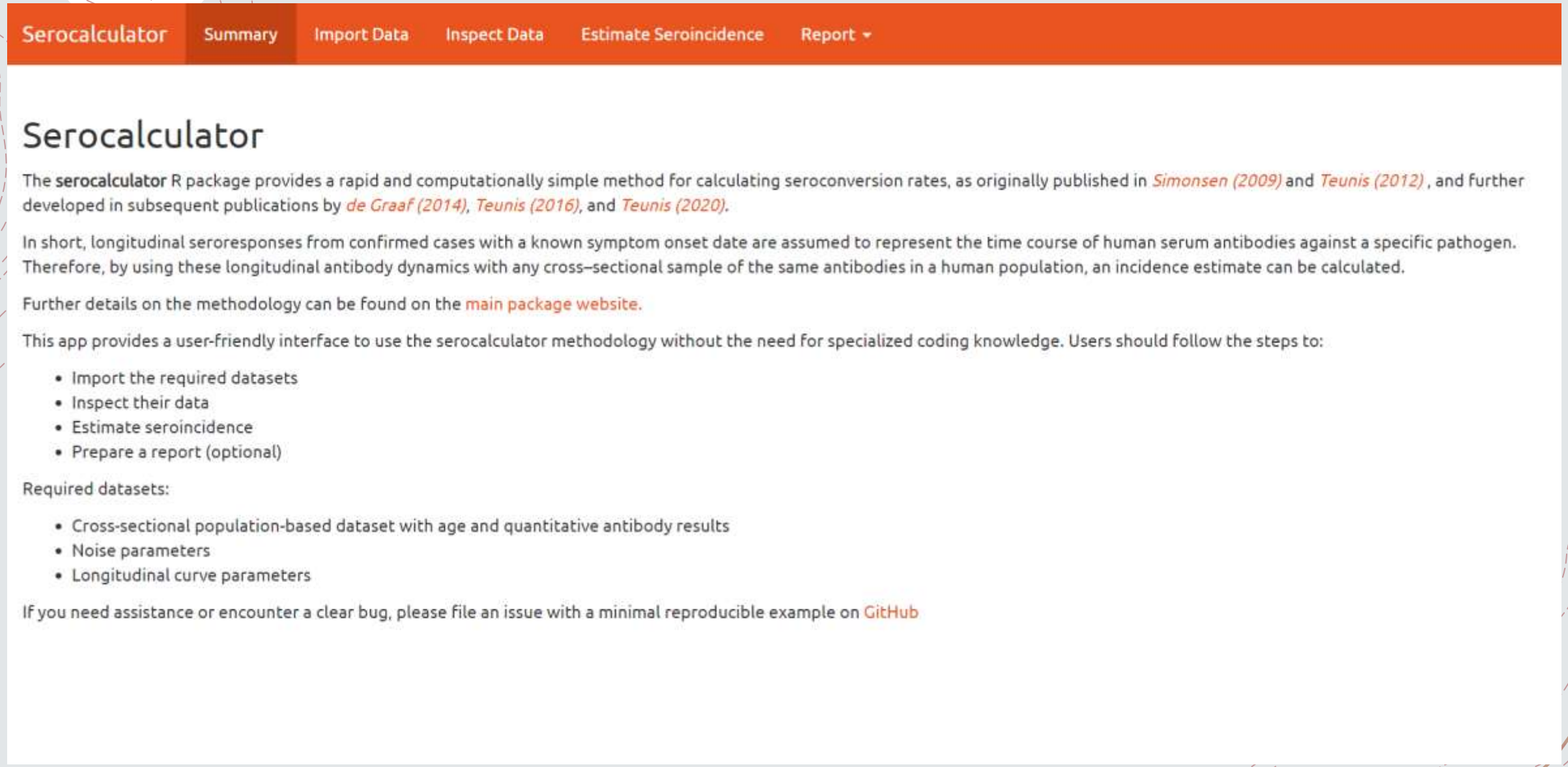
R code using serocalculator

```
# 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)
```

Results

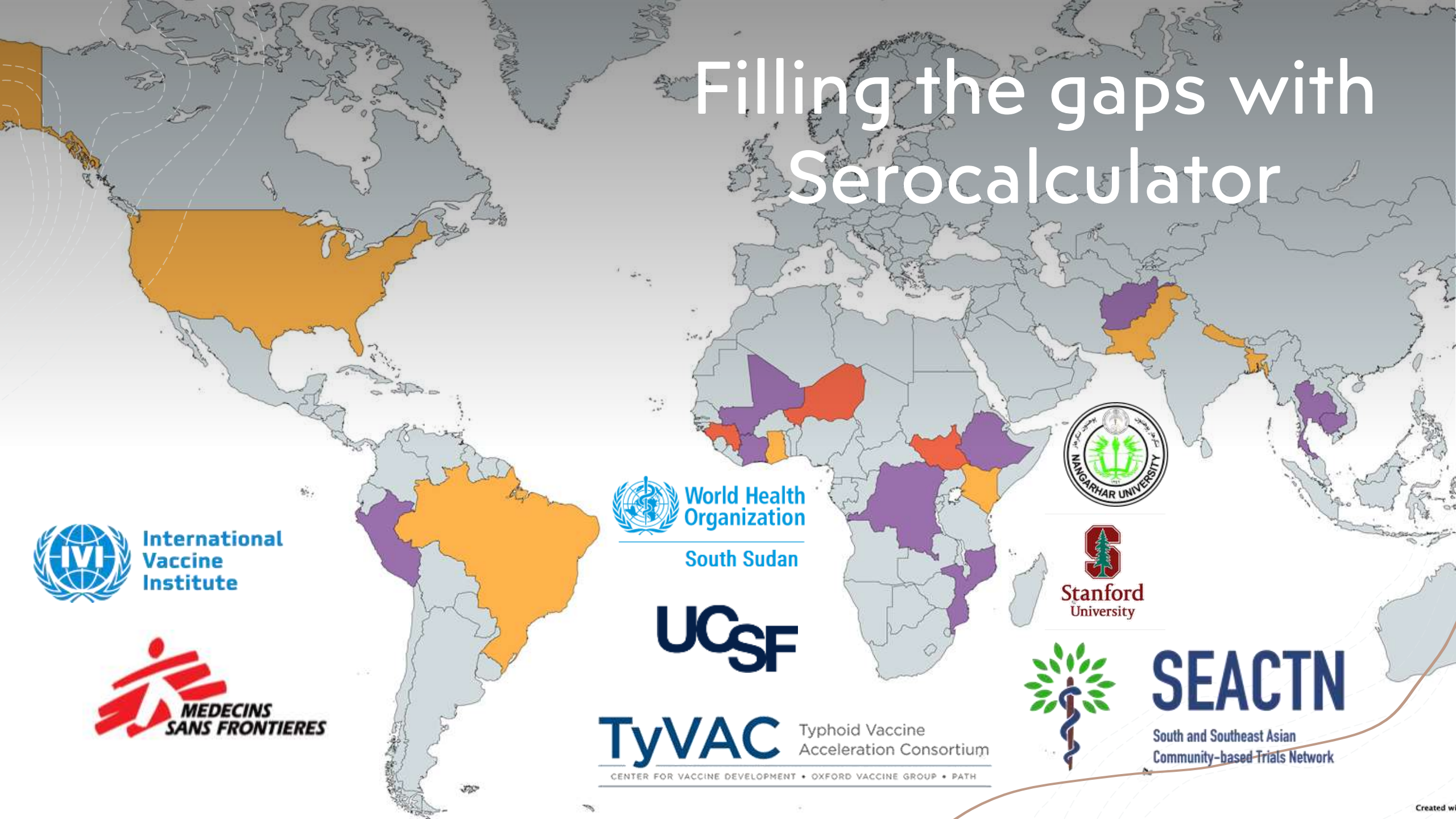
Stratum	Country	ageCat	n	est.start	incidence.rate	SE	CI.lwr	CI.upr
Stratum 1	Bangladesh	<5	101	0.1	0.39998293	0.0395	0.3297	0.4853
Stratum 2	Bangladesh	5-15	256	0.1	0.47701125	0.032	0.4183	0.544
Stratum 3	Bangladesh	16+	44	0.1	0.44929893	0.0763	0.3221	0.6267
Stratum 4	Nepal	<5	171	0.1	0.02026628	0.0044	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





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

Filling the gaps with Serocalculator



International Vaccine Institute

World Health Organization
South Sudan

UCSF

Nangarhar University

Stanford University

MEDECINS SANS FRONTIERES

TyVAC Typhoid Vaccine Acceleration Consortium
CENTER FOR VACCINE DEVELOPMENT • OXFORD VACCINE GROUP • PATH

SEACTN
South and Southeast Asian Community-based Trials Network

Created wi

ACKNOWLEDGEMENTS

National Institute of Allergy and Infectious Diseases

Bill & Melinda Gates foundation

Fogarty International Center

CHRF

Estd. 1996
DHULIKHEL HOSPITAL

THE AGA KHAN UNIVERSITY

UHN
Toronto General
Toronto Western
Princess Margaret
Toronto Rehab
Michener Institute

Mahidol University
Wisdom of the Land

International Vaccine Institute

UCSF

NANGARHAR UNIVERSITY

MORU
Tropical Health Network

MGH
1811

Stanford University

SABIN
VACCINE INSTITUTE

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:

Country

Age

Typhi/Paratyphi A