serocalculator, an R package for estimating seroincidence from cross-sectional serological data

Authors: Kristina W. Lai[1*], Chris Orwa[2], Jessica C. Seidman[3], Denise O. Garrett[3], Samir K. Saha[4], Dipesh Tamrakar[5], Farah Naz Qamar[6], Richelle C. Charles[7,8], Jason R. Andrews[9], Kristen Aiemjoy[1,10†], Douglas Ezra Morrison[1†]

Affiliations:

1. Department of Public Health Sciences, University of California Davis School of Medicine, Davis, CA, USA
2. SkyeHi Technologies, Nairobi, Kenya
3. Sabin Vaccine Institute, Washington, DC, USA
4. Child Health Research Foundation, Dhaka, Bangladesh
5. Dhulikhel Hospital, Kathmandu University Hospital, Dhulikhel, Nepal
6. Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan
7. Harvard Medical School, Harvard University, Boston, MA, USA
8. Department of Immunology and Infectious Diseases, Harvard TH Chan School of Public Health, Harvard University, Boston, MA, USA
9. Division of Infectious Diseases and Geographic Medicine, Stanford University School of Medicine, Stanford, CA, USA
10. Department of Immunology and Microbiology, Mahidol University Faculty of Tropical Medicine, Bangkok, Thailand

***Corresponding author:**

Kristina W. Lai
Medical Sciences 1C
One Shields Avenue
Davis, CA 95616
kwlai@ucdavis.edu

[†]Joint senior authors

Word Count: 1996 words

Abstract: (150/150 words)

Motivation: Seroincidence—the rate of new infections in a population—is a key measure for understanding pathogen transmission dynamics and informing public health action. Estimating seroincidence from cross-sectional data is complicated by antibody waning, cross-reactivity, and individual heterogeneity in antibody responses.

Implementation: serocalculator is an open-source R package that uses a likelihood-based framework incorporating antibody decay models, biological variability, and measurement noise to estimate seroincidence rates from cross-sectional serological data.

General features: The package supports overall and stratified seroincidence estimation using single or multiple biomarkers. It requires three inputs: (1) a pre-estimated seroresponse model characterizing post-infection antibody waning; (2) noise parameters capturing biological and assay-related variability; and (3) quantitative antibody responses from a cross-sectional survey. It is computationally efficient, well-documented, and includes a point-and-click R Shiny interface. These features promote usability across research and public health.

Availability: The package **serocalculator** is freely available on CRAN, with development versions on GitHub.

Introduction

Serological surveys can be used to measure population-level antibody responses to characterize infection occurrence.(1,2) Following pathogen exposure, the adaptive immune system generates antibodies against specific antigens. These antibodies are produced rapidly after infection and gradually wane over time.(1,3,4) The concentration of antibodies against a given pathogen reflects the time since last infection.(2–4) In cross-sectional surveys, a large proportion of individuals with high antibody concentrations suggests high burden of recent or frequent infection, while a smaller proportion indicates lower burden.

Population-level antibody responses are typically characterized using two key epidemiological measures: seroprevalence, the proportion of individuals with antibody levels above a defined threshold; and seroincidence, the rate at which new infections occur in the population.(1,2) The magnitude and speed of an individual's immune response to infection is complex and dynamic, shaped by factors such as infectious dose, age, disease severity, prior exposures, and vaccination history. Most analytic methods and software used to estimate seroprevalence and seroincidence dichotomize quantitative antibody levels into binary outcomes—seropositive or seronegative. While simple and widely used, this approach ignores temporal changes in antibody concentrations, individual-level heterogeneity, and immunological cross-reactivity between pathogens. Moreover, seroprevalence reflects the cumulative history of prior exposures in a population and is influenced by the age distribution of the sample, making it an imperfect proxy for recent transmission. In contrast, seroincidence rates quantify the rate at which new infections occur and are increasingly favored by epidemiologists, modelers, and public health professionals.(3,4) These rates can help determine whether a disease has been eliminated,(5,6) identify risk factors for transmission, and locate areas where interventions are most needed. (3,7)

Traditional methods to assess seroincidence require resource- and time-intensive longitudinal studies that are often infeasible. However, new statistical methods estimate seroincidence rates from cross-sectional serosurveys by incorporating models of antibody decay dynamics from confirmed cases.(2,8,9) Seroincidence can be estimated as a function of the peak antibody response and decay rate from infected individuals while accounting for heterogeneity in immunity, responses to multiple antigens, age-dependence, non-specific binding of the assay probes (noise), and laboratory assay variability.(1,8–10) This approach assumes that exposure follows a Poisson process, meaning all individuals in the sampled population are subject to the same constant infection risk over time. Using this assumption, maximum likelihood estimation (MLE) is used to estimate seroincidence from cross-sectional serologic data.(2,8–11) We also assume that seroresponses from each antibody isotype are independent of the others, allowing us to sum each of the likelihoods from multiple antibodies and antigens, thereby increasing precision in the seroincidence estimate.(1,9) Here, we introduce **serocalculator**, a statistical software package that implements recent methodological advancements for estimating seroincidence rates from cross-sectional serosurveys.(1,8–10) **serocalculator** offers a validated and accessible framework for estimating seroincidence by integrating antibody kinetics with noise correction to produce dynamic infection rate estimates. It includes intuitive tools such as multi-biomarker support, stratified analysis, and an R Shiny interface for non-coders. This package supersedes the now-defunct *seroincidence* package (12) and is designed for use across a range of skill levels in public health practice.

Implementation

serocalculator can be downloaded from the Comprehensive R Archive Network (CRAN). Over 30 exported functions guide users from data import and cleaning to analysis and visualization. All functions in **serocalculator** version 1.4.0 are described on the package website. The general

workflow is as follows, 1) import data, 2) inspect and visualize data, and 3) estimate seroincidence rate.

Import Data

Three inputs are required: 1) Longitudinal seroresponse parameters (separately pre-estimated) from a seroresponse model characterizing the longitudinal response of the selected serum antibodies; 2) Noise parameters that encompass biological noise (individual variation in response) and measurement noise (laboratory assay variability); and 3) Quantitative antibody levels from a cross-sectional population-based serosurvey sample that is calibrated on the same scale as the longitudinal seroresponse parameters. For testing purposes, users can access small example files within the package, or they can import larger sample files for specific diseases from the publicly available Serocalculator Data Repository on Open Science Framework. Example input datasets with formatting guidance are provided in Supplement S1.

Longitudinal Seroresponse Parameters

Longitudinal seroresponse parameters describe how antibodies rise and fall after infection. They are calculated by fitting two-phase within-host models to observed quantitative antibody responses measured in a cohort of confirmed cases. The responses are then modeled using a Bayesian hierarchical framework to obtain predictive posterior samples using Markov chain Monte Carlo (MCMC) sampling.(1,8–10) This results in a joint sample of five parameters (further described in Supplement S2). Seroresponse parameters per antigen isotype are assumed to describe a generic infection process and generalize to any sampled population. As improved longitudinal data become available, the validity of this assumption can be tested. Seroresponse parameters can be used for stratified estimation, such as age- or country-specific seroresponse models.

Users can import their own set of seroresponse parameters in either CSV or RDS format and then apply `as_sr_params()`, which will assign the imported dataset as a `sr_params` object and ensure required variables are present and in the correct format. If using online files from a URL, users can apply the function `load_sr_params()`, which will perform necessary column and formatting checks. The longitudinal parameter sets must include the following columns: one or more disease-specific antigen and antibody isotype pairs that matches those included in the cross-sectional population data (`antigen_iso`), the baseline antibody concentration (`y0`), the peak antibody concentration (`y1`), the time to peak antibody concentration (`t1`), the antibody decay rate (`alpha`), and the antibody decay shape (`r`).

Further details on the longitudinal seroresponse parameters can be found in Supplement S2.

Noise Parameters

The noise parameters object defines assumptions about the measurement process.(1) Biological noise (ν) reflects potential non-specific antibody binding from other exposures, such as cross-reactivity with a related pathogen. It is defined as the 95th percentile of the antibody response distribution, assumed to be normal, in a population with no prior exposure to the pathogen.(1) Measurement noise (ϵ) represents laboratory assay variability from the laboratory testing process. It is defined by a coefficient of variation (CV), or the ratio of the standard deviation to the mean for replicates across plates. The serocalculator package can also accommodate censored data: lower and upper limits of reliable quantification may be specified as y_{low} and y_{high}, respectively, and any observations recorded outside of those limits will be treated as censored.

As above, users can import their own noise parameters into R (e.g., from CSV or RDS files) as a data.frame and then apply `as_noise_params()`, which will convert the data.frame into a

`noise_params` object (a subclass of `tibble::tbl_df` with added metadata attributes) and ensure required variables are present and in the correct format. If using online files from a URL, users can apply the function `load_noise_params()` and then follow the process described above for the seroresponse parameters . The required noise parameter variables are: one or more disease-specific antigen and antibody isotype that match those included in the cross-sectional population data (`antigen_iso`), the biological noise (`nu`), the measurement noise (`eps`), and the lower and upper limits of detection of the assay (`y.low` and `y.high` respectively). Further details on the noise parameter calculations are in Supplement S3.

Cross-sectional population data

Once again, users can import their own cross-sectional population data in either CSV or RDS format and apply the function `as_pop_data()`, which assigns the imported dataset as a `pop_data` object and ensures correct formatting. The required variables for this dataset are: age measured in years (`age`), one or more specified antigen isotype pairs (`antigen_iso`), quantitative antibody response (`value`), and a unique participant id (`id`). The `as_pop_data()` function detects the required variables or allows users to specify them. If using online data from a URL for an online repository such as Open Science Framework, users can apply the function `load_pop_data()` and then once again follow the process described above for the seroresponse parameters. Population datasets can include additional variables for stratification.

Estimating seroincidence

Once all parameters and datasets are imported, users can estimate either overall or stratified seroincidence rates using the functions `est_seroincidence()` and `est_seroincidence_by()`, respectively. Users begin by specifying the three required data inputs and selecting one or more antigen-isotype pairs for inclusion in the estimate. Optional parameters allow customization of

visualizations, warning messages, parallel computing, and the estimation process. Stratified seroincidence estimates can be generated by specifying one or more covariates in the cross-sectional dataset, and optionally in the longitudinal seroresponse parameters.

Use

We will demonstrate the use of **serocalculator** with a reproducible example of enteric fever seroincidence using a subset of data from a series of serosurveys conducted in Bangladesh, Nepal and Pakistan in 2016-2021.(13,14) Additional details on these data are available in Supplement S4. Data for this analysis are stored in the Serocalculator Data Repository on Open Science Framework. This example is also available as an article on the **serocalculator** website.

First, we load each of our three inputs (cross-sectional population data, noise parameters, and seroresponse parameters) and then create descriptive plots for both cross-sectional data and modeled seroresponse parameters using built-in `autoplot()` visualization methods. In this example, we import noise parameters calculated at each study site laboratory from analysis using enzyme-linked immunosorbent assays (ELISA).

[Figure 1]

We can estimate overall and stratified seroincidence rates using the `est_seroincidence()` and `est_seroincidence_by()` functions, respectively. In both cases, we combine IgA and IgG seroresponses to the hemolysin E (HlyE) antigen (Figure 2a). For the overall seroincidence rate, we choose to filter to only sites in Pakistan and find a seroincidence rate of 128 per 1000 person-years (95% CI). To compare across countries and ages, we specify the strata variables and find that 5-15 year olds in Bangladesh experience the highest seroincidence of enteric fever, with a

rate of 477 per 1000 person-years (Figure 2). Model outputs from `est_seroincidence()` and `est_seroincidence_by()` can be summarized and visualized using the `summary()` and `autoplot()` methods, respectively.

[Figure 2]

Discussion

serocalculator provides a rapid and computationally simple method for calculating population-level seroincidence rates that account for antibody decay dynamics, biologic variability, and measurement noise and can integrate information from multiple biomarkers. Building upon the prior *seroincidence* package,(12) the ***serocalculator*** package implements updated analytic methods that account for biological and measurement noise, new function names, new summary and visualization options, and cross-sectional data simulations.

Several groups have developed R packages to support seroepidemiologic analysis. The *serosv* package estimates seroprevalence and force of infection (FOI) using built-in datasets, but it is not intended for user-supplied data.(15) *serosim* simulates serological study data to support study design and evaluate within-host dynamics and new analytic approaches.(16) While **serocalculator** also includes a simulation function, its primary purpose is to estimate seroincidence rates from user-collected data. *Rsero* provides Bayesian tools for comparing serocatalytic models, assessing FOI under various assumptions (e.g., transmission mode, seroreversion), and evaluating sampling strategies.(17) *serojump* focuses on infection timing and antibody kinetics using Bayesian reversible jump MCMC and individual-level data.(18) Collectively, these packages expand access to seroepidemiologic tools. Notably, **serocalculator** is the only package to incorporate individual-level heterogeneity, combine multiple antigen isotypes, and account for multiple sources of noise.

Future enhancements to **serocalculator** will address current methodological limitations and expand usability. While the package supports interval censoring for some assays, its application is constrained by the availability of longitudinal seroresponse parameters for specific pathogens and assays. To increase flexibility across study settings, we are developing a complementary R package that will enable users to model seroresponse parameters from their own data.

Another key development area is improving performance in high-burden settings, where repeated exposures can bias seroincidence estimates. Current methods do not fully address this challenge, and we are actively working to incorporate Approximate Bayesian Computation techniques to enhance estimation in areas with high force of infection.(11)

The **serocalculator** package has been applied to studies of enteric fever and scrub typhus,(14,19,20) and its methods are currently being adopted by research teams and public health institutions to inform vaccine introduction decisions and surveillance strategies in endemic regions. **serocalculator** provides a free, open-source, and user-friendly tool for estimating seroincidence rates from cross-sectional quantitative serology data. Additional documentation and tutorials are available at https://ucd-serg.github.io/serocalculator/.

Funding:

This work was supported by the National Institute of Allergy and Infectious Diseases at the National Institutes of Health [grant number R21AI176416 to KA and DEM] and the Fogarty International Center at the National Institutes of Health [grant number K01TW012177 to KA].

Conflict of interest:

None declared.

Supplementary data:

Supplementary data are available at IJE online.

Data availability:

Source code for the **serocalculator** package is available on GitHub: https://github.com/UCD-SERG/serocalculator. Deidentified data from the sample analysis is hosted on Open Science Framework: https://osf.io/ne8pc/files/osfstorage.

References

1. Teunis PFM, van Eijkeren JCH. Estimation of seroconversion rates for infectious diseases: Effects of age and noise. Stat Med [Internet]. 2020 Sep 20;39(21):2799–814. Available from: https://onlinelibrary.wiley.com/doi/10.1002/sim.8578
2. Simonsen J, Mølbak K, Falkenhorst G, Krogfelt KA, Linneberg A, Teunis PFM. Estimation of incidences of infectious diseases based on antibody measurements. Stat Med [Internet]. 2009 Jun 30;28(14):1882–95. Available from: https://onlinelibrary.wiley.com/doi/10.1002/sim.3592
3. Aiemjoy K, Seidman JC, Charles RC, Andrews JR. Seroepidemiology for Enteric Fever: Emerging Approaches and Opportunities. Open Forum Infect Dis. 2023 Jun 2;10(Supplement_1):S21–5.
4. Hay JA, Routledge I, Takahashi S. Serodynamics: A primer and synthetic review of methods for epidemiological inference using serological data. Epidemics. 2024 Dec;49:100806.
5. Kamau E, Ante-Testard PA, Gwyn S, Blumberg S, Abdalla Z, Aiemjoy K, et al. Characterizing trachoma elimination using serology. 2024.
6. Tedijanto C, Solomon AW, Martin DL, Nash SD, Keenan JD, Lietman TM, et al. Monitoring transmission intensity of trachoma with serology. Nat Commun. 2023 Jun 5;14(1):3269.
7. Cutts FT, Hanson M. Seroepidemiology: an underused tool for designing and monitoring vaccination programmes in low- and middle-income countries. Tropical Medicine & International Health. 2016 Sep;21(9):1086–98.

8. de Graaf WF, Kretzschmar MEE, Teunis PFM, Diekmann O. A two-phase within-host model for immune response and its application to serological profiles of pertussis. Epidemics. 2014 Dec;9:1–7.

9. Teunis PFM, van Eijkeren JCH, Ang CW, van Duynhoven YTHP, Simonsen JB, Strid MA, et al. Biomarker dynamics: Estimating infection rates from serological data. Stat Med. 2012 Sep 10;31(20):2240–8.

10. Teunis PFM, van Eijkeren JCH, de Graaf WF, Marinović AB, Kretzschmar MEE. Linking the seroresponse to infection to within-host heterogeneity in antibody production. Epidemics [Internet]. 2016 Sep;16:33–9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1755436516300135

11. Teunis PFM, Wang Y, Aiemjoy K, Kretzschmar M, Aerts M. Estimating seroconversion rates accounting for repeated infections by approximate Bayesian computation. Stat Med. 2023 Dec 10;42(28):5160–88.

12. Teunis P, van Eijkeren J. Estimating Infection Rates from Serological Data. R package version 2.0.0. Archived May 23, 2022 [Internet]. 2018. Available from: https://CRAN.R-project.org/package=seroincidence

13. Garrett DO, Longley AT, Aiemjoy K, Yousafzai MT, Hemlock C, Yu AT, et al. Incidence of typhoid and paratyphoid fever in Bangladesh, Nepal, and Pakistan: results of the Surveillance for Enteric Fever in Asia Project. Lancet Glob Health. 2022 Jul;10(7):e978–88.

14. Aiemjoy K, Seidman JC, Saha S, Munira SJ, Islam Sajib MS, Sium SM Al, et al. Estimating typhoid incidence from community-based serosurveys: a multicohort study. Lancet Microbe. 2022 Aug 1;3(8):e578–87.

15. Phan Truong Quynh A, Pham Nguyen The N, Bui Thanh L, Huynh T, Ong T, Choisy M. serosv: Model Infectious Disease Parameters from Serosurveys. [Internet]. 2025 [cited 2025 Jan 29]. Available from: https://oucru-modelling.github.io/serosv/

16. Menezes A, Takahashi S, Routledge I, Metcalf CJE, Graham AL, Hay JA. serosim: An R package for simulating serological data arising from vaccination, epidemiological and antibody kinetics processes. PLoS Comput Biol. 2023 Aug 14;19(8):e1011384.

17. Hozé N, Pons-Salort M, Metcalf CJE, White M, Salje H, Cauchemez S. RSero: A user-friendly R package to reconstruct pathogen circulation history from seroprevalence studies. PLoS Comput Biol. 2025 Feb 3;21(2):e1012777.

18. Hodgson D, Hay J, Jarju S, Jobe D, Wenlock R, de Silva TI, et al. serojump: A Bayesian tool for inferring infection timing and antibody kinetics from longitudinal serological data. 2025.

19. Aiemjoy K, Katuwal N, Vaidya K, Shrestha S, Thapa M, Teunis P, et al. Estimating the Seroincidence of Scrub Typhus using Antibody Dynamics after Infection. Am J Trop Med Hyg. 2024 Aug 7;111(2):267–76.

20. Aiemjoy K, Rumunu J, Hassen JJ, Wiens KE, Garrett D, Kamenskaya P, et al. Seroincidence of Enteric Fever, Juba, South Sudan. Emerg Infect Dis. 2022 Nov;28(11).

Figure captions

Figure 1: Code and visualization of log antibody responses from three study sites. a) Code for importing the three required inputs. b) Code for visualizing cross-sectional antibody data and longitudinal seroresponse. c) Plot for cross-sectional curve parameters. d) Plot for longitudinal seroresponse parameters.

Alt Text: R code and output depicting data import and visual inspection in subfigures labeled from a to d.

Figure 2: Estimation of seroincidence rates stratified by country and age category. a) Code for executing and summarizing results of the `est_seroincidence_by()` function. b) Output from `summary()` for stratified seroincidence estimates. c) Code for plotting seroincidence rate estimates by country and age category as a bar plot. d) Output from `autoplot()` for visualizing seroincidence by strata.

Alt Text: R code and results showing seroincidence estimation, results summary, and visualization in subfigures labeled from a to d.

Supplementary material for: '**serocalculator**, an R package for estimating seroincidence from cross-sectional serological data'

Table of Contents

S1. Dataset Formatting Requirements 2
S2. Further Details on Longitudinal Seroresponse Parameters 4
S3. Further Details on Noise Parameters 9
S4. Description of Use Case Dataset 10

S1. Dataset Formatting Requirements

The table shells below are examples of formatted input datasets for each of the three required inputs for **serocalculator**. Any of the inputs may include variables to stratify by, but only strata in the cross-sectional population data will produce stratified seroincidence estimates. Noise parameters and longitudinal seroresponse parameters may be used by strata or as overall values applied equally to each cross-sectional strata.

Table S1.1 Cross-sectional population data

id	age	antigen_iso	value	strata1
1001		A		1
1001		B		1
1002		A		1
1002		B		1
1003		A		2
1003		B		2

- **id** = Unique identifier for each individual
- **age** = Numeric age in years (can include decimals)
- **antigen_iso** = Antigen and antibody isotype pair(s)
- **value** = Quantitative antibody response value
- **strata variables** (optional) = Any demographic features or covariate relevant to your analysis (must match those in other inputs to be used in analysis)

Table S1.2: Noise parameters

antigen_iso	y.low	y.high	eps	nu	strata1
A					1
A					2
B					1
B					2

- **antigen_iso** = Antigen and antibody isotype pair(s)
- **y.low** = Lower limit of detection of the assay used to produce the cross-sectional survey data set (a)
- **y.high** = Upper limit of detection of the assay
- **eps** = Measurement noise of the assay; CV%, range 0-1 (See Supplement S3)
- **nu** = Biologic noise of the assay; 95% of the distribution among those never exposed (See Supplement S3)
- **strata variables** (optional) = any demographic features or covariates relevant to your analysis (must match or be a subset of the stratifying variables included in the cross-sectional data)

Table S1.3: Longitudinal seroresponse parameters

antigen_iso	y0	y1	t1	alpha	r	strata1
A						1
A						2
B						1
B						2

- **antigen_iso** = Antigen and antibody isotype pair(s) measured
- **y0** = Baseline antibody concentration (U/mL)
- **y1** = Peak antibody concentration (U/mL)
- **t1** = Time to peak antibody concentration (days)
- **alpha** = Antibody decay rate (1/days)
- **r** = Antibody decay shape
- **strata variables** (optional) = any demographic features or covariate relevant to your analysis (must match or be a subset of the stratifying variables included in the cross-sectional data)

S2. Further Details on Longitudinal Seroresponse Parameters

The longitudinal seroresponse parameters provided by the user in Table S1.1 are components of the within-host seroresponse model proposed in de Graaf et al. 2014 and extended in Teunis et al. 2016 and Diekmann et al. 2018.[1-3] The model consists of two phases: an active-infection/antibody growth phase and a post-infection/antibody decay phase. The active infection phase of the model is a modified version of the Lotka-Volterra model,[4,5] which is typically used to describe predator-prey dynamics in ecological systems. In **serocalculator**, this model is adapted to represent the competitive relationship between the host's adaptive immune system cells (predators) and the pathogen (prey), where the current rate of change in the concentration of each population depends on the current concentration of the other population. The general form of the Lotka-Volterra model is shown below:

Rate of change in Population X (prey): $x'(t) = \alpha x(t) - \beta x(t)y(t)$
Rate of change in Population Y (predator): $y'(t) = -\gamma y(t) + \delta x(t)y(t)$

Parameters:
$x(t)$ = size of prey population at time t
$y(t)$ = size of predator population at time t
α = prey's maximum per capita growth rate
β = effect of predators on prey death rate
γ = predator's per capita death rate
δ = effect of prey on predator's growth rate

The seroresponse model simplifies the Lotka-Volterra model to represent the immune response during the active infection period.[1,3] In the seroresponse model, the following differential equations represent the dynamics of pathogen growth, antibody production, and antibody decay (waning) after the end of the active infection, when the pathogen has been eliminated:

Pathogen growth during infection: $b'(t) = \mu_0 b(t) - cy(t)$
Antibody growth during infection: $y'(t) = \mu_1 y(t)$
Antibody decay after infection: $y'(t) = -\alpha y(t)^r$

Parameters:
$b(t)$ = pathogen concentration at time t
$y(t)$ = antibody concentration at time t
μ_0 = pathogen growth rate
μ_1 = antibody growth rate
c = rate of pathogen elimination
α = Antibody decay rate (1/days)
r = Antibody decay shape

The model switches from the antibody growth (active infection) phase to the antibody decay (post-infection) at the time point when the pathogen concentration $b(t)$ first reaches 0. This time point is denoted t_1, and can be calculated as:

$$t_1 = \frac{1}{\mu_1 - \mu_0} \log(1 + \frac{(\mu_1 - \mu_0) b(0)}{c\, y(0)}$$

During the growth phase, the antibody concentration grows exponentially and can be calculated for any given time t as:

$$y(t) = y(0) e^{\mu_1 t}$$

In the Teunis et al. 2016 model, the parameters $\mu_0, c, \mu_1, d,$ and r and initial conditions $b(0)$ and $y(0)$ are allowed to vary between individuals, and potentially between repeated infections of the same individuals. After recruiting a longitudinal cohort of infected individuals and collecting longitudinally-repeated blood specimens over time from each individual, we can estimate each participant's individual-specific parameters as well as the underlying population distributions of those parameters, using a multilevel (a.k.a. "hierarchical") random-effects longitudinal modeling framework[6] and a Bayesian inferential approach.[7]

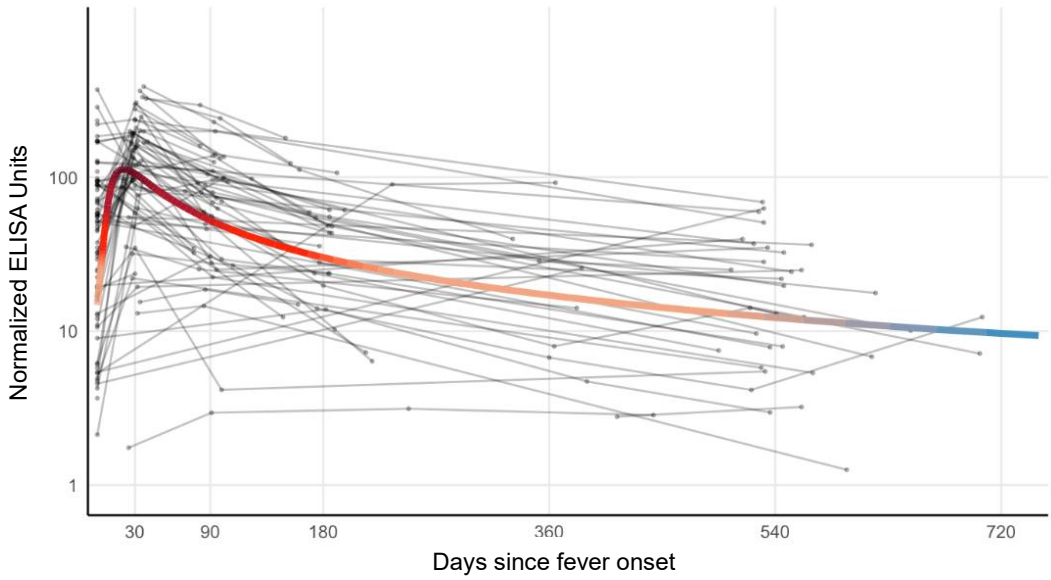
The result of the longitudinal model is a two-phase response curve of antibody dynamics (Figure S2.1). Beginning at time t_0, or the initial time of infection, there is an initial exponential growth phase until the peak antibody response (t_1). After t_1, we model antibody decay using a power function, with the parameters listed below:

Growth phase: $y'(t) = \mu y(t), \;\; t \in [0, t_1]$

Decay phase: $y'(t) = -y_1^r \alpha [1 + (r+1) y_1^{r-1} \alpha (t - t_1)]^{-\frac{r}{r-1}}, \;\; t \in [t_1, \infty)$

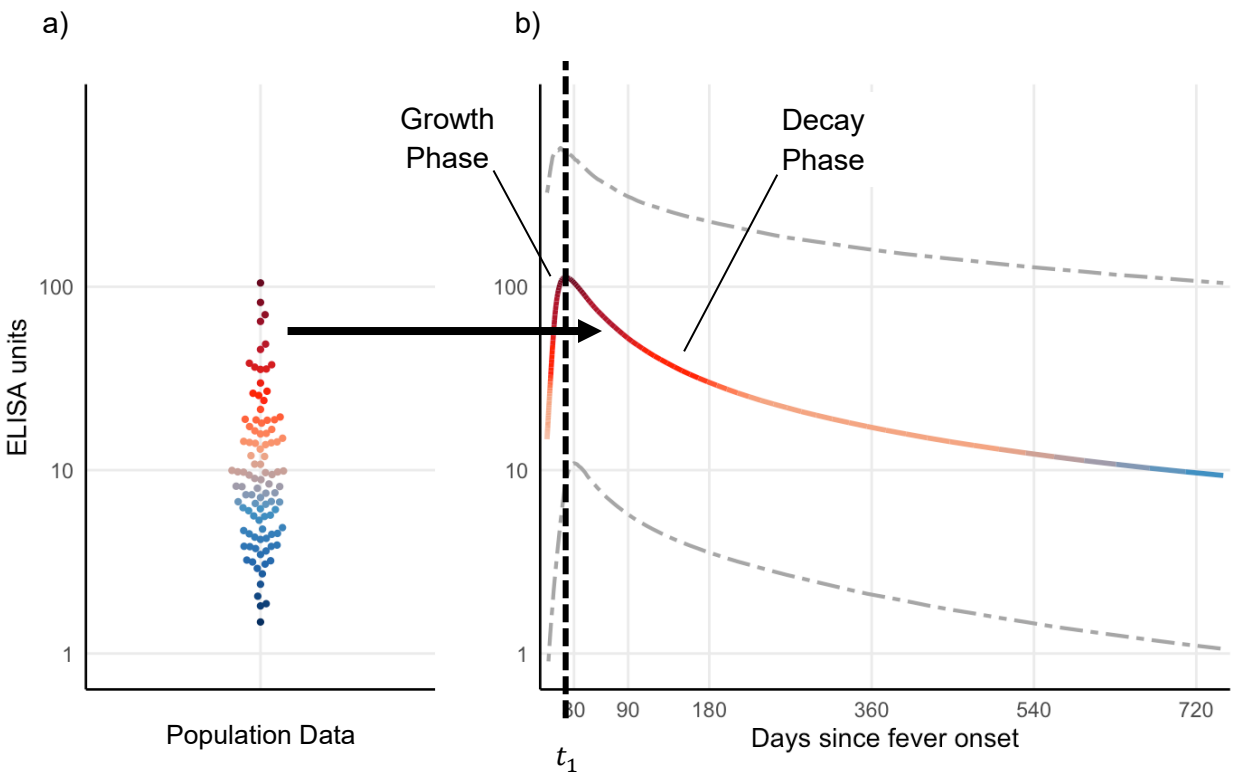
Parameters:
y_0 = Baseline antibody concentration (U/mL); i.e., $y_0 = y(0)$
y_1 = Peak antibody concentration (U/mL); i.e., $y_1 = y(t_1)$
μ = Antibody growth rate
t_1 = Time to peak antibody concentration (days)
α = Antibody decay rate (1/days)
r = Antibody decay shape

Figure S2.1: Grey dots and lines depict observed longitudinal antibody responses measured from confirmed enteric fever patients, each dot is a measured value and the lines connect individuals. The colored line is the median modeled antibody responses, red indicates higher quantitative antibody responses and blue indicates lower responses. Antibody responses are net (minus) blank and normalized to plate-specific positive controls (Normalized Elisa Units).



When we subsequently use the fitted seroresponse model to estimate seroincidence from a cross-sectional population survey sample, each participant in the cross-sectional survey contributes a single observation at a random point in the infection process (Figure S2.2). Each individual's seroresponse at that point in time is determined by their particular seroresponse parameters. When computing the seroincidence estimate's standard error, **serocalculator** incorporates both the between-person variation in seroresponse and the estimation error from the pre-estimated longitudinal model.

Figure S2.2: a) Sample cross-sectional population data with low (blue) to high (red) antibody responses. b) Two-phase modeled longitudinal seroresponse curve with colors corresponding to distant vs. recent infection.



References and Further Reading:

1. de Graaf WF, Kretzschmar MEE, Teunis PFM, Diekmann O. A two-phase within-host model for immune response and its application to serological profiles of pertussis. Epidemics. 2014;9:1-7. doi:10.1016/j.epidem.2014.08.002
2. Teunis PFM, van Eijkeren JCH, de Graaf WF, Marinović AB, Kretzschmar MEE. Linking the seroresponse to infection to within-host heterogeneity in antibody production. Epidemics. 2016;16:33-39. doi:10.1016/j.epidem.2016.04.001
3. Diekmann O, de Graaf WF, Kretzschmar MEE, Teunis PFM. Waning and boosting: on the dynamics of immune status. J Math Biol. 2018;77(6-7):2023-2048. doi:10.1007/s00285-018-1239-5
4. Lotka AJ. Analytical Note on Certain Rhythmic Relations in Organic Systems. Proceedings of the National Academy of Sciences. 1920;6(7):410-415. doi:10.1073/pnas.6.7.410
5. Volterra V. Fluctuations in the Abundance of a Species considered Mathematically. Nature. 1926;118(2972):558-560. doi:10.1038/118558a0
6. Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. Biometrics. 1982;38(4):963. doi:10.2307/2529876
7. Bayes T. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. Philos Trans R Soc Lond. 1763;53:370-418. doi:10.1098/rstl.1763.0053
8. Aiemjoy K, Seidman JC, Saha S, et al. Estimating typhoid incidence from community-based serosurveys: a multicohort study. Lancet Microbe. 2022;3(8):e578-e587. doi:10.1016/S2666-5247(22)00114-8
9. Simonsen J, Mølbak K, Falkenhorst G, Krogfelt KA, Linneberg A, Teunis PFM. Estimation of incidences of infectious diseases based on antibody measurements. Stat Med. 2009;28(14):1882-1895. doi:10.1002/sim.3592

S3. Further Details on Noise Parameters

Seroincidence estimates can be biased by both systematic and random error (e.g. noise).[1] **serocalculator** is designed to incorporate two sources of noise: biological noise from cross-reactivity or non-specific binding, and measurement noise during laboratory analysis. In the package, biologic noise is denoted by ν ("nu") and measurement noise is denoted by ε ("epsilon").[1]

The biological noise, ν, represents error from cross-reactivity to other antibodies. Users of **serocalculator** should estimate ν using the 95th percentile of the distribution of antibody responses to the antigen-isotype in a population with no exposure. Biological noise is assumed to be random and is always positive.

Measurement noise, ε, represents laboratory assay variability from the laboratory testing process. Users of **serocalculator** should estimate ε using the CV (coefficient of variation; i.e., the ratio of the standard deviation to the mean) for replicate measurements of the same biological samples.[2] Note that the CV should ideally be estimated using replicates across plates rather than within the same plate. Measurement noise can be positive or negative and can be reduced by running laboratory assays in duplicate and taking the mean.

The full details of these noise parameters were published by Teunis and Eijkeren in 2020 and specific noise estimation procedures were described by Aiemjoy et al. in 2022.[1,2]

References:

1. Teunis PFM, van Eijkeren JCH. Estimation of seroconversion rates for infectious diseases: Effects of age and noise. *Stat Med*. 2020;39(21):2799-2814. doi:10.1002/sim.8578
2. Aiemjoy K, Seidman JC, Saha S, et al. Estimating typhoid incidence from community-based serosurveys: a multicohort study. *Lancet Microbe*. 2022;3(8):e578-e587. doi:10.1016/S2666-5247(22)00114-8

S4. Description of Use Case Dataset

Enteric fever remains a major cause of morbidity and mortality in south Asia, southeast Asia, and sub-Saharan Africa, particularly for children.[1] It is caused by systemic infection with *Salmonella enterica* serovars Typhi and Paratyphi, which typically produce non-specific symptoms of fever, headache, and malaise. Complications can include intestinal perforation, gastrointestinal bleeding, altered mental state, septic shock, and death.[2,3] Although vaccines are available, many countries do not have the evidence of enteric fever burden to justify vaccine introduction. Estimates of seroincidence rates can provide the necessary data for vaccine introductions, while also specifying the population groups that may experience the highest risks of infection.

In the use case, we elect to estimate our seroincidence based on both IgA and IgG antibody responses to the Hemolysin E (HlyE) antigen (in column 'antigen_iso'), which have been shown to reliably distinguish enteric fever from other invasive bacterial infections.[4] Data for the use case comes from the SeroEpidemiology and Environmental Surveillance (SEES) for enteric fever study, which was conducted in Bangladesh, Nepal, and Pakistan.[5] It includes quantitative antibody responses from blood culture-confirmed enteric fever cases from the prospective clinical surveillance study, Surveillance for Enteric Fever in Asia Project (SEAP) that ran from 2016 to 2019 across sites in Bangladesh, Nepal, and Pakistan,[6] as well as population-based cross-sectional serologic samples collected between 2019-2021 from the same catchment areas.[5] The study sought to evaluate whether new diagnostic serological markers for enteric fever could reliably estimate population-level incidence.

In the population-based arm, households were selected using a single-stage, cluster random sampling method.[7] A grid was laid onto a map of each catchment area, dividing each area into 1000-2500 cells. Densely populated areas were further broken down into more cells of smaller size. Cells were then randomly selected from each catchment area and a field team member approached every household in that cell. Individuals were randomly selected using an age-stratified sample (age 0–4 years, age 5–9 years, age 10–15 years, and age 16–25 years). Participants were followed up to three times approximately 6 months apart.

The SEES study demonstrated that when combined with longitudinal antibody dynamics models, cross-sectionally collected antibody responses to HlyE and LPS could reliably estimate seroconversion rates and capture differences in seroconversion that correlated with clinical incidence estimates.[5] The available data from this study include quantitative antibody responses (IgA and IgG) to *S.* Typhi and *S.* Paratyphi antigens HlyE and LPS from both confirmed enteric fever patients and population-based participants from the same catchment areas. Full details of the SEAP and SEES studies have been previously elsewhere.[5-7]

References:

1. Stanaway JD, Reiner RC, Blacker BF, et al. The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Infect Dis. 2019;19(4):369-381. doi:10.1016/S1473-3099(18)30685-6
2. Pegues D, Miller S. Salmonellosis. In: Loscalzo J, Fauci A, Kasper D, Hauser S, Longo D, Jameson J, eds. Harrison's Principles of Internal Medicine, 21e. 21st ed. McGraw-Hill Education; 2022. Accessed September 17, 2024. https://accessmedicine.mhmedical.com/content.aspx?bookid=3095&sectionid=265422529
3. Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ. Typhoid Fever. New England Journal of Medicine. 2002;347(22):1770-1782. doi:10.1056/NEJMra020201
4. Andrews JR, Khanam F, Rahman N, et al. Plasma Immunoglobulin A Responses Against 2 Salmonella Typhi Antigens Identify Patients With Typhoid Fever. Clinical Infectious Diseases. 2019;68(6):949-955. doi:10.1093/cid/ciy578
5. Aiemjoy K, Seidman JC, Saha S, et al. Estimating typhoid incidence from community-based serosurveys: a multicohort study. Lancet Microbe. 2022;3(8):e578-e587. doi:10.1016/S2666-5247(22)00114-8
6. Garrett DO, Longley AT, Aiemjoy K, et al. Incidence of typhoid and paratyphoid fever in Bangladesh, Nepal, and Pakistan: results of the Surveillance for Enteric Fever in Asia Project. Lancet Glob Health. 2022;10(7):e978-e988. doi:10.1016/S2214-109X(22)00119-X
7. Andrews JR, Vaidya K, Saha S, et al. Healthcare Utilization Patterns for Acute Febrile Illness in Bangladesh, Nepal, and Pakistan: Results from the Surveillance for Enteric Fever in Asia Project. Clin Infect Dis. 2020;71(Suppl 3):S248-S256. doi:10.1093/cid/ciaa1321