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

Authors: Kristina W. Lai, Chris Orwa, Jessica C. Seidman, Denise O. Garrett, Samir K. Saha, Dipesh Tamrakar, Farah Naz Qamar, Richelle C. Charles, Jason R. Andrews, Peter Teunis, Kristen Aiemjoy*, Douglas Ezra Morrison*

**Co-senior authors*

Funding Statement:

The **serocalculator** software development was supported by the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) to KA and EM under award number R21AI176416. KA was supported by Fogarty International Center (FIC) of the NIH under award number K01TW012177.

Abstract:

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 error 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 exposure to a pathogen, the adaptive immune system generates antibodies that bind to 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 an individual was last infected.[2–4] In cross-sectional surveys, a large proportion of individuals with high antibody concentrations suggests a high burden of recent or frequent infection, while a smaller proportion indicates a 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 this approach is simple and widely used, it 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]

New statistical methods have been developed to estimate seroincidence rates from cross-sectional serosurveys 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, antibody isotype responses to multiple antigens, age-dependence, non-specific binding of the assay probes (noise), and measurement error.[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 a statistical software package, **serocalculator**, that implements the methodological advancements for estimating seroincidence rates from cross-sectional serosurveys.[1,8–10] Our goal is to provide a clear, validated, and user-friendly implementation of these methods, enabling users with varying skillsets to apply them in public health practice. This package supersedes the now defunct *seroincidence* package.[12] Simplified functionality is also available as an R Shiny App, a graphical user interface that walks users through the process of data import, data inspection, and seroincidence rate estimation without the need for coding in R. **serocalculator** offers a robust and accessible framework for estimating seroincidence from cross-sectional data. By integrating antibody kinetics with noise correction, it enables dynamic infection rate estimates using intuitive tools, including multi-marker support, stratified analysis, and an R Shiny interface for non-coders.

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 in the supplement and on the package website. The general workflow is as follows, 1) import data, 2) inspect and visualize data, and 3) calculate 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 (https://osf.io/ne8pc/) 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 that are 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 more or better longitudinal data become available, the validity of this assumption can be tested. Seroresponse parameters can be used for stratified estimation, such as in age- or country-specific seroresponse models.

Once again, 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() and then once again follow the process described above for the population data. 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 measurement error 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 population data. 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

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(), which will perform necessary column and formatting checks. 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 will define datasets for each of the three data inputs, and then they will specify one or more antigen isotype pair(s) to be included in the estimate. Users may also specify additional options for visualizations, warning messages, parallel computing, and the estimation process as desired. Stratified seroincidence rate estimates are produced by specifying one or more covariates in the cross-sectional population dataset, and, optionally, in the longitudinal seroresponse parameters.

Use Case

We will demonstrate the use of **serocalculator** with an example of enteric fever using a subset of data from a series of serosurveys conducted in Bangladesh, Nepal and Pakistan in 2016-2021.[12] 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 use case, we import noise parameters calculated at each study site laboratory.

R code

a) Import data

```
# Import sample cross-sectional data
xs_data <- "https://osf.io/download//n6cp3/" |>
  load_pop_data()

# Import longitudinal seroresponse parameters
curves <- "https://osf.io/download/rtw5k/" |>
  load_sr_params()

# Import sample noise parameters
noise <- url("https://osf.io/download//hqy4v/") |>
  load_noise_params()
```

b) Visualize data

```
# Visualize cross-sectional antibody data
xs_data |>
  autoplot(strata = "Country", type = "density",
log = TRUE)

# Visualize seroresponse parameters
curves |>  filter(antigen_iso %in% c("HlyE_IgA",
"HlyE_IgG")) |>
  autoplot()
```

Output

c) Cross-sectional antibody data

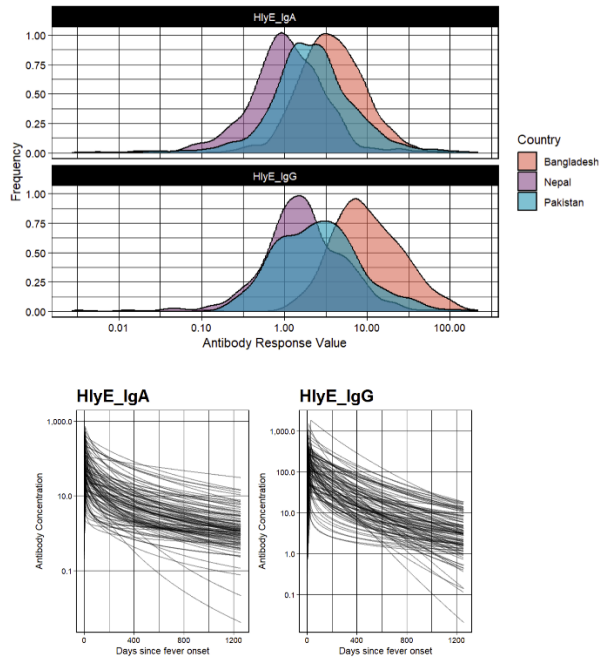

Figure 1: Code and visualization of log antibody responses from three study sites. a) Code for importing the three required inputs. b) Code for visualizing cross-sectional antibody data and longitudinal seroresponse. c) Plot for cross-sectional curve parameters. d) Plot for longitudinal seroresponse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

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.

R code

a) Estimate stratified seroconversion rate

```
# Estimate seroconversion rate
est_country_age = estimate_scr_by(
   strata = c("Country", "ageCat"),
   pop_data = xs_data,
   curve_params = curves,
   noise_params = noise,
   antigen_isos = c("HlyE_IgG", "HlyE_IgA")
)

# Summarize results and save as dataframe
est_country_agedf <-
summary(est_country_age)
```

b) Visualize stratified seroconversion rate

```
# Plot seroconversion rates as a bar plot
autoplot(
   est_country_agedf,
   type = "bar",
   yvar = "ageCat",
   color_var = "Country",
   CIs = TRUE
)
```

Output

c) Abbreviated summary output

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

d) Visualization

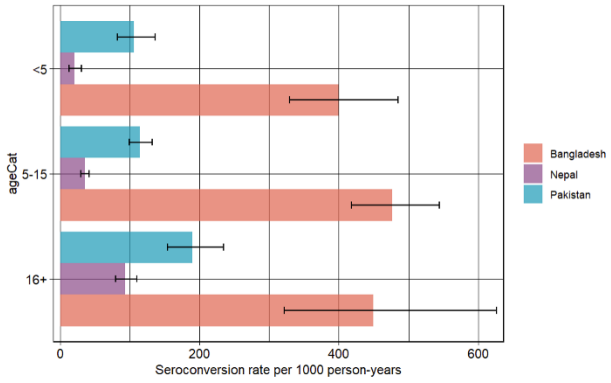

Figure 2: Estimation of seroincidence rates stratified by country and age category. a) Code for executing and summarizing results of the `est_seroincidence_by()` function. b) Code for plotting seroincidence rate estimates by country and age category as a bar plot.

Discussion

serocalculator provides a rapid and computationally simple method for calculating population-level seroincidence rates that account for antibody decay dynamics, cross-reactivity and measurement error 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.[13] *serosim* simulates serological study data to support study design and evaluate within-host dynamics and new analytic approaches.[14] 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.[15] *serojump* focuses on infection timing and antibody kinetics using Bayesian reversible jump MCMC and individual-level data.[16] 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 measurement noise.

Future enhancements to **serocalculator** will address methodological limitations and broaden its usability. Although the package currently accommodates interval censoring from some serological assays, its use is limited by the availability of longitudinal seroresponse parameters for specific pathogens and assays. To address this, we are developing a complementary R package that will allow users to model seroresponse parameters from their own data, enabling greater flexibility across study settings.

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 already been applied in seroepidemiologic studies of enteric fever and scrub typhus,[17–19] 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/.

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. 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

3. Aiemjoy K, Seidman JC, Charles RC, Andrews JR. Seroepidemiology for Enteric Fever: Emerging Approaches and Opportunities. *Open Forum Infect Dis*. 2023;10(Supplement_1):S21-S25. doi:10.1093/ofid/ofad021

4. Hay JA, Routledge I, Takahashi S. Serodynamics: A primer and synthetic review of methods for epidemiological inference using serological data. *Epidemics*. 2024;49:100806. doi:10.1016/j.epidem.2024.100806

5. Kamau E, Ante-Testard PA, Gwyn S, et al. Characterizing trachoma elimination using serology. Published online September 24, 2024. doi:10.1101/2024.09.20.24313635

6. Tedijanto C, Solomon AW, Martin DL, et al. Monitoring transmission intensity of trachoma with serology. *Nat Commun*. 2023;14(1):3269. doi:10.1038/s41467-023-38940-5

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;21(9):1086-1098. doi:10.1111/tmi.12737

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;9:1-7. doi:10.1016/j.epidem.2014.08.002

9. Teunis PFM, van Eijkeren JCH, Ang CW, et al. Biomarker dynamics: Estimating infection rates from serological data. *Stat Med*. 2012;31(20):2240-2248. doi:10.1002/sim.5322

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*. 2016;16:33-39. doi:10.1016/j.epidem.2016.04.001

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;42(28):5160-5188. doi:10.1002/sim.9906

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

13. Phan Truong Quynh A, Pham Nguyen The N, Bui Thanh L, Huynh T, Ong T, Choisy M. serosv: Model Infectious Disease Parameters from Serosurveys. Published online 2025. Accessed January 29, 2025. https://oucru-modelling.github.io/serosv/

14. 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;19(8):e1011384. doi:10.1371/journal.pcbi.1011384

15. 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;21(2):e1012777. doi:10.1371/journal.pcbi.1012777

16. Hodgson D, Hay J, Jarju S, et al. serojump: A Bayesian tool for inferring infection timing and antibody kinetics from longitudinal serological data. Published online March 5, 2025. doi:10.1101/2025.03.04.25323335

17. 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

18. Aiemjoy K, Katuwal N, Vaidya K, et al. Estimating the Seroincidence of Scrub Typhus using Antibody Dynamics after Infection. *Am J Trop Med Hyg*. 2024;111(2):267-276. doi:10.4269/ajtmh.23-0475

19. Aiemjoy K, Rumunu J, Hassen JJ, et al. Seroincidence of Enteric Fever, Juba, South Sudan. *Emerg Infect Dis*. 2022;28(11). doi:10.3201/eid2811.220239

20. 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

21. 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

22. 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

23. 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

24. 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

25. 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

26. 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

Supplement

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.

a) 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)

b) 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)

c) 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 Supplement S1 Table c 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 derives from 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:

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

Parameters: 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 prior studies by de Graaf et. al. proposed a simplified version of the Lotka-Volterra model to represent the immune response.[1,3] In that model, differential equations represent dynamics of the pathogen growth antibody production, and antibody decay (waning) following infection. General equations by de Graaf et al.:

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

Parameters: $b(t)$ = pathogen concentration
$y(t)$ = antibody concentration
μ_0 = pathogen growth rate
μ_1 = antibody growth rate
c = rate of pathogen elimination
d = rate of antibody decay

In the de Graaf et al. model, the parameters $\mu_0, c, \mu_1, d,$ and r and initial conditions $b(0)$ and $y(0)$ are assumed to vary between individuals, and potentially between repeated infections of the same individuals. By recruiting a longitudinal cohort of infected individuals and collecting repeated blood specimens, we are able to 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 growth phase that follows an exponential distribution until the peak antibody response (t_1), when antibody decay follows a power function with the parameters below:

Growth phase: $d'y_g(t) = \mu y_g(t), \ [0 \leq t < t_1]$

Decay phase: $d'y_d(t) = -y_1^r \alpha[1 + (r+1)y_1^{r-1}\alpha(t - t_1)]^{-\frac{r}{r-1}}, \ [t_1 \leq t < \infty]$

Parameters:
y_0 = Baseline antibody concentration (U/mL)
y_1 = Peak antibody concentration (U/mL)
μ = Antibody growth rate
t_1 = Time to peak antibody concentration (days)
α = Antibody decay rate (1/days)
r = Antibody decay shape

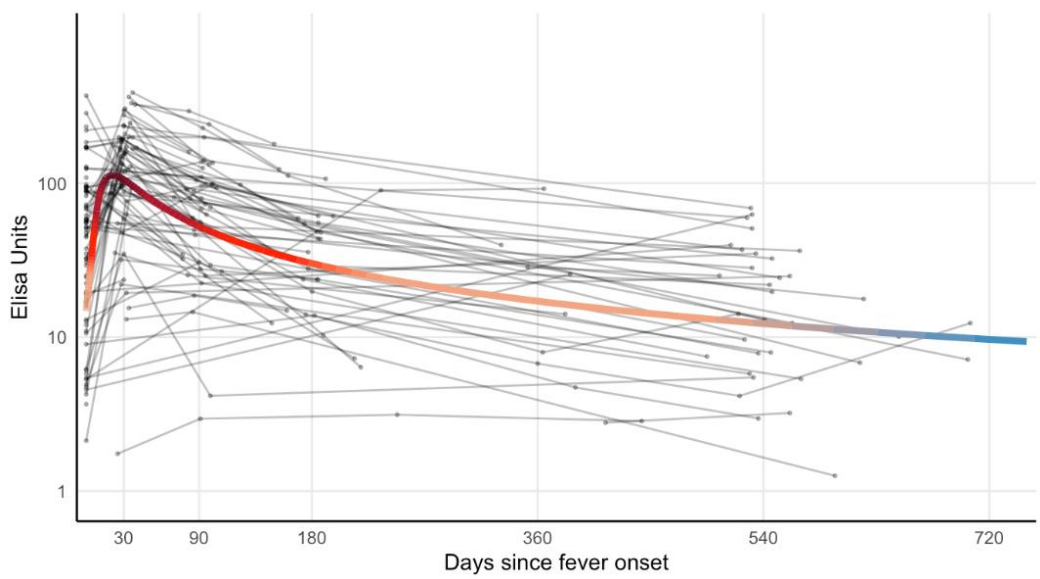

Figure S2.1: Sample longitudinal case data along with the estimated population mean of the seroresponse model.

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 S#.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.

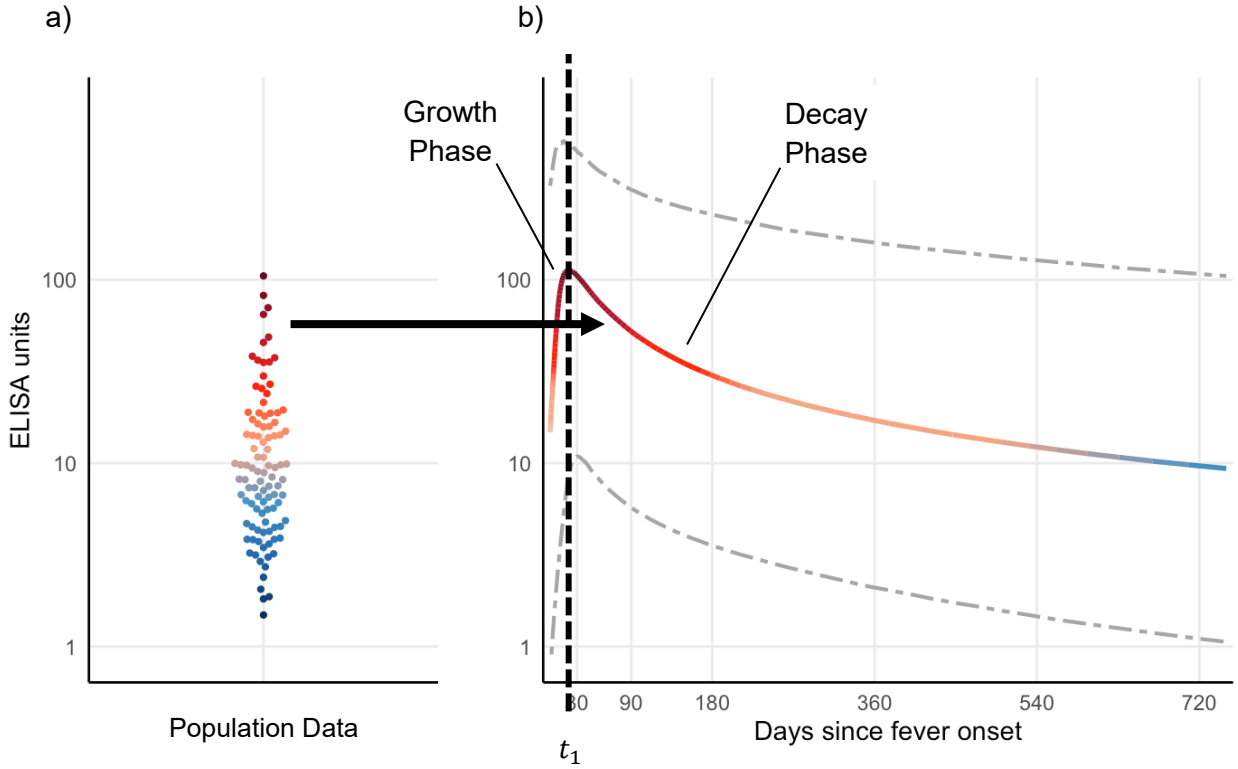

Figure S2.2: a) Sample cross-sectional population data with low (blue) to high (red) antibody responses. b) Two-phase modeled longitudinal seroresponse curve with colors corresponding to distant vs. recent infection.

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 measurement error 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

539 Education; 2022. Accessed September 17, 2024.
540 https://accessmedicine.mhmedical.com/content.aspx?bookid=3095&sectionid=2654225
541 29

542 3. Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ. Typhoid Fever. New England
543 Journal of Medicine. 2002;347(22):1770-1782. doi:10.1056/NEJMra020201

544 4. Andrews JR, Khanam F, Rahman N, et al. Plasma Immunoglobulin A Responses Against
545 2 Salmonella Typhi Antigens Identify Patients With Typhoid Fever. Clinical Infectious
546 Diseases. 2019;68(6):949-955. doi:10.1093/cid/ciy578

547 5. Aiemjoy K, Seidman JC, Saha S, et al. Estimating typhoid incidence from community-
548 based serosurveys: a multicohort study. Lancet Microbe. 2022;3(8):e578-e587.
549 doi:10.1016/S2666-5247(22)00114-8

550 6. Garrett DO, Longley AT, Aiemjoy K, et al. Incidence of typhoid and paratyphoid fever in
551 Bangladesh, Nepal, and Pakistan: results of the Surveillance for Enteric Fever in Asia
552 Project. Lancet Glob Health. 2022;10(7):e978-e988. doi:10.1016/S2214-109X(22)00119-
553 X

554 7. Andrews JR, Vaidya K, Saha S, et al. Healthcare Utilization Patterns for Acute Febrile
555 Illness in Bangladesh, Nepal, and Pakistan: Results from the Surveillance for Enteric
556 Fever in Asia Project. Clin Infect Dis. 2020;71(Suppl 3):S248-S256.
557 doi:10.1093/cid/ciaa1321