Serodynamics: A primer and synthetic review of methods for epidemiological inference using serological data

22 May 2025

Seroanalytics Training
Blantyre, Malawi

Reference

Epidemics 49 (2024) 100806

Contents lists available at ScienceDirect

Epidemics

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

Check for updates

Serodynamics: A primer and synthetic review of methods for epidemiological inference using serological data

James A. Hay [a,*,1], Isobel Routledge [b,*,1], Saki Takahashi [c,*,1]

[a] Pandemic Sciences Institute, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
[b] Department of Medicine, University of California San Francisco, San Francisco, CA, USA
[c] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Seroanalytics Hub

Infectious Disease DYNAMICS

Premise of this review

- This is a review and primer of methods to understand epidemiological dynamics and identify past exposures from serological data, referred to as **serodynamics**.

- Discussion of **processing and interpreting serological data** prior to fitting serodynamical models.

- Review of approaches for estimating epidemiological trends and past exposures, ranging from:
  - **serocatalytic models** applied to binary serostatus data, to
  - more **complex models** incorporating quantitative antibody measurements and immunological understanding.

- Discussion of key areas for methodological development to improve scientific discovery and public health insights in seroepidemiology.

Key public health and research questions typically addressed using serological data

1. What is an individual's **serostatus**?
    - In the majority of sero-epidemiological studies, serostatus is used to estimate attack rates, monitor progress towards elimination, and to estimate population susceptibility.

2. What can we learn about **immunological mechanisms** from measured biomarkers?
    - Immune responses are highly dynamic and variable, but can be described using mathematical models.
    - Serological data can be used to parameterize and predict the modeled kinetics of individual-level immunity and biomarkers.

3. Can serological data provide **correlates of protection**?
    - Serological measurements indicate pre-existing immunity and are correlated with the probability of infection or disease, and may explain more variation in outcomes than presence/absence of prior vaccination or infection.
    - Correlates of protection are not well characterized for many pathogens.

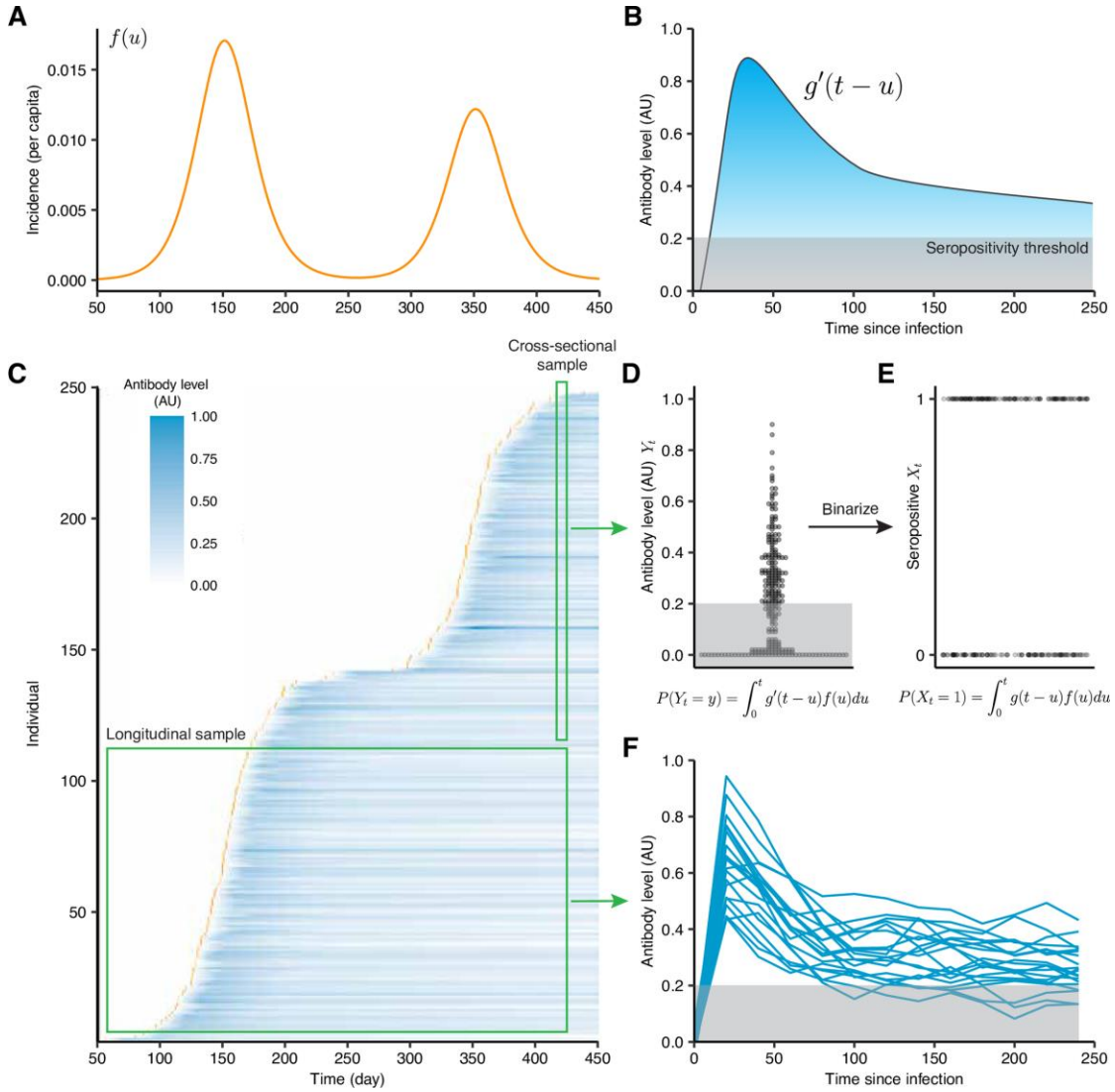
Key public health and research questions typically addressed using serological data

4. What is the **time-since-infection** of a sample?
    - Antibody levels follow predictable trajectories following infection, and thus the timing of infection can be projected backwards from one or more antibody measurements.

5. What were the **historical epidemiological dynamics** which led to the current serological landscape?
    - The distribution of antibody levels in the population at a given time reflects the convolution of past infection and vaccination dynamics in the population, with within-host immunological kinetics.

6. What is the **population immune landscape** for a particular pathogen?
    - The probability and size of an outbreak depends on both the transmissibility of an invading pathogen and the susceptibility of the population.

The data-generating process for serological data



Sources of biological variation impacting serological data

1. Pathogen factors
    - **Different pathogens** and **host immunological responses** exhibit different within-host kinetics (e.g., measles vs. tetanus antibody waning rates; response to acute vs. chronic infections).

2. Exposure type and dose
    - The **composition and dose of exposure** to an antigen determines the immunological response e.g., which antigens are included in a vaccine.
    - The **route of exposure** determines the location of antigenic stimulation (e.g., an intranasal vaccination may trigger a strong mucosal response whereas an intramuscular vaccination may not).

3. Host factors
    - **Host factors** such as age, sex, comorbidities, coinfection and being immunocompromised can affect biomarker kinetics.
    - **Genetic factors** such as human leukocyte antigen (HLA) class have also been shown to explain variation in humoral immunity.

Sources of biological variation impacting serological data

4. Individual-level variation
    - **Within-host kinetics are stochastic** such that two identical exposure events might lead to different biomarker kinetics.
    - Between-individual variation in kinetics **may be greater than can be explained** by measurable host factors.

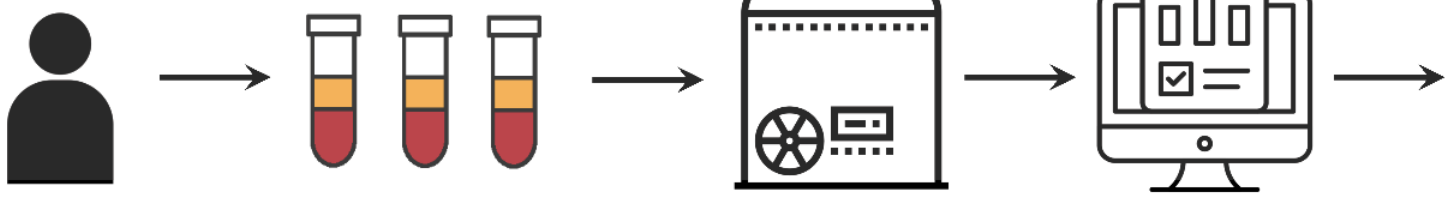
5. Infection history
    - Within-host kinetics depend on **pre-existing immunity**; e.g., IgG titers are boosted more quickly upon re-exposure, whereas IgM titers are only elevated following primary exposure.
    - Antibody responses following re-infection involve both de novo plasmablasts and memory B cells targeting multiple epitopes, which can lead to effects such as **immune imprinting**, **titer-ceiling effects**, and **antigenic seniority**.

6. Target biomarker
    - Antibodies to different antigens, different immunoglobulins, and different types of biomarkers (e.g., avidity) exhibit **different degrees of boosting and waning**.
    - This can provide **orthogonal information on time-since-infection**, but is also a potential pitfall when comparing kinetics parameters using different assays or targets.

Sources and pre-processing of serological data

Process

Collected specimens → **Serologic assay** → **Raw lab data** → **Quantitative data**



Sample	Titer (or MFI)
A	1:32
B	1:4
C	1:2

→ **Binary data**

Sample	Serostatus
A	1
B	0
C	0

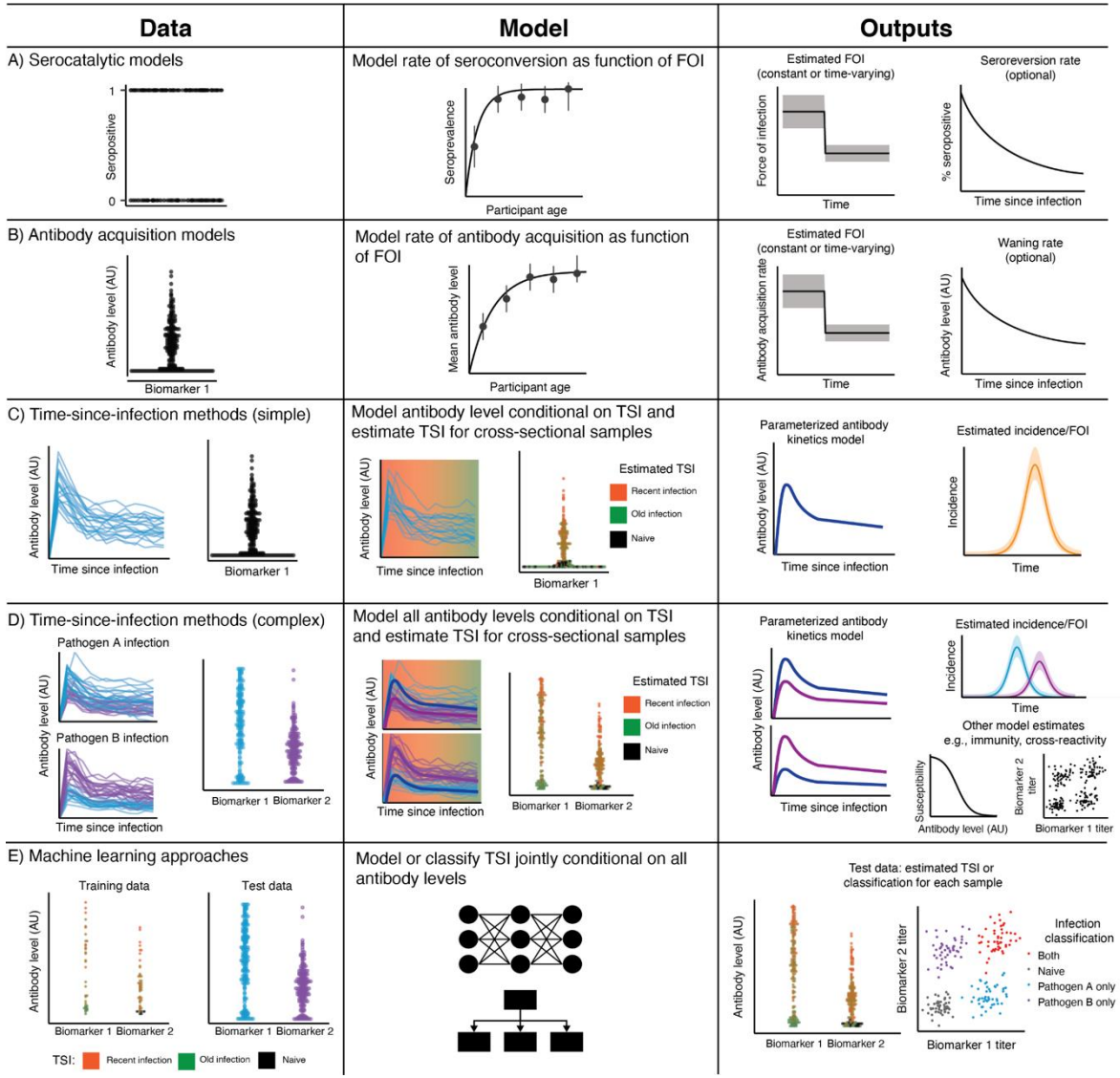
Considerations

Testing Procedures
Specimen type and storage conditions
Choice and biological/epidemiological interpretation of biomarker(s)
Choice of assay (test performance characteristics, assay noise)

Normalization & Standardization of Lab Data
Background noise correction (i.e., use normalization pipelines to adjust)
QC & within-lab batch effects (i.e., use controls to standardize)
Measurement agreement between replicates
Censoring (i.e., upper and lower limit of detection; interval-based titers)

Classification of Serostatus
Choice of control samples (positives, negatives; relevance to epidemiological characteristics in sample population)
Classification algorithm and choice of threshold

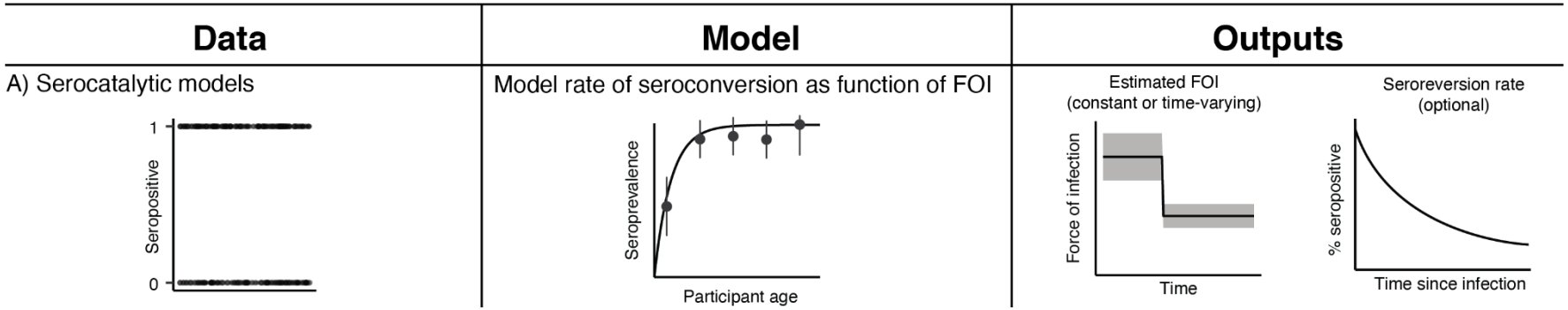
Serodynamics methods for reconstructing transmission dynamics



Binary data

Quantitative data

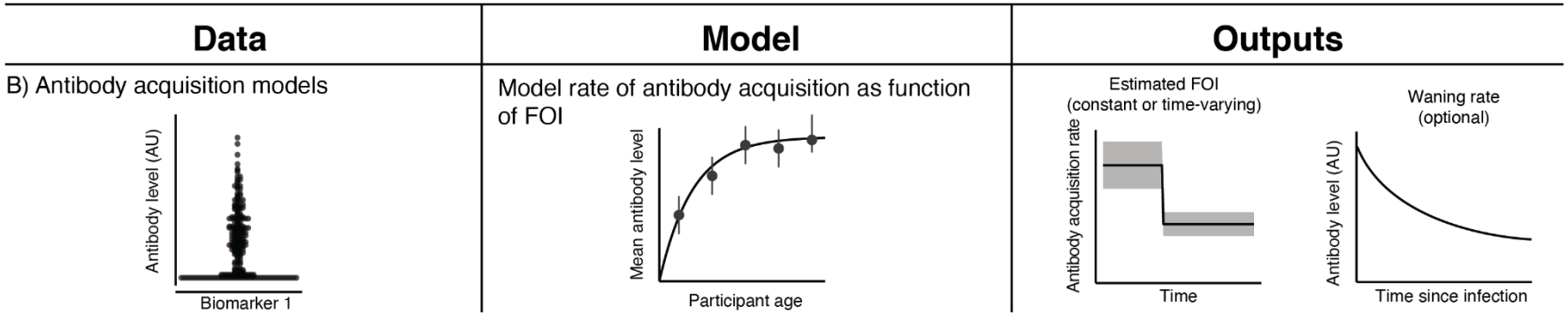
1. Serocatalytic models



- For pathogens that typically cause infrequent infections followed by durable or well-characterized immune responses, the per capita rate of new infections – the **force of infection (FOI)** – can be estimated by fitting a serocatalytic model to age-stratified seroprevalence data.

- Preview for Lecture 5 & Lab 5 ☺

2. Antibody acquisition models



- Antibody acquisition models assume that **antibody levels increase over time proportional to the force of infection**, without assigning serostatus.
- Acquisition models are mainly used in pathogen systems where **individuals experience multiple serological boosts over their life**, and thus seroprevalence becomes saturated over time, particularly when antibody decay rates are low.
- However, these models can be difficult to interpret and results between studies may not be generalizable.

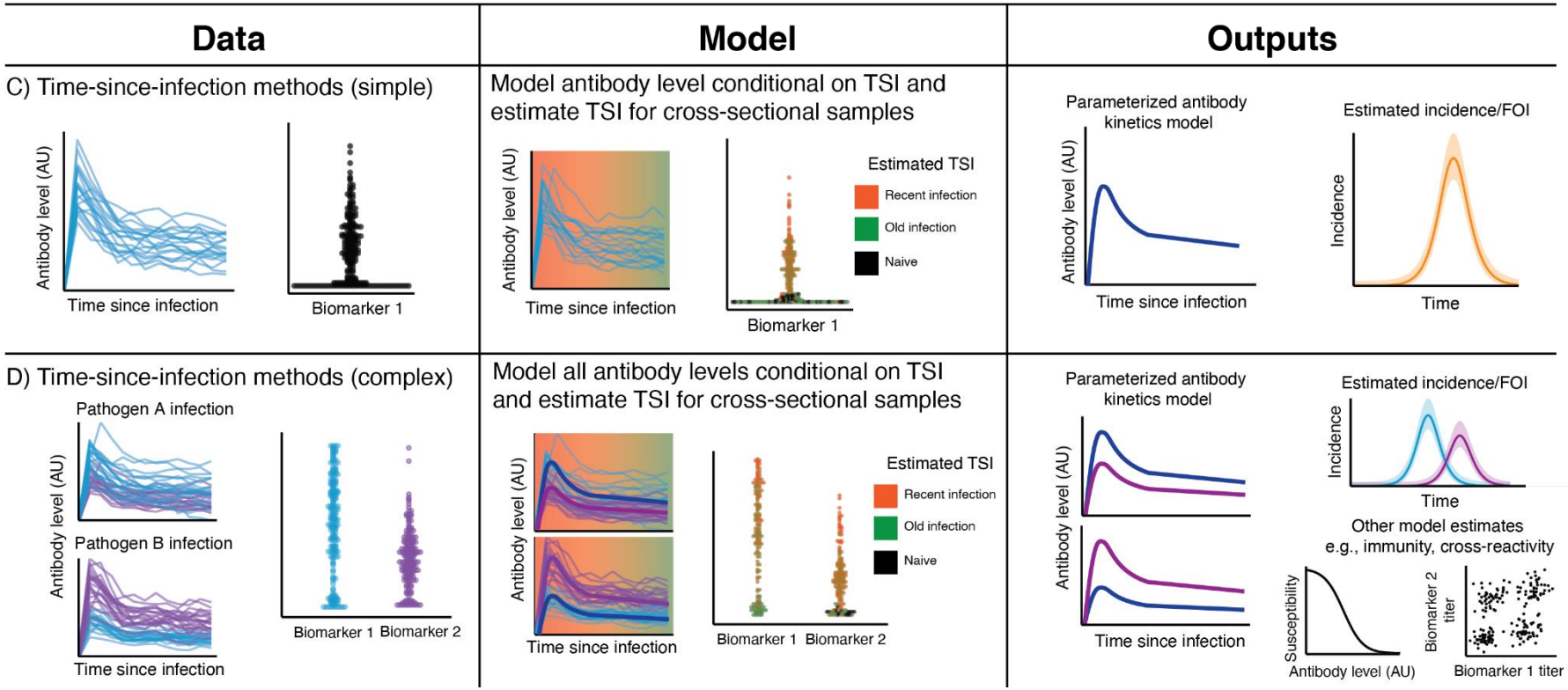
3. Time-since-infection methods

- An alternative approach to estimating epidemiological trends from quantitative serological data exploits the fact that **individual-level antibody levels change predictably over time following exposure**.

- When the within-host processes governing changes in biomarker levels over time-since-infection or time-since-vaccination are well-characterized, the **timing of an individual's infection or vaccination can be back-calculated** conditional on their observed biomarker levels from cross-sectional or longitudinal samples.

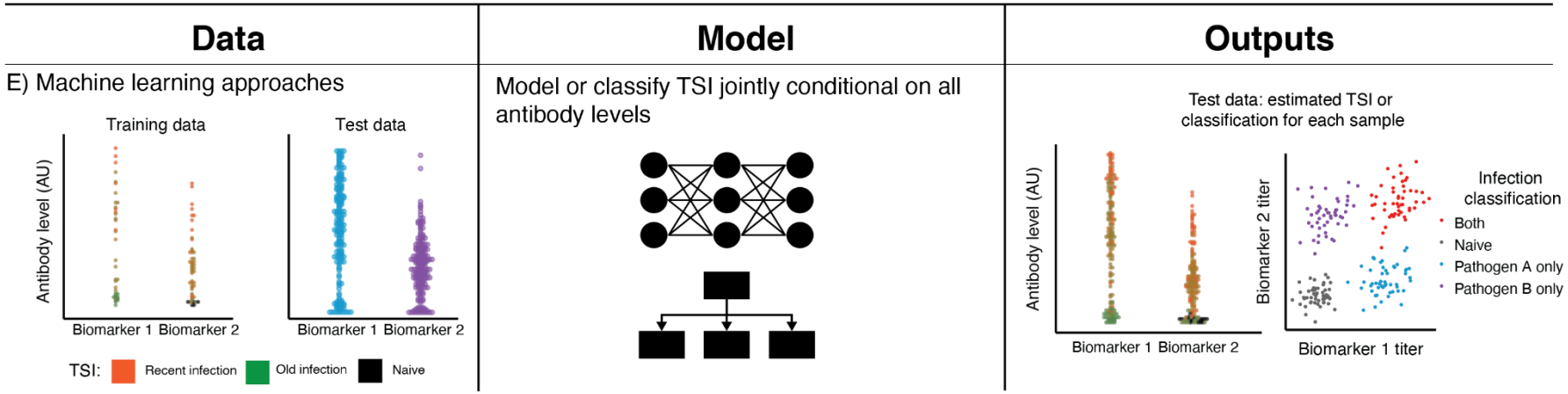
- Individual infection time estimates can then be combined to estimate **population-level trends**.

3. Time-since-infection methods



- Packages exist to implement these models; e.g., `serosolver`

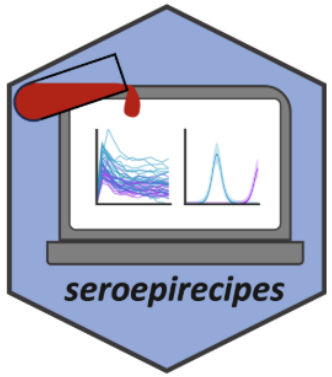
4. Machine learning approaches



- Many pathogens lack validated biomarkers for exposure recency, and thus which biomarkers to measure to back-calculate exposure timing may be unknown.
- In these cases, **supervised machine learning approaches** are useful for identifying minimal sets of biomarkers to predict or classify past infections and vaccinations.
- This is particularly important for pathogens where cross-reactivity with off-target pathogens is likely, or where substantial variation in serological signatures are expected, making serocatalytic models and time-since-infection methods very difficult to parameterize.
- Machine learning methods are also useful for analyzing **high-dimensional serological data**.

seroepirecipes 0.1.0 Reference Articles

seroepirecipes: implementations and tutorials of common models in seroepidemiology



seroepirecipes implements and links to R packages implementing commonly used mathematical and statistical models for analyzing serological data. The package implements a range of methods in R vignettes, from fitting antibody kinetics models to longitudinal antibody titer data, estimating the force of infection using serocatalytic models, and inferring infection histories using time-since-infection methods. This codebase accompanies a literature review of analytical methods for seroepidemiology.

All of the tutorials use either simulated datasets from the serosim R-package or publicly available datasets.

https://seroanalytics.org/seroepirecipes/