Lecture 1
Introduction to the Seroanalytics Workshop

21 May 2025

Seroanalytics Training
Blantyre, Malawi

Welcome from the instructor team!

Sonia Hegde (JHU)

Sophie Berube (Florida)

Sarah Lapidus (JHU)

Alyssa Sbarra (JHU)

Megan O'Driscoll (Geneva)

Mika Malouf (LSHTM)

Upendo Mseka (MLW)

Cebile Lekhuleni (NICD)

Goals of the course

- To learn the **core concepts** to pre-process, visualize, analyze, and interpret serological data.
- To learn the **applied techniques** to conduct the above steps using R.
- To apply learnings to participants' own **research questions** and **data sets**.
- To build a **network with other participants** from around the world who are also conducting serological data analyses.

Course Overview

Part 1: Introduction to the Seroanalytics Workshop

Part 2: Introduction to serological data analyses in R

Part 3: Pre-processing serological data

Part 4: Visualizing and standardizing serological data

Part 5: Determining serostatus and estimating seroprevalence

Part 6: Inferring transmission dynamics from seroprevalence data

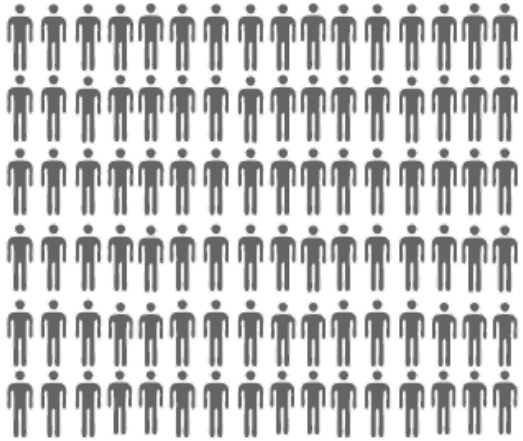
+ Time to work on **individual projects**
+ **Extra lectures** on advanced topics

Traditional surveillance data may provide an incomplete picture of transmission and disease burden

- Existence of **asymptomatic** or **sub-clinical** infections
- **Incomplete understanding** of factors leading to symptomatic disease
- Differences in **testing and reporting practices** between and within locations, and over time
- These may be complicated by **immunity**

Disease burden estimation largely relies on clinical surveillance

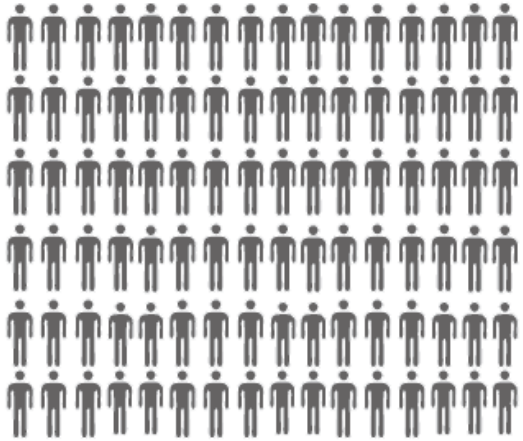
Infections



Serologic Cohort Data

Disease burden estimation largely relies on clinical surveillance

Infections



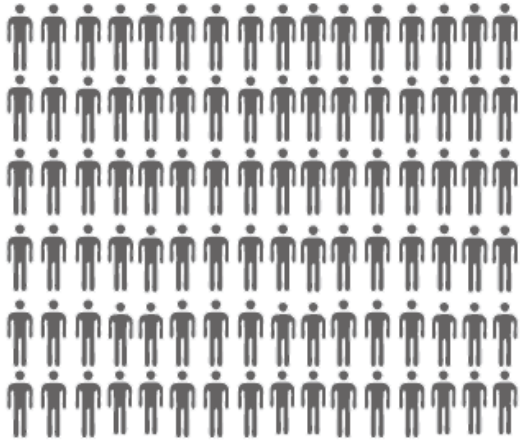
Serologic Cohort Data

Symptomatic

Surveillance Data

Disease burden estimation largely relies on clinical surveillance

Infections



Serologic Cohort Data

Symptomatic

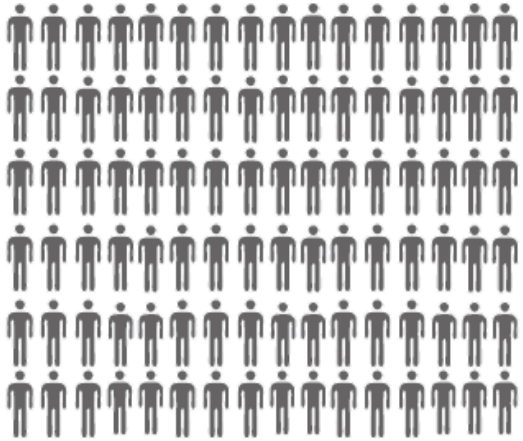
Surveillance Data

Seek care

Health Seeking Behavior from Surveys
Surveillance Data

Disease burden estimation largely relies on clinical surveillance

Infections



Serologic Cohort Data

Symptomatic

Surveillance Data

Seek care

Health Seeking Behavior from Surveys
Surveillance Data

Identified and reported

Surveillance Data

Use cases of serology at individual and population levels

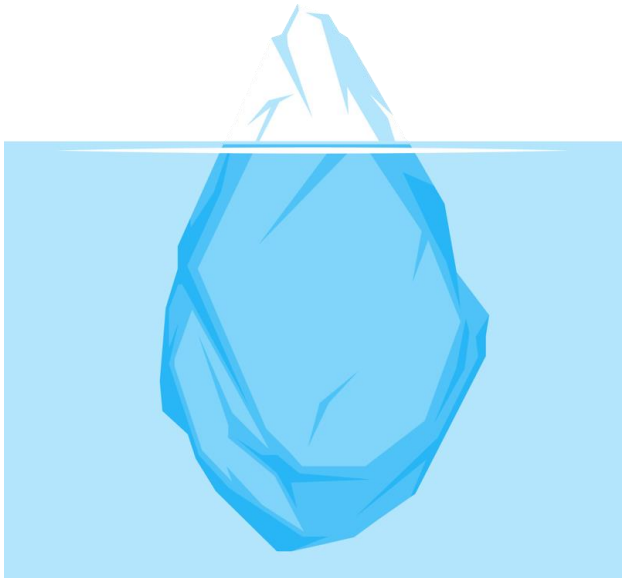
- At the **individual** level:
  - To indirectly diagnose recent or prior infection
  - To hypothetically ascertain who is protected
    - e.g., screening health care workers for measles antibody titers
  - To measure as a surrogate outcome in vaccine trials

- At the **population** level:
  - To quantify cumulative incidence or immunity in a population **(Part 5)**
  - To obtain biomarkers of additional metrics
    - e.g., recency or frequency of exposures (using quantitative responses)
  - To parametrize mechanistic transmission models

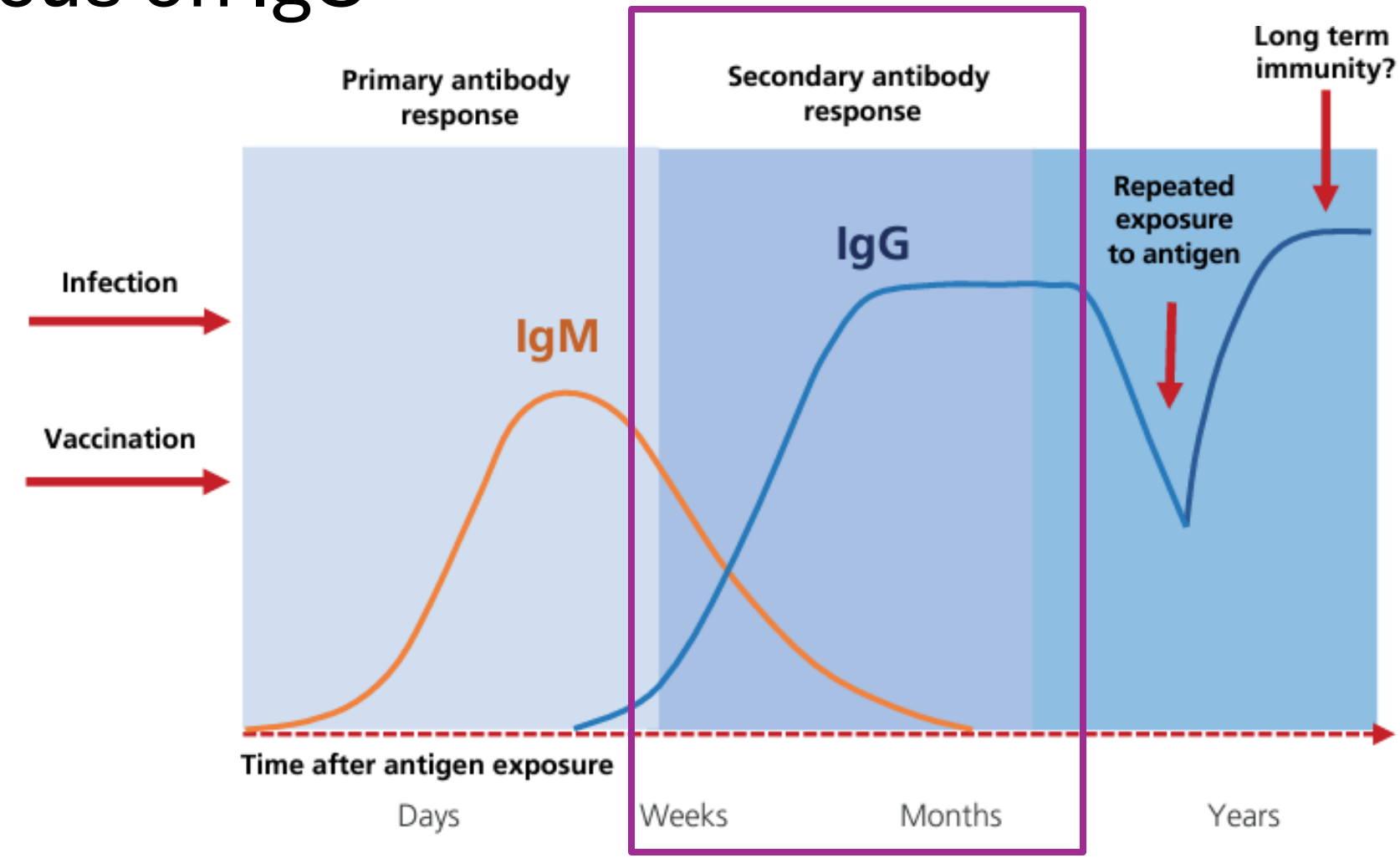
Serology is the gold standard to quantify infections in a population

- Antibodies allow for estimation of key disease transmission parameters
  - **Natural history**: case-ascertainment ratio, infection-fatality ratio
  - **Epidemiology**: attack rate, R0, force of infection **(Part 6)**

- However, this requires having an **appropriate serological assay** and **adequate interpretation**.



Kinetics of the antibody response: for this course, we focus on IgG



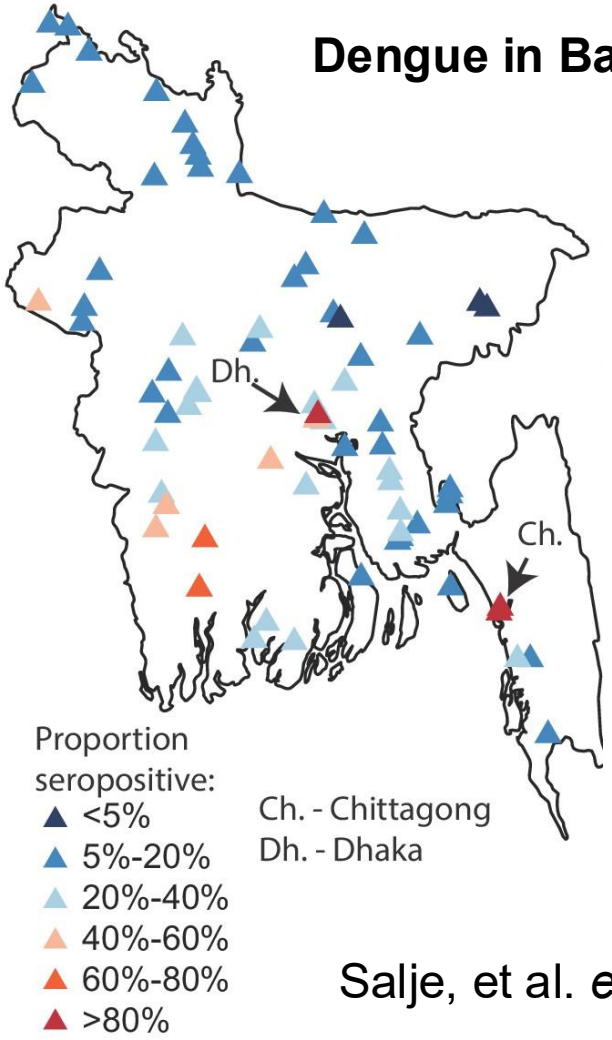
There are many public health use cases of sero-epidemiology

Use Case	Example Pathogens
1. Burden and distribution of infections	Campylobacter, Chagas, chikungunya, cholera, Cryptosporidium, cysticercosis, Giardia, neglected tropical diseases, Plasmodium species (some), strongyloidiasis, yaws, HIV
2. Identification of emerging and re-emerging infections	Ebola, Lassa, Marburg, Mpox, SARS-CoV-2, Zika
3. Identification of vaccine program reach or gaps	Measles, polio, rubella, SARS-CoV-2, yellow fever
4. Assessing infection changes due to behavioral, environmental, or pharmaceutical interventions or environmental changes	Chikungunya, dengue, malaria, PCV13 (must be able to distinguish between vaccine- and infection-derived immunity), Typhoid
5. Monitoring peri- and post-elimination surveillance settings	Guinea worm, human African trypanosomiasis, Lymphatic filariasis, malaria (sub-national levels), onchocerciasis, trachoma, visceral leishmaniasis, yaws
6. Research	Many!

1st Serosurveillance Summit meeting report, March 2023

Use case 1: Burden and distribution of infections

- Fills gaps in existing surveillance systems
- Identifies population levels of susceptibility or immunity
- Provides “true” number and distribution of infections
- Particularly useful for diseases with asymptomatic or mild infections, poor diagnostics for acute infections, and areas where there is poor access to care

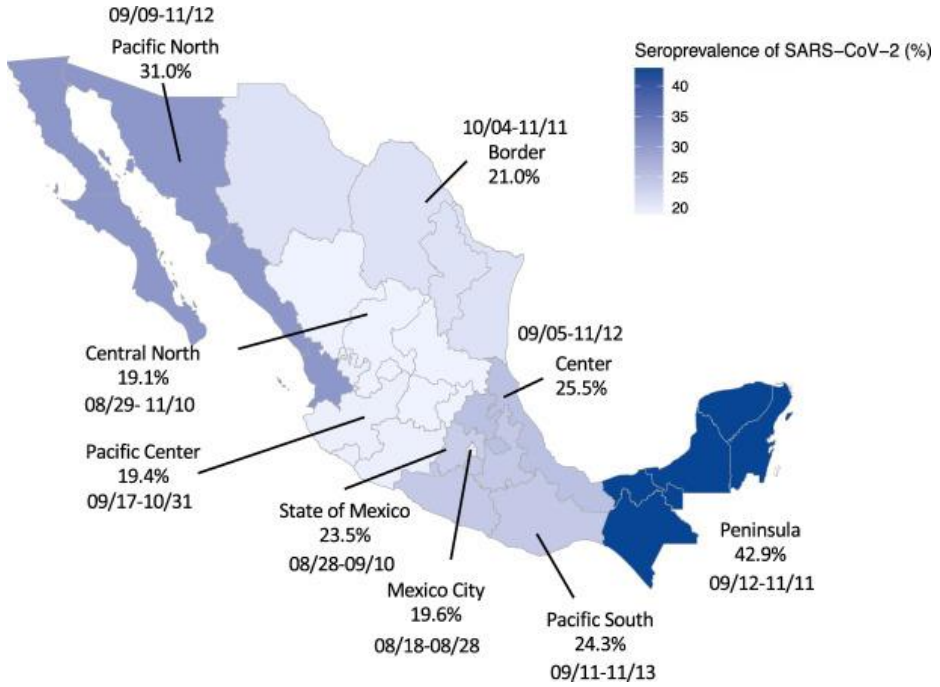


Salje, et al. *eLife* (2019)

Use case 2: Identification of emerging and re-emerging infections

- Allows identification of previously undetected infections
- Tracks outbreaks
- Monitors changes in types or magnitude of response required

SARS-CoV-2 in Mexico

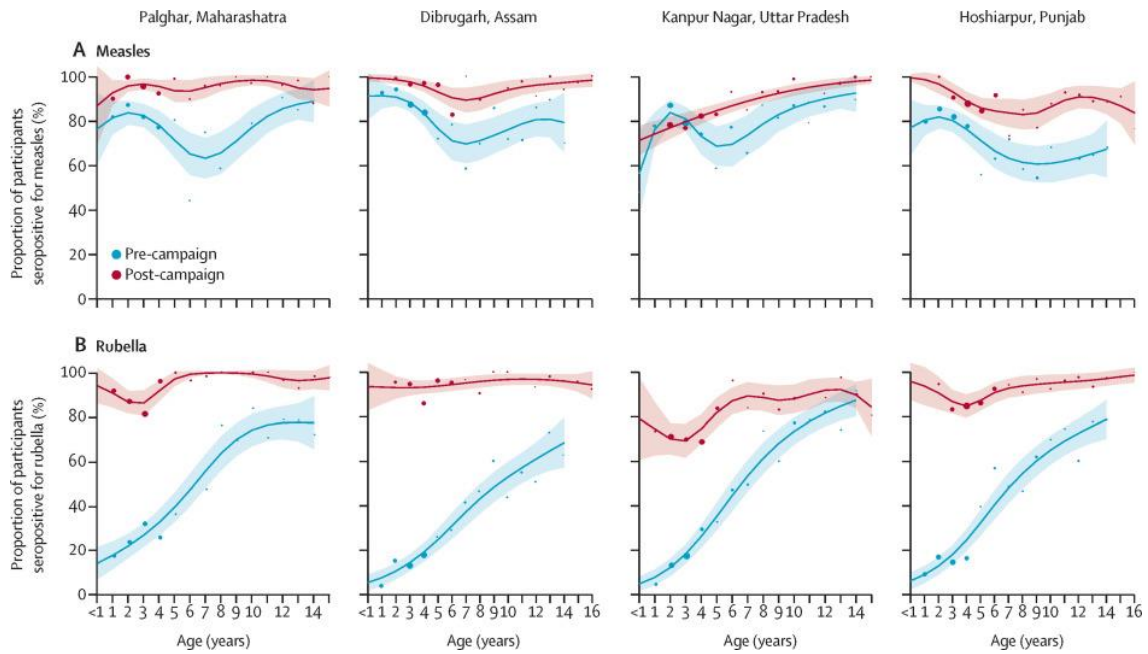


Basto-Abreu, et al. *Nat Comm* (2022)

Use case 3: Identification of vaccine program reach or gaps

- Geographic and demographic reach of vaccination or gaps in vaccination can be used:
  - To determine age of first routine vaccination
  - To estimate routine vaccination coverage
  - To identify the need for a vaccination campaign
  - To evaluate the impact of a vaccination campaign

Measles and rubella in India



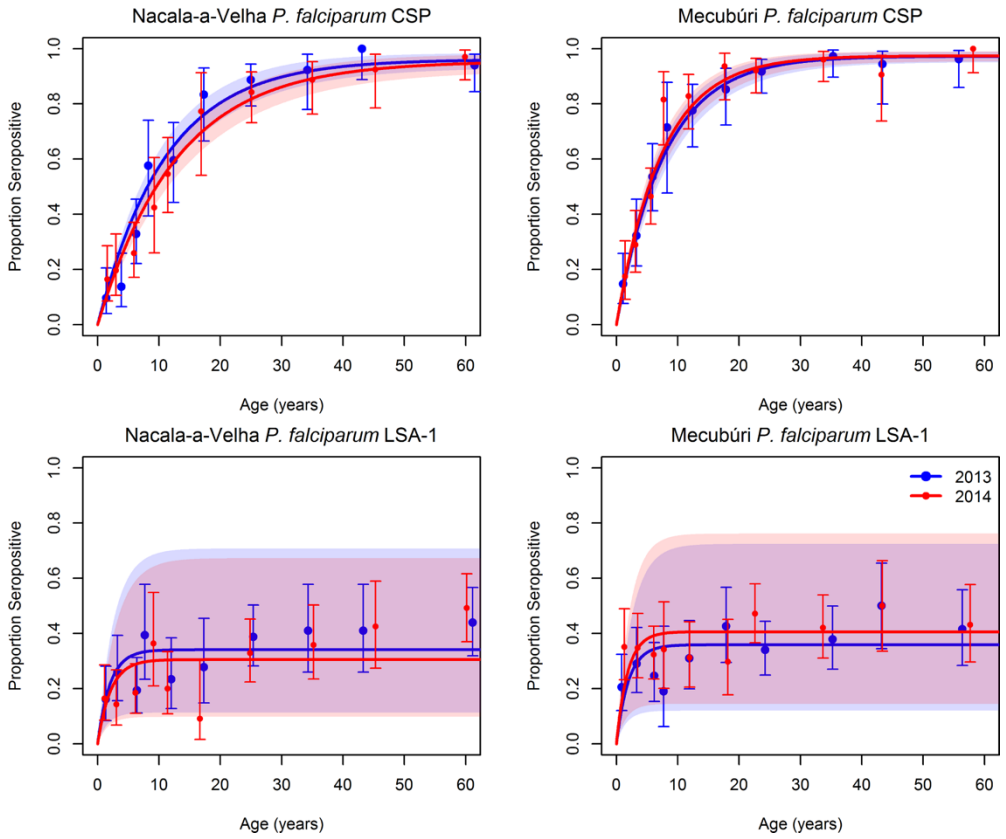
Murhekar, et al. *Lancet Glob Health* (2022)

Use case 4: Assessing effectiveness of behavioral, environmental or pharmaceutical interventions

- Evaluates the effect of interventions other than vaccination (e.g., bed nets, water, sanitation and hygiene interventions)

- However, IgG may not always be the best way to measure changes for some interventions – e.g., antibiotics

Malaria in Mozambique

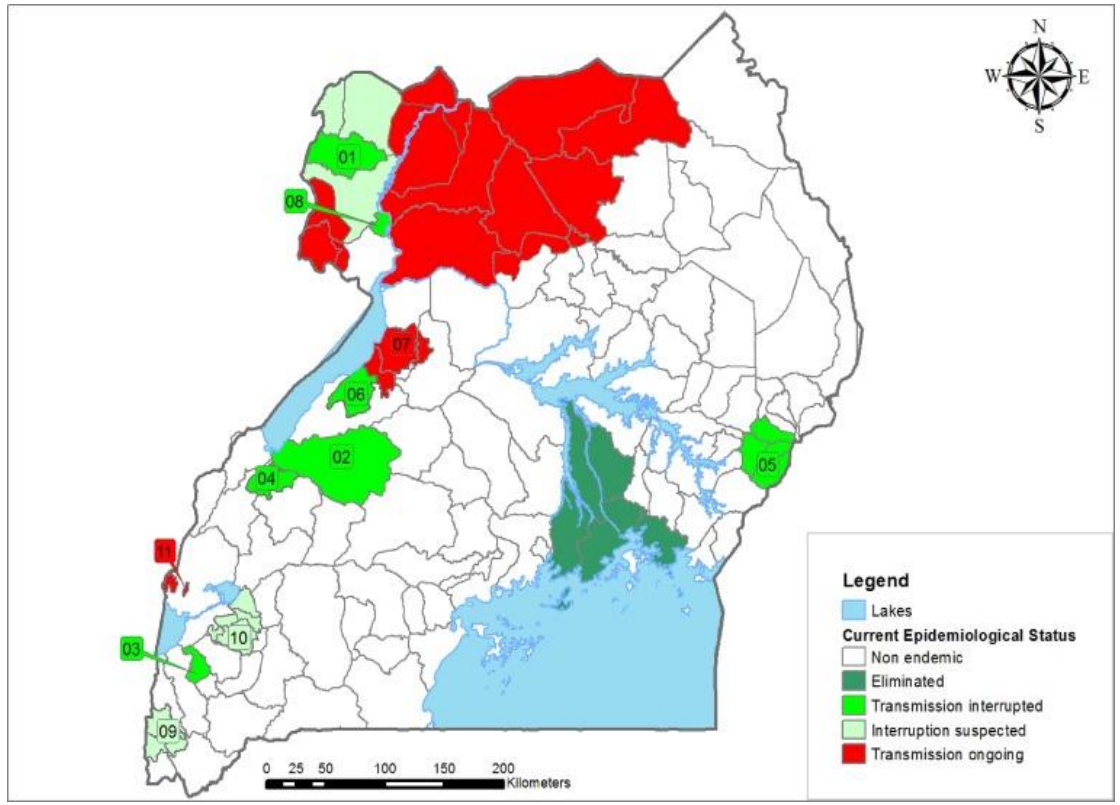


Plucinski, et al. *PLoS Negl Trop Dis* (2018)

Use case 5: Monitoring peri- and post-elimination surveillance settings

- Diseases where elimination of transmission is the target
- Can be used for validation of elimination as well, such as for NTDs
- Monitoring to ensure no re-emergence of disease

Onchocerciasis in Uganda



Oguttu, et al. *AJTMH* (2014)

What is the promise of **multiplex** serology?

Multiplex serology enables:

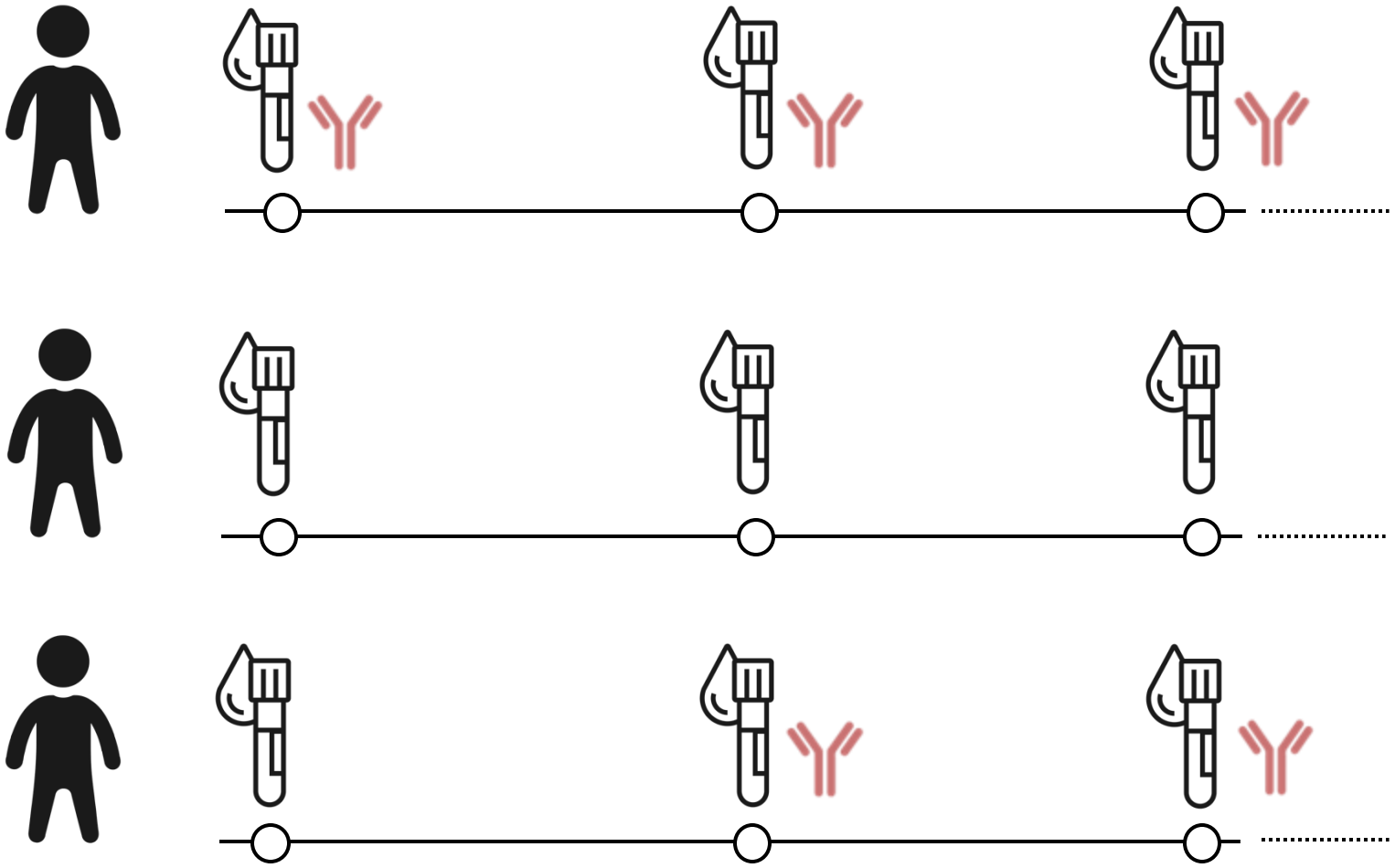
- Simultaneous estimation of antibody prevalence for **multiple pathogens** (i.e., maximizes specimen efficiency)
- **Lower marginal testing cost** and time (i.e., easy to include pathogens not routinely surveilled)

What are the use cases of **multiplex** serosurveillance?

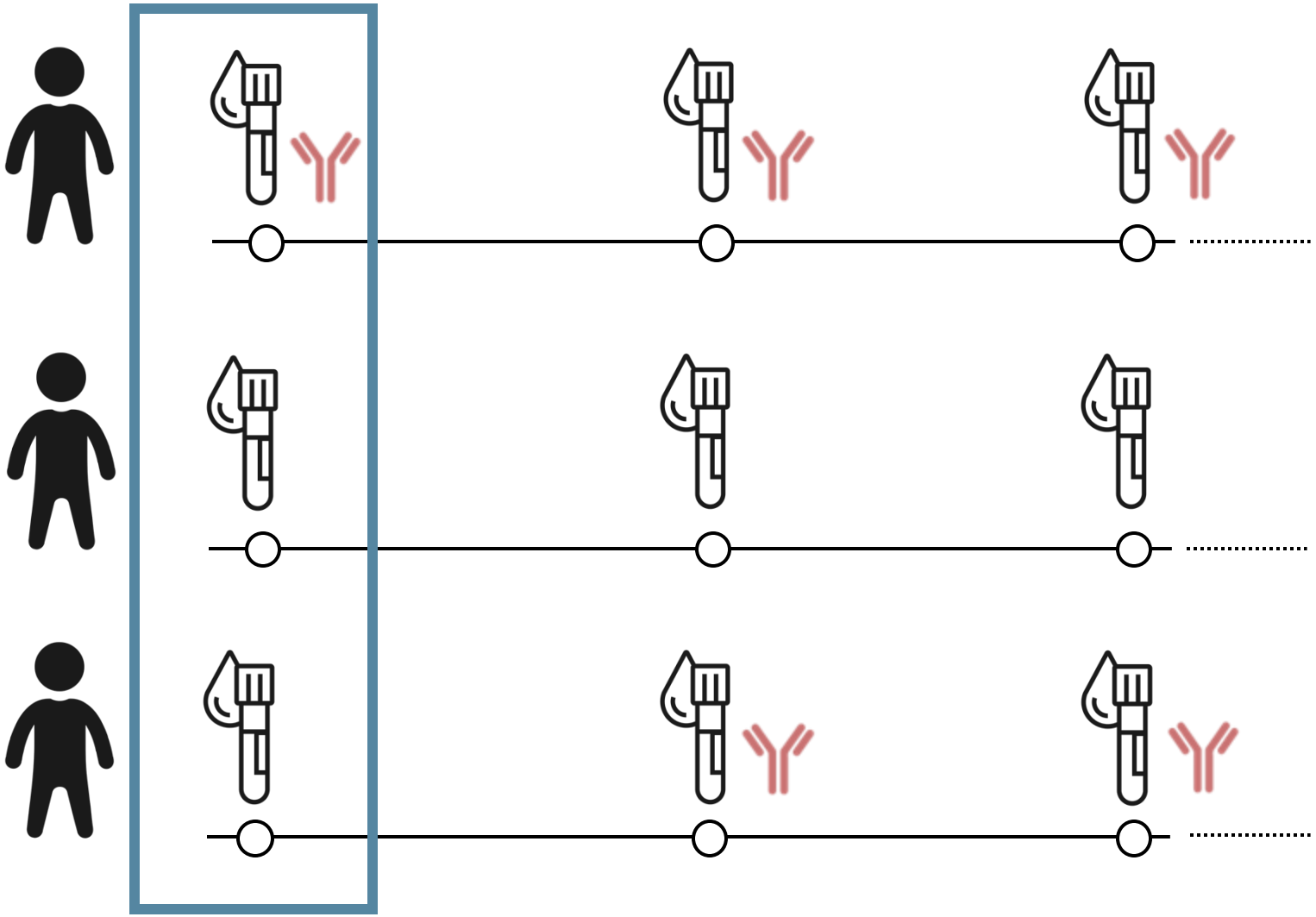
1. To identify geographic areas that have **high exposure to multiple pathogens or low coverage to multiple vaccine preventable diseases** (to inform surveillance systems, programmatic decision-making, and horizontally aligned interventions like WASH, vaccines)
2. To identify **vulnerable sub-populations** (e.g., immunity to measles among people living with HIV)
3. To more **accurately monitor intervention effectiveness** by measuring serological responses to multiple antigens from a single pathogen
4. To estimate **exposure history for pathogens that exhibit cross-reactivity** with increased specificity, by simultaneously measuring serological responses to antigens from cross-reacting pathogens

Study designs commonly used in seroepidemiology

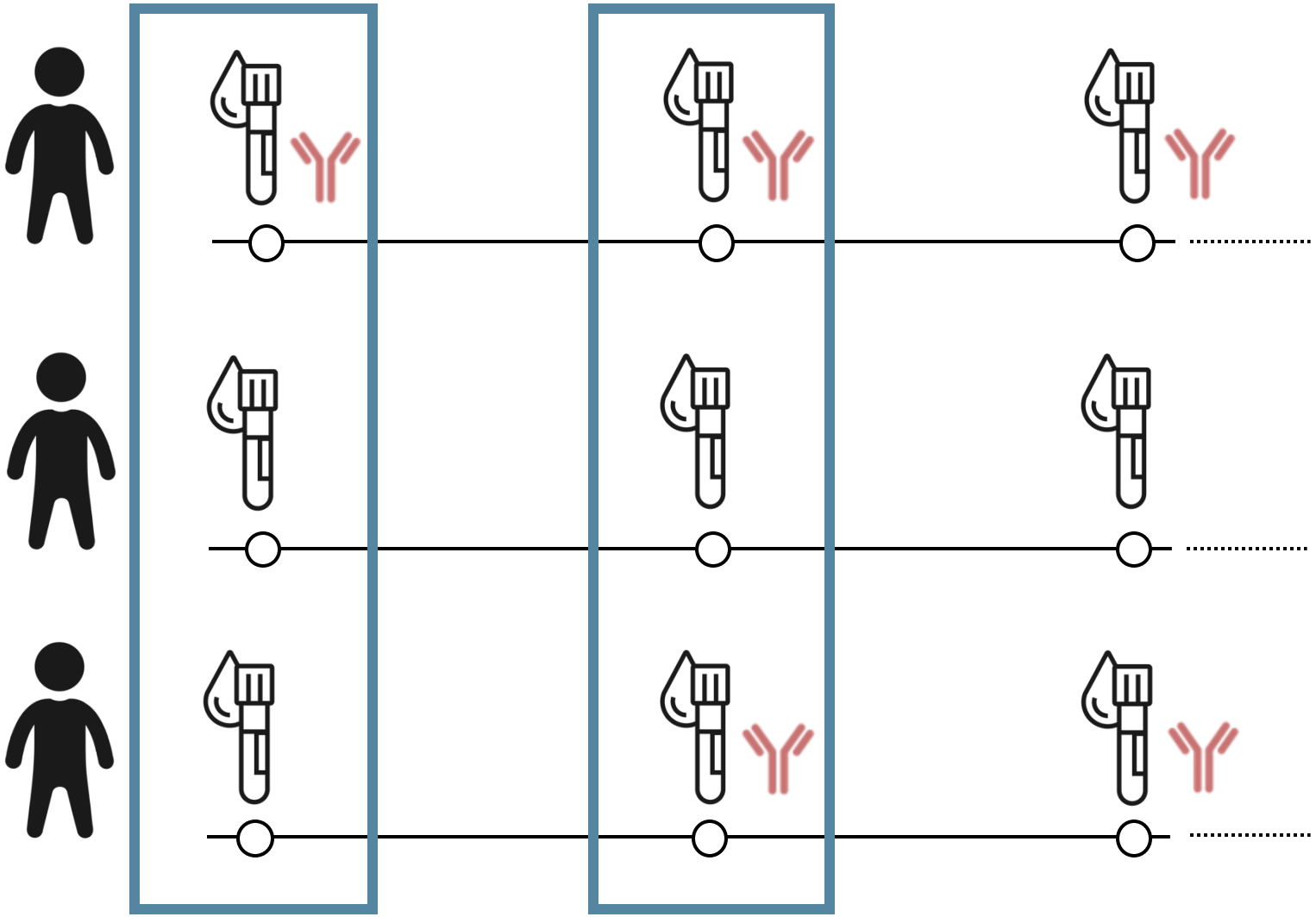
- Cross-sectional serosurvey (single or serial)
- Longitudinal cohort study



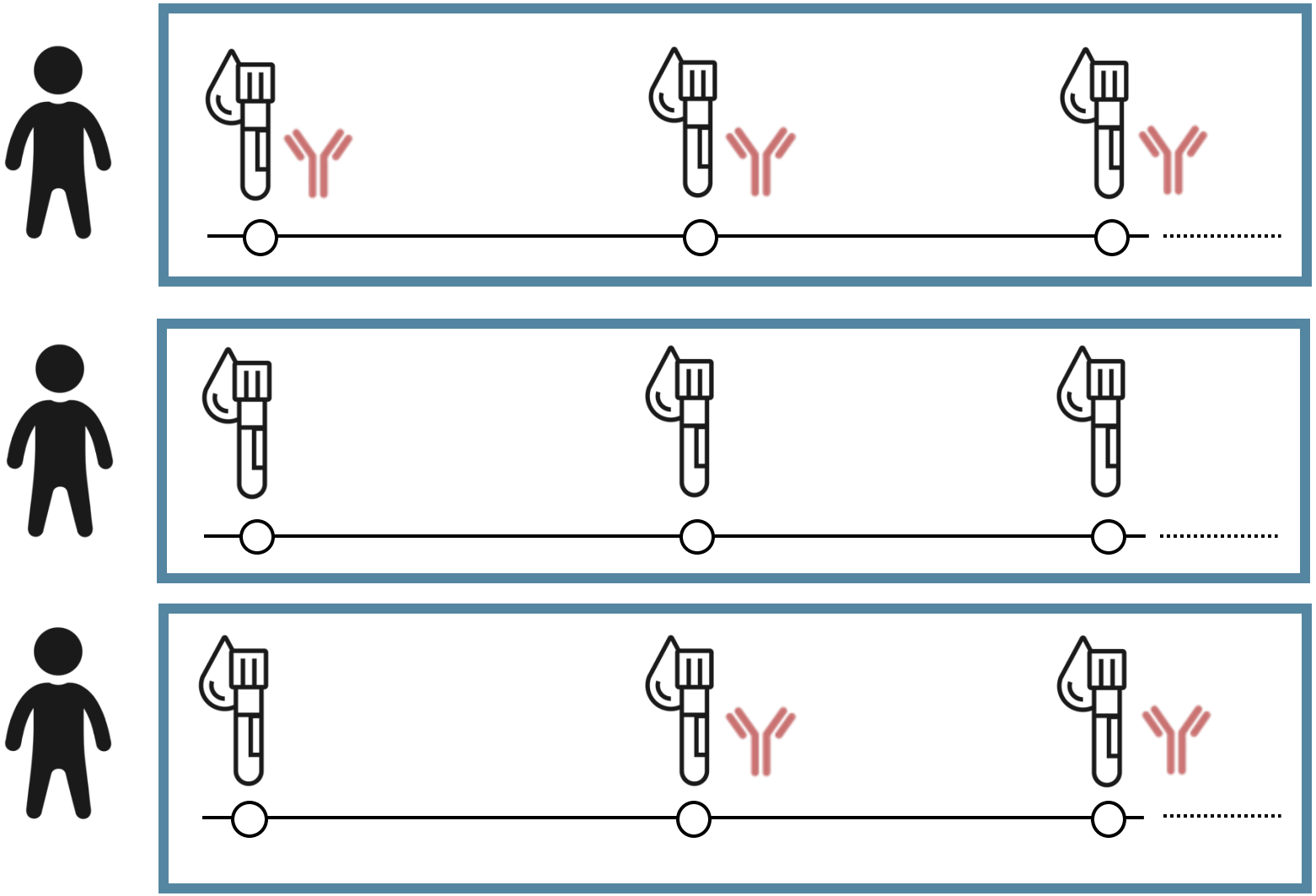
Sero-prevalence: single cross-sectional study



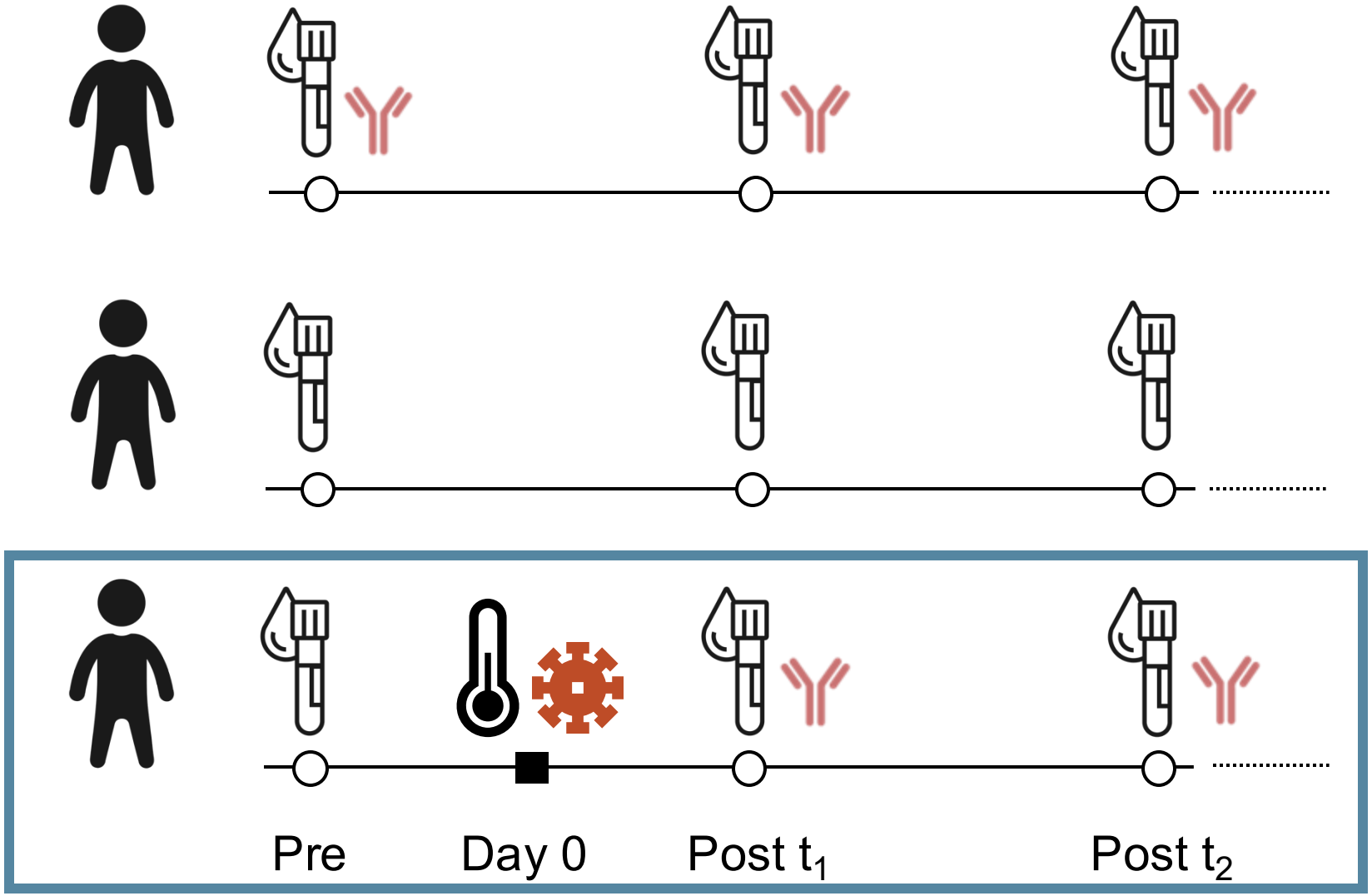
Attack rate (serial cross-sectional, or longitudinal cohort)



Longitudinal antibody dynamics



Pre- and post-infection antibody dynamics



Laboratory assays commonly used in seroepidemiology

Serological assays commonly used for **serosurveillance** and for **research**

- Enzyme immunoassays (EIA, ELISA)
- Multiplex bead assays (MBA; e.g., Luminex)
- Electrochemiluminescence assays (ECL; e.g., MSD)
- (Hem)agglutination (inhibition) assays (HI/HAI)**
- Neutralization assays** (e.g., vibriocidal for cholera)

***Functional assays*

Serological assays used for **research**

- Protein microarrays
- Phage display
- …

Laboratory assays commonly used in seroepidemiology

Serological assays commonly used for **serosurveillance** and for **research**

- Enzyme immunoassays (EIA, ELISA)
- Multiplex bead assays (MBA; e.g., Luminex)
- Electrochemiluminescence assays (ECL; e.g., MSD)
- (Hem)agglutination (inhibition) assays (HI/HAI)**
- Neutralization assays** (e.g., vibriocidal for cholera)

***Functional assays*

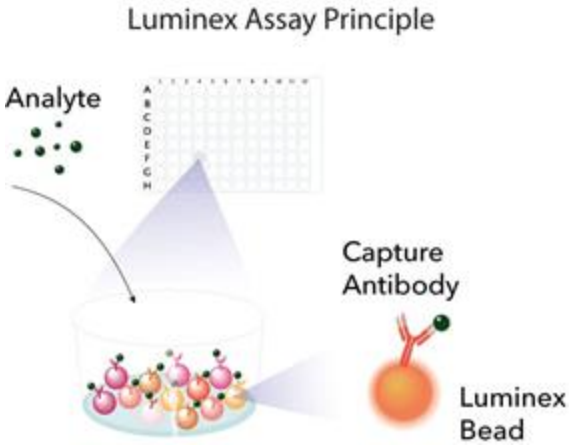
Serological assays used for **research**

- Protein microarrays
- Phage display
- …

Multiplex bead assays (MBAs) produce median fluorescence intensity (MFI) values

- MFI is a morphological measurement
- Raw MFIs contain biological and technical variation

Biological variability:
- Differences across samples (natural person-to-person variability)
- Differences across disease condition



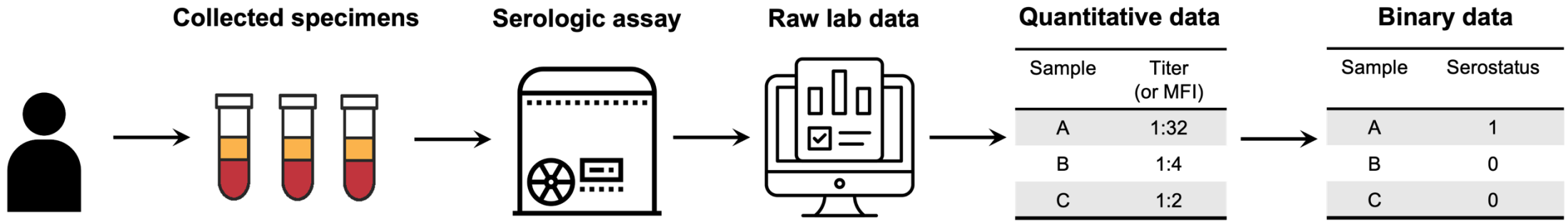
Laboratory variability:
- Sample and reagent storage
- Time since last machine calibration, maintenance, cleaning
- Pipetting precision

'Chemical' variability:
- Cross-reactivity or chemical interference with binding
- Presence of tags (e.g. GST) interfering with binding

We will discuss this in **Part 3**

From blood vials to seroepidemiological data

Process



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)

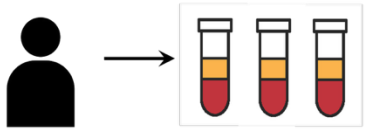
Classification of Serostatus
Choice of control samples (positives, negatives; relevance to epidemiological characteristics in sample population)
Classification algorithm and choice of threshold

Hay, Routledge, Takahashi. *Epidemics* (2024)

What we'll cover in this course

Process

Collected specimens



Serologic assay

Raw lab data

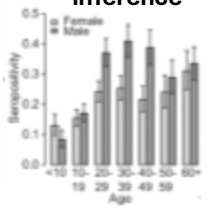
Quantitative data

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

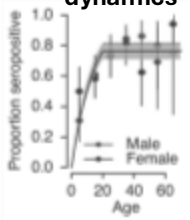
Binary data

Sample	Serostatus
A	1
B	0
C	0

Epidemiologic inference



Transmission dynamics



Data visualization & interpretation

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)

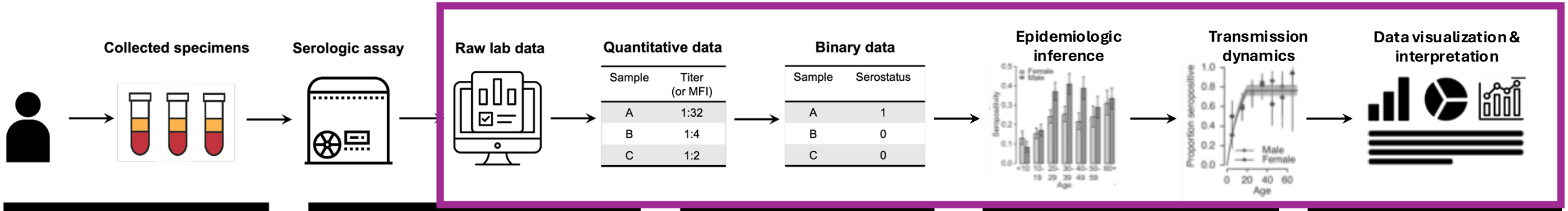
Classification of Serostatus
Choice of control samples (positives, negatives; relevance to epidemiological characteristics in sample population)
Classification algorithm and choice of threshold

Estimating relevant parameters
Estimating population seroprevalence (or seroincidence) and associated uncertainty
Inferring transmission dynamics (force of infection)

Data sharing
Visualize data to better inform inference
Present and interpret results for policy-relevant decision making

What we'll cover in this course

Process



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

Estimating relevant parameters
Estimating population seroprevalence (or seroincidence) and associated uncertainty
Inferring transmission dynamics (force of infection)

Communicating data
Visualize data to better inform inference
Present and interpret results for policy-relevant decision making

We will **not** be covering study design in the lectures (e.g., sample size calculation)

Course Overview

Part 1: Introduction to the Seroanalytics Workshop

Part 2: Introduction to serological data analyses in R

Part 3: Pre-processing serological data

Part 4: Visualizing and standardizing serological data

Part 5: Determining serostatus and estimating seroprevalence

Part 6: Inferring transmission dynamics from seroprevalence data

+ Time to work on **individual projects**
+ **Extra lectures** on advanced topics