Lecture 4
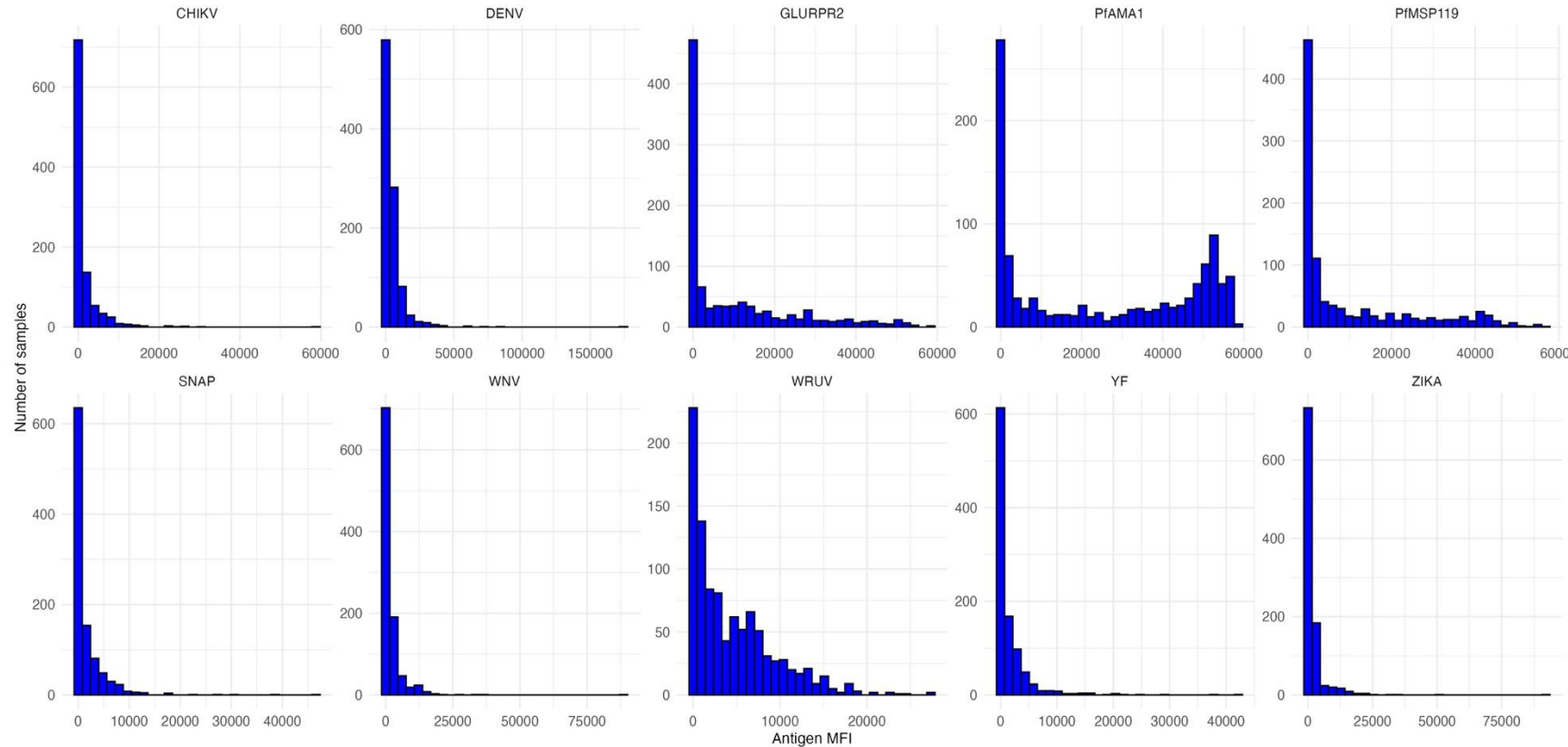# Visualizing & standardizing serological data

May 22, 2025

Seroanalytics Training
Blantyre, Malawi

Lecture outline

- Data visualization

- Why binarize serological data?

- Converting MFI values to standardized units (IU/mL)

- Applying an externally-defined cutoff to calculate seroprevalence

- Calculating confidence intervals for seroprevalence

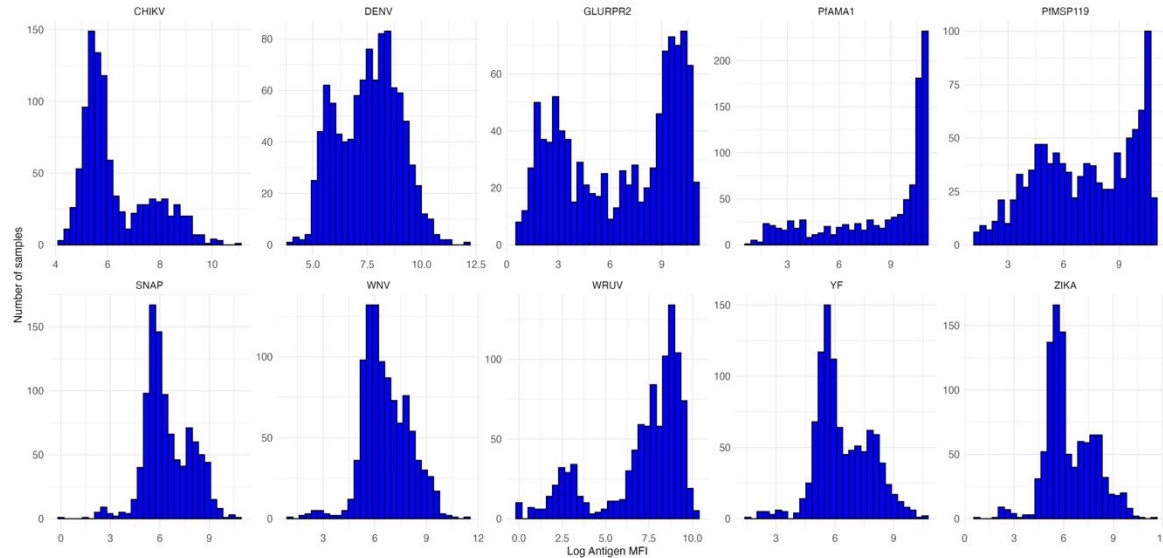# Visualizing serological data with histograms

How would you compare the different distributions of data (untransformed)?

Visualizing serological data with histograms
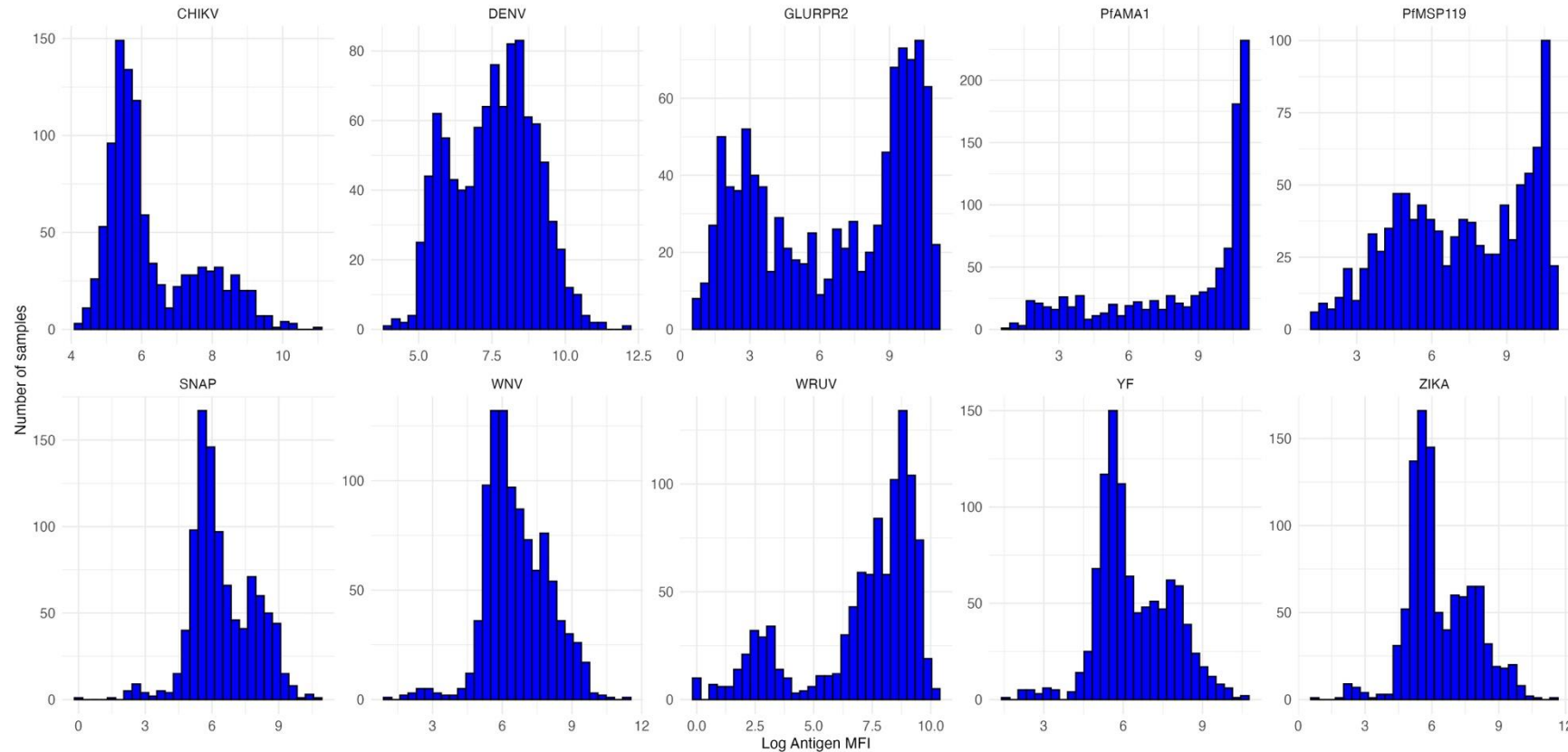
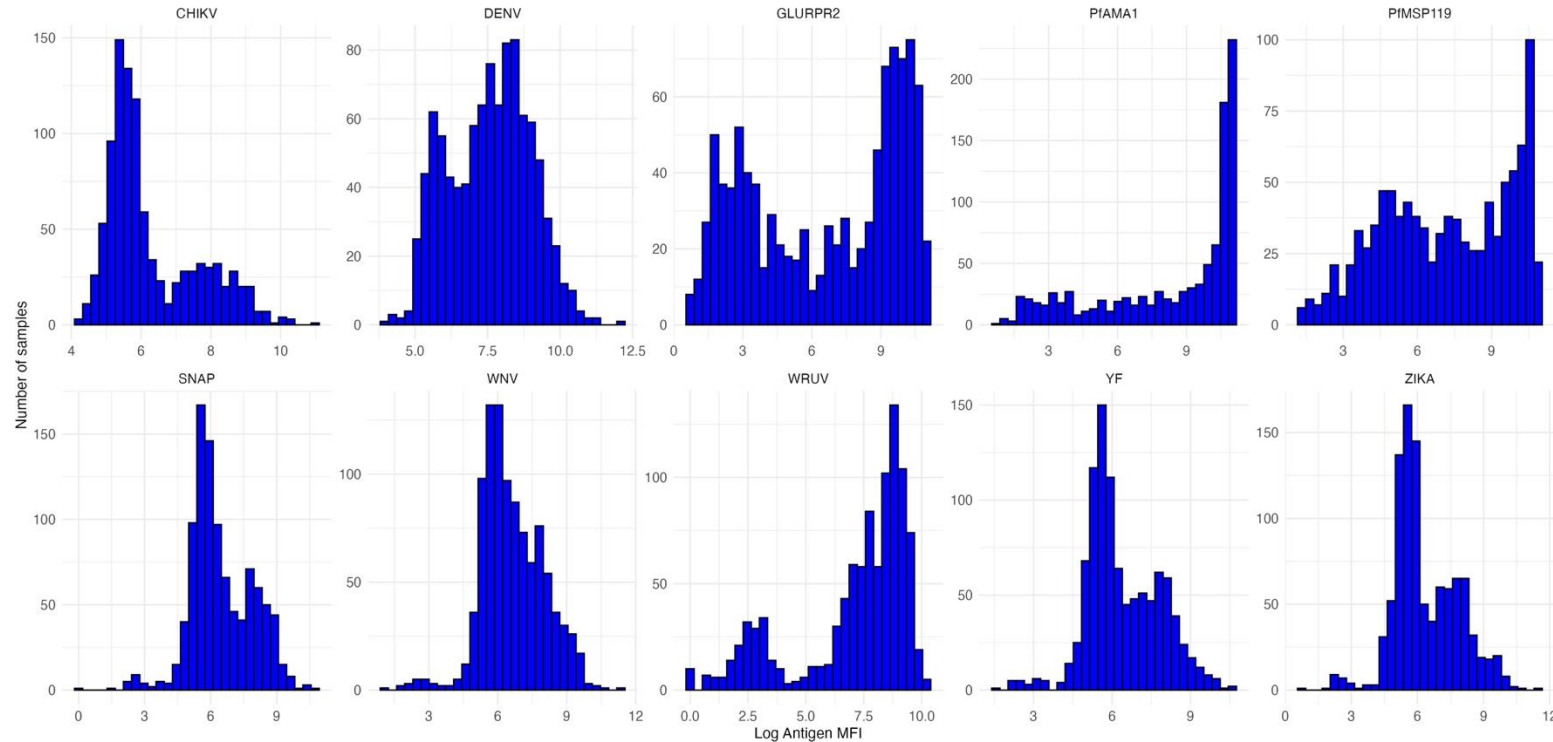How would you compare the different distributions of data (log transformed)?

Visualizing serological data with histograms

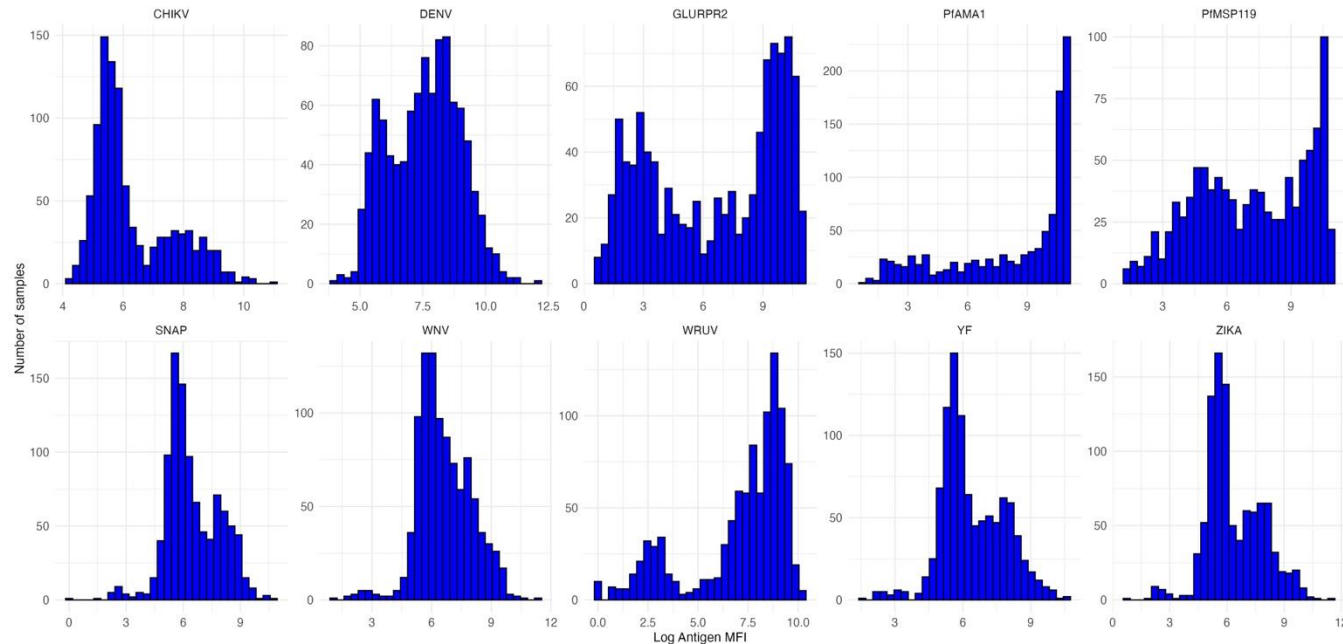How would you compare the different distributions of data?

- Skewed vs. non-skewed distributions
- Unimodal vs. bimodal vs. multimodal
- Narrow or wide distribution
- Are there outliers?

Visualizing serological data with histograms

What underlying differences might cause the different distributions of data?

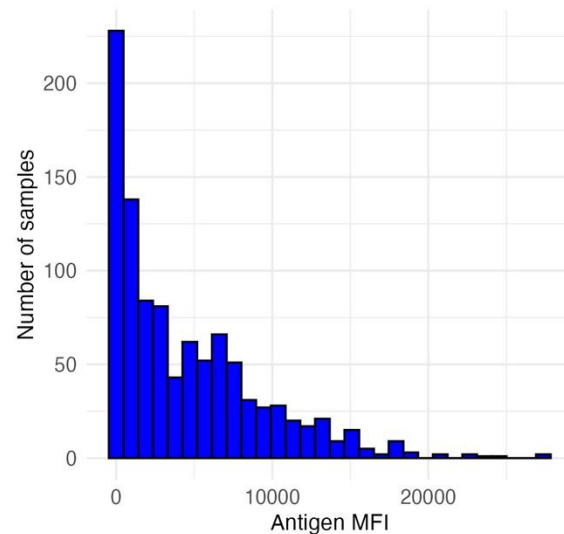# Visualizing serological data with histograms

What underlying differences might cause the different distributions of data?

- Mix of exposed and unexposed in population
- Mix of vaccinated / unvaccinated
- Waning antibody responses

Histograms of rubella antibody responses

Untransformed MFI Values

Log transformed MFI values
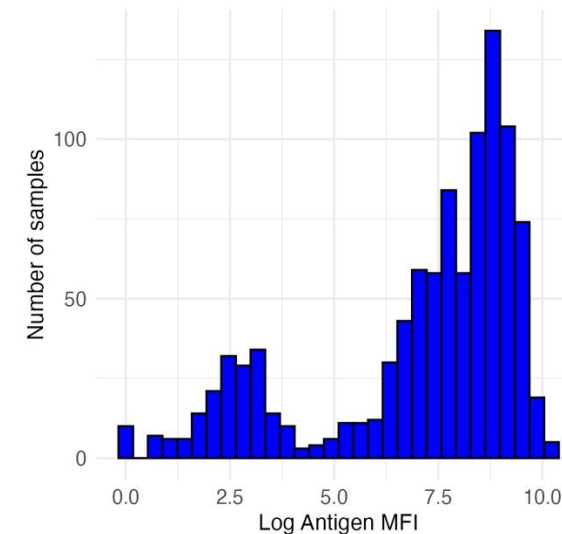
How do we get information from histograms like these?

- How do we compare the results of different histograms?

- What inferences can we make about pathogen exposure based on a histogram?

- A method of making inferences from distributions is **binarizing** data
  - Setting a cutoff, and everything above that cutoff is positive

Histograms of rubella antibody responses

Untransformed MFI Values
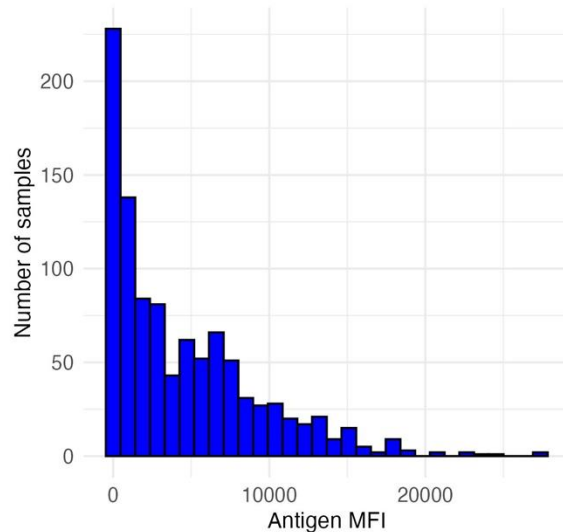
Log transformed MFI values
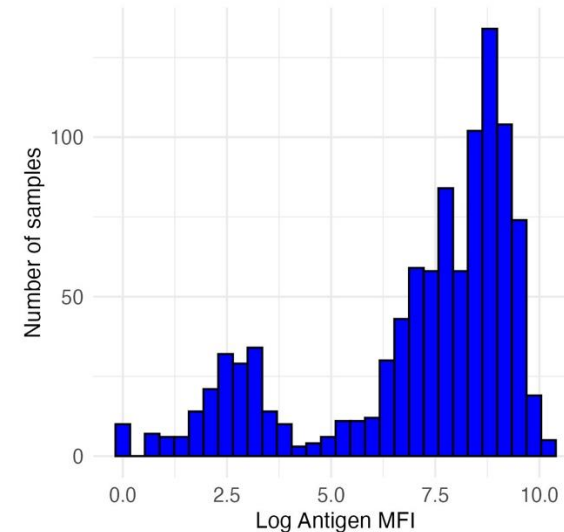
- Rubella has a **correlate of protection**
  - Individuals with antibody responses above a certain cutoff are expected to be protected from rubella infection
  - By applying a cutoff equivalent to the correlate of protection, we can calculate seroprevalence (here, also equals the immune proportion)

Histograms of rubella antibody responses

Untransformed MFI Values
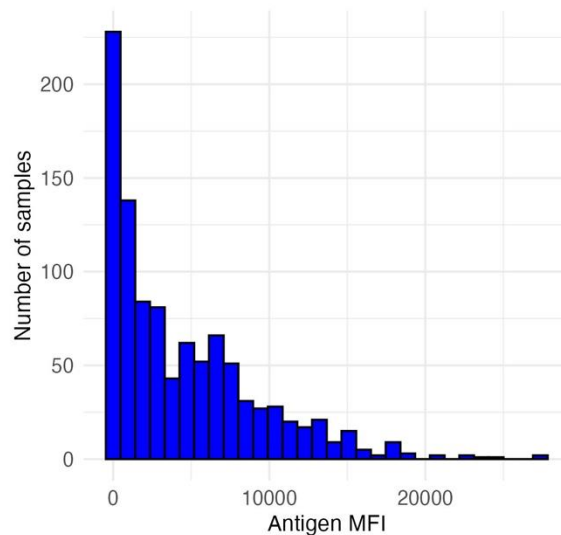
Log transformed MFI values
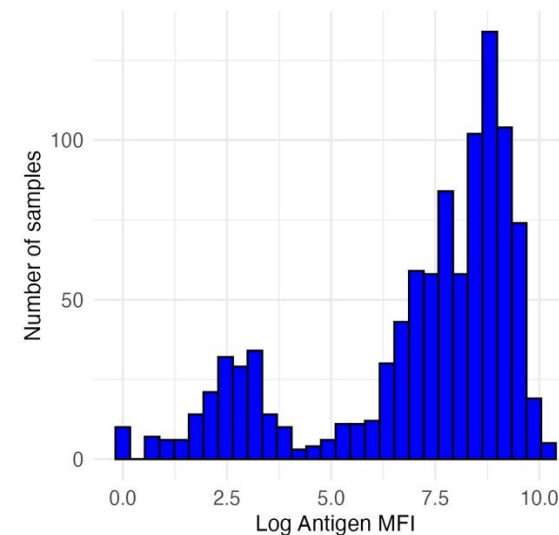
- Rubella correlate of protection is 9.36 IU/mL (international units per milliliter)

- However, the Luminex assay gives us antibody units in MFI (median fluorescence intensity)

- **We can convert from MFIs to IU/mL using a standard curve:** we convert all MFI values to IU/mL, and apply the 9.36 IU/mL cutoff

Converting MFI values to standardized units (IU/mL) from standard curves

- The standard curve is created using **known concentrations** of antibody standards (e.g., the WHO International Standard for rubella, below), ideally measured on the same plate(s) as the unknown samples.

Medicines & Healthcare products Regulatory Agency
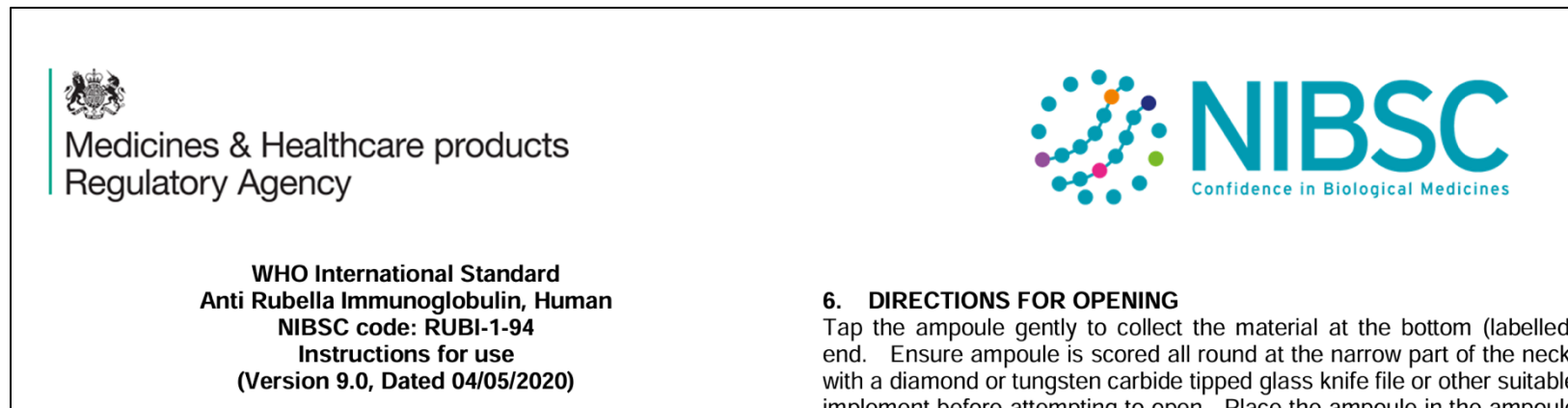
NIBSC
Confidence in Biological Medicines

**WHO International Standard
Anti Rubella Immunoglobulin, Human
NIBSC code: RUBI-1-94
Instructions for use
(Version 9.0, Dated 04/05/2020)**

6. DIRECTIONS FOR OPENING
Tap the ampoule gently to collect the material at the bottom (labelled) end. Ensure ampoule is scored all round at the narrow part of the neck, with a diamond or tungsten carbide tipped glass knife file or other suitable implement before attempting to open. Place the ampoule in the ampoule

https://nibsc.org/documents/ifu/RUBI-1-94.pdf

Seroanalytics Hub

Infectious Disease DYNAMICS
JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Converting MFI values to standardized units (IU/mL) from standard curves

(Recall we used standard curves to estimate relative antibody units.)

1. First, we fit a model to our standard curve data to establish the relationship between **MFI values** and known **IU/mL concentrations**.

2. Then, we can use this model to convert MFI values for the samples of interest to values in IU/mL.

3. Lastly, we apply the 9.36 IU/mL correlate of protection cutoff to the samples of interest to determine a binary serostatus for each sample.

Converting MFI values to standardized units (IU/mL) from standard curves

For rubella (the "WRUV" antigen), we can apply the cutoff (based on correlate of protection) of 9.36 IU/mL to determine individual serostatus.
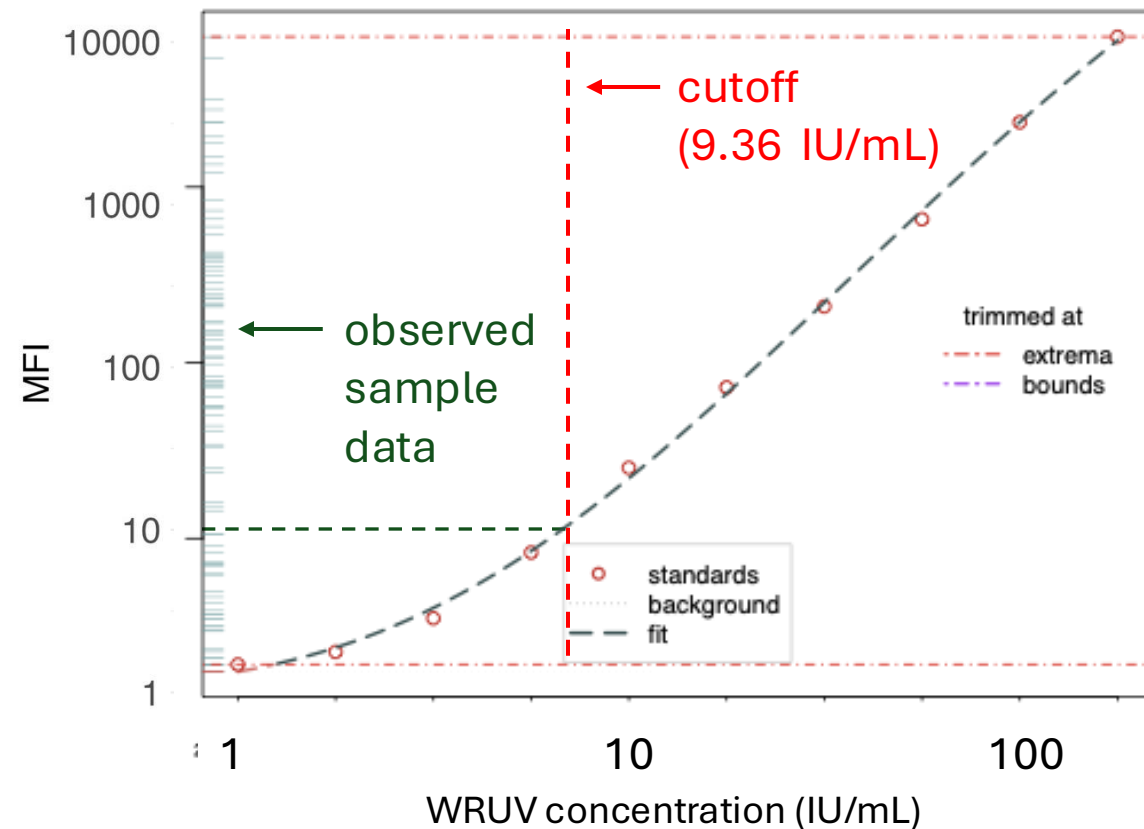
Determining serostatus from a cutoff

For each sample in the dataset, we convert the MFI value to a concentration value (IU/mL), and then classify each sample as seropositive (1) or seronegative (0)

Sample ID	WRUV MFI	WRUV IU/mL	Serostatus
1	MFI_1	conc_1	1
2	MFI_2	conc_2	1
3	MFI_3	conc_3	0
4	MFI_4	conc_4	1
5	MFI_5	conc_5	1
6	MFI_6	conc_6	0
7	MFI_7	conc_7	0

Seroanalytics Hub

Infectious Disease DYNAMICS
JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Calculating seroprevalence

For WRUV, with a cutoff of 9.36 IU/mL, we have:
- 767 seropositive samples
- 233 seronegative samples

$$Seroprevalence\ (\%\ Seropositive) = \frac{Number\ seropositive}{Total\ number\ tested} = \frac{767}{1000} = 76.7\%$$

Calculating seroprevalence – confidence intervals

- Estimating uncertainty is important for capturing true population seroprevalence; it typically accounts for variance around a mean estimate.

- Estimating a 95% confidence interval means that in 95 out of 100 calculations of seroprevalence from the same source population using 1000 individuals, the range will include the true population-level seroprevalence.

- This method uses a binomial distribution and accounts for the number of people sampled – luckily, we can use R to easily compute it:

$$\Pr(X \leq k) = \sum_{i=0}^{\lfloor k \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$$

Calculating seroprevalence – confidence intervals

With 767 seropositives and 1000 total samples:

Seroprevalence = 76.7% (95% CI: 74.0%, 79.3%)

Question: How do we interpret seroprevalence?

Interpretation strongly depends on how seroprevalence is determined, including controls used and selected cutoff.

For the seroprevalence of rubella (WRUV):

o Percentage of people who have been exposed to vaccine or natural infection.

o If we know there's no/little vaccination, we may assume seroprevalence is wholly due to natural infection, or vice versa.

o Since cutoff is a correlate as protection, seroprevalence could also indicate which individuals are susceptible to future infection, and whether there could be outbreaks in a region.

Question: How do we interpret seroprevalence?

How would we interpret seroprevalence if there is NOT a correlate of protection? (see Part 5)

Question: How do we interpret seroprevalence?

How would we interpret seroprevalence if there is NOT a correlate of protection? (see Part 5)

- For antigens in general, seropositivity could indicate:
    - Population who has ever been exposed to pathogen or vaccine
    - Population who has had recent infection
    - Population who has had symptomatic infection
    - Cross-reactive antibody responses

Conclusions

- You should visualize your data before conducting analysis.

- It can be useful to classify samples as seropositive and seronegative for direct interpretation of serological results.

- For some pathogens, the existence of **externally-defined correlates of protection** and the availability of **(international) antibody standards** make it straightforward to calculate and interpret seroprevalence. This is not the case for all pathogens!

- We can use a statistical model to produce confidence intervals around the mean seroprevalence estimate.