**Information about datasets:**

The simulated serosurvey data we’ll use during this training (**simulated\_sample\_wide\_training\_data.csv**, **Training demographics df.csv**, and **Training serostatus df.csv**) is modeled after real data. Therefore, the results and inferences of the analysis we’ll conduct are similar to those we may expect using real data.

**Methods:**

The simulated data are from a hypothetical, cross-sectional multiplex serological survey to assess population-level exposure to a range of VPDs, arboviruses, and malaria antigens. Dried blood spots from a total of 1,000 participants were selected using simple random sampling from a census list within a defined geographic catchment area. The survey included individuals of all ages (0–100 years). Sex, coded as 1 for male and 2 for female, was recorded for each participant. Assume that both measles and rubella vaccination is common in this region, although people may also be seropositive because of measles and rubella disease outbreaks.

Capillary blood samples were collected via finger prick onto Whatman 903 filter paper cards to create dried blood spots (DBS). DBS samples were dried overnight at ambient temperature, stored with desiccant in zip-sealed bags, and transported to the laboratory under cold chain conditions. Antibodies were eluted from DBS punches using Buffer B and diluted to a concentration equivalent to 1:400 serum dilution in the assay. Eluates were analyzed using a Luminex-based multiplex bead assay (MBA). Antigens were selected based on prior validation studies and were covalently coupled to uniquely labeled magnetic microspheres using standard carbodiimide chemistry.

Assays were performed in 96-well plates, with each well containing 50 µL of diluted eluate incubated with the bead panel, followed by biotinylated anti-human IgG detection and streptavidin-phycoerythrin labeling. Plates were read on a MAGPIX instrument, and median fluorescence intensity (MFI) values were recorded for each antigen.

This assay included 13 antigens from the following pathogens: WNV (West Nile virus), Yellow Fever (YF), Japanese Encephalitis Virus (JE3), Zika virus (ZIKA), Dengue virus (DENV), Chikungunya virus (CHIKV), *Plasmodium falciparum* malaria (GLURPR2, CSP, PfAMA1, and PfMSP119), Rubella (WRUV), and Measles (WMEV). Additionally, SNAP was included to as a control to assess technical consistency across samples.

Controls are available for many antigens (**simulated\_control\_long\_training\_data.csv**). Positive controls came from sera from clinically confirmed cases, while negative controls came from sera from healthy U.S. adults with no known travel to regions where the target pathogens are endemic. Control samples were not available for DENV or CHIKV antigens.

Samples were also run with three standard curves on all plates (raw plate data is provided for four plates: **Raw Plate 1 dataset.csv**, **Raw Plate 2 dataset.csv**, **Raw Plate 3 dataset.csv**, **Raw Plate 4 dataset.csv**, **Plate maps plates1 to 4.csv**) to allow for quantitative interpretation of antibody levels. These standard curves are for WRUV, WMEV, and a “pan control” standard curve that is valid for the remaining 10 antigens (all except WRUV, WMEV, and SNAP) (**simulated\_standard\_curve\_training\_data.csv**). While in practice, every sample should have an MFI value for every antigen, in this toy data set, the rubella standard curve samples only have MFIs for WRUV; the measles standard curve samples only have MFIs for WMEV, and the “pan control” standard curve samples do have not MFIs for WRUV and WMEV.