



## Data Analysis in R Training: Microbial Pathogen Surveillance Using TaqMan Array Cards (TAC)

Center for Infectious Disease Research and Surveillance  
Research and Development Division  
Dhulikhel Hospital Kathmandu University Hospital, Dhulikhel, Kavre

Department of Global, Environmental, and Occupational Health  
University of Maryland School of Public Health

### Deliverables

By the end of the workshop, each participant will have:

- R scripts for cleaning, analysis, and visualization
- Clean, merged dataset (survey + culture + qPCR)
- Publication-ready figures and tables
- A GitHub repo with their workflow for reproducibility

### Agenda

#### Day 1: Processing Multi-Source Data

**Objectives:** Import, clean, and integrate survey, culture, and qPCR data

#### Topics

- **Survey data:** Clean variable names, recode factors
- **Culture data:** Organize by isolate, sample, and pathogen
  - E.g., presence/absence of *E. coli* and antibiotic-resistant *E. coli*
- **qPCR data:** Join all qPCR output files, clean variables, define positive/negative

#### Hands-on

- Access project repository using GitHub
- Clean data:
  - Survey: demographic & other metadata
  - Culture: isolate-level pathogen presence & AMR
  - TAC: Ct → presence/absence or load
- Push cleaned and merged data to GitHub



## Day 2: Descriptive Analysis of Pathogen and AMR Data

**Objectives:** Summarize prevalence, burden, and AMR patterns

### Topics

- **TAC data:**
  - Calculate prevalence of each target
  - Summarize pathogen burden (number of detections/sample)
- **Culture data:**
  - Prevalence of bacterial isolates
  - AMR profile frequency tables (e.g., ESBL, MDR)
- **Comparisons:**
  - **Merge** by sample ID or household ID
  - Concordance between TAC and culture for overlapping organisms

### Hands-on

- Join data from multiple sources (culture and TAC data)
- Generate:
  - Prevalence tables (overall + by group)
  - Barplots/heatmaps of pathogen detection
  - AMR profile tables
  - TAC vs. culture agreement table

## Day 3: Statistical Comparison & Visualization

**Objectives:** Compare groups and produce publication-ready outputs

### Topics

- Hypothesis testing:
  - Prevalence (chi-square/Fisher's)
  - Load/ $\Delta$ Ct (Wilcoxon/t-tests)
- Regression models:
  - Logistic: e.g., diarrhea ~ pathogen presence + WASH
  - Poisson/Negative binomial: burden ~ exposures
- Visualizations:
  - Boxplots, prevalence barplots, forest plots, stacked AMR profiles



## Hands-on

- Model associations:
  - WASH factors → pathogen detection (TAC or culture)
  - AMR burden ~ age, recent antibiotics
- Make:
  - Forest plots (with CIs)
  - Heatmaps of AMR and pathogen frequency
  - Clean summary tables and formatted figures for a manuscript