

# Analysis of infectious disease outbreaks using TranStat

Jonathan D. Sugimoto, PhD  
[jons@fredhutch.org](mailto:jons@fredhutch.org)

Lecture Website:  
<http://www.cidid.org/transtat/>

## Lecture 6: Outline

- Introduction to TranStat
- Case study: Pandemic influenza A(H1N1) 2009 outbreak in Western Washington State
- Summary

# Introduction to TranStat

# Motivation

To enable field personnel and researchers to analyze data from local outbreaks of infectious diseases, with the aim of...

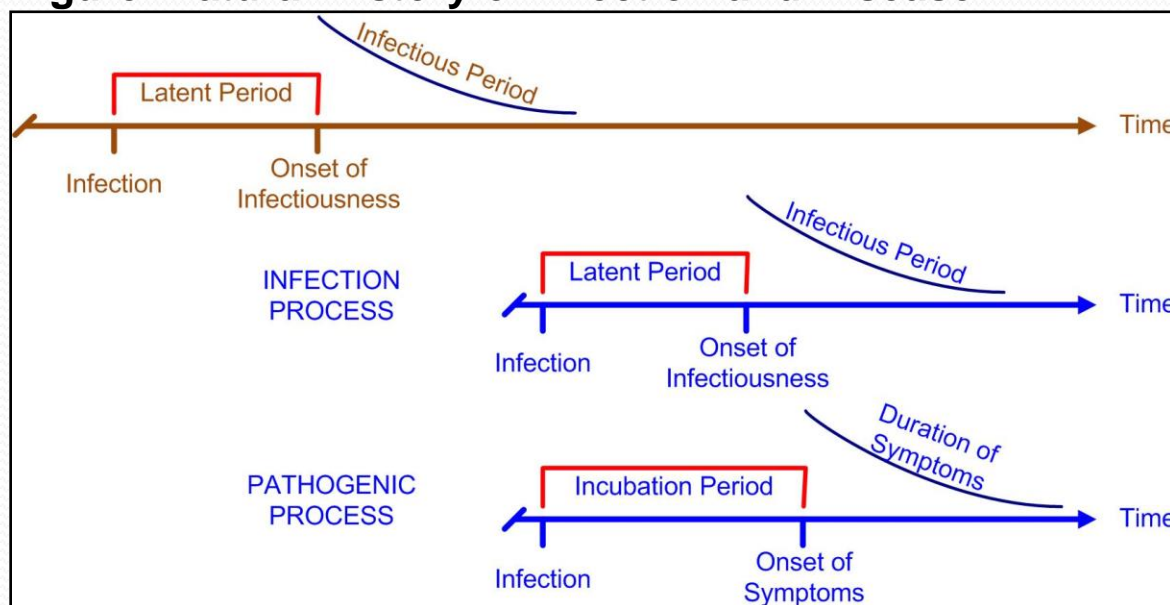
- Evaluating the transmissibility of pathogens
- Evaluating the effects of interventions and risk factors on transmission
- Performing simulation studies,
  - for example, in order to perform power calculations for study design purposes

# Basic Concepts: Natural History of Infection and Disease

**Figure. Natural History of Infection and Disease**

Infectious Individual

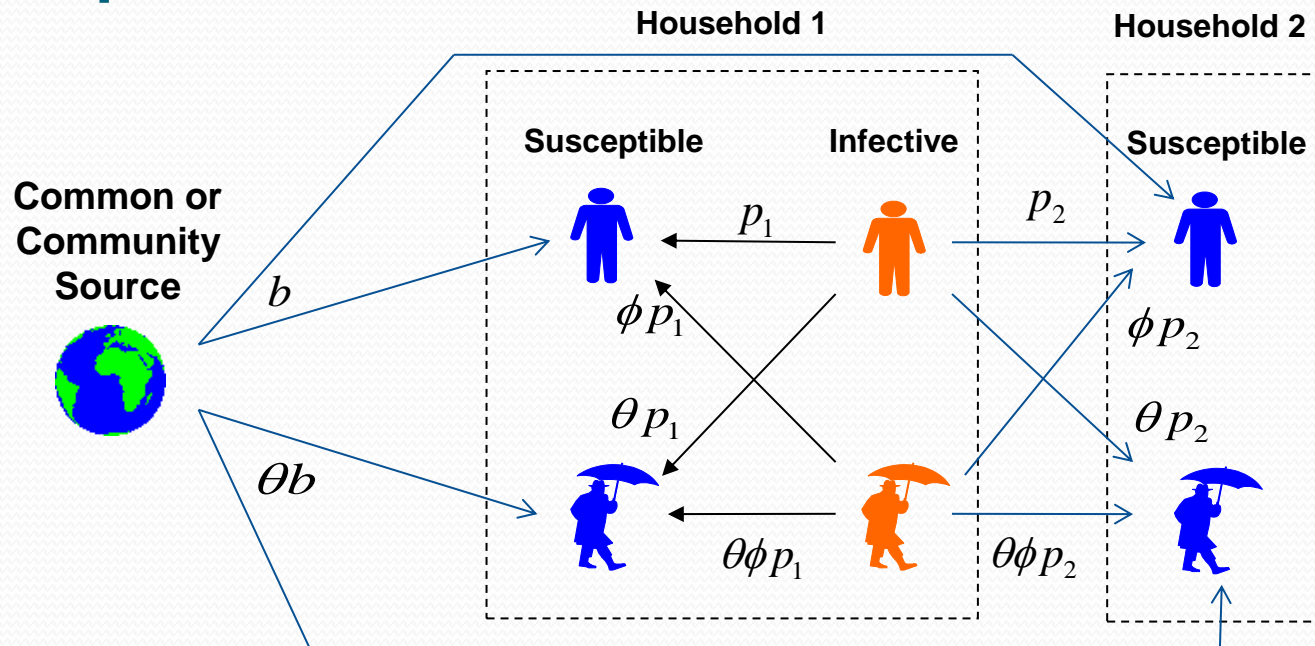
Susceptible Individual



- Infection depends upon exposure to an infectious individual (see next slide for more details)
- Simultaneously occurring processes, that are often (or are assumed to be) strongly correlated
- Individuals do not necessarily complete each entire process, e.g., an individual may become infectious, but never exhibit clinically-apparent symptoms (infectious asymptomatic infection)

# Basic Concepts: Exposure to Infection

**Figure.  
Population and  
Contact  
Structure**



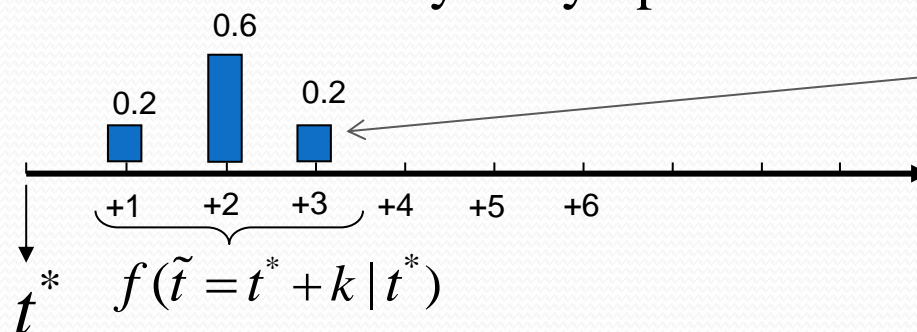
- Contact = exposure to a specific source of infection for a defined period of time
- 'Household' = general term for clusters of individuals who are more likely to mix with each other than with other members of the population. Multiple types of households may be defined.
- Types of contact and associated transmission probabilities
  - P2P, or person-to-person, exposure to a specific individual: within household,  $p_1$ , and between household (for example, household in the same neighborhood),  $p_2$
  - C2P, or community-to-person exposure to non-specific sources of infection:  $b$
- $\theta$  and  $\phi$  denote covariate effects (risk-factors or interventions) on susceptibility and infectiousness, respectively.

# Model: Incubation/Latent and Infectious Period Distributions (assumed known)

$t^*$ : day of infection

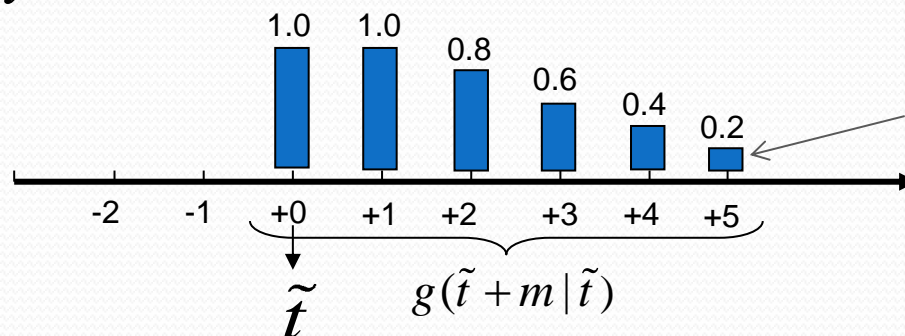
$\tilde{t}$ : day of symptom onset

**Incubation /  
Latent  
Period**



Probability of onset of infectiousness on day  $k$  since  $t^*$

**Infectious  
Period**



Probability of still being infectious on day  $m$  since  $\tilde{t}$

- These are sample distributions for the incubation/latent and infectious periods.
- This example assumes that onset of symptoms indicates onset of infectiousness, *i.e.*, incubation=latent periods.
- TranStat inputs: Minimum and maximum values for  $k$  and  $m$  and the daily probability distributions (blue bars)

# Likelihood

- $T = \begin{cases} \text{onset of infection,} & \text{infected} \\ \text{end of follow – up,} & \text{otherwise} \end{cases}$
- Probability that  $j$  infects  $i$  during day  $t$ :  

$$\text{logit}(p_{ijt}) = \text{logit}(p) + \mathbf{X}_i\beta_S + \mathbf{X}_j\beta_I + \mathbf{X}_{ij}'\beta_{SI}, j \in \mathcal{H}_i$$
- An important example of interaction:
  - Let  $r_i$  be the vaccination status and the only covariate for person  $i$
  - $\text{logit}(p_{ijt}) = \text{logit}(p) + r_i\theta + r_j\phi + r_ir_j\psi$
  - $VE_S = 1 - \theta, VE_I = 1 - \phi, VE_T = 1 - \psi$



## Likelihood (continued)

- Probability that the common/community source infects  $i$  on day  $t$ :

$$\text{logit}(b_{it}) = \text{logit}(b) + \mathbf{X}_i \alpha_S$$

- Probability of  $i$  escaping infection on day  $t$ :

$$e_{it} = (1 - b_{it}) \prod_{j=1}^N (1 - p_{ijt} g(t|\tilde{t}_j))$$

- Probability of escaping infection up to day  $t$ :

$$Q_{it} = \prod_{\tau=1}^t e_{i\tau}$$

- Likelihood contribution by  $i$ :

$$L_i = \begin{cases} Q_{iT}, & \text{infected} \\ \sum_t f(\tilde{t}_i|t) Q_{i(t-1)} (1 - e_{it}), & \text{otherwise} \end{cases}$$

# Some Statistical Adjustments Implemented by TranStat

- **Selection bias**: a household's probability of being included in the study sample is conditioned on the ascertainment of an index case among its members

- Probability of no symptom onset on day  $\tilde{t}_{idx}$ :

$$L_i^m = \begin{cases} L_i, & i \text{ is index} \\ Q_{i\tilde{t}_{idx}} + \sum_{t < \tilde{t}_{idx}} \Pr(\tilde{t}_i > \tilde{t}_{idx} | t) Q_{i(t-1)} (1 - e_{it}), & \text{not index} \end{cases}$$

- Maximize the conditional likelihood,  $\prod_i L_i / L_i^m$
- **Right censoring**: showing no symptoms by day  $T$  does not necessarily mean that  $i$  escaped infection.

$$L_i = Q_{iT} + \sum_{t < T} \Pr(\tilde{t}_i > T | t) Q_{i(t-1)} (1 - e_{it}), \quad \text{not index}$$

## Other Statistical Features

- Goodness of fit: comparing observed with expected frequency of symptom onset per person-day
- Permutation test to detect person-to-person transmission (Yang et al. Annals of Applied Stat, 2006)
  - $H_0: p = 0$  vs.  $H_1: p \neq 0$
  - Test statistic:  $\lambda = -2 \log \frac{\sup_{\mathbf{b}} L_o(\mathbf{b}|\mathbf{t})}{\sup_{\mathbf{b}, \mathbf{p}} L(\mathbf{b}, \mathbf{p}|\mathbf{t})}$
  - Under  $H_0$ , permute the symptom onset dates.

# Measures of Transmissibility for Close Contact Cluster

- **SAR** = secondary attack rate

- probability (%) that during his/her infectious period an infectious individual will infect a close contact



$$SAR = 1 - \prod_{t=0}^{L-1} (1 - g(t)p)$$

$g(t)$  is the probability that a case remains infective on day  $t$  of an infectious period with a maximum length of  $L$

- **Local  $R$**  = local reproductive number

- Average number of secondary cases expected to develop in a larger setting, such as a camp or school, when a typical infectious case is introduced
- Estimated as the product of the average size of the cluster and SAR for the setting

## TranStat Version

- Any number of  *b*'s and  *p*'s
- Covariate adjustment
- Flexible contact structure
- Accounts for unobserved pre-existing immunity and/or asymptomatic infection
- Accounts for missing data related to infection or symptomatic status, and missing onset times.
- Permutation test available to evaluate  $H_0: p = 0$
- Command line interface

# Model: Data Inputs

- Individual-level information
  - Household (cluster) membership
  - Covariates: for example, age or vaccination status
  - Outcome-related information: infection and symptomatic status, onset times, and laboratory test results.
  - Information about levels of existing immunity to infection
  - Indication of whether or not data are missing for each of the outcome and pre-existing immunity related data inputs
- Household or Cluster level information
  - Population and/or contact structure
  - Beginning and end of observation period for each cluster

# Case Study: Western Washington State Youth Camp and Associated Households

Determinants of the Transmissibility of Pandemic Influenza A  
(H1N1) 2009 in Community Settings

# Background

- First major outbreak of pandemic influenza A(H1N1) 2009 (pH1N1) in Washington State
- Occurred not long after the first cases were described in Mexico City
- 6<sup>th</sup>-grade students, teachers, and staff from 4 schools in Western Washington State attended a week-long camp. Camp staff were involved in the outbreak.
- Around 50% of those at the camp fell ill with symptoms consistent with influenza-like illness
- The camp closed early due to the high burden of illness.
- Symptomatic influenza subsequently reported among households members of ill camp attendees
- Public Health - Seattle & King County conducted an outbreak investigation with the assistance of the CDC's Epidemic Intelligence Service





# Data Collection

- Study design: Retrospective cohort study
- Data collection: May 18 – June 9, 2009
  - Public Health – Seattle & King County AND Centers for Disease Control and Prevention (CDC)
  - Retrospective interviews: multiple modes
  - Data:
    - symptom histories, onset dates, attendance, demographic
    - Camp participants and households of ill participants
- Determined to be public health response by the relevant IRBs

# Study Setting and Context

## **Person:**

- Camp population: 96 participants (66% of attendees)
  - 72 6<sup>th</sup>-grade students
  - 24 teachers and camp staff
- Household members (primary case definition)
  - 42 camp participants (index cases)
  - 136 household contacts

## **Place:** Western Washington State

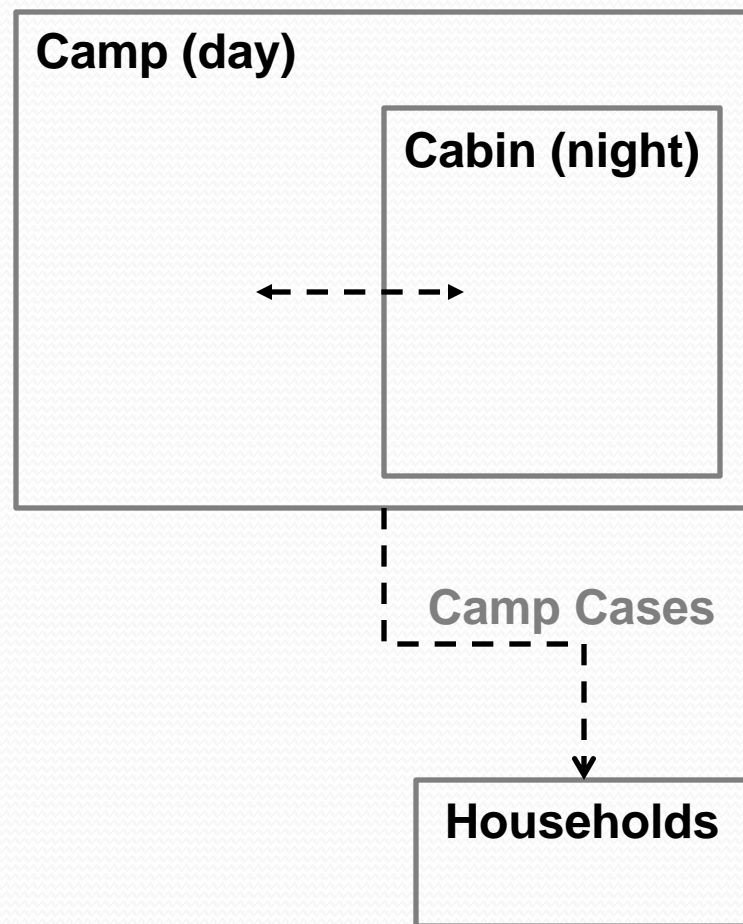
- youth camp
- 41 households of ill camp participants

## **Time:** Spring 2009

- Camp: April 25 – May 7 (closed April 30)
- Households: April 30 – May 12

# Study Objectives

- Estimate the transmission potential for symptomatic pH1N1 in a “school-like” camp and associated households
- Estimate the relative level of susceptibility to symptomatic pH1N1 among children versus adults
- Estimate a ...
  - Daytime Camp Local R
  - Nighttime Cabin SAR
  - Household SAR
  - Effect of age category on susceptibility to symptomatic pH1N1



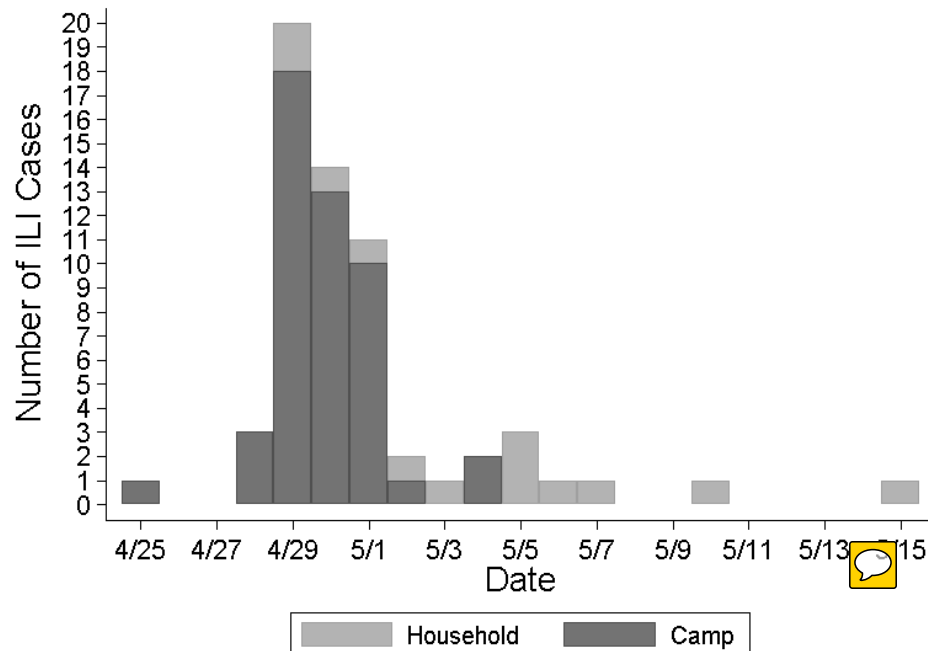
# ILI Case Definitions

- During the early phase of the pandemic, there was uncertainty with regard to the optimal case definition for ILI, so we considered a set of case definitions.
- Outcome: Symptomatic pH1N1
  - 6 case definitions
  - Primary ~ CDC's influenza-like illness (ILI)
- Predictor: Age
  - Children =  $\leq 17$  years
  - Adults =  $\geq 18$  years

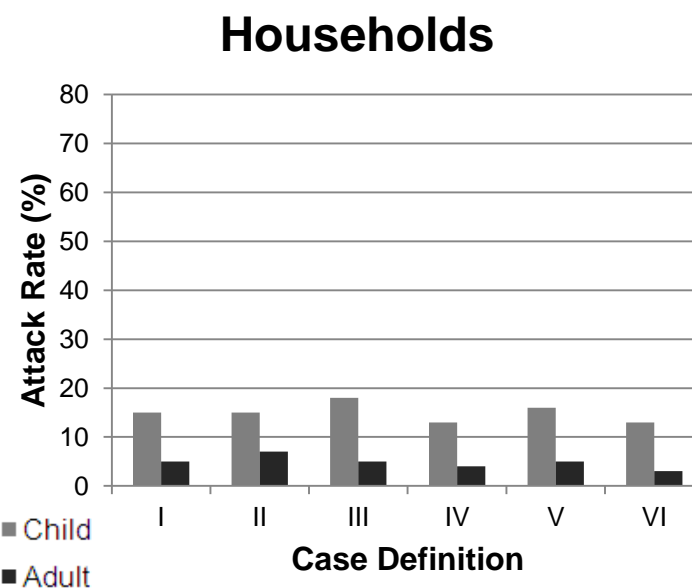
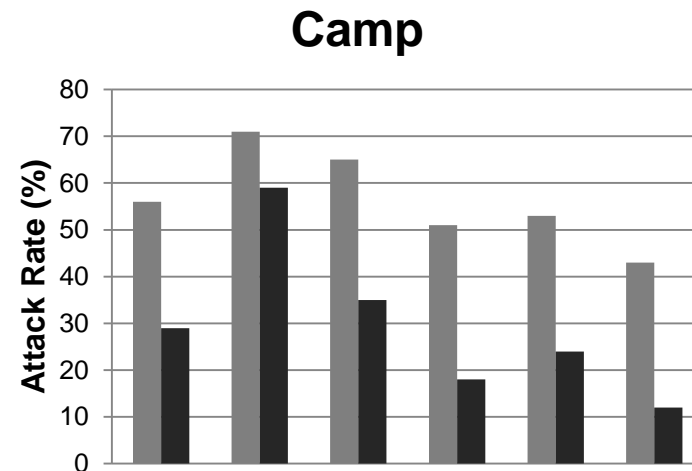
Case Definition	Symptoms
I (ILI)	- Reported Fever or Feverishness and - Cough or Sore throat
II	At least one of the following symptoms: Reported Fever, Feverishness, Cough, Sore throat, Diarrhea, Difficulty breathing, Runny nose, or Vomiting
III	Reported Fever or Feverishness
IV	Reported Fever with measured temperature $\geq 100.4^{\circ}\text{F}$ ( $38^{\circ}\text{C}$ )
V	- Reported Fever and - Cough or Sore throat
VI	- Reported Fever with measure temperature $\geq 100.4^{\circ}\text{F}$ ( $38^{\circ}\text{C}$ ) and - Cough or Sore throat

# Descriptive Statistics for the Primary Case Definition (I)

Characteristic	Camp Participants (N = 96)	Household Contacts (N = 136)
<i>No. male (%)</i>	38 (40%)	63 (28%)
<i>Age (years)</i>		
Children ( $\leq 17$ years): No. (% of all individuals)	79 (82%)	48 (35%)
Adult ( $\geq 18$ years): No. (% of all individuals)	17 (18%)	88 (65%)
Mean (SD: Range)	16 (12: 10, 59)	34 (18: 0.5, 74)
<i>Number of cabins or households</i>	13	41
<i>Individuals per cabin or household: Mean (SD: Range)</i>		
Children	7.2 (2.1: 4, 10)	1.2 (0.8: 0, 3)
Adults	3.0 (2.0: 1, 5)	2.1 (0.7: 1, 5)
All individuals	6.3 (2.8: 1, 10)	3.3 (1.3: 1, 8)

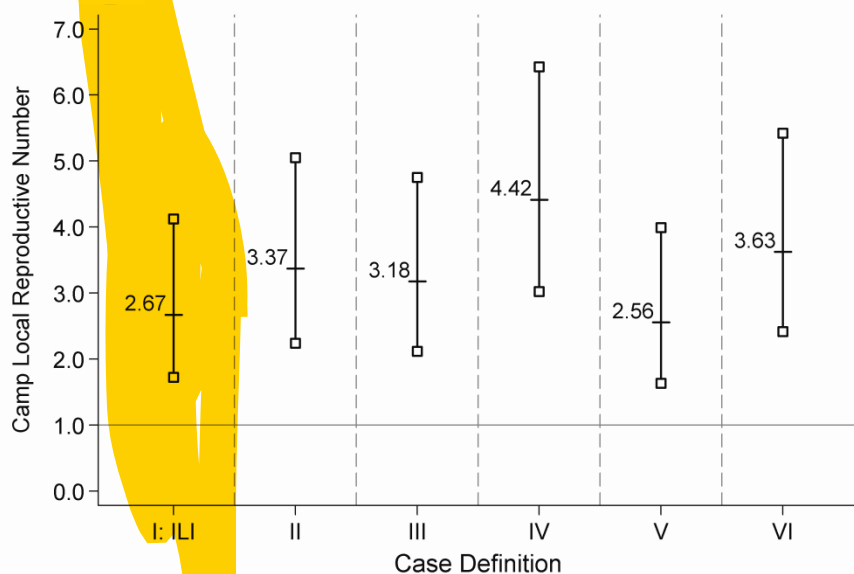


- ILI attack rate
  - Camp: 51% (N = 49)
  - Household contacts: 8% (N=11)
- Camp: 5 cases were laboratory-confirmed

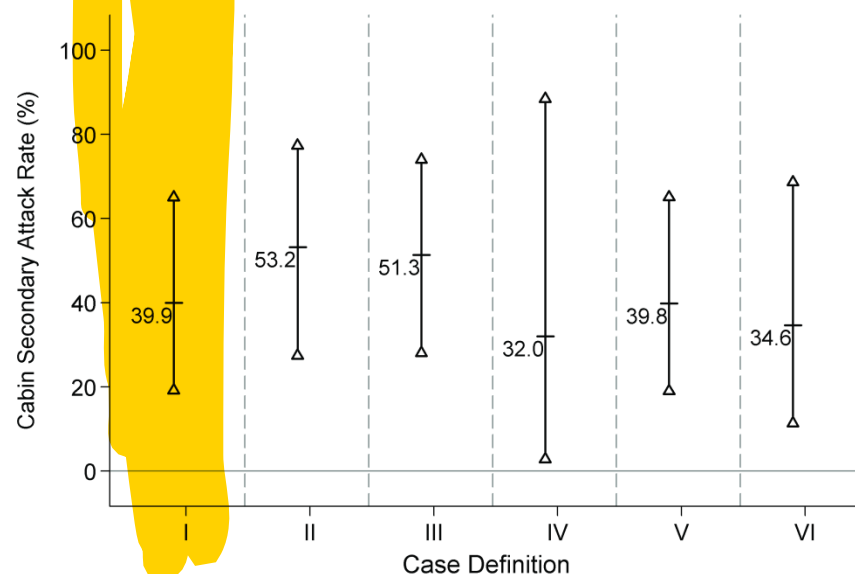


# Results: Camp Transmission

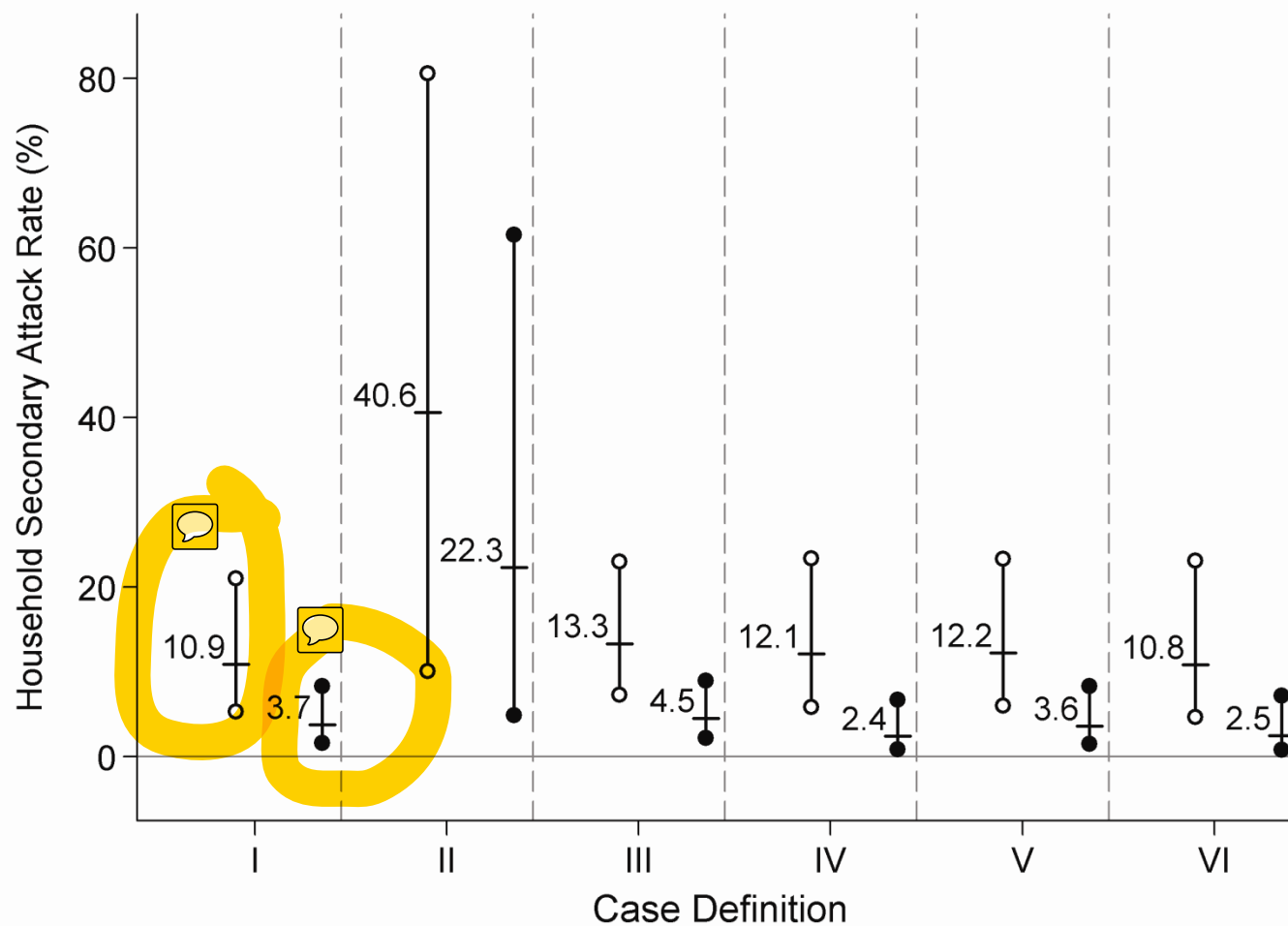
## Camp Local R: Daytime



## Cabin SAR: Nighttime

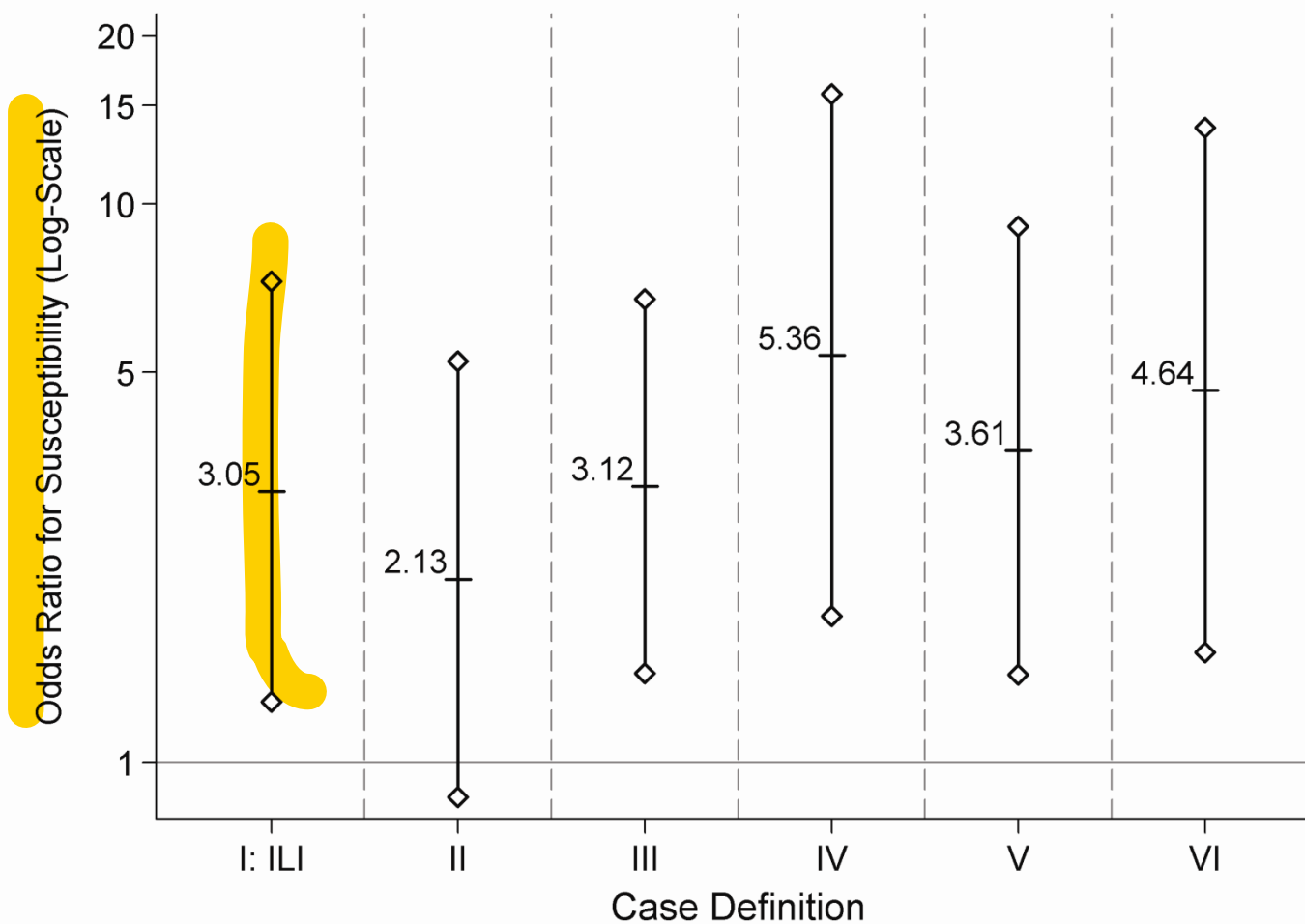


# Results:





# Results:



# Limitations

- Low survey response rate: 66%
  - Selection bias: differential response for case vs. non-case
  - If **all** non-respondents had been ...
    - Non-cases: camp ILI attack rate = 34%
    - Cases: camp ILI attack rate = 68%
  - Households: condition out the camp-attending index cases
- Limited laboratory confirmation:
  - 5 of 49 camp cases
  - Multiple case definitions: sensitivity analysis
- Small sample size: limited number of age groups

# Summary

- Observed ...
  - Children are significantly more susceptible than adults to symptomatic pH1N1
  - Elevated transmission in the camp, which is similar to levels reported for schools
  - Lower-than-expected transmission in households, which is similar to other published estimates
- SAR's and local R were not sensitivity to assumptions about the incubation/latent and infectious period distributions

# Toy Analyses to Demonstrate the Range of TranStat's Capabilities

- Note: All of these analyses have been adapted from Sugimoto *et al.* (2011)
- Analysis of
  - Independent clusters of individuals: only using data collected for household contacts
  - Dependent or interconnected clusters of individuals: only using data collected for camp attendees
  - Multiple types of clusters: using all data that was collected
  - Accounting for missing information: the missing laboratory-confirmation status for all illness camp participants and associated households

# Toy Analysis #1

## Independent Clusters

- In some circumstance it might be reasonable to assume that there was/is no meaningful interaction between each of the clusters/households included in a study sample.
- For this outbreak, we could assume that once sick camp attendees returned home they did not come into contact with the members of the other 40 households in our sample.
- Under this somewhat dubious assumption, we would have 41 independent close contact groups.
- We can estimate a
  - Household SAR, and possibly the
  - Effect of age on susceptibility to symptomatic pH1N1

# Pause to Demonstrate Toy Example #1

# Toy Analysis #2

## Dependent Clusters

- Our understanding of the typical camp setting in the US would suggest that there was likely a high degree of mixing between students, school staff, and camp staff during the day time, but that they stay in their cabins/dormitories/homes at night.
- Our assumed daytime mixing patterns introduces dependence between the otherwise independent cabins.
- Therefore, we assign each individual to a cabin for the night time (*i.e.*, they only come into contact with other cabin members during that 12 hour period), and then they are allowed to contact all other camp attendees during the 12-hours of daytime activities.
- For the purposes of this toy analysis, we ignore the data collected on household contacts.
- We can estimate a
  - Daytime Camp Local R
  - Nighttime Cabin SAR
  - Effect of age category on susceptibility to symptomatic pH1N1

# Pause to Demonstrate Toy Example #2



# Toy Analysis #3

## Multiple Cluster Types

- We jointly model the camp and the households.
- One index case among the 96 camp attendees
- Each of the 41 households contained at least one index case, who/whom had attended and developed ILI at the camp
- Estimate a ...
  - Daytime Camp Local R
  - Nighttime Cabin SAR
  - Household SAR
  - Odds ratio: Effect of age on susceptibility to symptomatic pH1N1
- Since index cases are not counted as secondary cases, we are effectively estimating separate transmission parameters for the camp and household settings, but we model age category as having the same effect on susceptibility in both settings.
- Similar to the analysis reported in the original publication

# Pause to Demonstrate Toy Example #3

# Toy Analysis #4

## Multiple Cluster Types

- As with Toy Analysis #3, we jointly model transmission and age effects in the camp and households.
- Again, we estimate the same parameters as for Toy Analysis #3.
- The available data and original analysis suffer from the potential for bias associated with lab-confirmation of pH1N1 infection for only 5 of the 49 symptomatic pH1N1 cases among camp participants.
- Since an individual's risk of symptomatic pH1N1 infection is largely dependent on the infection/infectiousness status of other individuals with whom they have close contact, we would ideally like to account for this uncertainty in true infection status.
- More-recent versions of TranStat have included facilities for implementing a hybrid EM-MCEM algorithm to integrate over missing information about outcome status, outcome onset time, and/or level of existing immunity.
- The amount of computation time required to complete this integration increases quickly with the level of missingness, leading to the need for the MCEM portion of the hybrid algorithm.
- We use the EM-MCEM algorithm to integrate over the missing pH1N1 infection status of all ILI cases reported in the original dataset.

# Pause to Demonstrate Toy Example #4

# Lecture Summary

- TranStat is designed to..
  - Estimate transmission parameters from clustered infectious disease surveillance data
  - Estimate covariate effects on transmission
  - Provide real-time estimates of these parameters
- The data input format and transmission model are quite flexible, making TranStat useful for analyzing a wide range of potential situations involving transmission of an acute infection within clusters/groups of individuals
- Additional example analyses and updated versions of will be available via [www.cidid.org/software-development/](http://www.cidid.org/software-development/).