Lecture 6: Real-time analysis of infectious disease outbreaks using TranStat

Jonathan D. Sugimoto, PhD
jons@fredhutch.org

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

Lecture 6 Outline

- Overview of TranStat
- Basic description of the statistical model implement by TranStat
- Case Studies
  - Case study 1: Illustrative example
  - Case study 2: Dependent cluster data
  - Case study 3: Independent cluster data
  - Case study 4: Accounting for missing outcome information
  - Case study 5: Multiple types of clusters
- Summary

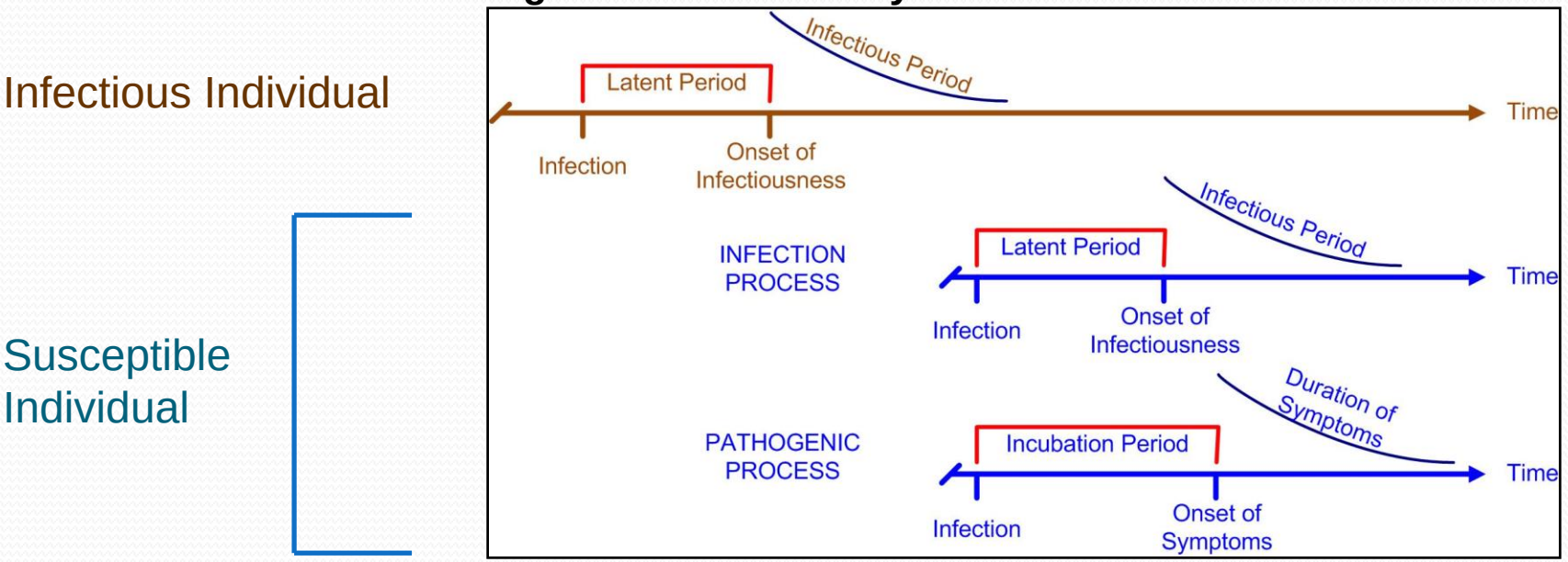
Motivation

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

- Detecting individual-to-individual (person-to-person) transmission of pathogens
- Evaluating the transmissibility of pathogens
- Evaluating the effects of risk factors and interventions on transmission
- Performing simulation studies, for example, to perform power calculations for study design purposes

Basic Concepts: Natural History of Infection and Disease

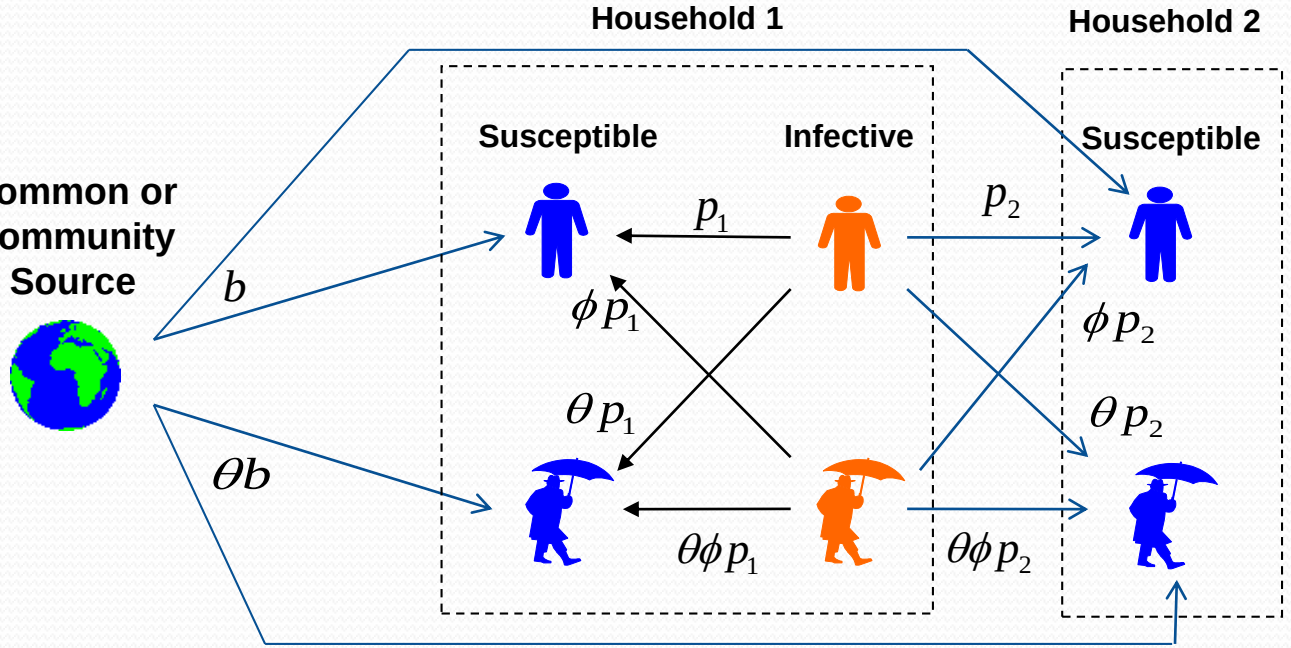
Figure. Natural History of Infection and Disease



- Infection depends upon exposure to an infectious individual (see next slide for more details)
- Both concurrently occurring processes are often (or are assumed to be) strongly correlated, for example, onset of symptoms may be assumed to indicate onset of infectiousness
- Individuals do not necessarily complete each process in its entirety, *e.g.*, an individual may become infectious, but never exhibit clinically-apparent symptoms (infectious asymptomatic infection)

Basic Concepts: Exposure

Figure. Population and Contact Structure



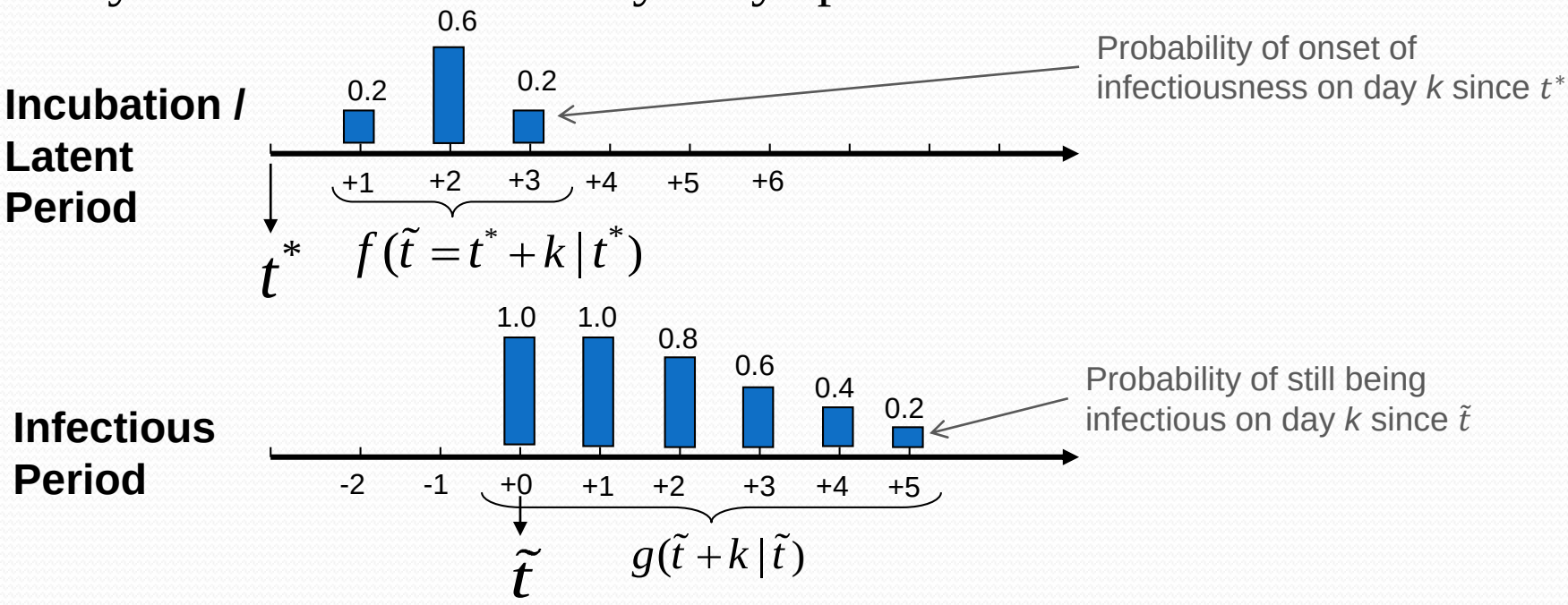
- Contact = exposure to a specific source of infection for a defined period of time (typically, a day)
- 'Household' = general term for clusters of individuals who are more likely to mix with each other than with other members of the population. Multiple types of households may be defined.
- Types of 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
- θ and ϕ denote covariate effects (risk-factors or interventions) on susceptibility and infectiousness, respectively.

Model: Data Inputs

- Individual-level information
  - Household (cluster) membership
  - Covariates: *e.g.*, age or vaccination status
  - Outcome-related information: infection and symptomatic status, onset times, and laboratory test results.
  - Information about any pre-existing immunity to infection
  - Indication of whether or not data is 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

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

t^*: day of infection $\tilde{t}$: day of symptom onset



- 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 period.
- TranStat inputs: Minimum and maximum values for *k* and the daily probability distribution (blue bars)

Likelihood

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

  $$logit(p_{ijt}) = logit(p) + \boldsymbol{X_i}\beta_S + \boldsymbol{X_j}\beta_I + \boldsymbol{X_{ij}}'\beta_{SI}, j \in \mathcal{H}_i$$

- An important example of interaction:
  - Let r_i be the vaccination status and the only covariate for person *i*
  - $logit(p_{ijt}) = logit(p) + r_i\theta + r_j\phi + r_i \ \ r_j\psi$
  - $VE_S = 1 - \theta, VE_I = 1 - \phi, VE_T = 1 - \psi$

Likelihood (continued)

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

$$logit(b_{it}) = logit(b) + \boldsymbol{X_i}\alpha_S$$

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

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

- 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

- Selection bias: a household is observed only upon ascertainment of an index case
  - 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.

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

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: $\lambda = -2log \frac{sup_{\boldsymbol{b}} L_o(\boldsymbol{b}|\boldsymbol{t})}{sup_{\boldsymbol{b},\boldsymbol{p}} L(\boldsymbol{b},\boldsymbol{p}|\boldsymbol{t})}$
  - Under H_0, permute the symptom onset dates.

Note about previous Version 1

- Can fit simple models with b, p_1, and p_2, but no covariates
- GUI available
- Data input and basic editing functions available
- Sample datasets provided
- No longer under development, so bugs are still present

TranStat Version 3

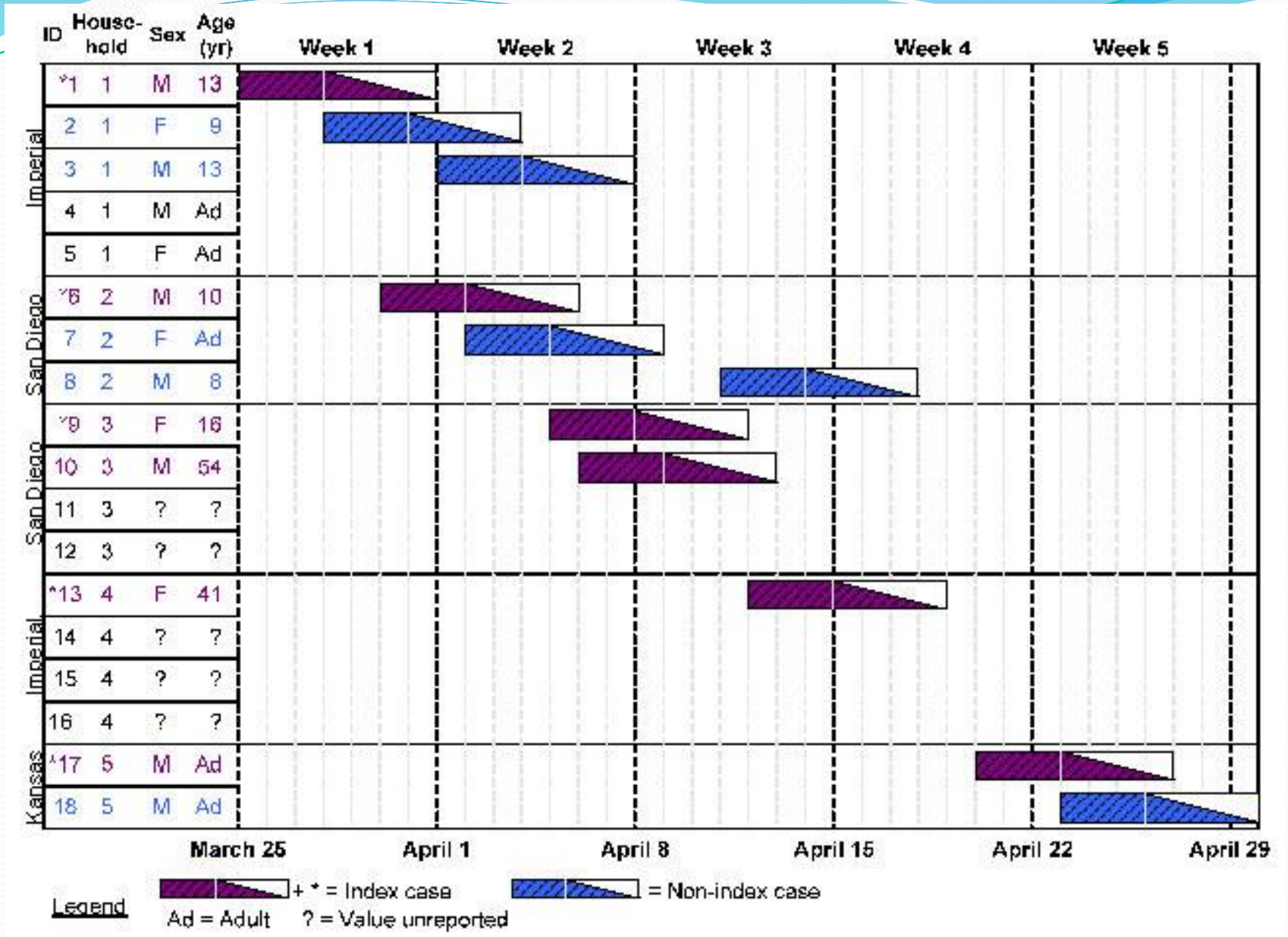
- Any number of *b*'s and *p*'s
- Covariate adjustment
- Flexible contact structure
- Accounts for unobserved pre-existing immunity and/or asymptomatic infection
- Accounts for missing data related to infection or symptomatic status, and missing onset times.
- Permutation test available to evaluate $H_0: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

Case Studies 1 and 2

Novel Influenza Strains

Case Study 1: US household outbreaks of Influenza A(H1N1) 2009

- Household structure is known => can model within-household transmission.
- Households not in the same neighborhood => can not model inter-household transmission.
- Households can be regarded as independent mini-communities.
- People in the same households share the same history of contact and exposure.


ID
House-hold
Sex
Age (yr)
Week 1
Week 2
Week 3
Week 4
Week 5
Imperial
San Diego
San Diego
Imperial
Kansas
*1 1 M 13
2 1 F 9
3 1 M 13
4 1 M Ad
5 1 F Ad
*6 2 M 10
7 2 F Ad
8 2 M 8
*9 3 F 16
10 3 M 54
11 3 ? ?
12 3 ? ?
*13 4 F 41
14 4 ? ?
15 4 ? ?
16 4 ? ?
*17 5 M Ad
18 5 M Ad
March 25
April 1
April 8
April 15
April 22
April 29
Legend
+ * = Index case
= Non-index case
Ad = Adult
? = Value unreported

Case Study 1: Influenza A(H1N1) 2009 outbreak in Mexico

- People: 2,895 confirmed cases
- Time: March 11–? We use the data up to May 15.
- Case numbers are aggregated by day.
- Contact structure is unknown.
- R_0 is estimable:
  - Distribution of serial interval based on all possible transmission networks
  - Chain binomial model

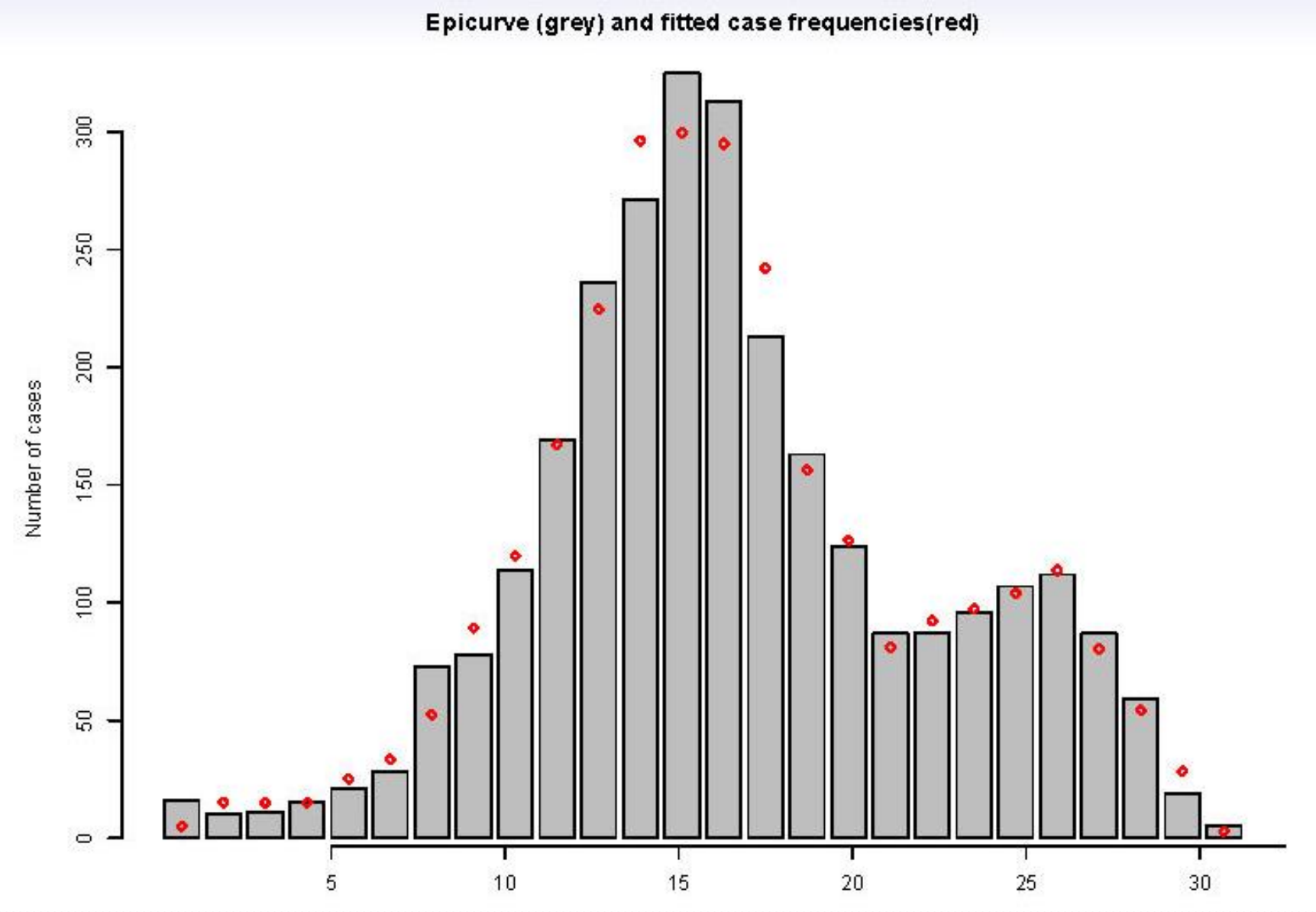
Case Study 1 (continued)

- For a large population, the chain binomial model converges to the Poisson distribution
- On day t, observe number of susceptibles *S(t)*, infectives *I(t)*, and new infections *X(t)*.

$$\binom{S(t)}{X(t)}\{1-(1-p)^{I(t)}\}^{X(t)}(1-p)^{I(t)S(t+1)}$$

$$\rightarrow \frac{(\lambda I(t))^{X(t)}}{X(t)!}\exp\{-\lambda I(t)\}$$

- $\hat{\lambda}=\frac{\sum_{t=1}^{T}X(t)}{\sum_{t=1}^{T}I(t)}\rightarrow\lambda$ a.s., and $\hat{R}_0=D\hat{\lambda}\rightarrow R_0$ a.s.
- To use TranStat, create D-1 uninfected people for each observed case. D = 100 is sufficiently large.


Epicurve (grey) and fitted case frequencies(red)
Number of cases
0
50
100
150
200
250
300
5
10
15
20
25
30

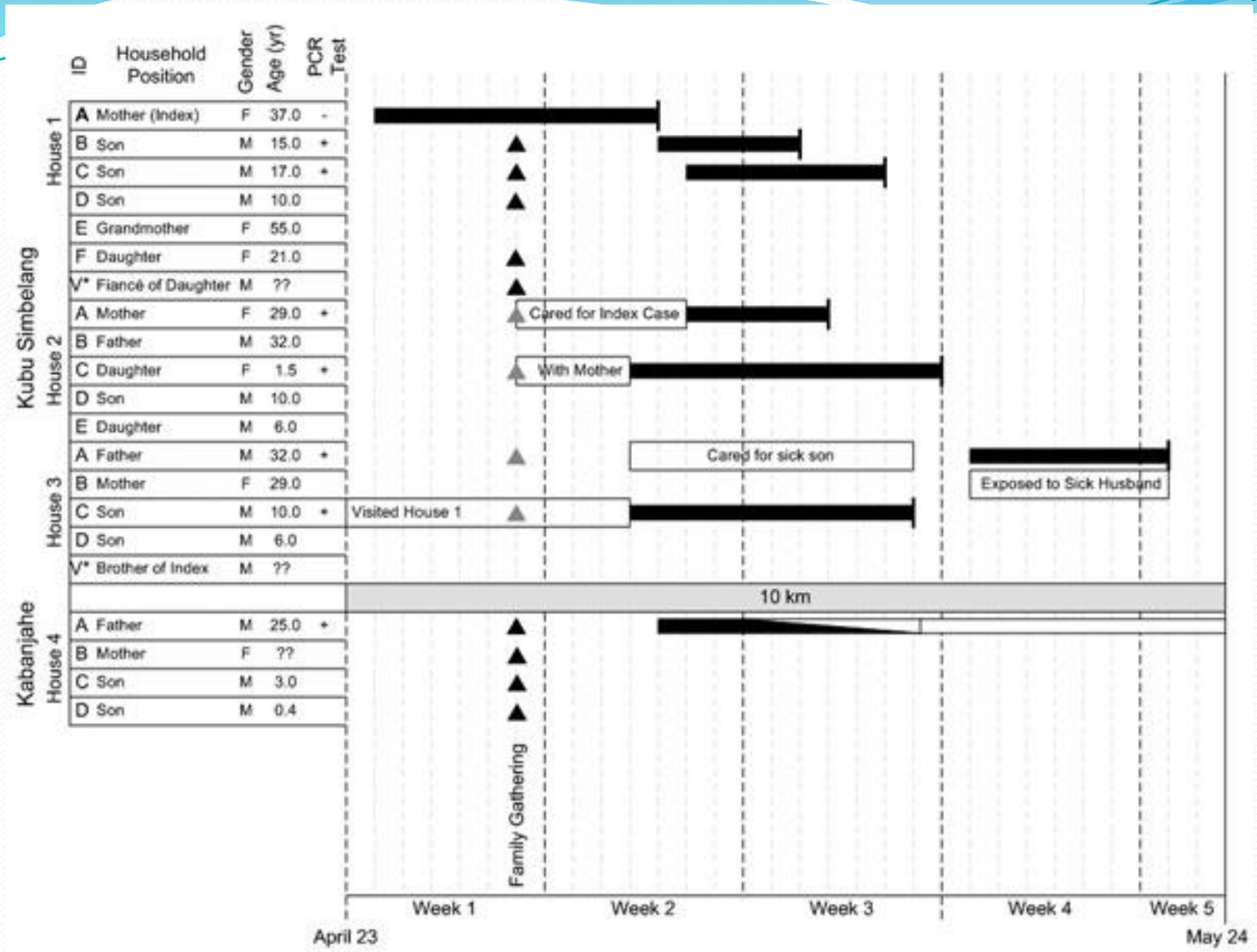
Case Study 1: Analysis Results

- Fixed community-to-person probability at an expected value based upon external data sources
- Household Secondary Attack Rate:
  - 20.5% (95% CI: 7.1%, 46.4%)

- Similar analysis using these data (Yang et al. 2009):
  - School local R = 2.4 (95% CI: 1.8, 3.2)
  - R_0: ranged from 1.3 to 1.7

Case Study 2: Indonesian household outbreaks of avian influenza A(H5N1)

- An outbreak caused by a family gathering of multiple households.
- Transmission occurred both within and between households.
- In TranStat, clusters that have cross-transmission should be considered as a single community.
- Individual level contact and risk history.


ID
Household Position
Gender
Age (yr)
PCR Test
Kubu Simbelang
House 1
A Mother (Index) F 37.0 -
B Son M 15.0 +
C Son M 17.0 +
D Son M 10.0
E Grandmother F 55.0
F Daughter F 21.0
V* Fiancé of Daughter M ??
House 2
A Mother F 29.0 +
B Father M 32.0
C Daughter F 1.5 +
D Son M 10.0
E Daughter M 6.0
House 3
A Father M 32.0 +
B Mother F 29.0
C Son M 10.0 +
D Son M 6.0
V* Brother of Index M ??
Kabanjahe
House 4
A Father M 25.0 +
B Mother F ??
C Son M 3.0
D Son M 0.4
Cared for Index Case
With Mother
Cared for sick son
Exposed to Sick Husband
Visited House 1
10 km
Family Gathering
Week 1
Week 2
Week 3
Week 4
Week 5
April 23
May 24

Analysis Results

- Fixed community-to-person probability at an expected value based upon external data sources
- Household Secondary Attack Rate:
  - 20.6% (95% CI: 6.4%, 49.6%)
- Community-to-person Probability of Infection:
  - 17.1% (95% CI: 3.0%, 67.6%)
- Local Reproductive Number:
  - 0.82 (95% CI: 0.26, 2.64)

Input File Formats

Input Files for TranStat 3

DO NOT INCLUDE COLUMN TITLES IN ANY TRANSTAT INPUT FILE!

- Household / Cluster profile: “community.dat”

Household / Cluster ID	Start Observation	End Observation
1	1	45
…	…	…
H	34	56

- Population profile: “pop.dat”

Person ID	Cluster ID	Pre-existing Immune Status	Infection Status	Symptomatic Status	Onset Time	Index Case Indicator	Disease Outcome	Disease Outcome Time	Pathogenicity Type	Pre-existing Immunity Type	Ignore Indicator
1	1	0	1	1	34	1	1	39	0	0	0
…	…	…	…	…	…	…	…	…	…	…	…
N	C	1	0	0	-1	0	0	-1	0	0	0

Input Files for TranStat 3 (continued)

- Time independent covariates: “time_ind_covariate.dat”
  - One line per individual
  - One column per covariate
  - No missing information

Person ID	Age	Vaccination Status	Gender
1	34	0	0
…	…	…	…
N	103	1	1

- Time dependent covariates: “time_dep_covariates.dat”
  - One line per individual per time period (a set of one or more contiguous time units)
  - One column per covariate
  - No missing information

Person ID	Start Time (day)	End Time (day)	Antiviral Prophylaxis
1	1	3	0
…	…	…	…
N	45	56	1

Input Files for TranStat 3 (continued)

- C2P contact file: “c2p_contact.dat”
  - C2P contacts can be indexed in three manners
    - no ID, which assumes the same contact history for all individuals

Start Time (day)	End Time (day)	Type of C2P Contact	Weight	Ignore C2P Contact Indicator
1	66	0	0	0
…	…	…	…	…
28	28	1	0.85	1

    - by cluster ID, which assumes the same contact history for all members of a cluster
    - by person ID, which specifies a separate contact history for each individual

Cluster or Person ID	Start Time (day)	End Time (day)	Type of C2P Contact	Weight	Ignore C2P Contact Indicator
1	1	66	0	0	0
…	…	…	…	…	…
C or N	28	28	1	0.85	1

  - Contact types are numbered using consecutive non-negative integers, beginning with 0

Input Files for TranStat 3 (continued)

- P2P contact file: “p2p_contact.dat”
  - P2P contacts can be indexed in three manners
    - by cluster ID, which assumes the same contact history between all members of a cluster (requires indexing c2p_contact.dat by cluster ID)

Cluster ID	Start Time (day)	End Time (day)	Type of P2P Contact	Weight	Ignore P2P Contact Indicator
1	1	66	0	0	0
…	…	…	…	…	…
C	28	28	1	0.85	1

    - by person ID, which specifies a separate contact history between each individual

Start Time (day)	End Time (day)	Person ID: Infective	Person ID: Susceptible	Type of P2P Contact	Weight	Ignore P2P Contact Indicator
1	66	1	4	0	0	0
…	…	…		…	…	…
28	28	N	66	1	0.85	1

  - Contact types are numbered using consecutive non-negative integers, beginning with 0

Input Files for TranStat 3 (continued)

- Imputation control file: “impute.dat”
  - Include one row per individual for whom at least one outcome or pre-existing immunity related value is missing

Person ID	Possible Pre-Existing Immunity	Possible Escape	Possible Symptomatic Infection	Start Time for Imputing Symptomatic Infection Onset Time	Stop Time for Imputing Symptomatic Infection Onset Time	Possible Asymptomatic Infection	Start Time for Imputing Asymptomatic Infection Onset Time	Stop Time for Imputing Asymptomatic Infection Onset Time
1	1	1	0	-1	-1	0	-1	-1
…	…	…	…	…	…	…	…	…
N	0	0	1	1	10	1	1	10

Configuration File

- Natural history of disease, *i.e.*, incubation and infectious periods.
- Profile of parameters to be estimated
  - Numbers of C2P and P2P contact types
  - Numbers of time-independent and time-dependent covariates
- Covariates effects on…
  - Susceptibility due to exposure through…
    - C2P contact
    - P2P contact
  - Infectiousness
  - Interaction between C2P and P2P transmission
- Define equivalence classes of parameters
- Specify which parameters have fix values

Configuration File (continued)

- How should TranStat handle C2P and P2P contact files.
  - Community members share contact history?
  - Community members share risk history (same covariates)?
  - Auto-generate the C2P/P2P contact files?
- Choose whether or not to adjust for selection bias and/or right censoring
- Choose whether or not to calculate/perform goodness-of-fit, case fatality ratio (deprecated), hypothesis test (under development), etc.
- Controlling optimization routine
- Controlling output
- Controlling data augmentation/multiple imputation procedures for missing data for variables related to outcome or pre-existing immunity

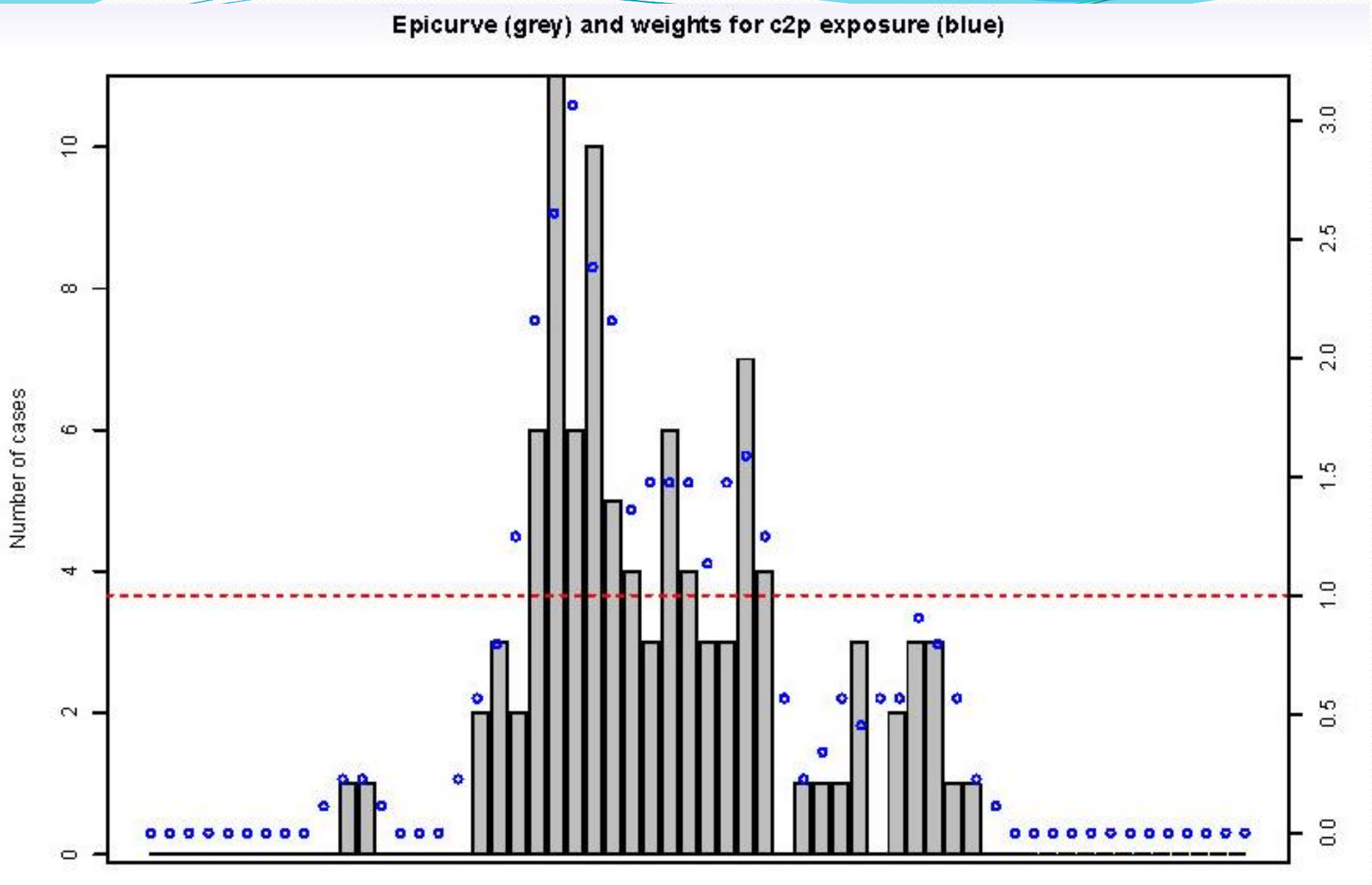
TranStat Output

- Estimates file: “estimates.txt”
  - Outputs estimates, standard errors, and 95% confidence intervals for *b*’s, *p*’s, covariate effects, CPI, SAR’s, R_0 (or local R), variance-covariance matrix
  - $CPI_x = 1-(1-b_x)^D$, where *D* is the average duration of exposure to the common source
  - $SAR_x = 1 - \sum_{t=0}^{Z}(1 - g(t|\tilde{t}_j)p_x)$, where *Z* denotes maximum length of the infectious period and $g(t|\tilde{t}_j)$ specifies the probability that infected individual j is infectious on day t given onset of symptoms on day $\tilde{t}_j$.
- Error file: “error.txt” – list of any errors encountered during the estimation process

Case Study 3

Case study 3: Influenza A(H1N1) 2009 household outbreaks in Los Angeles

- A total of 58 households with $\geq$ 1 cases, non-random sample.
- 60 index cases and 37 secondary cases.
- All index cases were laboratory confirmed with either pandemic H1N1 or seasonal influenza A.
- Outbreaks started from April 22 to May 19, 2009.
- Ages are known for all, and seasonal flu vaccine and antiviral treatment are known for part of the surveyed population.
- Missing information:
  - asymptomatic infection
  - pre-existing immunity: Assumed to be non-existent in this population, because this strain of influenza A was first described in humans during the spring of 2009.
- EM-MCEM (Yang et al., Biometrics 2012)

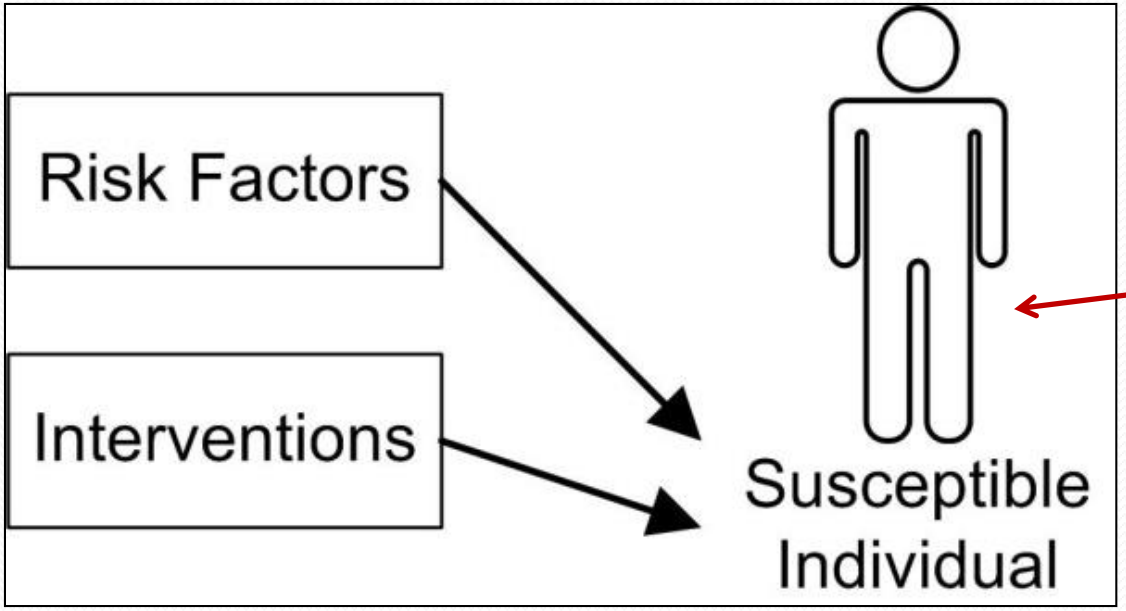

Epicurve (grey) and weights for c2p exposure (blue)
Number of cases
0
2
4
6
8
10
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Pause to Demonstrate Case 3

Case Study 4: Household Transmission of *Vibrio cholerae* O1/O139 in Bangladesh

Models for Infectious Disease Risk

Standard Epidemiologic Model



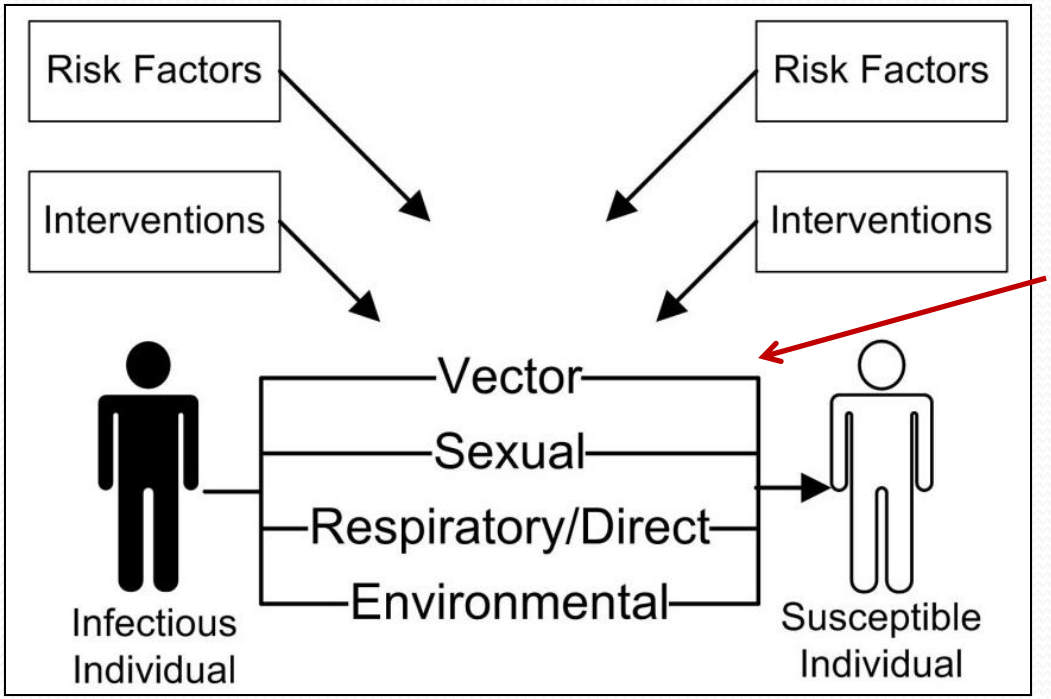
- **Assumes equal levels of exposure to infection within covariate strata**
- **Assumption is often invalid for infectious diseases**

Models for Infectious Disease Risk

General Transmission Model

Risk factors and Interventions:

- **Modify risk of transmission**
- **Differentiate effects on infectivity vs. susceptibility**



Susceptible exposed to an infectious individual

- **Accounts for variation in the level of exposure to infection**
- **CHALLENGE: Measuring the level of exposure to infection among susceptible individuals**

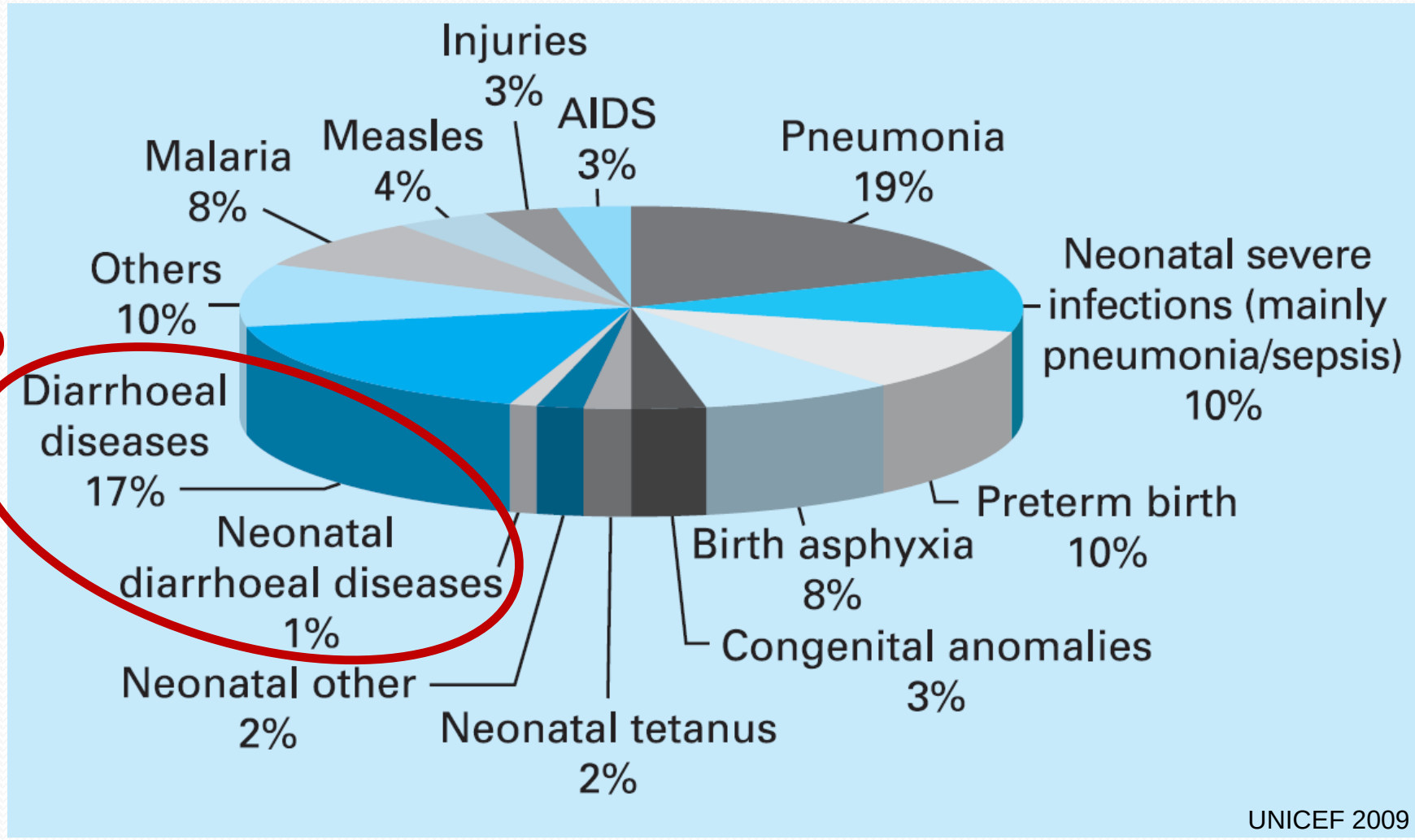
My Research Focuses…

1. **Design, implement, and analyze epidemiologic studies of infectious diseases transmission**

2. **Develop novel statistical methods and designs for transmission studies**

3. **Apply to infections of global health import**

Proximate Causes of Death Among

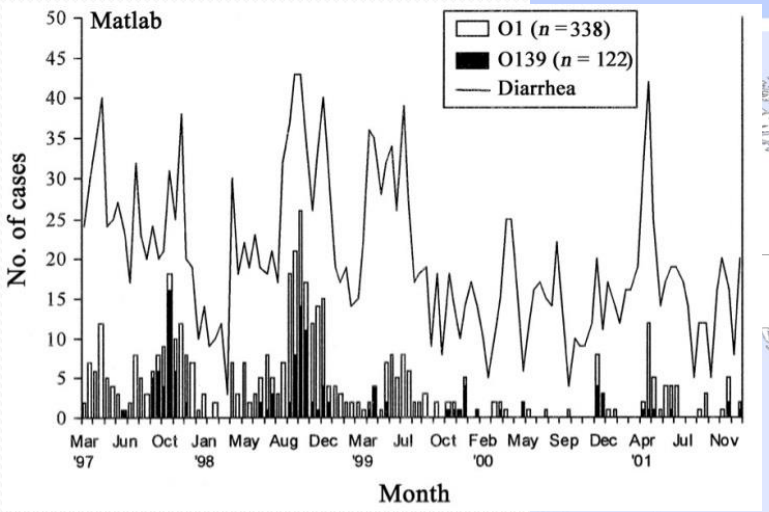


18%

Human Cholera

- *Vibrio cholerae*, primarily serogroups O1/O139
- Multiple bio- and sero- types of O1
- “Rice water” diarrhea +/- vomiting
- Vibrio is shed in stool

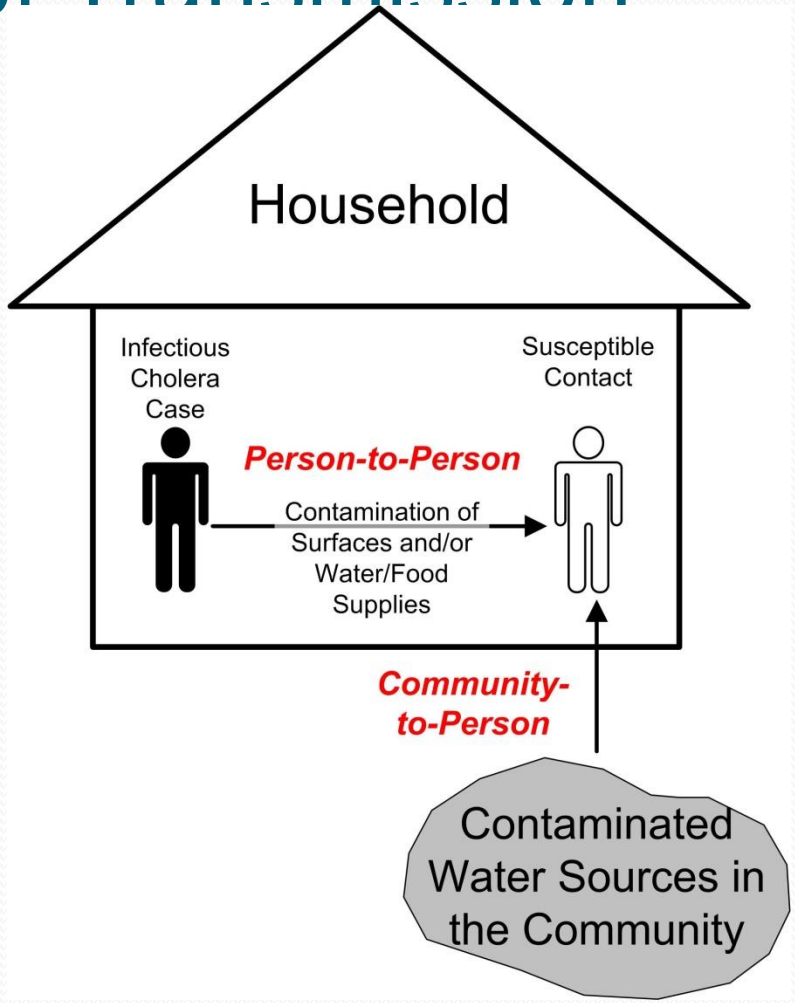
- 3-5 x 10^6 cases (WHO 2010)
- 100-130 x 10^3 deaths (WHO 2010)
- Seasonal outbreaks in endemic settings



Sack (2003)

Conceptual Model of Transmission

- Community-to-Person exposure
  - Substantial evidence [a]
  - Optimal interventions:
    - Clean drinking water technologies
    - Targeted pre-epidemic vaccination of high risk groups [b]
- Person-to-Person exposure
  - Indirect evidence [c]
  - Optimal interventions:
    - Promotion of better personal hygiene practices
    - Pre-epidemic vaccination of entire households
- Ongoing debate: Relative contribution of person-to-person exposure to endemic transmission [d]



[a] Huq 1990, Islam 1997, Sack 2003, Constantin 2008
[b] Chao 2011
[c] Clemens 1990, Mosley 1965, Kendall 2010, others
[d] Sack 2004, Pollitzer 1959

Goal

Characterize the role of person-to-person exposure in the endemic transmission *V. cholerae*, by serogroup-serotype

=> inform the selection of cholera prevention/control strategies

Scientific Objectives

- Test for the presence of person-to-person transmission within households
- Estimate the transmissibility of cholera through…
  - person-to-person exposure within households
  - community-to-person exposure
- Estimate the effects of potential risk factors on transmission
  - Age, Sex, and ABO blood group
- Describe aspects of the natural history of endemic cholera

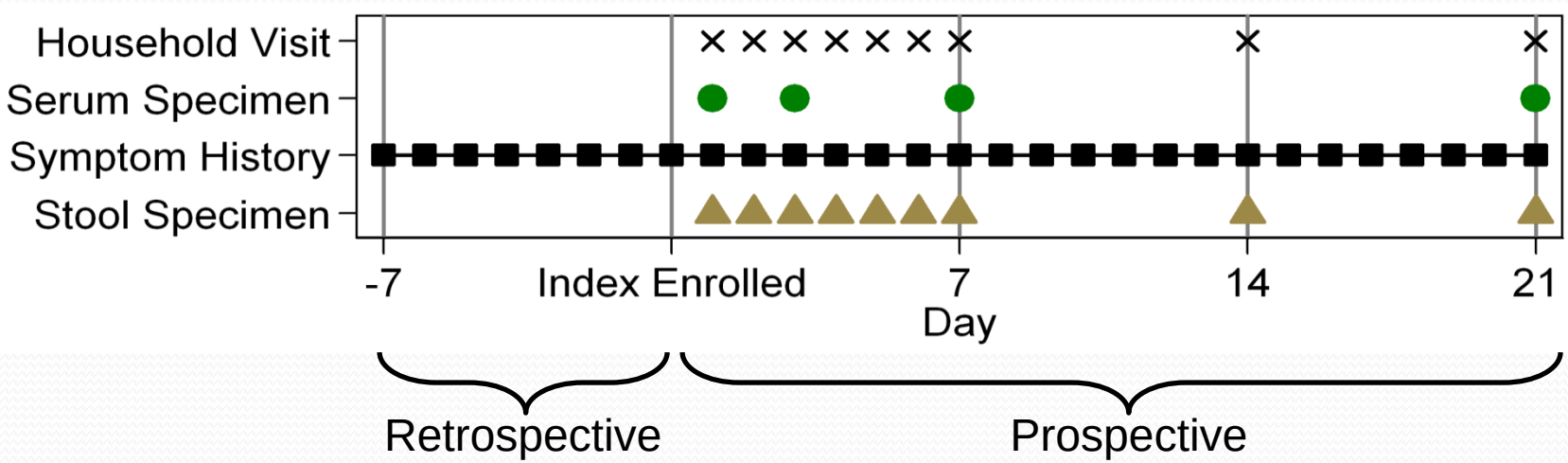
Study Design

- **Design**: Prospective follow-up of the households of hospital-ascertained cholera cases [a,b]
- **Time**: 01/2002 to 05/2006
- **Place**: 364 households in urban Dhaka, Bangladesh
  - Index cases are hospital-ascertained
    - Acute watery diarrhea (≥3 watery stools per day)
    - Stool culture positive for *V. cholerae* infection
  - Members of the household enrolled after receipt of informed consent
- **Person**:
  - 364 index case
  - 1050 household contacts

[a] Harris 2005, Weil 2009, Kendall 2010
[b] Example of case-ascertained study design (Yang 2006)

Data

Data collected for EACH member of an enrolled household



Laboratory Tests Performed

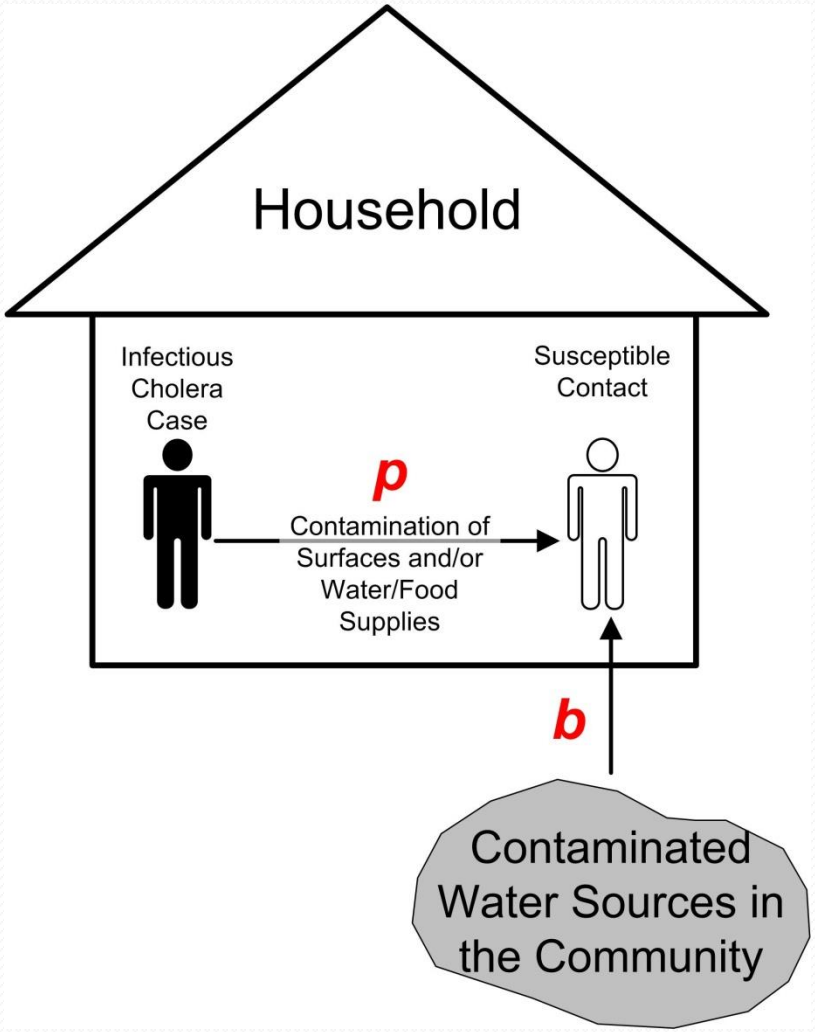
- Blood specimens: vibriocidal antibody titers
- Stool specimens: cultured for *V. cholerae* O1/O139, with serogroup-serotype determined
  - O1 El Tor Ogawa
  - O1 El Tor Inaba
  - O139

Outcome = cholera infection

- Infection = positive stool culture or ≥4-fold rise in vibriocidal antibody titer
- Infectious = ≥ 1 positive stool culture
- Onset of infectiousness = first stool specimen culture-positive for *V. cholerae*

Transmission Model - 1

- **Parameters**
  - ***b*** = infection probability per daily community-to-person exposure
  - ***p*** = transmission probability per daily person-to-person exposure



Transmission Model - 2

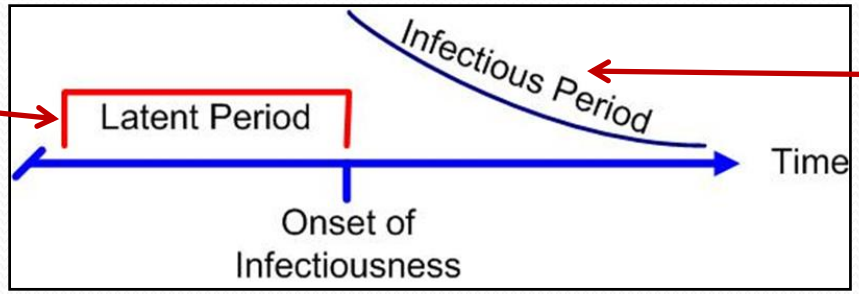
- **Extension of the chain-binomial model** [a]
  - Accounts for ascertainment bias in the enrollment process
  - Risk factors affect susceptibility to cholera infection
  - Missing onset and serotype information: ML EM algorithm [b]
  - Likelihood ratio test [c] of the null hypothesis of no person-to-person transmission within households: H_0: p=0

[a] Yang 2006
[b] Yang et al. *Biometrics* in press
[c] Yang et al. 2007

Transmission Model - 3

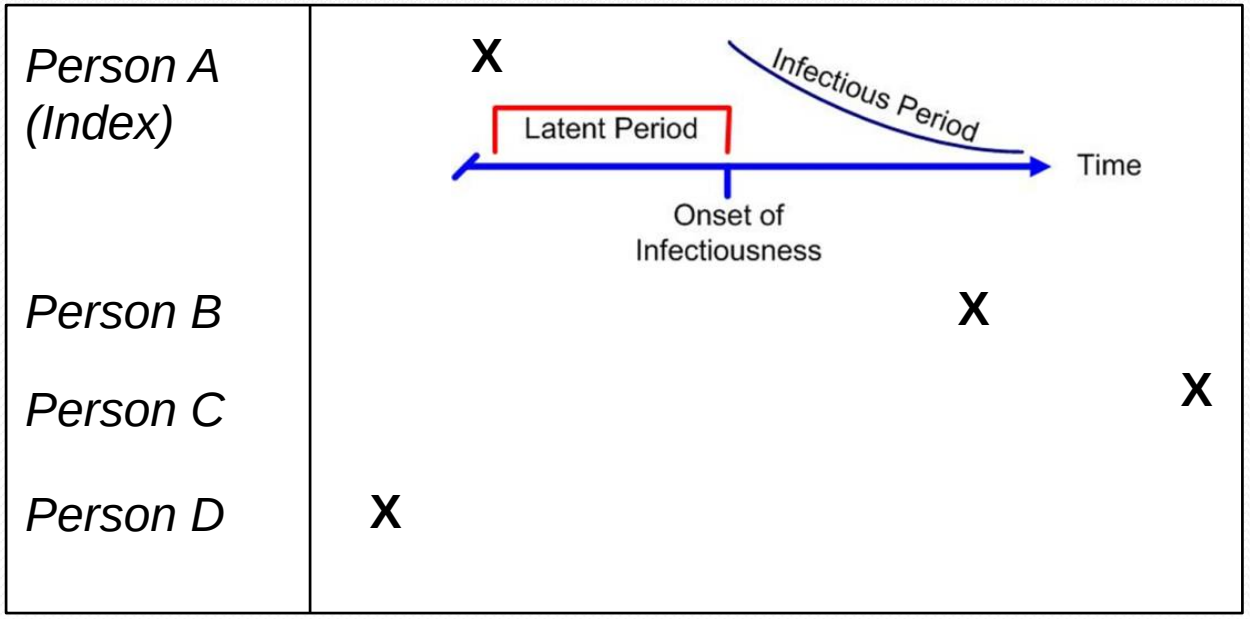
- Information in the data: Relative timing of the onset dates within a household

1-5 days uniform

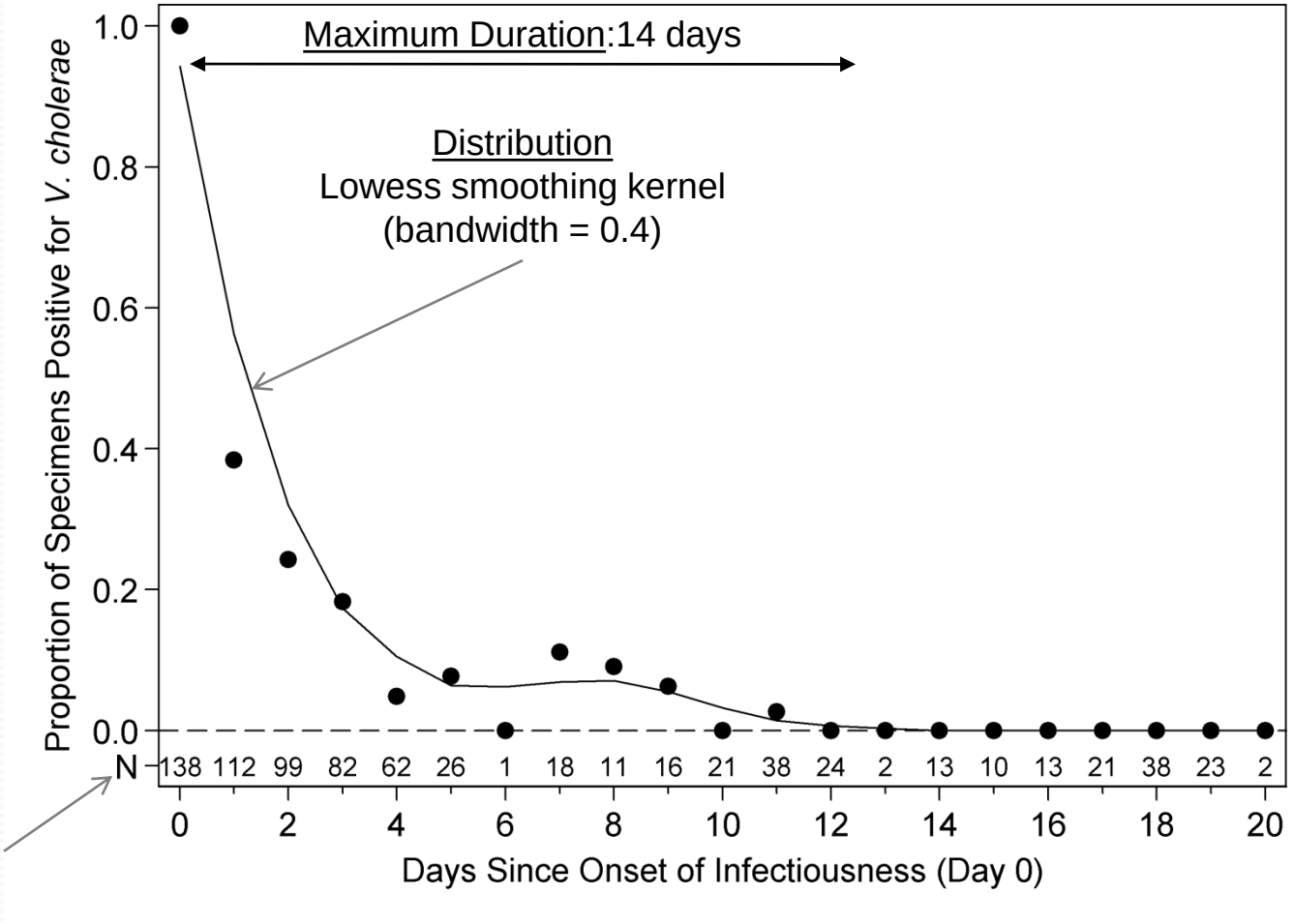


Derived from the data (next slide)

Example: Simplified Transmission Scenario within a Household



Empirical Infectious Period Distribution, *g(t)*



N = Number of Specimens Cultured

Epidemiologic Summary Measures

- ***SAR*** = household secondary attack rate
  - probability (%) that during his/her infectious period an infected individual will infect a household contact through within-household person-to-person exposure

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

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

- ***CPI*** = community probability of infection
  - probability (%) that a household contact will be infected through exposure to a community-based source of infection

during a 14-day period

$$CPI = 1 - (1 - b)^{14}$$

Descriptive Statistics

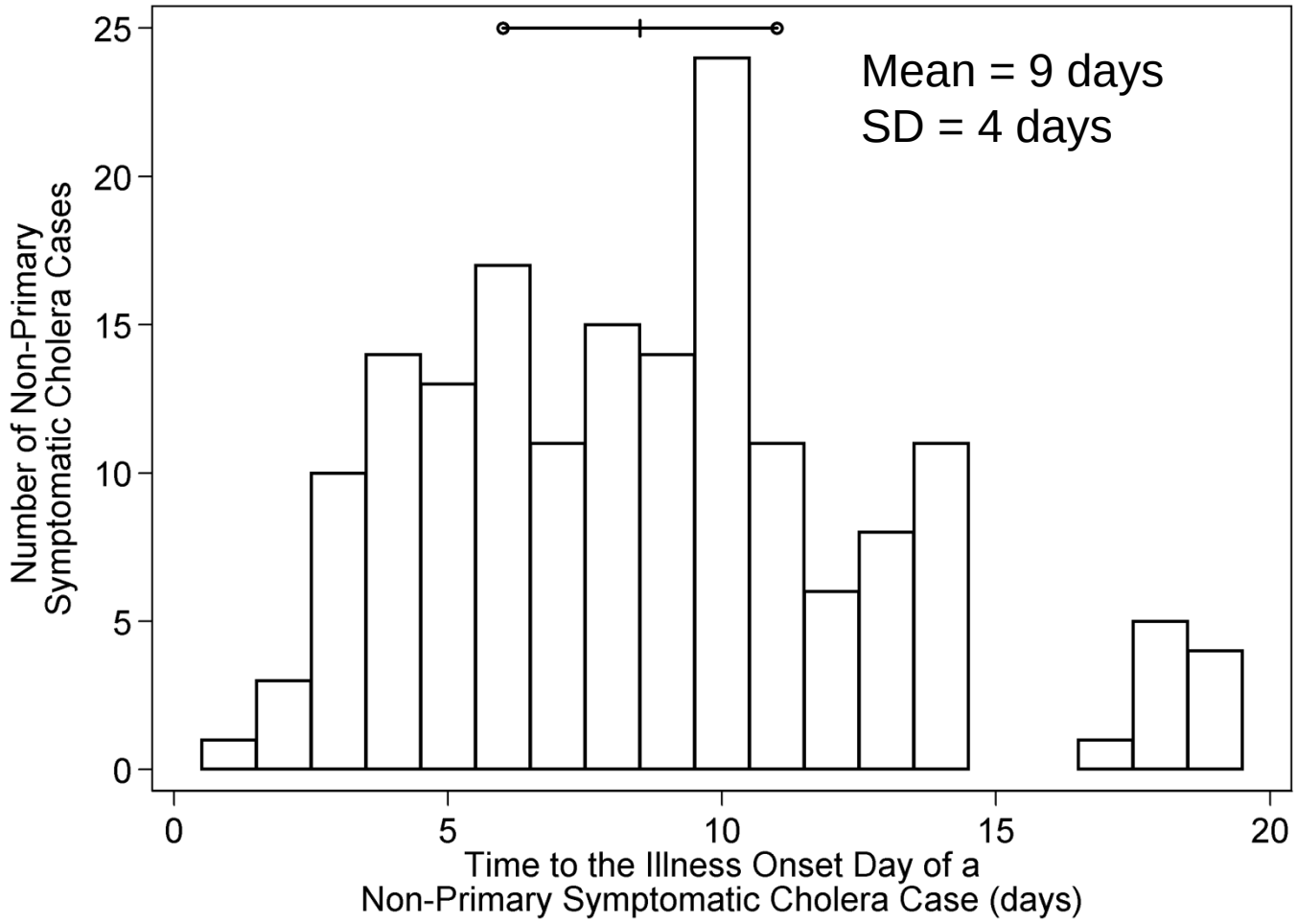
Covariate	All Members	Index Cholera Infections	Household Contacts	
			Non-Index Cholera Infections	Non-Infections
Number of individuals : (% of All Members)	1414	364 (26%)	318 (23%)	732 (51%)
***Age* (years):**				
Mean (SD)	22 (15)	24 (14)	19 (15)	23 (15)
Median	20	23	15	20
Male sex: % (SE)	49% (1.3%)	44% (2.6%)	51% (2.8%)	51% (1.8%)
Rice water diarrhea: %	54%	100%	57%	29%
≥1 Culture-positive stool specimen: %	42%	100%	70%	0%
O1 Ogawa : O1 Inaba : O139 : Unknown: %		34 : 49 : 17 : 0	22 : 30 : 18 : 30	

- 43% of non-index cholera infections were asymptomatic (no rice water diarrhea)
- 70% of non-index cholera infections had a positive stool culture (*i.e.*, infectious)

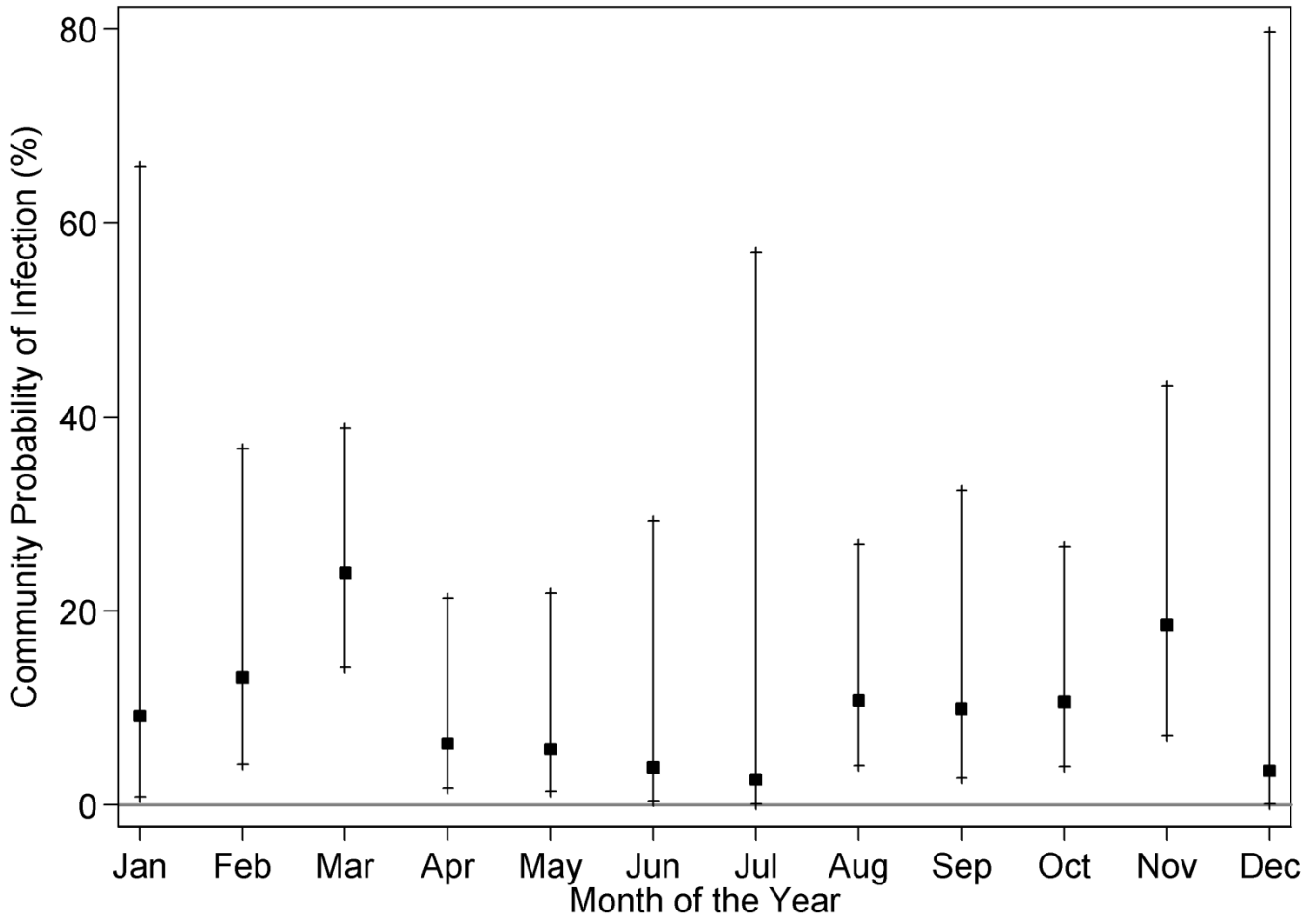
Results

Parameter	Serogroup-Serotype		
	O1 Ogawa	**O1 Inaba**	**O139**
Transmission			
SAR	**6.93%** (5.03%-9.47%)	**7.80%** (5.87%-10.31%)	**12.60%** (8.98%-17.41%)
CPI	**0.15%** (0.04%-0.60%)	**0.44%** (0.18%-1.03%)	**0.42%** (0.20%-0.91%)
Risk Factor (univariate) – odds ratios			
Age: 0-4 vs. >/= 18 years	2.3 (1.0-5.4)	1.4 (0.7-3.0)	1.4 (0.5-3.7)
Age: 5-17 vs. >/= 18 years	0.9 (0.4-2.0)	1.3 (0.7-2.3)	0.7 (0.4-1.5)
Sex: Male vs. Female	1.5 (0.8-2.8)	0.8 (0.5-1.4)	1.1 (0.6-2.0)
ABO blood group: O vs. non-O	0.5 (0.2-1.2)	0.7 (0.4-1.3)	2.2 (1.2-4.3)

Natural History: Observed Serial Interval Distribution



Natural History:



Summary

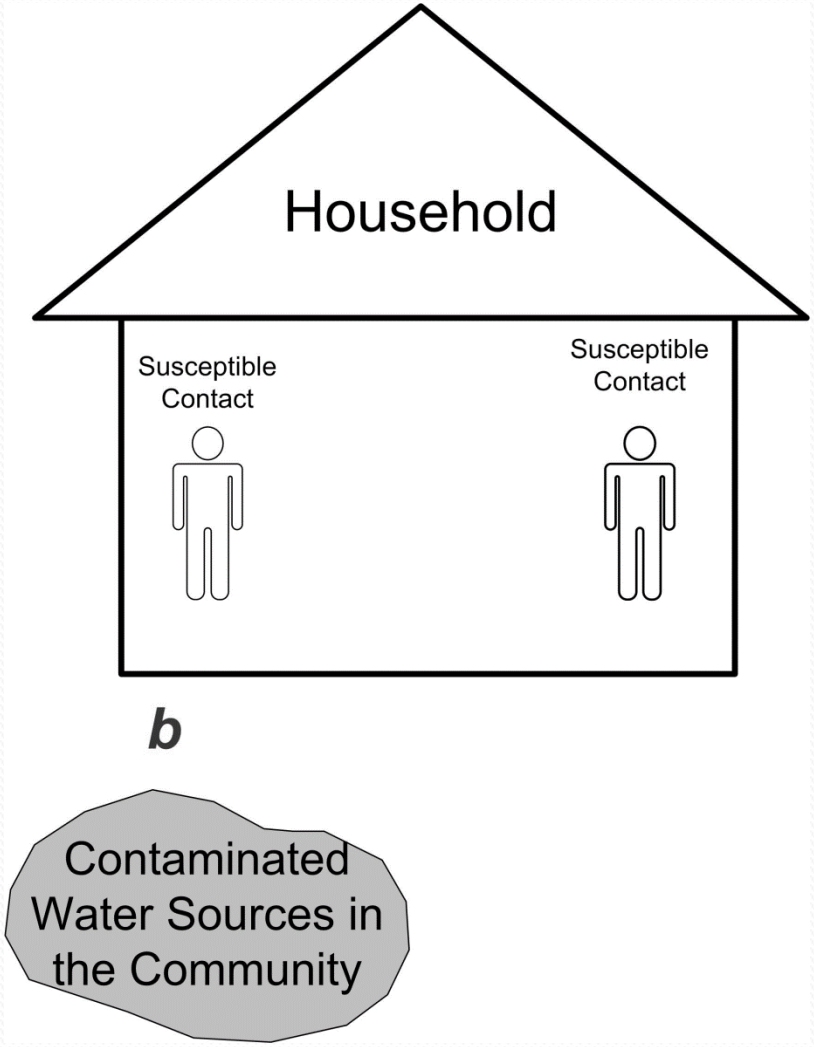
- Significant person-to-person transmission of *Vibrio cholerae* occurred in households ($p<0.0001$ for each serogroup-serotype)
- First direct estimates of the transmissibility of *V. cholerae* through person-to-person exposure in households
- Pre-school aged children are the most susceptible to cholera infection
- O blood group appears to significantly elevate susceptibility to O139 infection
- Our results replicate the previously-reported seasonality of transmission in Bangladesh

Limitations

- Time-constant community-to-person infection probability, *b*, for the study period
  - Community-to-person exposure may take other forms, for example, a point-source in time and space
- Spatial confounding of estimates of *b*
  - Households that cluster in time and space are likely to be similar with respect to *b*
  - Estimating a single *b* for all households likely introduces some confounding to the estimates of both this parameter and *p*

Conclusions - 1

- Our high estimates for the SAR relative to the CPI suggests the following transmission scenario for endemic settings
  - A low-level and persistent risk of the cholera infection being introduced into the household via community-to-person exposure
  - Once introduced into the household, then spread of the infection is comparative explosive between household members through person-to-person exposure



Conclusions - 2

- Control interventions should place more emphasis on interrupting person-to-person transmission of cholera within households
  - For example, promotion of better personal or sanitary hygiene

© Gary Larson, *The Far Side*

Acknowledgements

Collaborators

- Center for Statistics and Quantitative Infectious Diseases, Hutchinson Research Center (Seattle, WA)
  - Amanda L. Allen
  - Eben E. Kenah
  - M. Elizabeth Halloran
  - Ira M. Longini, Jr.
- International Centre for Diarrhoeal Disease Research in Dhaka, Bangladesh
  - Fahima Chowdhury
  - Ashraful I. Khan
  - Firdausi Qadri
- Massachusetts General Hospital and Harvard Medical School (Boston, MA)
  - Regina C. LaRocque
  - Edward T. Ryan
  - Stephen B. Calderwood
  - Jason B. Harris

Funding Sources

- NIH grants:
  - Models of Infectious Disease Agent Study (U01-GM070749).
  - NIGMS: F32-GM085945
  - U01-AI058935
  - U01-AI077883
  - K01-TW07409
  - K01-TW07144
  - R03-A1063079
  - Fogarty/Ellison Fellowship in Global Health (D43-TW005572)
- International Centre for Diarrhoeal Disease Research, Bangladesh
- Charles H. Hood Foundation Child Health Research Award
- Swedish SIDA (HN01AV)
- Physician-Scientist Early Career Award from the Howard Hughes Medical Institute

Bibliographic References

Chao DL, Halloran ME, Longini IM, Jr. Vaccination strategies for epidemic cholera in Haiti with implications for the developing world. *Proc Natl Acad Sci U S A* 2011;**108**(17)**:**7081-5.

Clemens JD, Sack DA, Chakraborty J, Field trial of oral cholera vaccines in Bangladesh: evaluation of anti-bacterial and anti-toxic breast-milk immunity in response to ingestion of the vaccines. *Vaccine* 1990;**8**(5)**:**469-72.

Constantin de Magny G, Murtugudde R, Sapiano MR, Environmental signatures associated with cholera epidemics. *Proc Natl Acad Sci U S A* 2008;**105**(46)**:**17676-81.

Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Boca Raton, Florida: Chapman & Hall/CRC 1993.

Harris JB, Khan AI, LaRocque RC, Blood group, immunity, and risk of infection with *Vibrio cholerae* in an area of endemicity. *Infect Immun* 2005;**73**(11)**:**7422-7.

Huq A, Colwell RR, Rahman R, Detection of *Vibrio cholerae* O1 in the aquatic environment by fluorescent-monoclonal antibody and culture methods. *Appl Environ Microbiol* 1990;**56**(8)**:**2370-3.

Islam M, Drasar B, Albert M, Sack R, Huq A, Colwell R. Toxigenic *Vibrio cholerae* in the environment: a minireview. *Trop. Dis. Bull.* 1997;**94:**R1-R11.

Kendall EA, Chowdhury F, Begum Y, Relatedness of *Vibrio cholerae* O1/O139 isolates from patients and their household contacts, determined by multilocus variable-number tandem-repeat analysis. *J Bacteriol* 2010;**192**(17)**:**4367-76.

Lindenbaum J, Greenough WB, Islam MR. Antibiotic therapy of cholera. *Bull World Health Organ* 1967;**36**(6)**:**871-83.

Mosley WH, Alvero MG, Joseph PR, Studies of cholera El Tor in the Philippines. 4. Transmission of infection among neighbourhood and community contacts of cholera patients. *Bull World Health Organ* 1965;**33**(5)**:**651-60.

Pollitzer R, Swaroop S, Burrows W. Cholera. *Monogr Ser World Health Organ* 1959;**58**(43)**:**1001-19.

Sack RB, Siddique AK, Longini IM, Jr., A 4-year study of the epidemiology of *Vibrio cholerae* in four rural areas of Bangladesh. *J Infect Dis* 2003;**187**(1)**:**96-101.

Sack DA, Sack RB, Nair GB, Siddique AK. Cholera. *Lancet* 2004;**363**(9404)**:**223-33.

Tamayo JF, Mosley WH, Alvero MG, Studies of cholera El Tor in the Philippines. 3. Transmission of infection among household contacts of cholera patients. *Bull World Health Organ* 1965;**33**(5)**:**645-9.

Weil AA, Khan AI, Chowdhury F, Clinical outcomes in household contacts of patients with cholera in Bangladesh. *Clin Infect Dis* 2009;**49**(10)**:**1473-9.

WHO. Cholera vaccines: WHO position paper. *Wkly Epidemiol Rec* 2010;**85**(13)**:**117-28.

UNICEF. State of the World's Children, 2008. United Nations Children's Fund. New York, New York.

Yang Y, Ira M. Longini J, Halloran ME. Design and evaluation of prophylactic interventions using infectious disease incidence data from close contact groups. *Applied Statistics* 2006;**55:**317-30.

Yang Y, Longini IM, Halloran ME. A resampling-based test to detect person-to-person transmission of infectious disease. *Ann Appl Stat* 2007;**1**(1)**:**211-228.

Pause to Demonstrate Case 4

Case Study 5: Western Washington State Youth Camp and Associated Households

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

Study Objectives

- Transmission of symptomatic pH1N1 in a “school-like” camp and associated households
- Estimate a …
  - Daytime Camp Local R
  - Nighttime Cabin SAR
  - Households SAR
  - Odds ratio: Effect of age on susceptibility to symptomatic pH1N1

Study Setting and Context

Person:

- Camp population: 96 participants (66% of attendees)
  - 72 6th-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

Methods
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

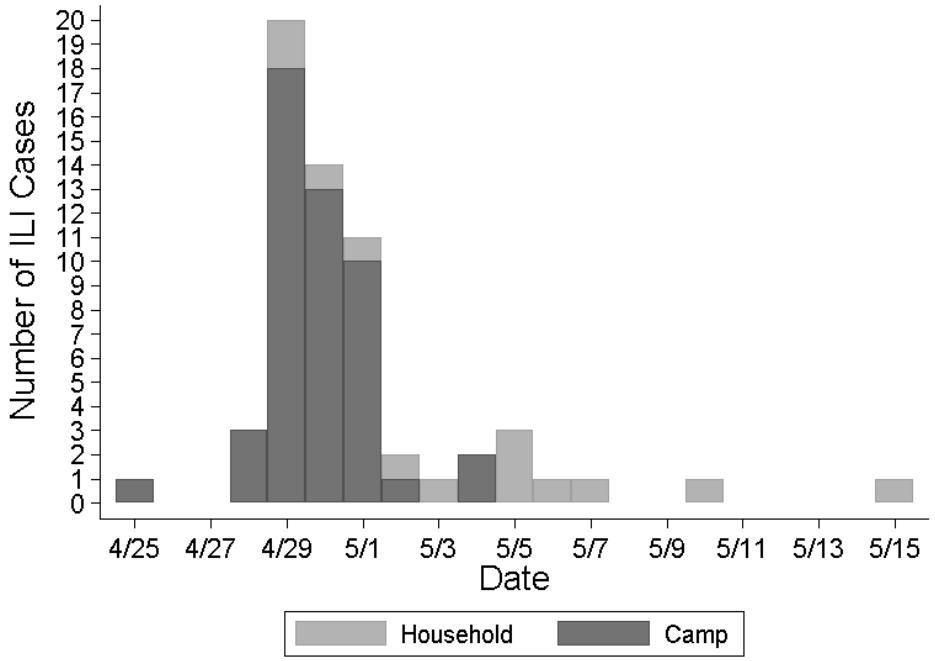
Methods Definitions

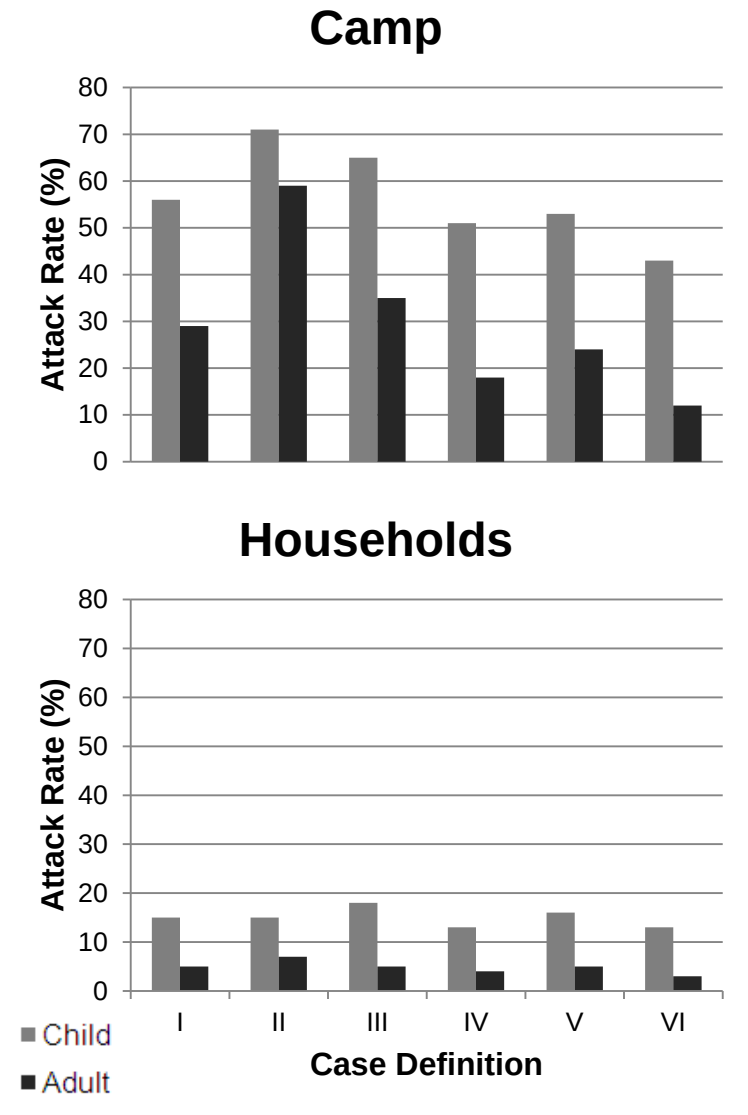
- Outcome: Symptomatic pH1N1
  - 6 case definitions
  - Primary ~ CDC's influenza-like illness (ILI)
- Predictor: Age
  - Children = ≤17 years
  - Adults = ≥18 years

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 ≥ 100.4°F (38°C)
V	- Reported Fever and - Cough or Sore throat
VI	- Reported Fever with measure temperature ≥ 100.4°F (38°C) and - Cough or Sore throat

Descriptive Statistics for the Primary Case Definition (I)

Characteristic	Camp Participants (N = 96)	Household Contacts (N = 136)
No. male (%)	38 (40%)	63 (28%)
Age (years)		
Children (≤17 years): No. (% of all individuals)	79 (82%)	48 (35%)
Adult (≥18 years): No. (% of all individuals)	17 (18%)	88 (65%)
Mean (SD: Range)	16 (12: 10, 59)	34 (18: 0.5, 74)
Number of cabins or households	13	41
Individuals per cabin or household: Mean (SD: Range)		
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)



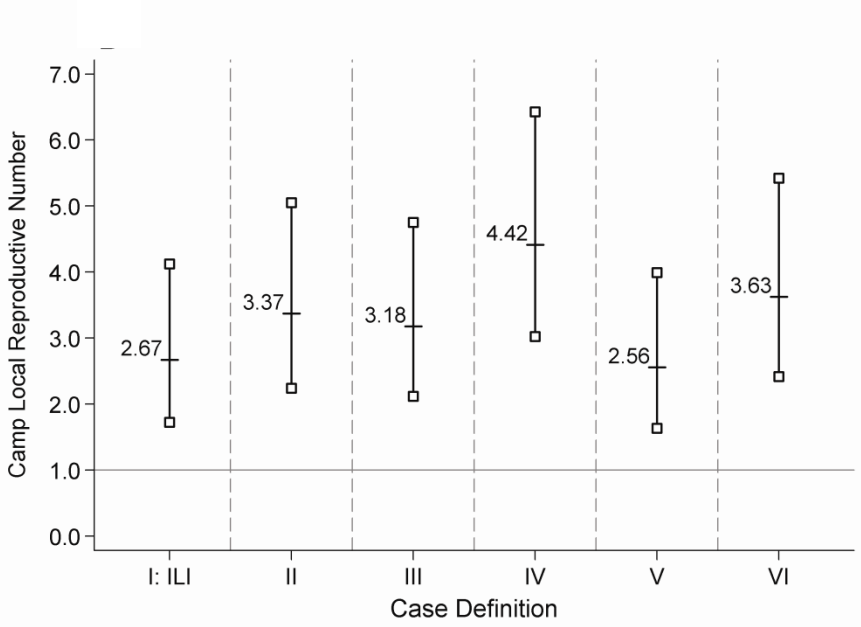


- ILI attack rate
  - Camp: 51% (N = 49)
  - Household contacts: 8% (N=11)
- Camp: 5 cases were laboratory-confirmed

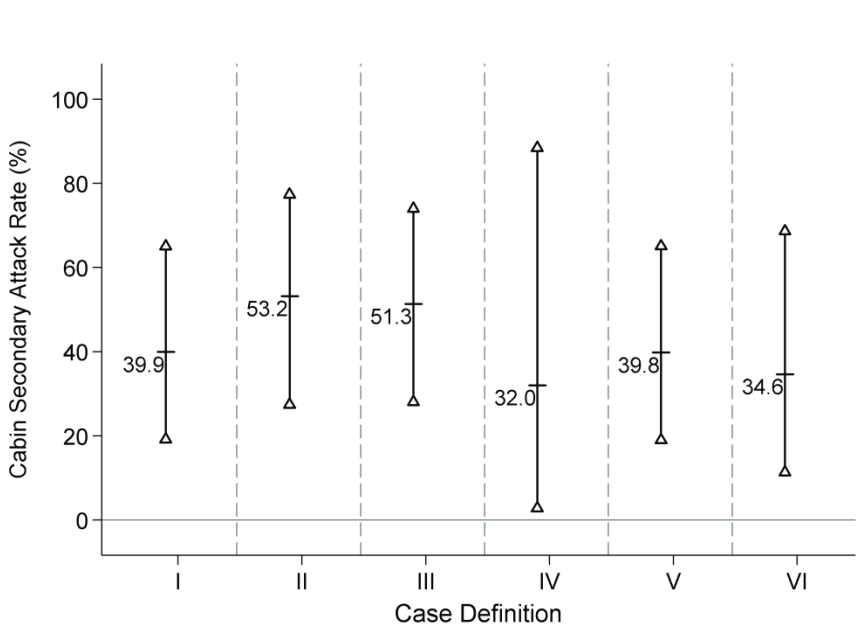
Pause to Demonstrate Case 5

Results: Camp Transmission

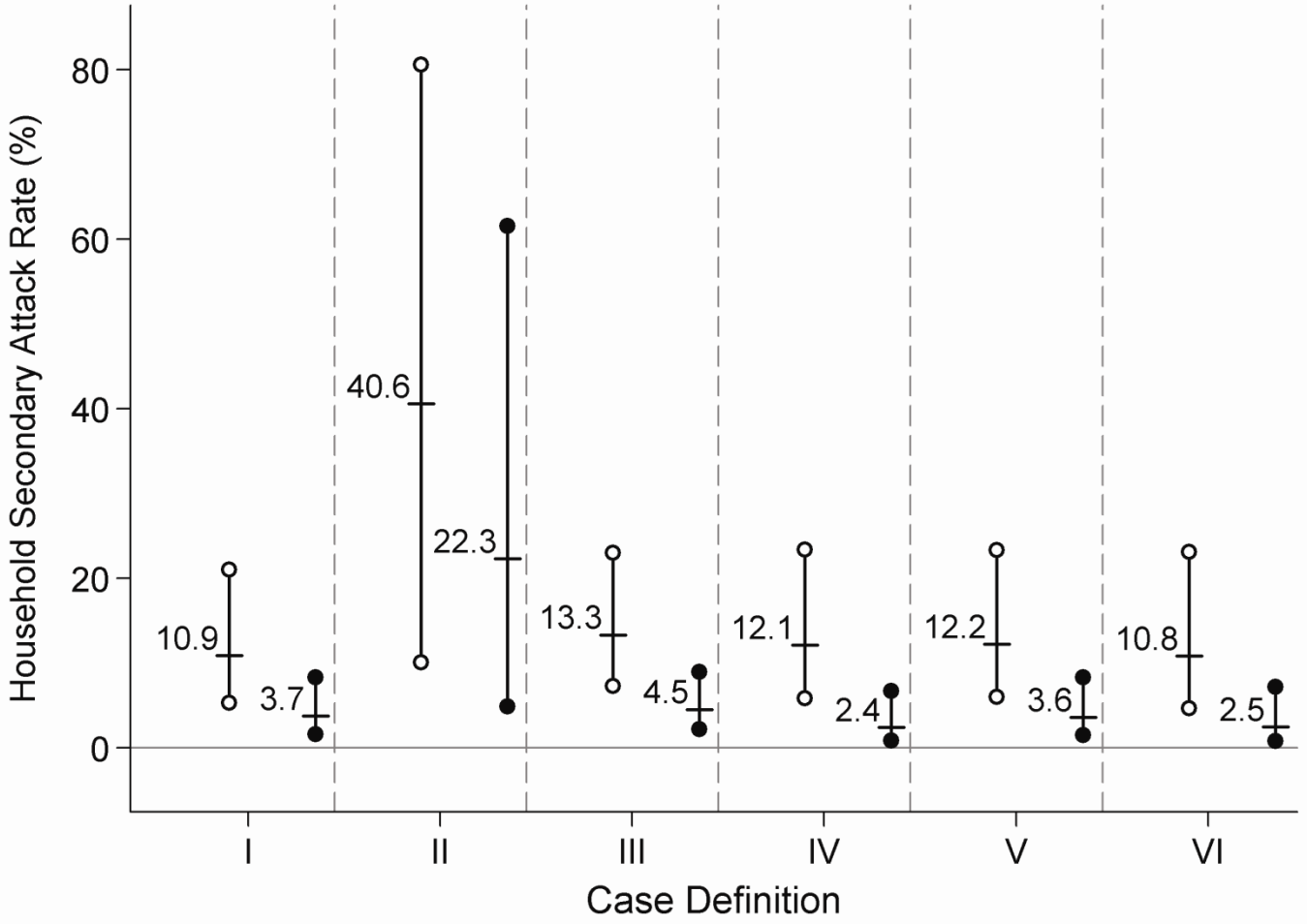
Camp Local R: Daytime



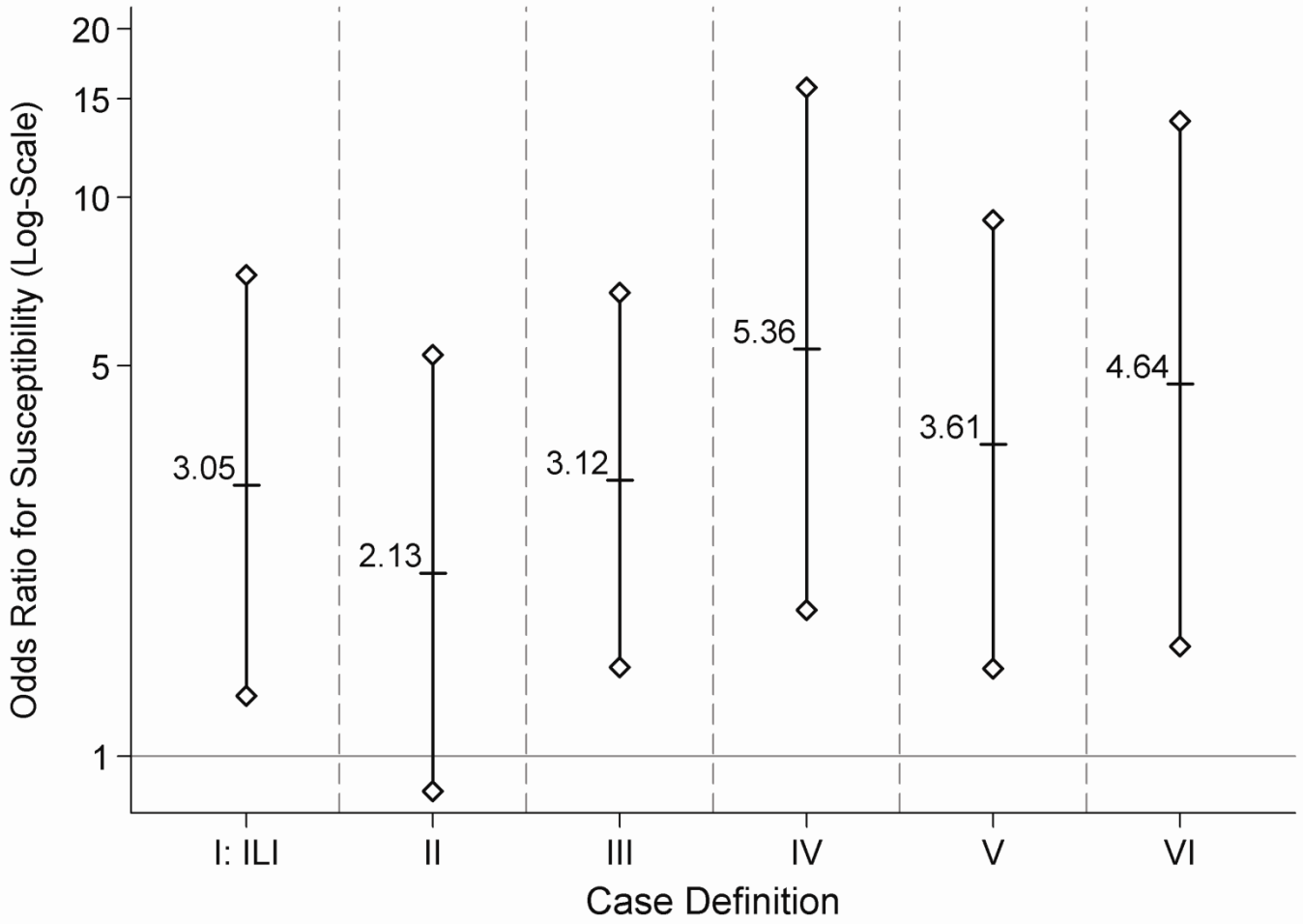
Cabin SAR: Nighttime



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 R were not sensitivity to assumptions about the incubation/latent and infectious period distributions

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
- TranStat will continue to be updated, new features will be added, and these will freely-available through www.cidid.org/software-development/ .
- A graphical user interface is currently being developed, with a target completion date of July 30, 2015. The GUI version of TranStat and associated documentation will be made available on the CIDID website, soon thereafter.