

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

- Note: All of these analyses have been adapted from the full analysis presented in 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

# Formatting Files For Toy Analysis #1

- TranStat requires several input files for this analysis.
- pop\_1.dat
- time\_ind\_covariate.dat
- community\_1.dat
- c2p\_contact.dat
- p2p\_contact.dat
- config.file
- We will consider each in detail for this 1<sup>st</sup> Toy Example.
- None of the data files contain column headers.

# Ex1: pop\_1.dat

- This file records one row for each individual member of the study population.
- There are 13 columns that are tab or space delimited.
- Columns (Allowable values are 0=No,1=Yes, unless indicated otherwise)
  1. Unique person identifier; Values: 0..N
  2. Community identifier; Values: 0..C
  3. Pre-existing immune status; Values:
  4. Infection Indicator
  5. Symptomatic Disease Indicator
  6. Symptom Onset Time: Values: Integer time step or, if not symptomatic, -1
  7. Disease State Exit Indicator
  8. Disease State Exit Time: Values: Integer time step or, if not symptomatic, -1
  9. Cluster Index Case Indicator
  10. Disease Severity category/level; Values: 0..Number of levels
  11. Pre-existing immune state category/level; Values: 0..Number of levels
  12. Individual's population weight; Values: 1...Positive Infinity [allows for one row to represent multiple individual members of the same cluster/household]
  13. Ignore Indicator [tells TranStat to ignore the record]

# Household 0 from eTable 24 of the online supplement to Sugimoto et al. 2011

HHID	ID	Age <sup>b</sup> (yr)	Sex <sup>b</sup>  1=Male 2=Female	Case Definition 1	
				Onset Day <sup>c</sup>	Index Case <sup>d</sup> 1=Yes 0=No
0	0	12	1	28	1
0	1	45	1	31	1
0	2	39	2		0
0	3	10	2	12	0
0	4	16	2	40	0

[FULL FILE](#)

Household  
Index =  
Symptomatic  
Camp  
Attendee

HHID	ID	Age	Sex	Onset Day	Index Case	Other 1	Other 2	Other 3	Other 4	Other 5	Other 6	Other 7	Other 8
0	0	12	1	28	1	0	-1	1	0	0	1	0	
0	1	45	1	31	1	0	-1	0	0	0	1	0	
0	2	39	2		0	0	-1	0	0	0	1	0	
0	3	10	2	12	0	0	-1	0	0	0	1	0	
0	4	16	2	40	0	0	-1	0	0	0	1	0	

# Ex1: time\_ind\_covariate\_1.dat

- This file records one row for each individual member of the study population.
- The first column lists each individual's unique person identifier, followed by one additional column per time-independent covariate
- Columns are tab or space delimited.
  
- Columns for this analysis
  1. Unique person identifier; Values: 0..N
  2. Age category;  
Values:           0=under-18 years,  
                  1=18 years and older

# Household 0 from eTable 24 of the online supplement to Sugimoto et al. 2011

HHID	ID	Age <sup>b</sup> (yr)	Sex <sup>b</sup>  1=Male 2=Female	Case Definition 1	
				Onset Day <sup>c</sup>	Index Case <sup>d</sup> 1=Yes 0=No
0	0	12	1	28	1
0	1	45	1	31	1
0	2	39	2		0
0	3	10	2	12	0
0	4	16	2	40	0

[FULL FILE](#)

time\_ind\_covariate\_1.da

File	Edit	Format	View
0			0
1			1
2			1
3			0
4			0

# Ex1: Community.dat

- This file records 1 row for each cluster/household.
- There are 3 columns that are space or tab delimited
- Columns
  1. Community identifier; Values: 1..C [Unique value for each row]
  2. Start of observation time for cluster/household; Values: 1..T
  3. End of observation time for cluster/household; Values: 1..T

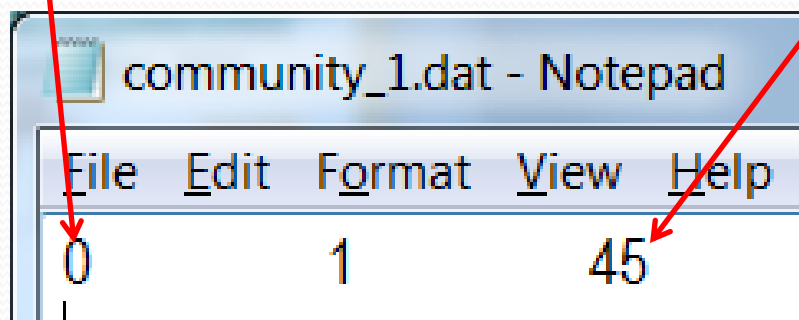


# Household 0 from eTable 24 of the online supplement to Sugimoto et al. 2011

HHID	ID	Age <sup>b</sup> (yr)	Sex <sup>b</sup>  1=Male 2=Female	Case Definition 1	
				Onset Day <sup>c</sup>	Index Case <sup>d</sup> 1=Yes 0=No
0	0	12	1	28	1
0	1	45	1	31	1
0	2	39	2		0
0	3	10	2	12	0
0	4	16	2	40	0

[FULL FILE](#)

We chose an end day that occurred some time after the last onset day for the household/cluster.



# Ex1: c2p\_contact\_1.dat

- This file allows the user to provide detailed information about the study population's exposures to 'community' sources of infection (C2P contact).
- For the user's convenience three formats are permitted for this file. The user can specify the same C2P contact history for all members of the study population (Version A), or unique histories by cluster (Version B) or individual (Version C).
- Version A involves 4 columns, and Versions B and C involve 6. All columns are tab or space delimited. *All of the toy examples use the most flexible Version C.*
- Columns:
  1. Version A: do not include this column; Version B: Community identifier; and Version C: Unique person identifier
  2. Start time for C2P contact; Values: 1..T [for case-ascertained designs choose a C2P start time at least one full incubation period prior to earliest onset day of an cluster index case]
  3. End time for C2P contact; Values: 1..T
  4. Type of C2P contact; Values: 0..Number of types
  5. Contact weight: offset for each C2P contact's contribution to the likelihood on the natural logarithm scale
  6. Ignore indicator for C2P contact; Values: 0=No, 1=Yes [Version A, do not include this column]

# Household 0 from eTable 24 of the online supplement to Sugimoto et al. 2011

[FULL FILE](#)

c2p\_contact\_1.dat - Notepad

File	Edit	Format	View	Help				
0		1	46	0	0	0	0	0
1		1	46	0	0	0	0	0
2		1	46	0	0	0	0	0
3		1	46	0	0	0	0	0
4		1	46	0	0	0	0	0

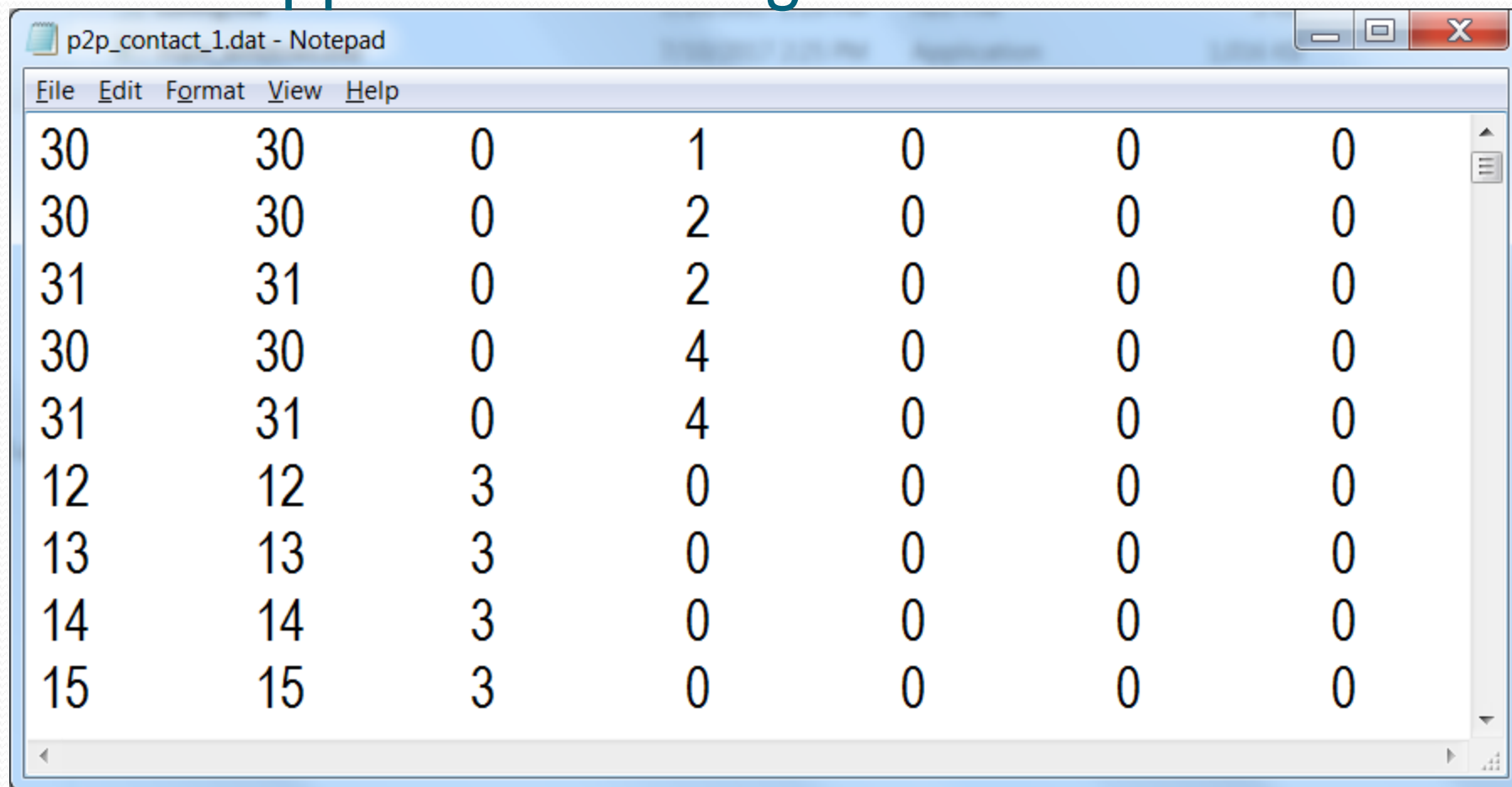
  

Person Identifier	C2P: Start day	C2P: End day	C2P: Type/Mode	C2P: Weight/Offset	C2P: Ignore Indicator
0	1	46	0	0	0
1	1	46	0	0	0
2	1	46	0	0	0
3	1	46	0	0	0
4	1	46	0	0	0

# Ex1: p2p\_contact\_1.dat

- This file allows the user to provide detailed information about the study population's exposures to other members (P2P contact).
- For the user's convenience two formats are permitted for this file: the same P2P contact history for all members of a cluster/household (Version A) OR unique histories by pairs of individuals (Version B).
- *All toy examples use the more flexible Version B.*
- Tab or space-delimited columns for Version B:
  1. Start time for P2P contact; Values: 1..T
  2. End time for P2P contact; Values: 1..T
  3. Individual A involved in a P2P contact: 1..N.
  4. Other individual involved in the P2P contact: 1..N
  5. Type of P2P contact; Values: 0..Number of types
  6. Contact weight: offset for each P2P contact's contribution to the likelihood on the natural logarithm scale
  7. Ignore indicator for P2P contact; Values: 0=No, 1=Yes

# Individual 0 from eTable 24 of the online supplement to Sugimoto et al. 2011 [FULL FILE](#)



P2P: Start day	P2P: End day	Person Identifier A	Person Identifier B	P2P: Type/Mode	P2P: Weight/Offset	P2P: Ignore Indicator
30	30	0	1	0	0	0
30	30	0	2	0	0	0
31	31	0	2	0	0	0
30	30	0	4	0	0	0
31	31	0	4	0	0	0
12	12	3	0	0	0	0
13	13	3	0	0	0	0
14	14	3	0	0	0	0
15	15	3	0	0	0	0

P2P: Start day    P2P: End day    Person Identifier A    Person Identifier B    P2P: Type/Mode    P2P: Weight/Offset    P2P: Ignore Indicator

## Ex1: config.file

- This file defines the model to be estimated and controls estimation and simulation related procedures.
- This file is long and contains a large number of possible settings and switches that the user can use to control the analysis and simulations.
- We will discuss the most common ones and those that are pertinent to each of the toy examples.
- [Config.file](#)

# Ex1: Analysis

- Run analysis
- Look at output

# 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



# Formatting Files for Toy Analysis #2

- We use a set of input files with the same name and format as for Toy Analysis #1, except that their contents are based upon the population of camp participants (eTable 23 of the online supplement to Sugimoto et al. 2011).
- Key differences from Toy Analysis #1
  - There is only 1 community, 'the camp'.
  - The nighttime cabin versus daytime camp P2P contact structure is built into how the p2p\_contact.dat is specified. Specifically, the latter type of P2P contact is assigned P2P contact type = 0 and the former is assigned type=1.
  - The self-contained nature of the camp environment and the fact that this outbreak occurred very early in the pandemic suggests that there was very little C2P exposure to pH1N1 during the camp. Therefore, we fix the value for  $b$  to a negligibly-low value.
- Analysis and Results

# 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

# Formatting Files for Toy Analysis #3

- We effectively merge the corresponding files from Toy Analyses 1 and 2, and then renumber unique person identifiers and community identifiers to ensure consistency with TranStat's input criteria.
- Important Notes
  - Camp participants who are also members of households appear twice, once in each type of cluster. Yet, they are only counted as susceptible to non-primary, symptomatic infections in the camp. In the household they expose their fellow housemates as index cases, who are modeled as not susceptible to pH1N1 infection.
  - As for Toy Analysis #2, the value for  $b$  is fixed to a negligibly-low value.
- Analysis and Results

# Toy Analysis #4

## Multiple Cluster Types and Missing Information

- As with Toy Analysis #2, we jointly model transmission and age effects in the camp and households.
- Again, we estimate the same parameters as for Toy Analysis #2.
- 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 9 ILI cases reported in the original dataset.

# Formatting Files for Toy Analysis #4

- We use the same input files as for Toy Analysis #2.
- We add one additional file, the impute.dat.
- This file contains 9 tab-delimited columns, with one row per member of the pop\_1.dat for whom there is ambiguity with regard to the his/her final infection status, i.e., was the individual pre-immune, a symptomatic infection, an asymptomatic infection, or did s/he escape infection.
- Columns (Values are 0=No and 1=Yes, unless otherwise indicated)
  1. Unique person identifier
  2. Possible pre-immune
  3. Possible escaped infection
  4. Possible symptomatic infection:
  5. Start time for possible onset day for symptomatic infection
  6. End time for possible onset day for symptomatic infection
  7. Possible asymptomatic infection:
  8. Start time for possible onset day for asymptomatic infection
  9. End time for possible onset day for asymptomatic infection

1	0	1	1	28	28	0	-1	-1
2	0	1	1	34	34	0	-1	-1

- [Analysis and Results](#)

# 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/).