BIOST/STAT 578 A
Statistical Methods in Infectious Diseases

Lecture 6

January 22, 2009

Estimating R_0 for emerging infectious diseases in real time

The Problem

- Observed illness onset times
  - Exposure information
- Is the disease infectious?
- If so, what are estimates of transmission parameters?
  - Secondary attack rates
  - Reproductive number
- How effective are interventions?
- Calibration of intervention models.

Information Needed

- Data
  - Illness onset times
  - Crude population connectivity
    - $\mathcal{R}_i$ – Risk set for person *i*
- Illness serial interval distribution
  - $\omega(t|\theta)$
  - Mean ≈ incubation period + ½ infectious period
  - Incubation and infectious period information

Estimation Methods

- fitting an exponential growth curve to the data
  - e.g., Lipsitch *et al., Science* 2003)

- fitting a random graph to the data
  - e.g., Wallinga and Teunis, *Am J Epidemiol* 2004

- fitting a more detailed transmission model to the data
  - e.g., Yang, Longini, Halloran, *Ann Apply Stat* 2007
  - e.g., Ph.D. thesis of Gail Potter, Stat Dept

Fitting Exponential Growth: Basic Idea

$$\frac{dI(t)}{dt} = cpS(t)\frac{I(t)}{n} - \gamma I(t)$$

$$\frac{dI(t)}{dt} = (cp\frac{S(t)}{n} - \gamma)I(t)$$

$$R_0 = \frac{cp}{\gamma},\ S(0) \doteq n$$

$$\frac{dI(t)}{dt} = (R_0\frac{S(t)}{n} - 1)\gamma I(t)$$

$$\frac{dI(t)}{dt}\ |_{t=0^+} = (R_0 - 1)\gamma I(t)$$

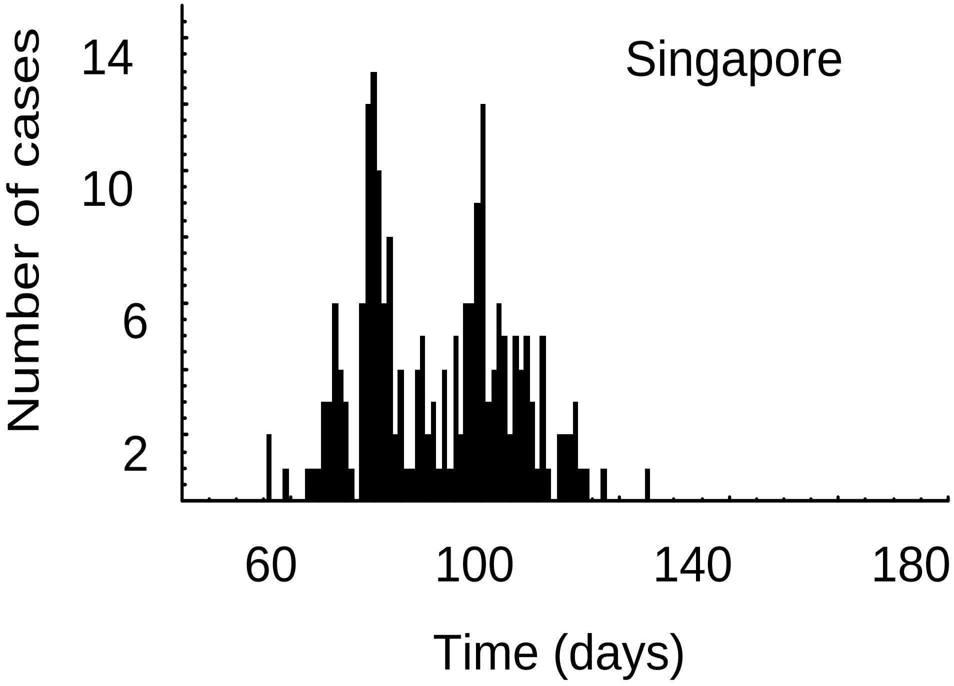
$I(t) = I(0)e^{(R_0-1)\gamma t}$, in the early exponential phase

$$\ln(\frac{I(t)}{I(0)}) = (R_0 - 1)\gamma t$$

Thus, we can fit early data and estimate R_0, if we know γ.

Available information: epidemic curve

- List with values t_i , date of onset of symptoms of the i^{th} case



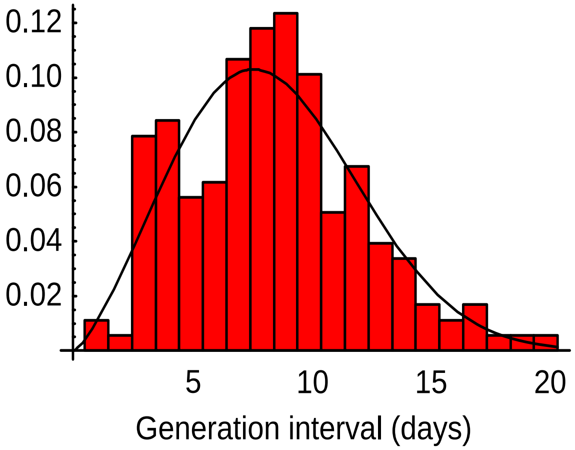
Random Graph

J Wallinga and P Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, **160**: 509 – 516, 2004.

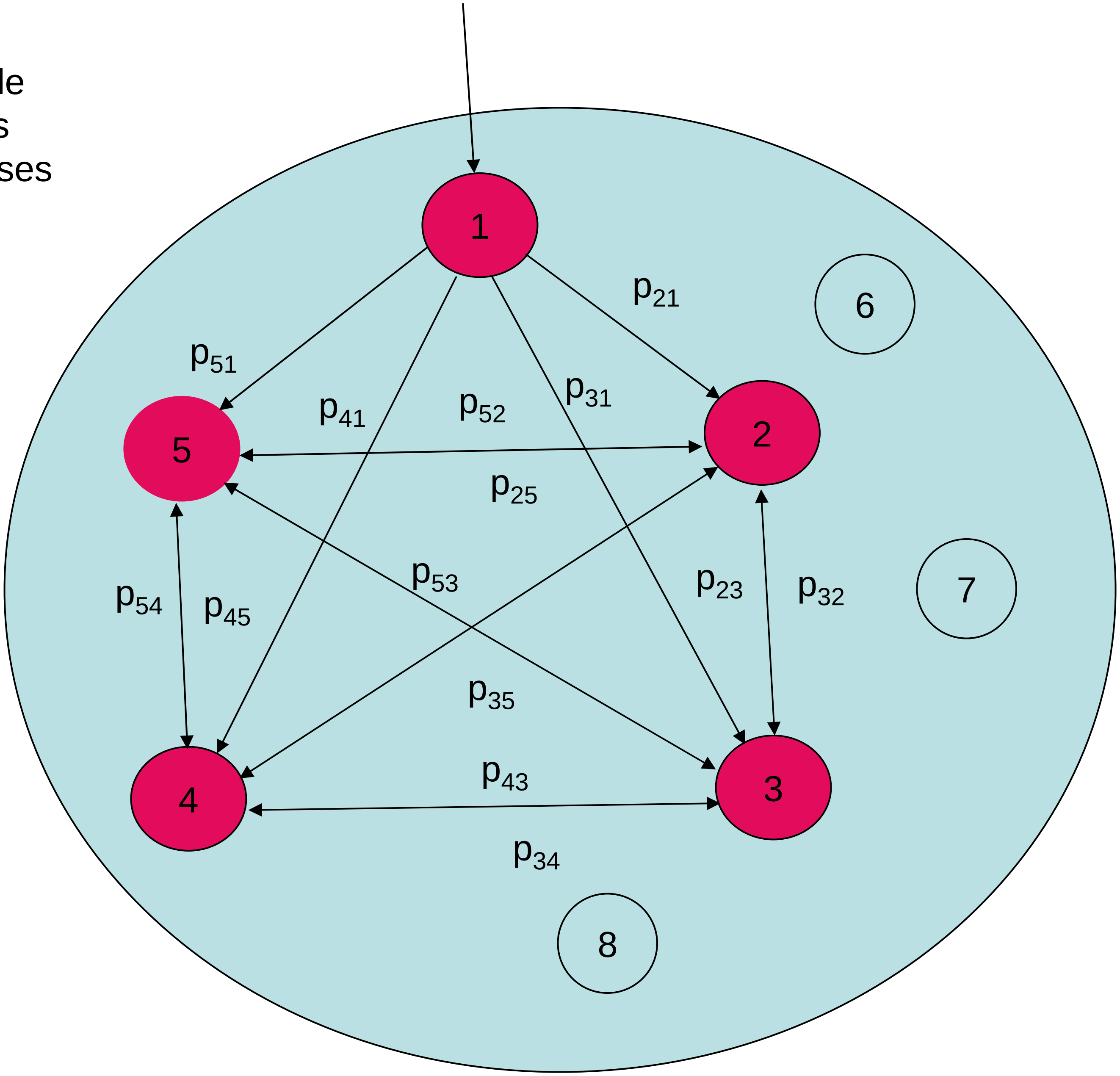
S Cauchemez, , PY Boelle, G Thomas, and AJ Valleron. Estimating in real-time the efficacy of control measures in emerging communicable diseases. *American Journal of Epidemiology*, **164**: 591 – 597, 2006.

Available information: generation interval

- the generation interval, $t_i - t_j$, is the duration between onset of symptoms of a secondary case t_i and its primary case t_j
- distribution of $t_i - t_j$ can be described by a Weibull distribution $w(t_i - t_j \mid \alpha,\beta)$



8 total people
n = 5 cases
q =1 index cases
red = case



Methods: from epidemic curve to reproduction number **(given the parameters α and β)**

- Likelihood of case *i* being infected by case *j*

$$L_{ij} = w(t_i - t_j \mid \alpha, \beta)$$

- Probability of case *i* being infected by case *j*

$$p_{ij} = \frac{L_{ij}}{\sum_{k=1, k \neq i}^{k=n} L_{ik}}$$

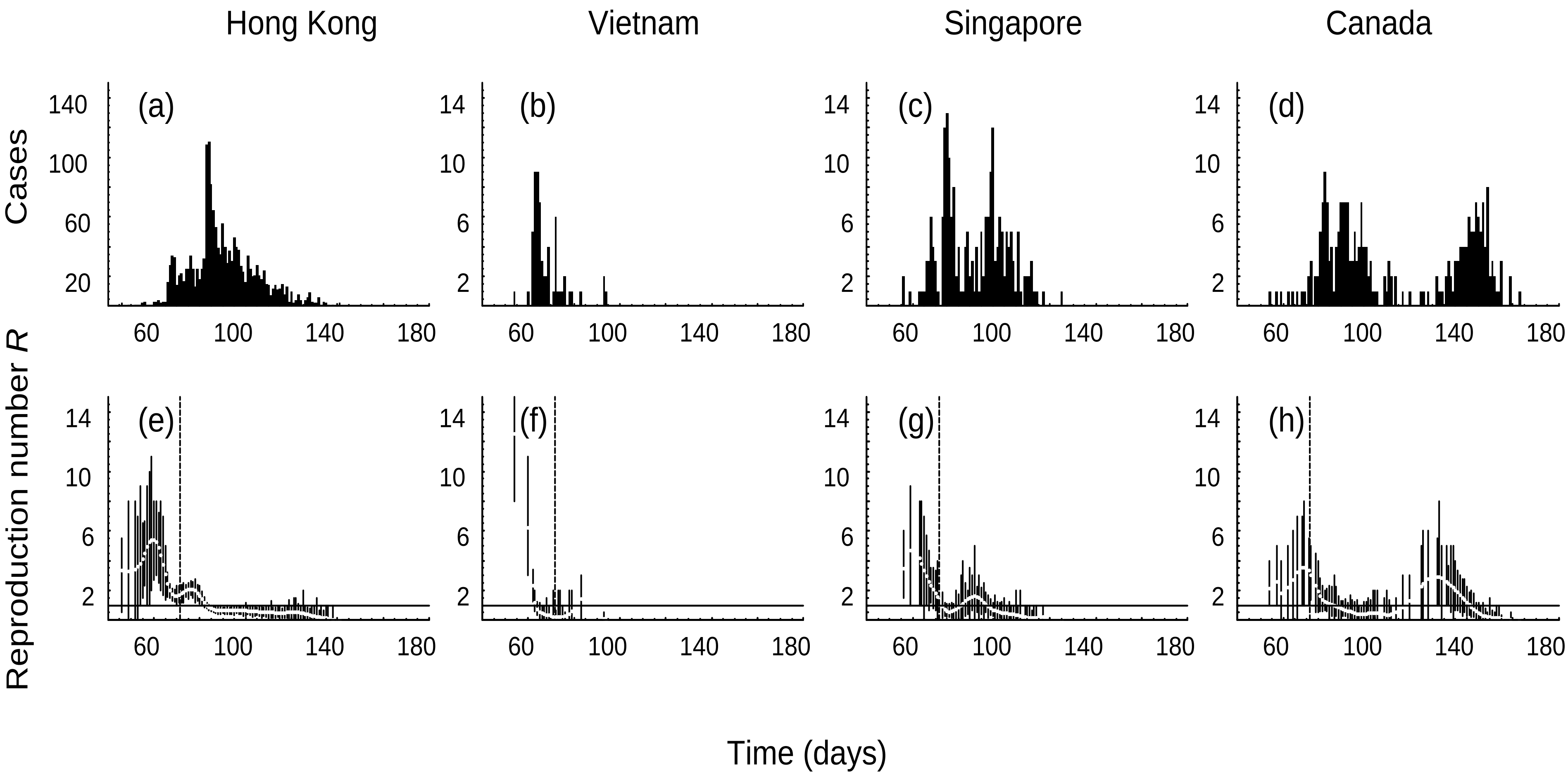
- Number of secondary cases attributable to case *j*

$$E(R_j) = \sum_{i=1}^{n-q} p_{ij}$$

- Reproductive number

$$E(R) = \frac{\sum_{j=1}^{n} E(R_j)}{n}$$

Results for SARS



Results for SARS

Average daily effective reproduction numbers *R* with 95 percent CI for cases with symptom onset date before and after the first global alert against SARS on 12 March 2003 for regions where infection was introduced late February 2003.

	Hong Kong		Vietnam		Singapore		Canada	
before alert	3.7	3.1, 4.2	2.5	1.8, 3.3	3.1	2.3, 4.0	2.7	1.8, 3.8
after alert	0.7	0.7, 0.8	0.3	0.1, 0.7	0.7	0.6, 0.9	1.1	0.9, 1.2

Conclusions for SARS

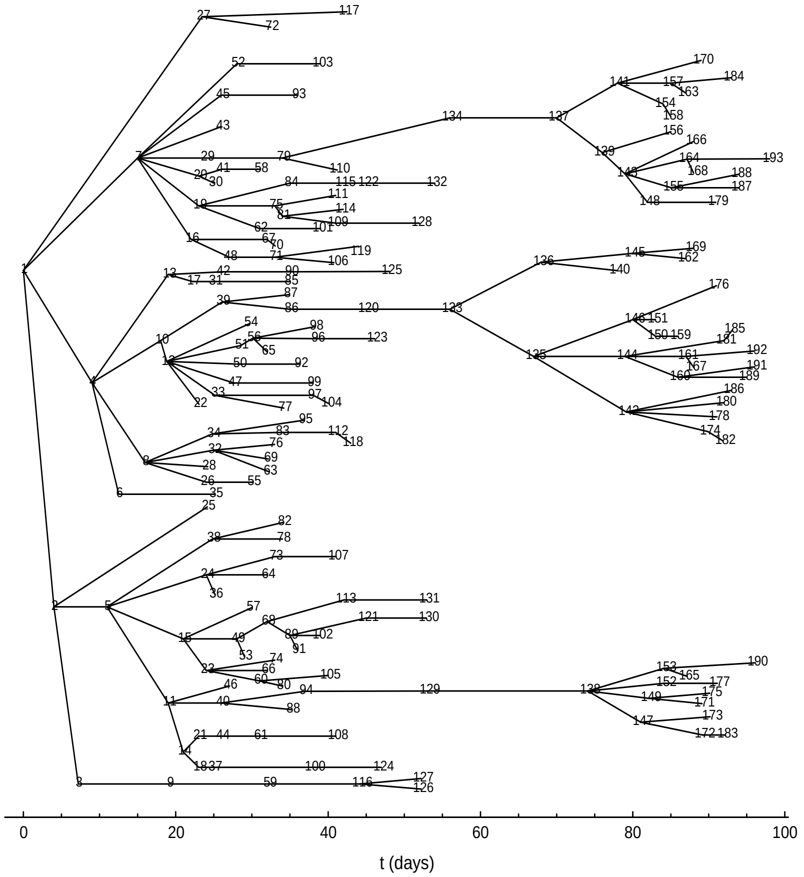
- SARS is a highly contagious disease
    - *R* in uncontrolled situation is around 3
    - *R* in controlled situation is around 0.7
- *R* decreased markedly after WHO alert on 12 March 2003
    - control measures have prevented 75% of potential secondary infections
    - which has been enough, but only barely, to stop the epidemics

Likelihood-based estimation procedure for the reproduction number *R*

- estimate of the reproduction number *R* is based on integrated likelihood of case *i* being infected by case *j*
- likelihood is integrated over all possible infection trees (genealogies) for the observed cases
- requires the assumption that all infectious contacts are independent

Likelihood-based estimation procedure for the reproduction number *R*

- estimate which infection tree is most likely
- Canadian SARS epidemic:



Model-based Estimation

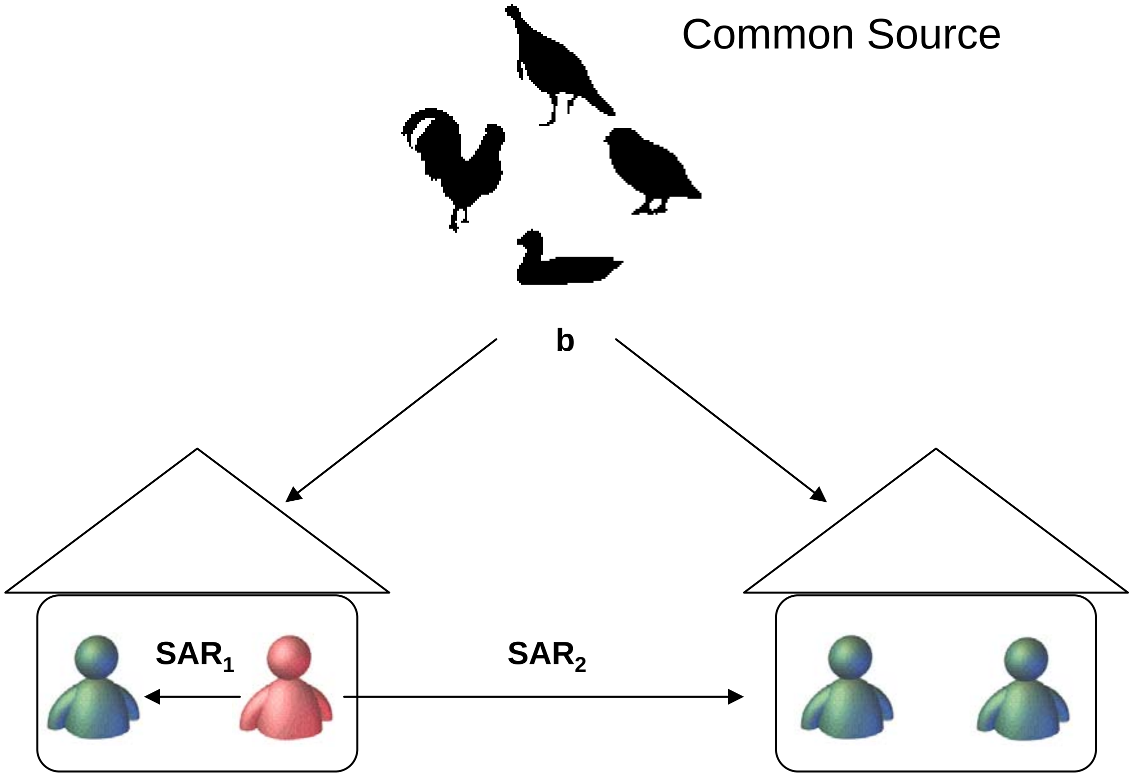
Yang, Y., Longini, I.M. and Halloran, M.E.: A resampling-based test to detect person-to-person transmission of infectious disease, *Annals of Applied Statistics* **1**, 211–28 (2007).

Yang, Y., Halloran, M.E., Sugimoto, J. and Longini, I.M.: Detecting human-to-human transmission of avian A (H5N1) influenza, *Emerging Infectious Diseases* **9**, 1348-1353 (2007).

Statistical Model

- Consider a community composed of households, and three channels through which infection may occur.
  - **Close contact within households**: the probability that a susceptible is infected by an infective in the same household in one day is p_1.
  - **Casual contact within community**: the probability that a susceptible is infected by an infective in the same community but different household in one day is p_2.
  - **Common source of infection** (e.g., zoonotic source or visiting infectives from outside of the community): the probability that a susceptible is infected by the common source in one day is b.

Statistical Model



Common Source

b

SAR_1

SAR_2

- Hypothesis to be tested.

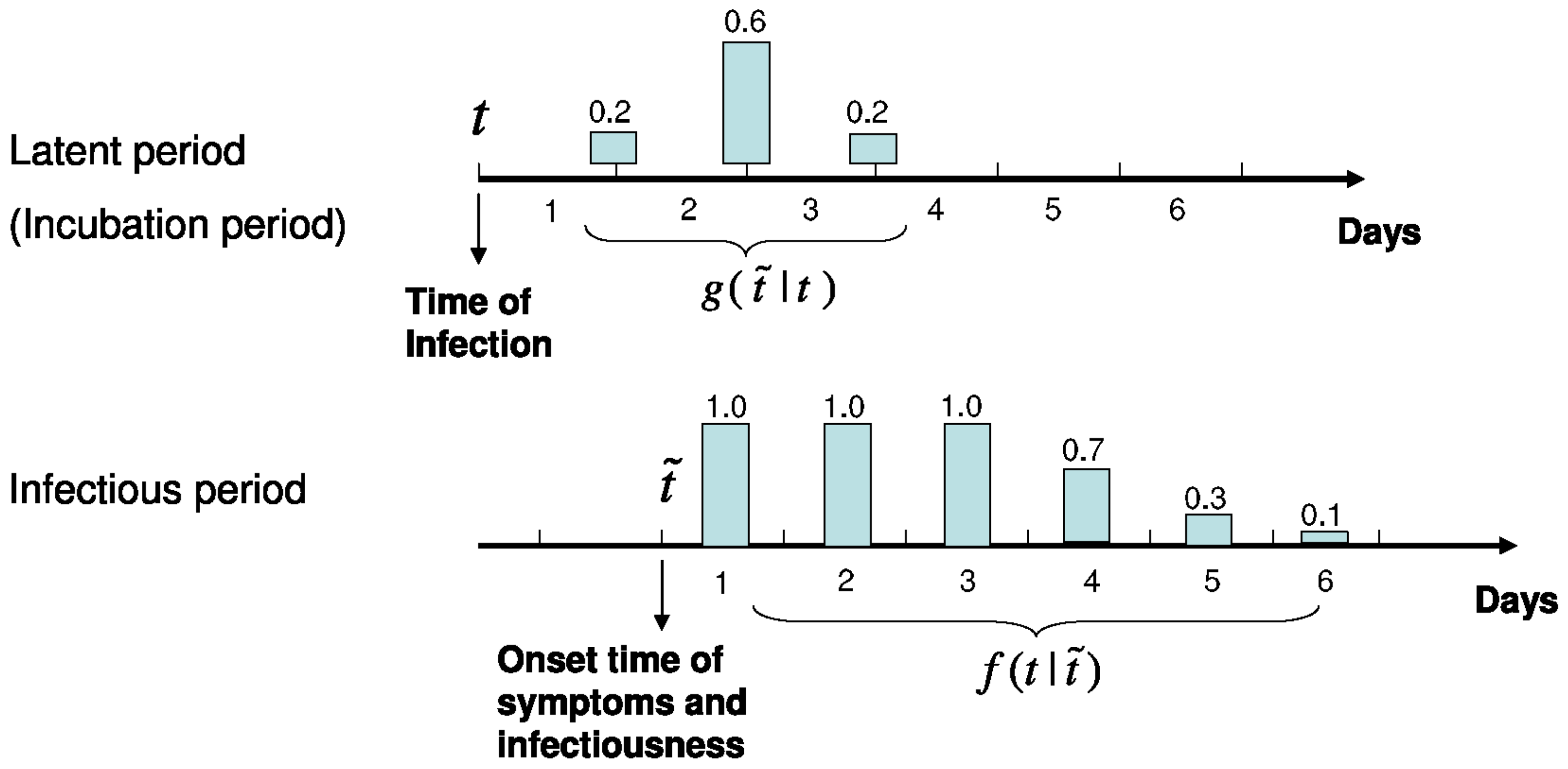
$$H_0 : p_1 = p_2 = 0 \;\; vs.$$
$$H_1 : p_1 > 0 \text{ or } p_2 > 0,$$

- Test statistic

$$\lambda = -2 \log \frac{\sup_b L_0(b|\tilde{t}_i, i = 1, \ldots, N)}{\sup_{b,p_1,p_2} L(b, p_1, p_2|\tilde{t}_i, i = 1, \ldots, N)},$$

where $\tilde{t}_i$ is the symptom onset time for person i, and N is the population size.

- More assumptions:
  - Random mixing in households and in the community.
  - The latent period coincides with the incubation period.
  - Distributions of the latent period (δ) and the infectious period (η) are known:
    - $\delta \sim g(l) = \Pr(\delta = l)$, $l = \delta_{min}, \delta_{min} + 1, \ldots, \delta_{max}$.
    - $\eta \sim f(l) = \Pr(\eta \geq l)$, $l = \eta_{min}, \eta_{min} + 1, \ldots, \eta_{max}$.
  - Observation starts from day 1 to day T, and exposure to the common source starts from day 1 to day $S \leq T$.
    - When $S < T - \delta_{min}$, asymptotic method does not work.



$g(\tilde{t}\mid t)$: The probability of symptom onset on day $\tilde{t}$ given infection on day t.

$f(t\mid\tilde{t})$: Probability that the host is infective on day t given symptom onset on day $\tilde{t}$.

- The probability that an infective j infects an susceptible i on day t is

$$p_{ji}(t) = p_1^{I(j \in H_i)} p_2^{I(j \notin H_i)} f(t - \tilde{t}_j), \tag{1}$$

  where H_i is the set of household members of person i.

- The probability that subject i escapes infection from all infective sources on day t is

$$e_i(t) = (1-b)^{I(t \le S)} \prod_{j=1}^{N} (1 - p_{ji}(t)). \tag{2}$$

- A likelihood for b, p_1 and p_2 contributed by person i is

$$L_i(b, p_1, p_2 | \tilde{t}_j, j = 1, \ldots, N)$$

$$= \begin{cases} \prod_{t=1}^{T} e_i(t), & \text{not infected,} \\ \sum_t g(\tilde{t}_i - t)\big(1 - e_i(t)\big) \prod_{\tau=1}^{t} e_i(\tau), & \text{otherwise,} \end{cases} \qquad (3)$$

- When $p_1 = p_2 = 0$, (2) reduces to

$$e_i(t) = (1 - b)^{I(t \leq S)}.$$

Important parameters

- Household $\mathrm{SAR}_1 = \sum_l f(l)\big(1-(1-p_1)^l\big)$
- Community $\mathrm{SAR}_2 = \sum_l f(l)\big(1-(1-p_2)^l\big)$
- Local $R_0 = (M-1)\times\mathrm{SAR}_1 + (N-M)\times\mathrm{SAR}_2$

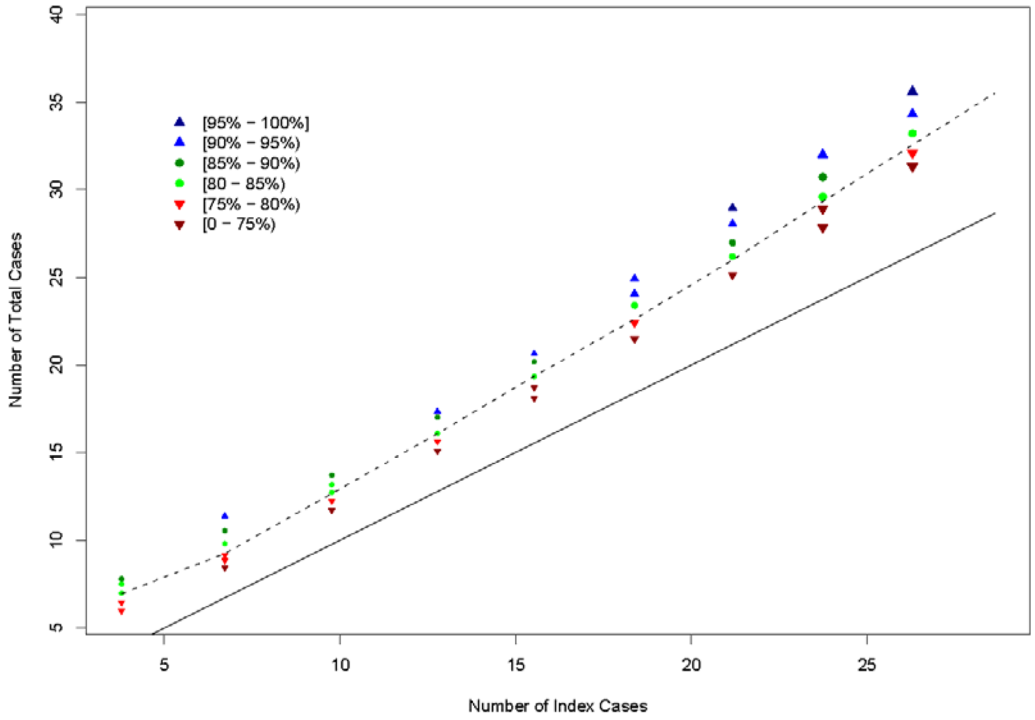
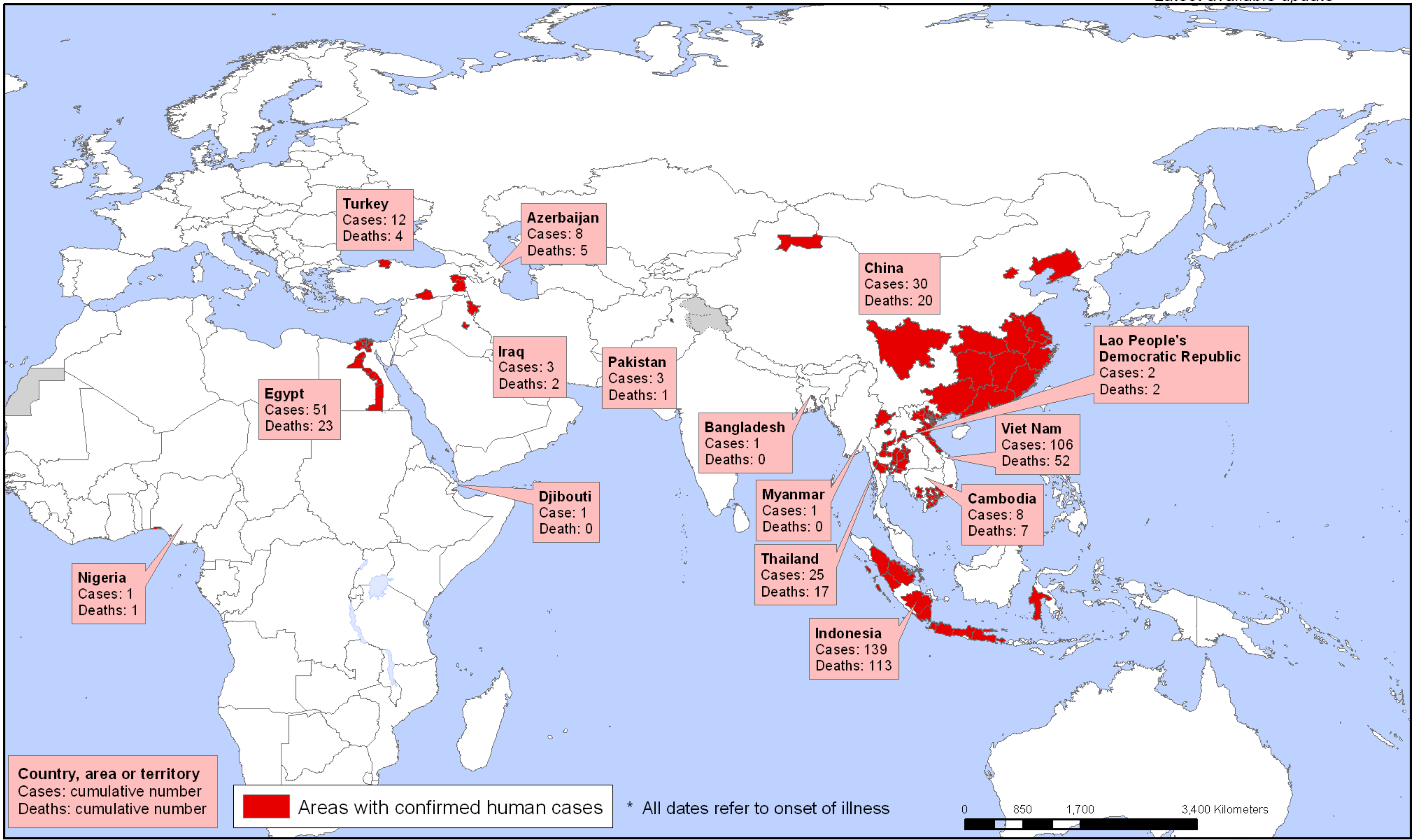


Figure 2: Power to detect person-to-person transmission plotted by the number of index cases versus the total number of cases. Results are based on 2000 simulations. 2000 permuted samples were drawn for each permutation test. The dashed line is the 80% power contour line obtained from Loess smoothing. The solid line is the lower bound (0) of power, where the number of index cases equals the total number of cases.

Cumulative 397 reported cases, 249 deaths, 63% case fatality ratio in 15 countries
31 documented family clusters

Areas with confirmed human cases of H5N1 avian influenza since 2003 *

Status as of 16 December 2008
Latest available update



* All dates refer to onset of illness

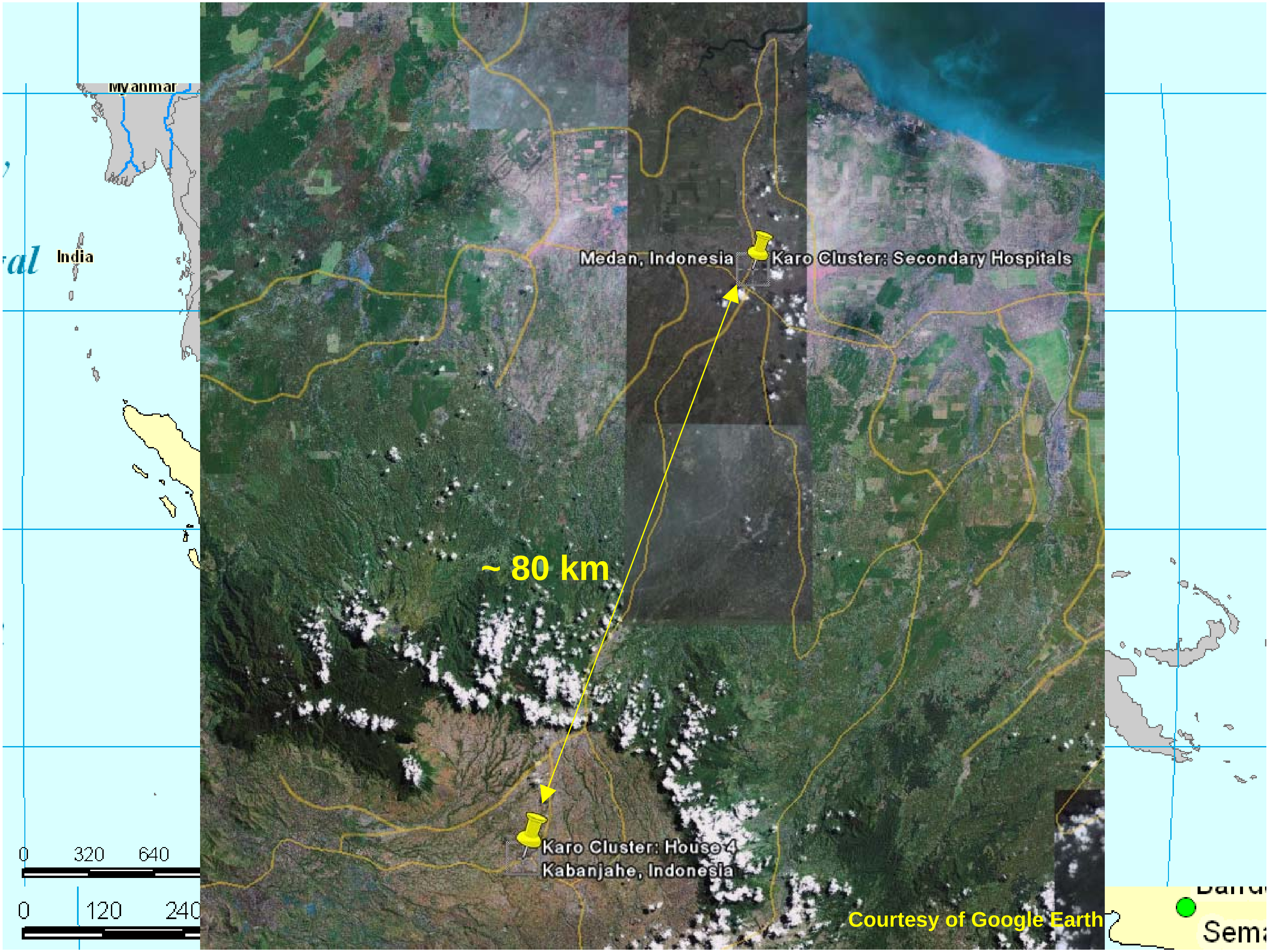
World Health Organization

Data Source: WHO
Map Production: Public Health Information and Geographic Information System (GIS)
World Health Organization

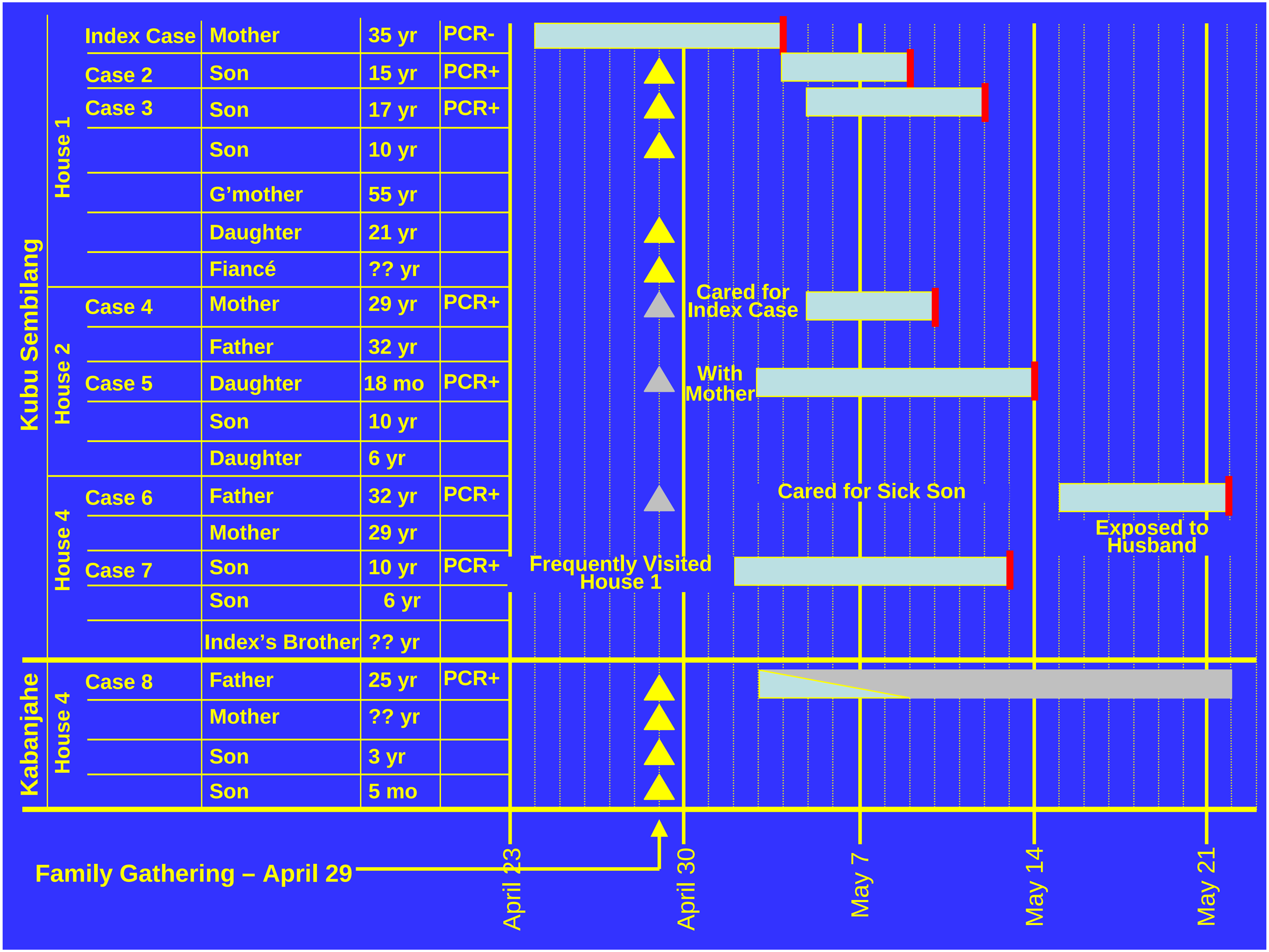
H5N1 Influenza in Family Cluster in North Sumatra, May 2006

Overview

- Suspected human-to-human of H5N1 in a familial cluster
- late April – May 2006
- Location
  - Kubu Sembilang and Kabanjahe
  - North Sumatra
  - ~80 km from Medan (largest city in the area)
- Situation
  - Group of 3 adjacent houses in the Kubu Sembilang
  - 1 house in Kabanjahe (~ 10 km from Kubu Sembilang)
  - 8 cases among 21 members of this extended family
    - 1 suspected – index case
    - 7 confirmed by WHO (PCR+)
  - 54 non-index family members and close contacts
    - Limited information
    - Received prophylactic Tamiflu (oseltamivir), unless contra-indicated
    - May have included some of the above 21 members
- Outcome
  - 7 Deaths
  - 1 Survivor

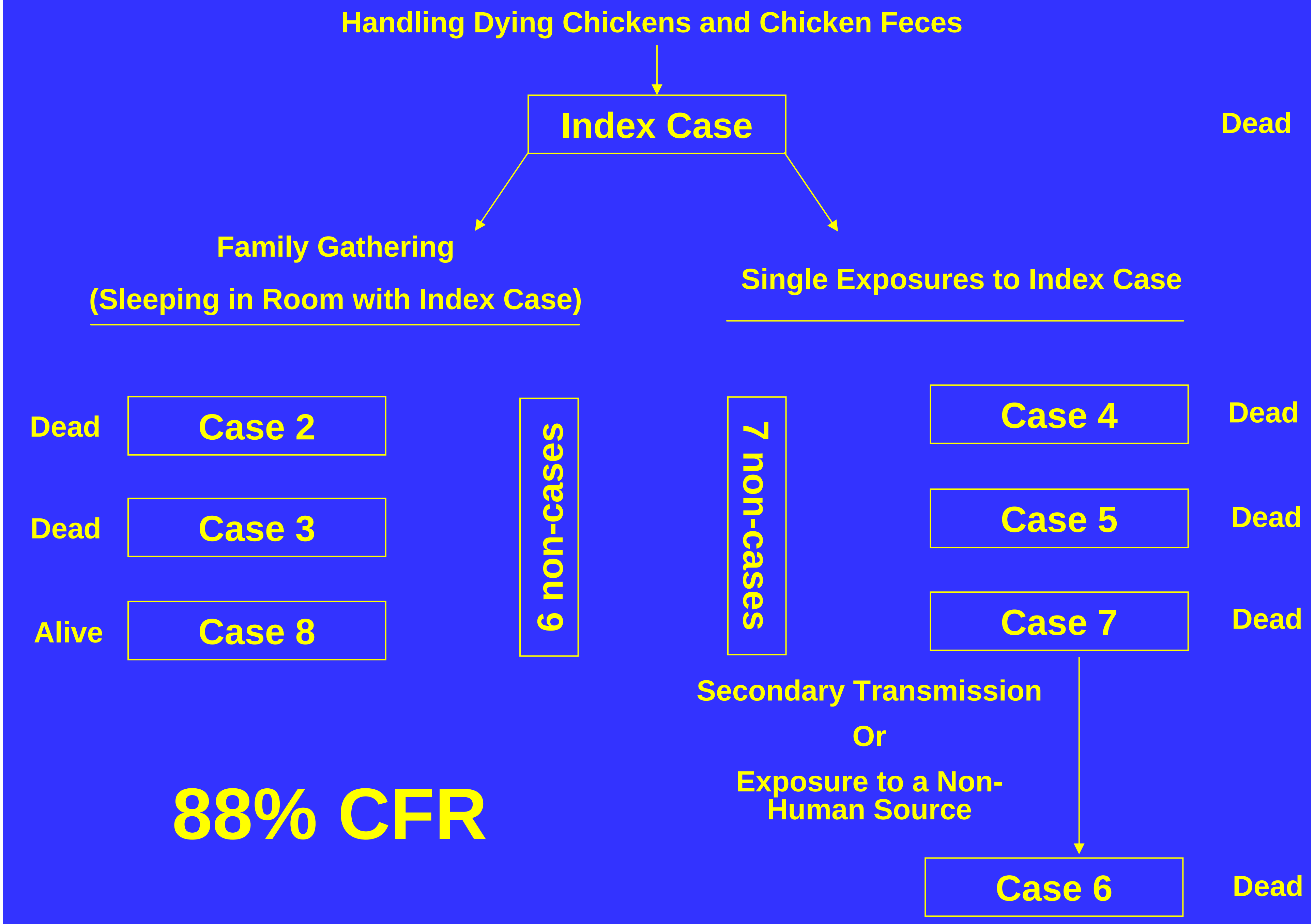


Courtesy of Google Earth



Kubu Sembilang	House 1	Index Case	Mother	35 yr	PCR-
		Case 2	Son	15 yr	PCR+
		Case 3	Son	17 yr	PCR+
			Son	10 yr	
			G’mother	55 yr	
			Daughter	21 yr	
			Fiancé	?? yr	
	House 2	Case 4	Mother	29 yr	PCR+
			Father	32 yr	
		Case 5	Daughter	18 mo	PCR+
			Son	10 yr	
			Daughter	6 yr	
	House 4	Case 6	Father	32 yr	PCR+
			Mother	29 yr	
		Case 7	Son	10 yr	PCR+
			Son	6 yr	
			Index’s Brother	?? yr	
Kabanjahe	House 4	Case 8	Father	25 yr	PCR+
			Mother	?? yr	
			Son	3 yr	
			Son	5 mo	

Family Gathering – April 29



Handling Dying Chickens and Chicken Feces

Index Case

Dead

Family Gathering

(Sleeping in Room with Index Case)

Dead Case 2

Dead Case 3

Alive Case 8

6 non-cases

Single Exposures to Index Case

7 non-cases

Case 4 Dead

Case 5 Dead

Case 7 Dead

Secondary Transmission

Or

Exposure to a Non-Human Source

Case 6 Dead

88% CFR

Tests and estimates

- Statistical evidence of person-to-person transmission ($p = 0.009$)

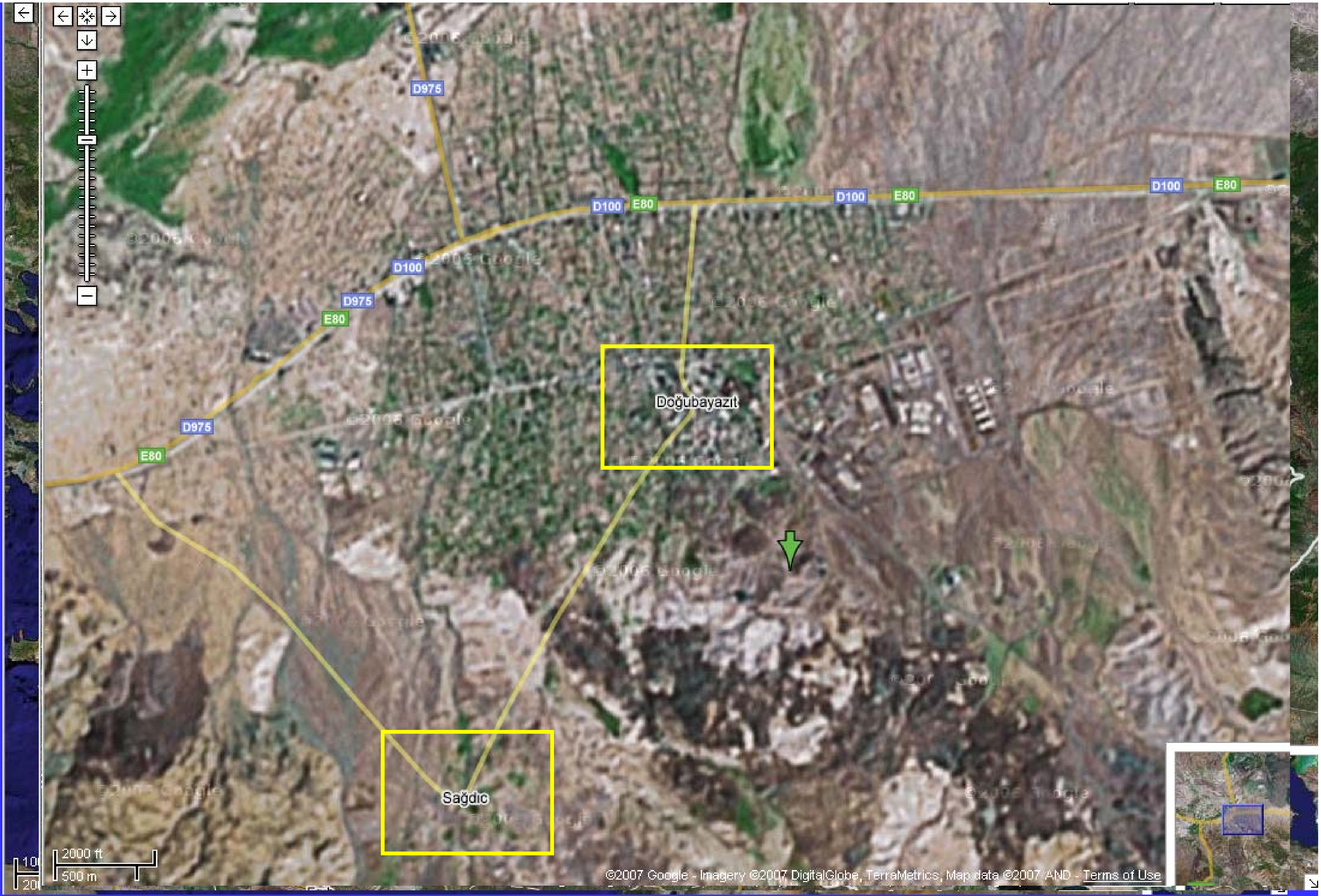
- Household SAR: 29% (95% CI, 15-51%)

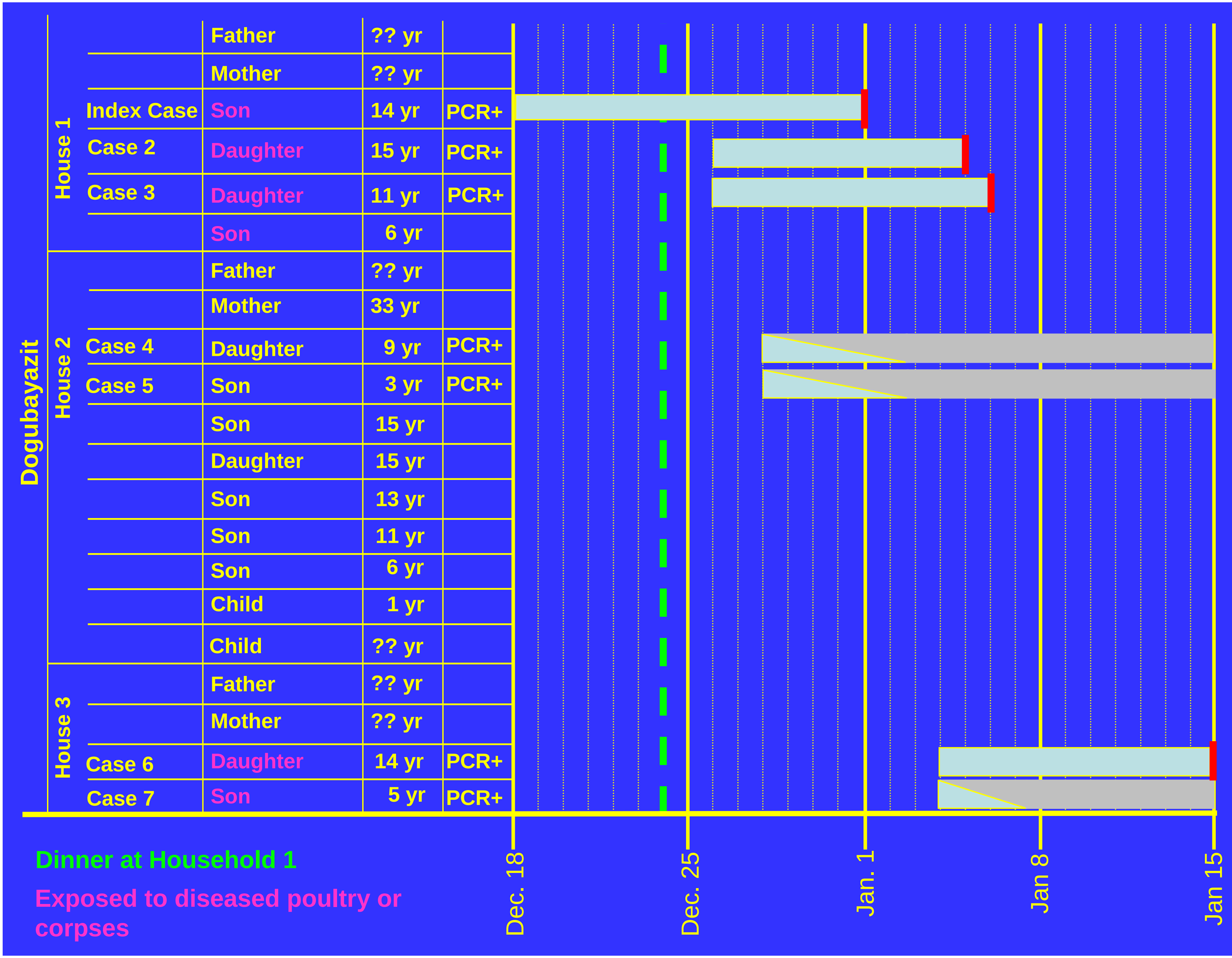
- Lower bound on the local R_0: 1.14 (95% CI, 0.61-2.14)
  - 12% chance of no further spread

Epidemiology of H5N1 Influenza Cluster in Dogubayazit, December 2005 - January 2006

Epidemiology of Turkey H5N1 Cluster

- Time: December 18, 2005 – January 15, 2006
- Place:
  - 2 villages
    - Sagdic
    - Sulucan
  - Dogubayazit, Agri province
  - Bordering Iran and Armenia
- People (N = 21):
  - 3 Households
    - Family 1: n = 4
    - Family 2: n = 11
    - Family 3: n = 6
  - 700 – 1500 meters apart
- Outcome
  - 7 confirmed cases
    - 4 deaths
    - 3 survivors
  - 14 non-cases
  - CFR = 57%


D975
D100
E80
D100
E80
D100
E80
D100
D975
E80
D975
E80
Doğubayazıt
Sağdic
2000 ft
500 m
©2007 Google - Imagery ©2007 DigitalGlobe, TerraMetrics, Map data ©2007 AND - Terms of Use



Dogubayazit	House 1		Father	?? yr	
			Mother	?? yr	
		Index Case	Son	14 yr	PCR+
		Case 2	Daughter	15 yr	PCR+
		Case 3	Daughter	11 yr	PCR+
			Son	6 yr	
	House 2		Father	?? yr	
			Mother	33 yr	
		Case 4	Daughter	9 yr	PCR+
		Case 5	Son	3 yr	PCR+
			Son	15 yr	
			Daughter	15 yr	
			Son	13 yr	
			Son	11 yr	
			Son	6 yr	
			Child	1 yr	
			Child	?? yr	
	House 3		Father	?? yr	
			Mother	?? yr	
		Case 6	Daughter	14 yr	PCR+
		Case 7	Son	5 yr	PCR+

Dinner at Household 1

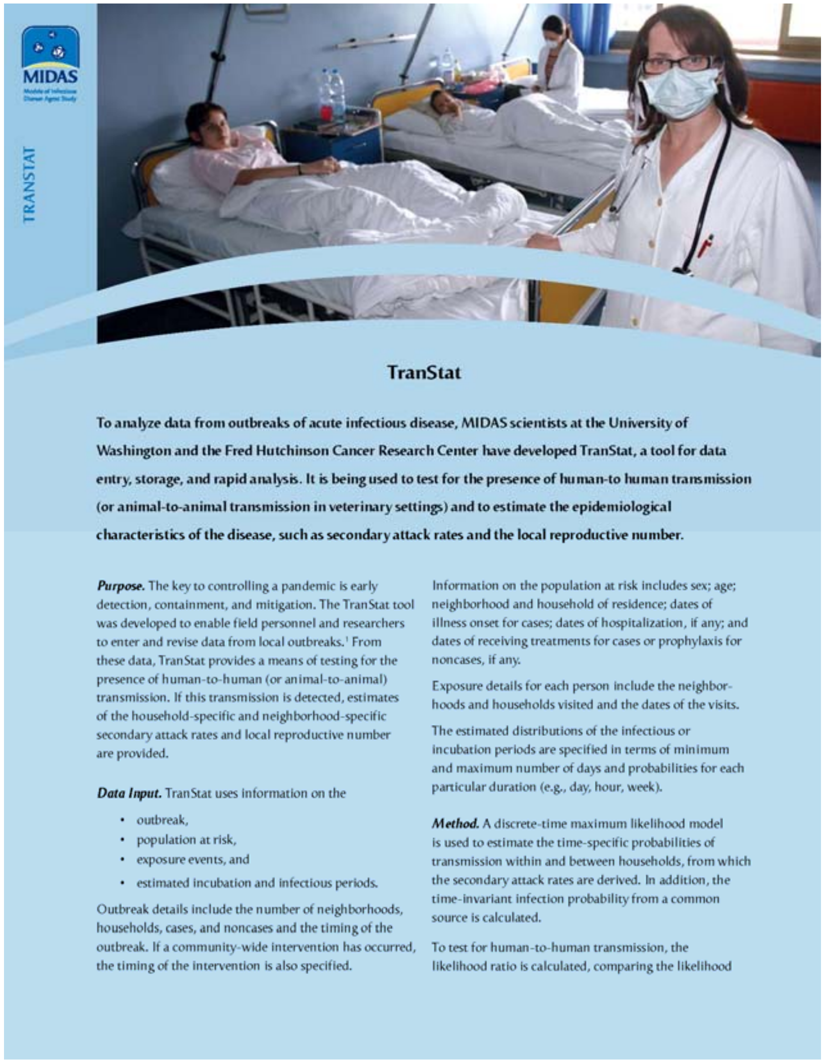
Exposed to diseased poultry or corpses

Tests and estimates

- No statistical evidence of person-to-person transmission (p = 0.114)

- Estimate of daily common source probability of infection, *b*: 0.011 (95% CI, *0.005 - 0.025*)

TranStat I



TranStat

To analyze data from outbreaks of acute infectious disease, MIDAS scientists at the University of Washington and the Fred Hutchinson Cancer Research Center have developed TranStat, a tool for data entry, storage, and rapid analysis. It is being used to test for the presence of human-to human transmission (or animal-to-animal transmission in veterinary settings) and to estimate the epidemiological characteristics of the disease, such as secondary attack rates and the local reproductive number.

Purpose. The key to controlling a pandemic is early detection, containment, and mitigation. The TranStat tool was developed to enable field personnel and researchers to enter and revise data from local outbreaks.[1] From these data, TranStat provides a means of testing for the presence of human-to-human (or animal-to-animal) transmission. If this transmission is detected, estimates of the household-specific and neighborhood-specific secondary attack rates and local reproductive number are provided.

Data Input. TranStat uses information on the

- outbreak,
- population at risk,
- exposure events, and
- estimated incubation and infectious periods.

Outbreak details include the number of neighborhoods, households, cases, and noncases and the timing of the outbreak. If a community-wide intervention has occurred, the timing of the intervention is also specified.

Information on the population at risk includes sex; age; neighborhood and household of residence; dates of illness onset for cases; dates of hospitalization, if any; and dates of receiving treatments for cases or prophylaxis for noncases, if any.

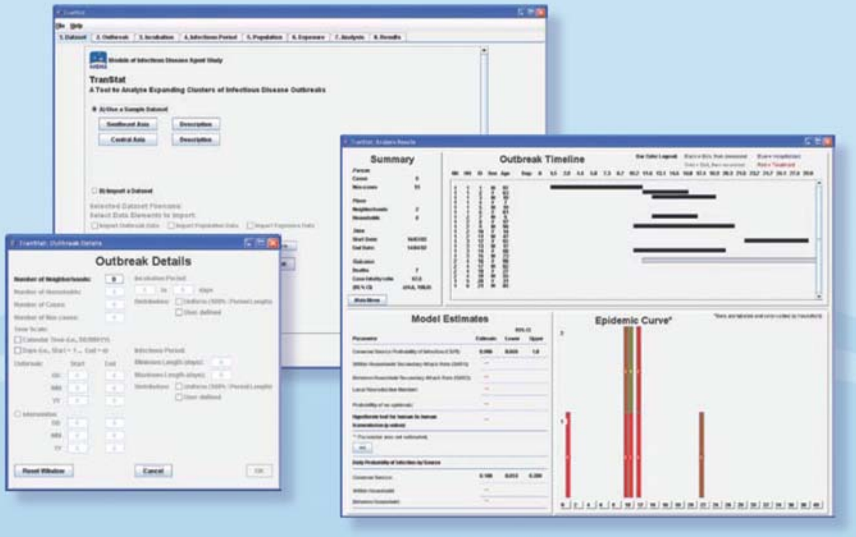
Exposure details for each person include the neighborhoods and households visited and the dates of the visits.

The estimated distributions of the infectious or incubation periods are specified in terms of minimum and maximum number of days and probabilities for each particular duration (e.g., day, hour, week).

Method. A discrete-time maximum likelihood model is used to estimate the time-specific probabilities of transmission within and between households, from which the secondary attack rates are derived. In addition, the time-invariant infection probability from a common source is calculated.

To test for human-to-human transmission, the likelihood ratio is calculated, comparing the likelihood

https://www.epimodels.org/midas/about.do



assuming that there is no transmission (i.e., the within- and between-household transmission probabilities are zero) with the likelihood when no assumption is made. In TranStat, the assumption of no human-to-human transmission reduces to a model with transmission from common sources only. Because this transmission does not vary between individuals, a simple permutation method was developed to estimate the distribution of likelihood ratio statistics under the assumption of no human-to-human transmission: The infection and symptom status of the observed cases are repeatedly reassigned to different groups of subjects. The observed likelihood ratio statistic can then be compared with the generated distribution of likelihood ratio statistics to obtain the *p*-value.[2,3]

Results. TranStat results include estimates (and 95% confidence intervals, as appropriate) for

- the case fatality rate,
- the daily probability of infection by common source (such as a zoonotic source),
- the within- and between-household daily transmission probabilities and secondary attack rates,
- local reproductive number, and
- the *p*-value for testing human-to-human transmission.

Availability. The TranStat tool can be downloaded from **http://www.midasmodels.org**. The version 0.1 Java code can be run on Windows platforms, and later versions can be run on any platform, including Linux.

For information about TranStat, send e-mail to

yang@scharp.org for questions about the tool methods and **tfarris@rti.org** for questions about operating TranStat.

For more information about MIDAS

MIDAS Web site: **http://www.midasmodels.org**

National Institute of General Medical Sciences (NIGMS): **http://www.nigms.nih.gov/Initiatives/MIDAS**

MIDAS Scientific Director at NIGMS:
Irene Eckstrand, **eckstrai@nigms.nih.gov**

---

[1] Yang, Y., Halloran, M.E., Sugimoto, J.D., & Longini, I.M. (2007). Detecting human-to-human transmission of avian influenza A (H5N1). *Emerging Infectious Diseases*, 13(9): 1348-1353. Retrieved from http://www.cdc.gov/eid/content/13/9/1348.htm.

[2] Yang, Y., Longini, I.M., & Halloran, M.E. (2006). Design and evaluation of prophylactic interventions using infectious disease incidence data from close contact groups. *Applied Statistics*, 55: 317-330.

[3] Yang, Y., Longini, I.M., & Halloran, M.E. (2007). A resampling-based test to detect person-to-person transmission of infectious disease. *Annals of Applied Statistics*, 1(1):211-228.

About MIDAS

MIDAS
Models of Infectious Disease Agent Study

Funded by the National Institute of General Medical Sciences (NIGMS), one of the National Institutes of Health, the MIDAS mission is to investigate computational and mathematical models of existing and emerging infectious diseases in order to prepare the nation to respond to disease outbreaks.

http://www.midasmodels.org

TranStat II

- Covariate adjustment
- Simulating epidemics with given parameters
- Multiple statistical methods