Challenges in estimating parameters and characterizing transmission dynamics from incomplete epidemic data
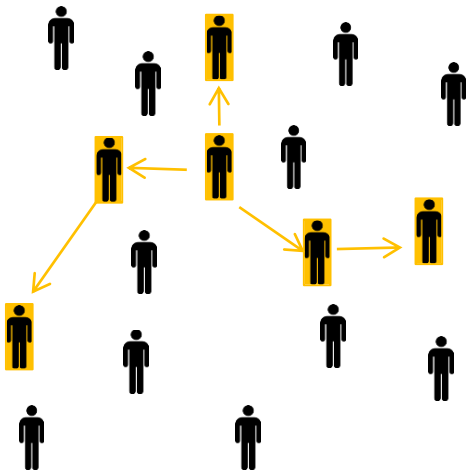
Simon Cauchemez

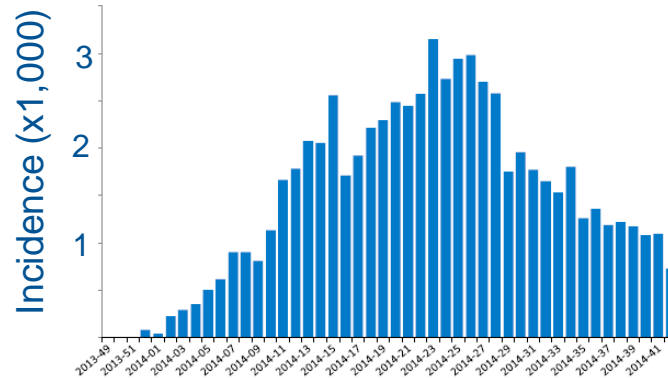Mathematical Modelling of Infectious Diseases Unit

Institut Pasteur, Paris

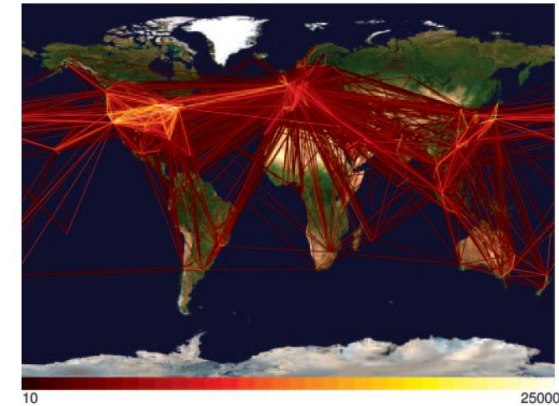Institut Pasteur

The multiple scales of spread

Local

Regional / national

International

- Analysis of detailed outbreak data is critical to estimate transmission parameters of infectious diseases
- But burdened by a number of challenges: missing data, measurement errors, censoring…
- Bayesian data augmentation have played a key role to address these challenges and maximize insight

Institut Pasteur

Studying transmission from incomplete data

- Household transmission studies.

- Progressively adding more complexity to the analysis:
  - ➤ Missing data issues.
  - ➤ Complex social structures.
  - ➤ Space.
  - ➤ Repeated measurements.

- Different pathogens:
  - ➤ Seasonal and pandemic influenza;
  - ➤ Chikungunya and dengue.

- Future directions.

Institut Pasteur

Study of influenza transmission in French households [Cauchemez et al, Stat Med, 2004; Ferguson et al, Nature 2005]

Follow-up of symptoms in 334 households for 15 days after onset in a confirmed index case
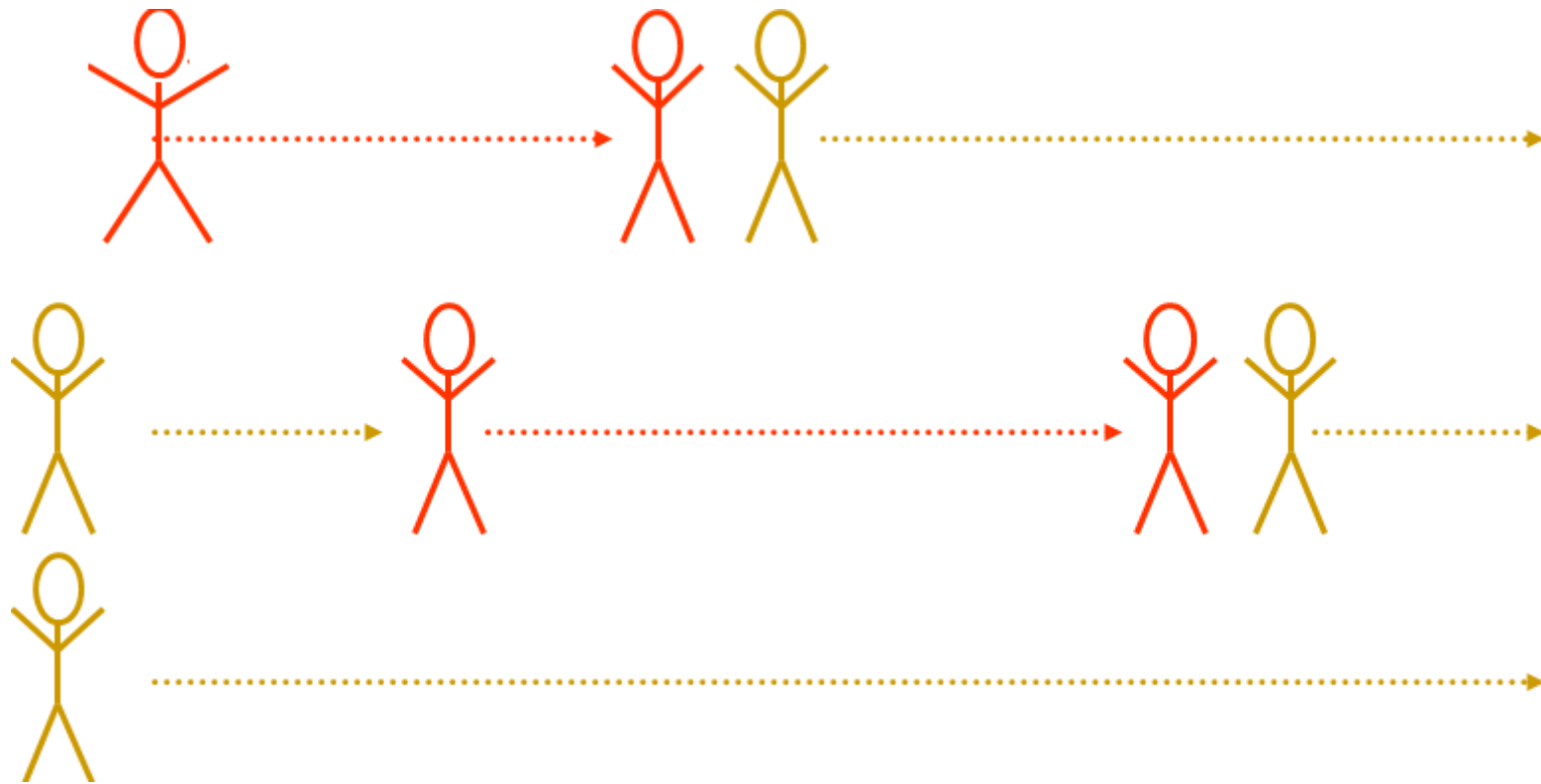
Transmission model and likelihood-based inference from complete data

Model for the hazard of infection of individual s at time t in the household:

$$\lambda_s(t) = \alpha_s + \sum_{i \in I(t)} \beta_i$$

Hazard of infection outside household

Force of infection exerted by infectious household members

Child 1

Child 2
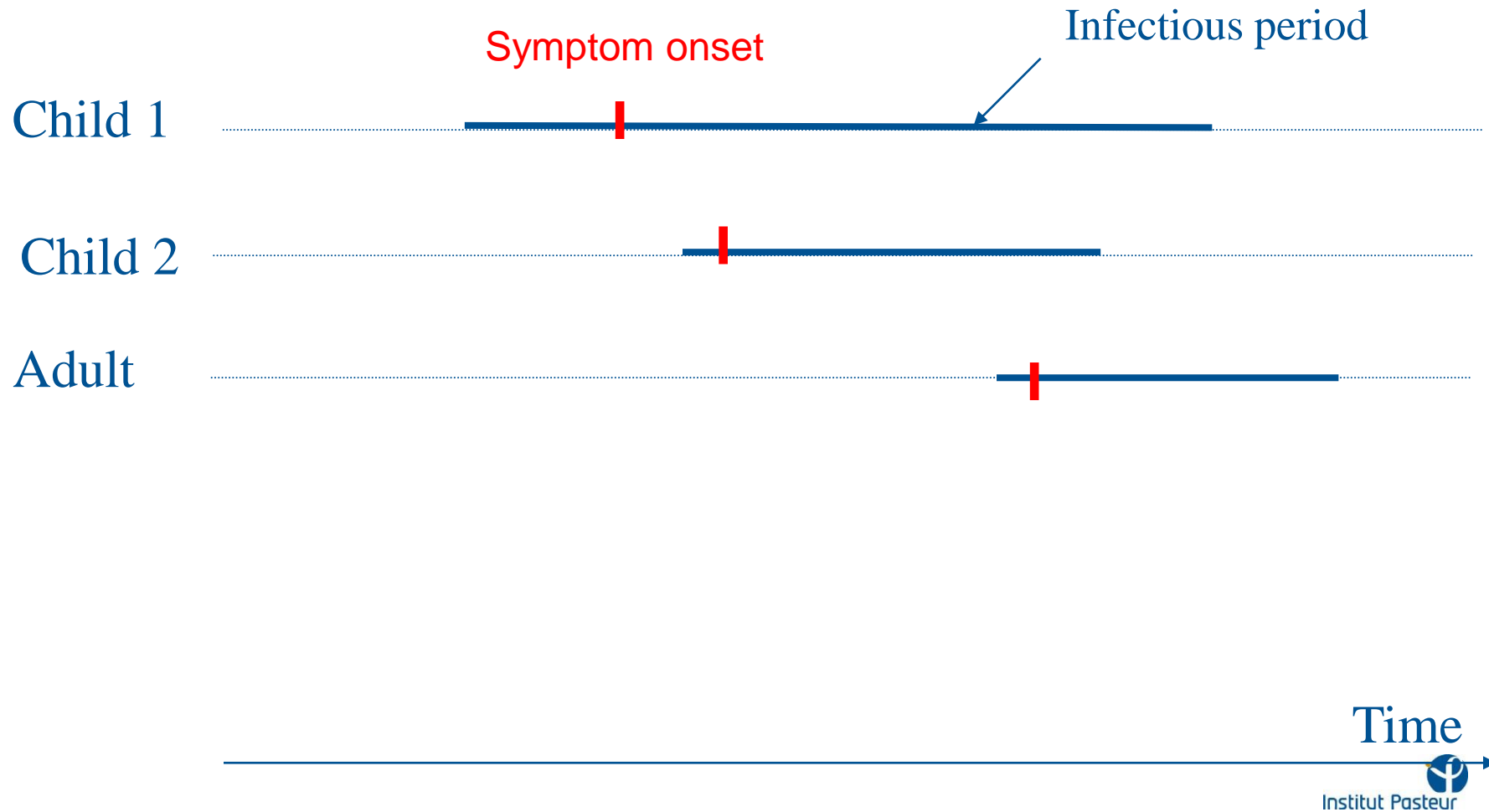
Adult

Infectious period

t_1 t_2
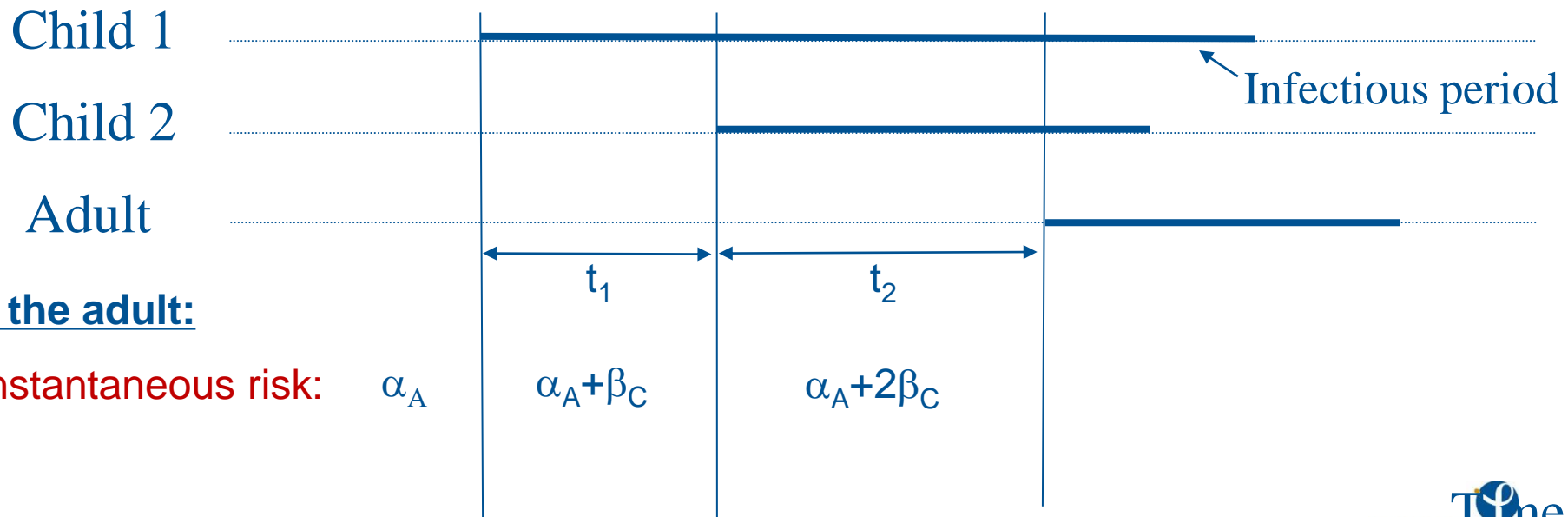
For the adult:

➤ Instantaneous risk: α_A $\alpha_A + \beta_C$ $\alpha_A + 2\beta_C$

Institut Pasteur

Inference from incomplete data (1)

- Infectious periods are unknown – we just observe dates of symptom onset;
- Explore trajectories consistent with data on symptoms onset.

Symptom onset

Infectious period

Child 1

Child 2

Adult

Time

Institut Pasteur

Inference from incomplete data (2)

- Notations:
  - ➤ Y: observed data – dates of symptom onset
  - ➤ Z: "missing" (augmented) data – dates when infectivity starts and when it ends;
  - ➤ θ: parameters

- Three-level hierarchical model:

$$P\left(Y, Z, \theta\right) = P\left(Y \mid Z\right) P\left(Z \mid \theta\right) P\left(\theta\right)$$

Observation level: ensures that augmented data consistent with observed data

Prior level

Transmission level: describes the latent transmission process

- Joint posterior distribution of augmented data and parameters explored via MCMC.

Institut Pasteur

Transmission of seasonal influenza in French households [Cauchemez et al, Stat Med, 2004; Ferguson et al, Nature 2005]

Household structure and transmission
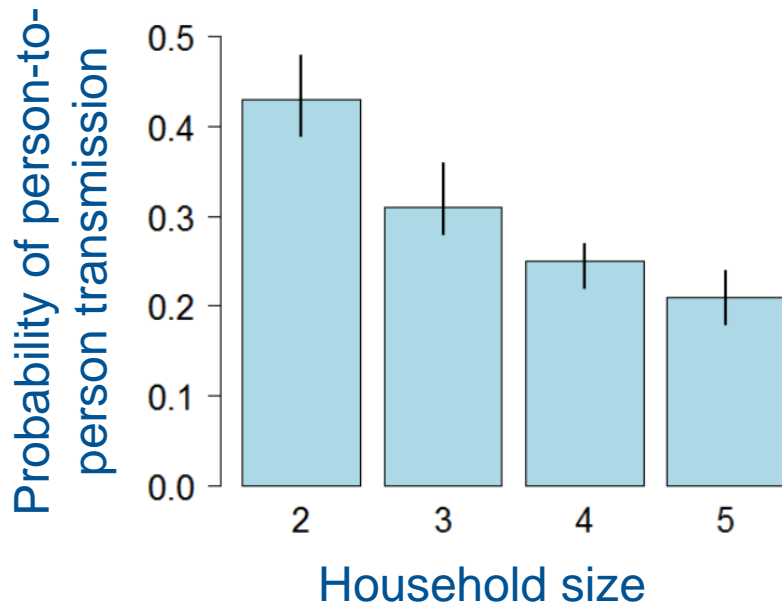
Key transmission parameters: the generation time
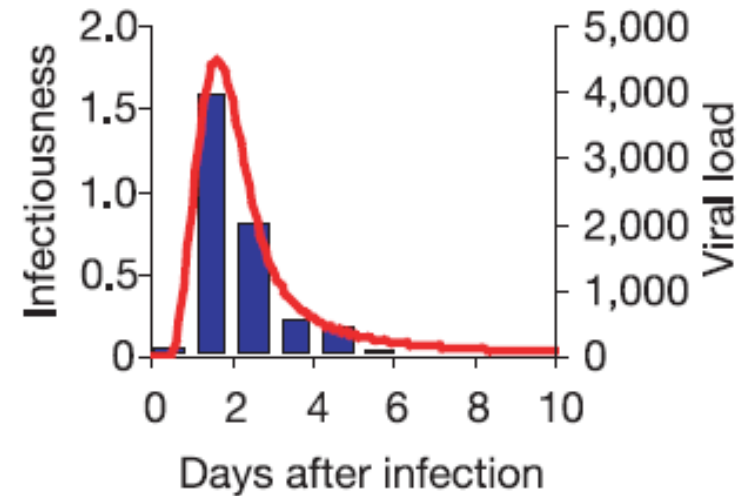
- Substantially shorter than initially thought (2.6 vs 4.1 days)
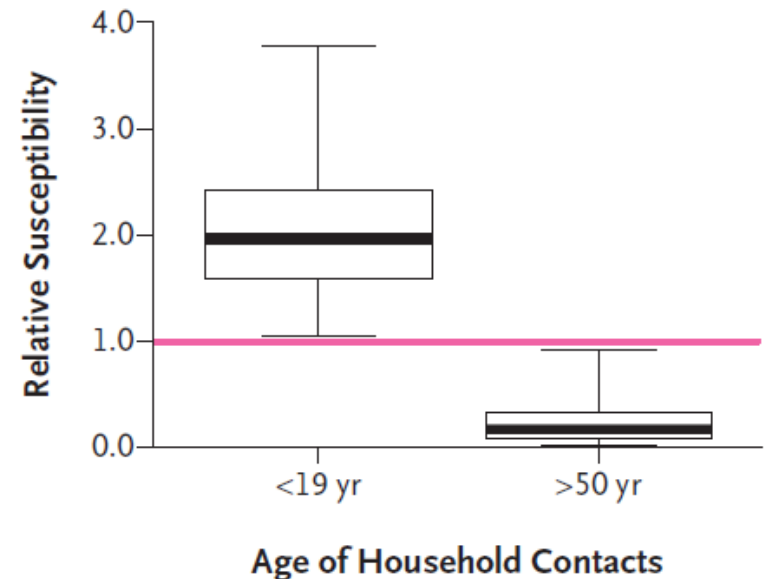- Important implications for control

Institut Pasteur

Household studies for early assessment during the 2009 H1N1 pandemic [Cauchemez et al, NEJM 2009]

Collaboration with CDC – Follow-up of 216 households for 7 days

Transmission risks within households
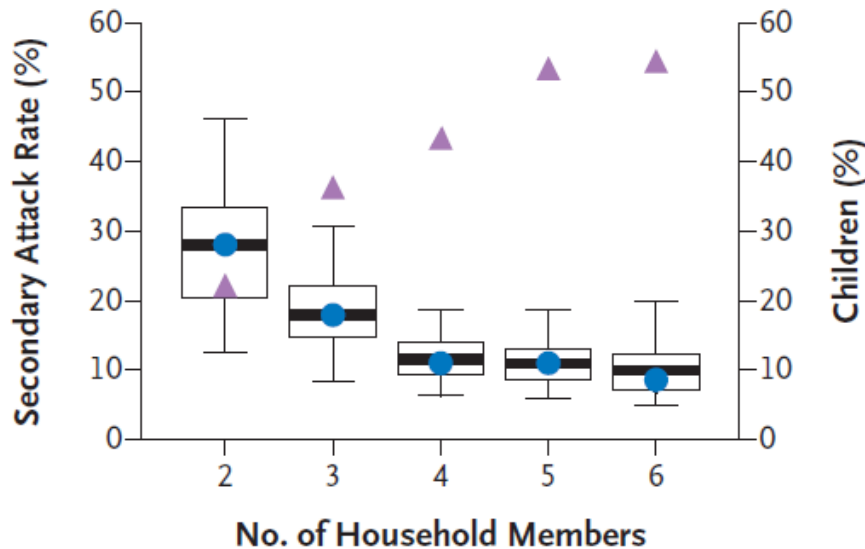
Children twice more susceptible than adults

Institut Pasteur

Need for more developments

**Households are not isolated
from the rest of the world!**
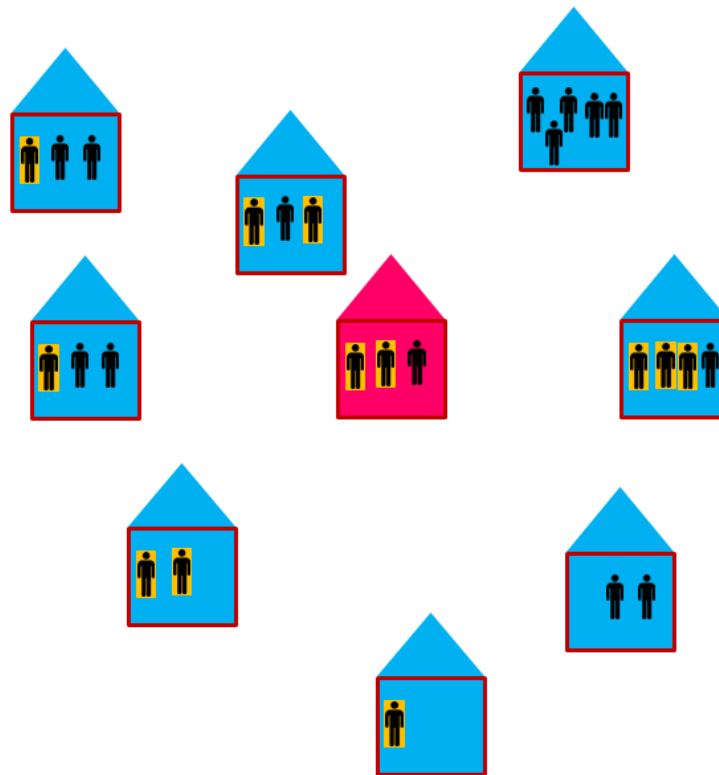
Transmission in other settings: A school outbreak of pandemic influenza in 2009 [Cauchemez et al, PNAS, 2011]

- Demographic & clinical information collected on students and their family members – 2 phone interviews:
  - 370 students,
  - 899 household members.

- Surveys in school for 4th graders:
  - Activities,
  - Seating charts,
  - Playmates.

Epidemic curve
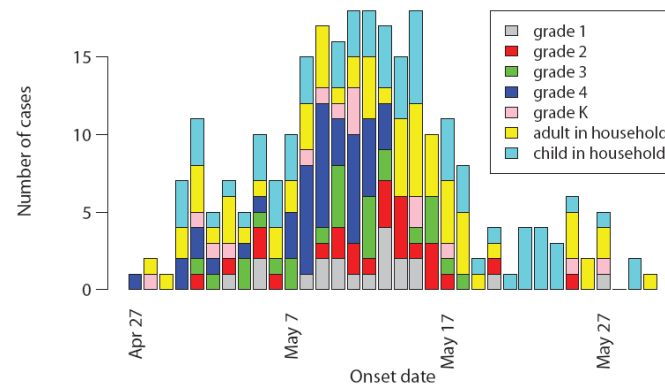
Seating charts

Social networks - Who are your playmates?
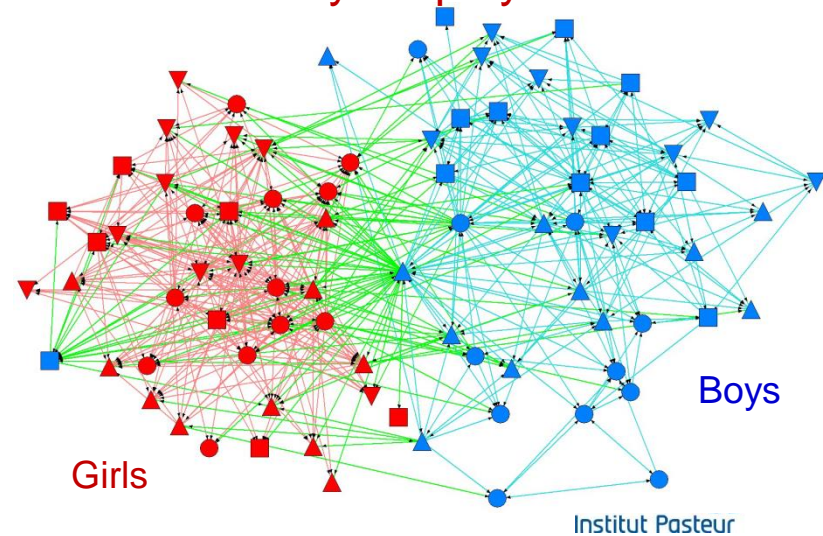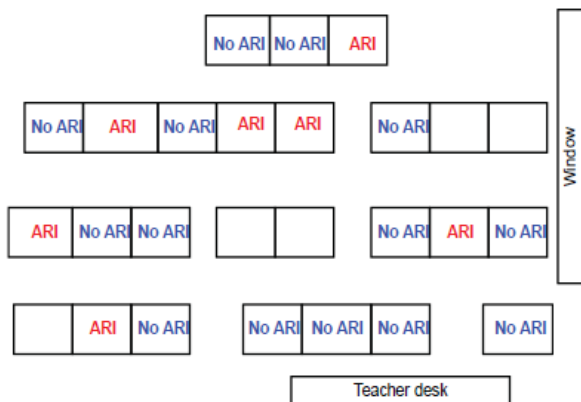
Statistical model
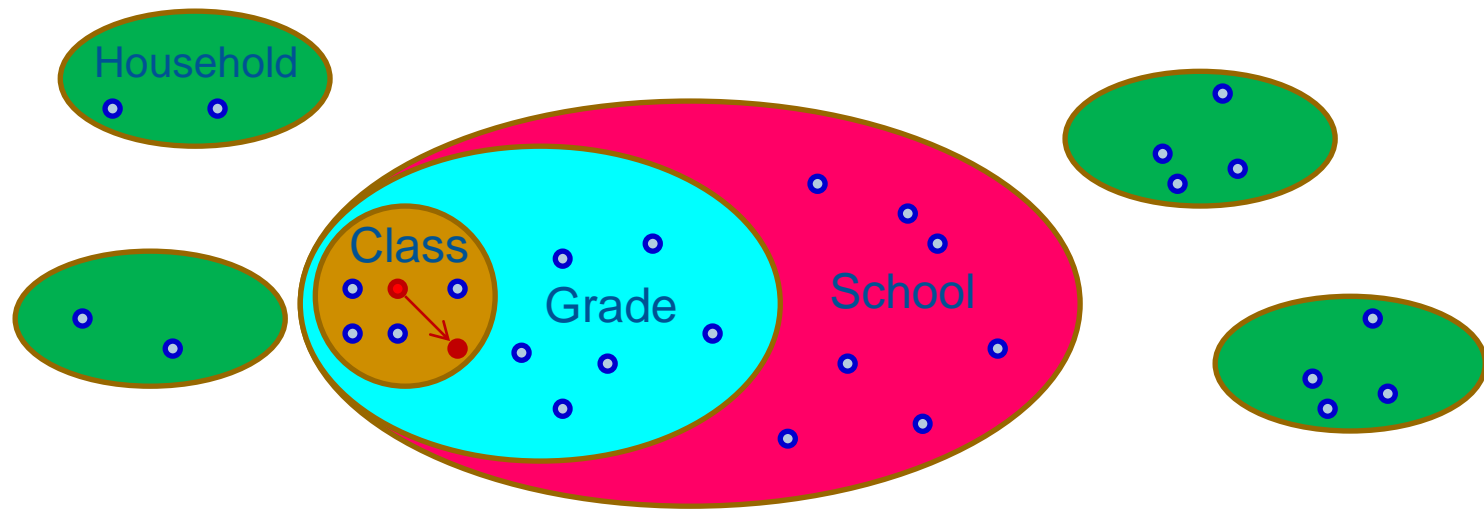
Transmission rates in the school and in the household
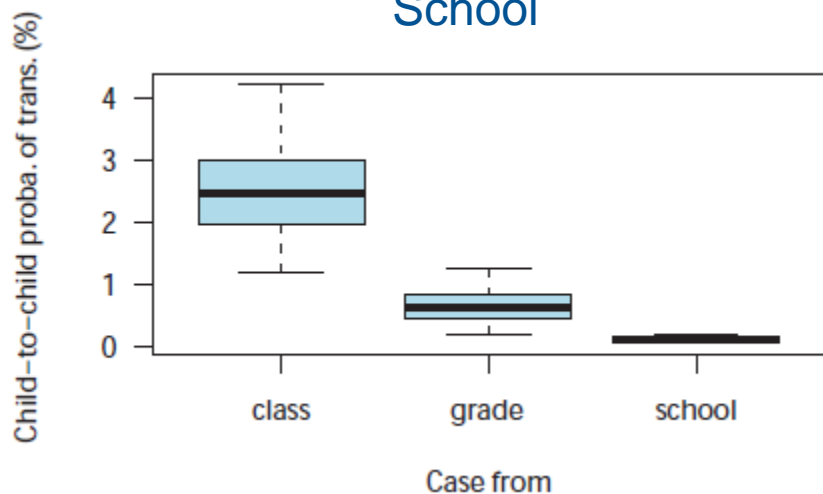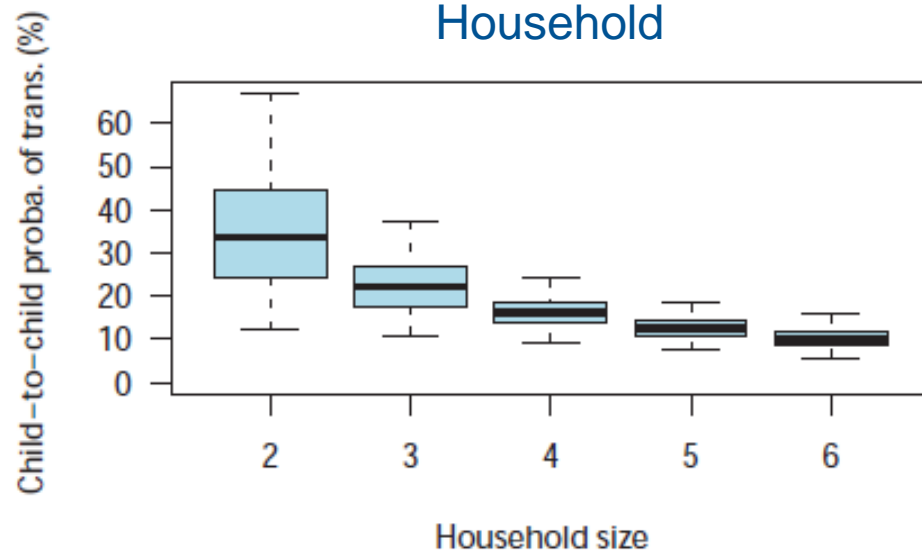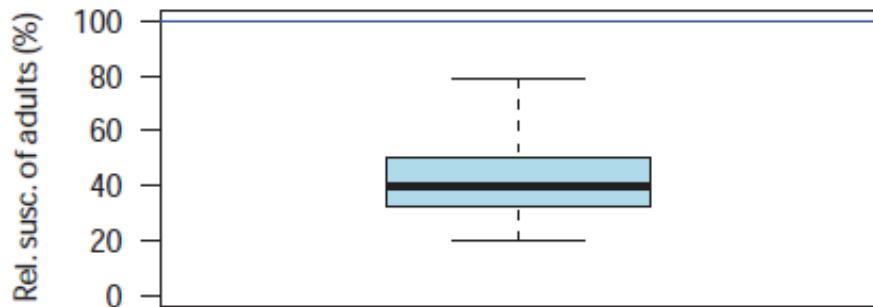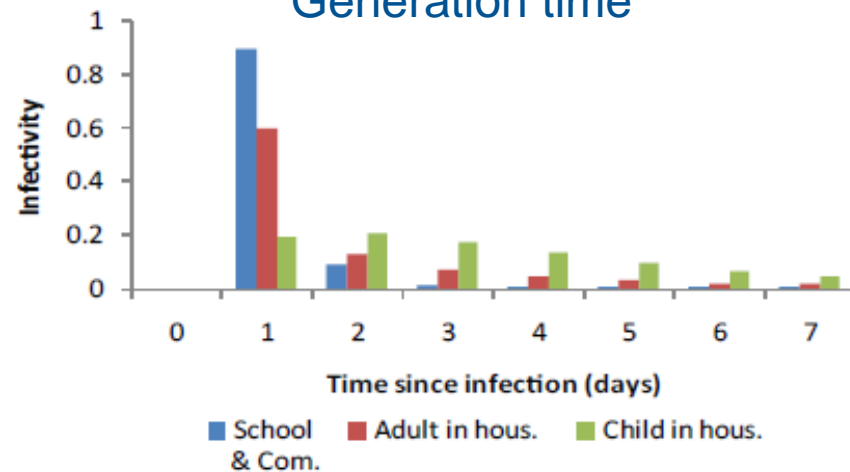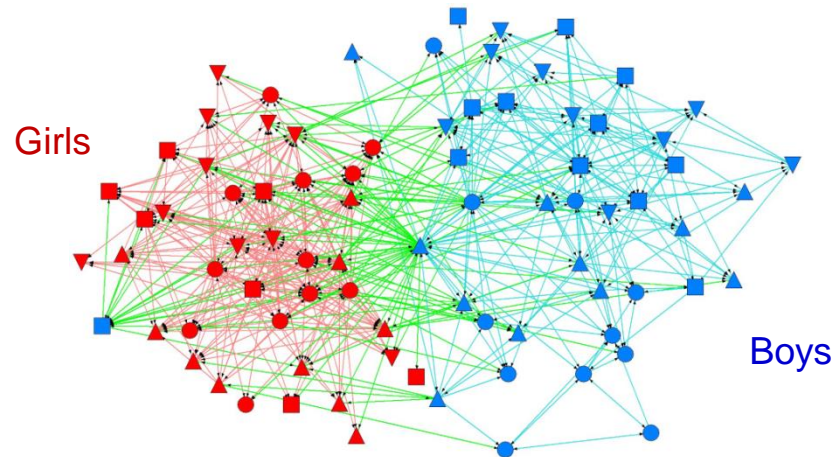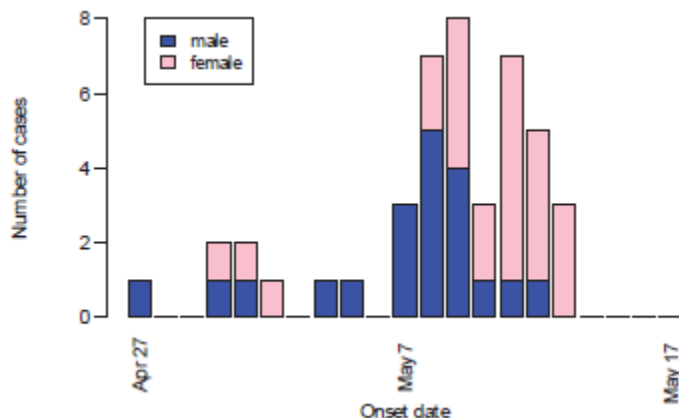
Gender-related mixing & transmission patterns

Social networks: Students are 4 times more likely to play with students of the same gender

Girls

Boys
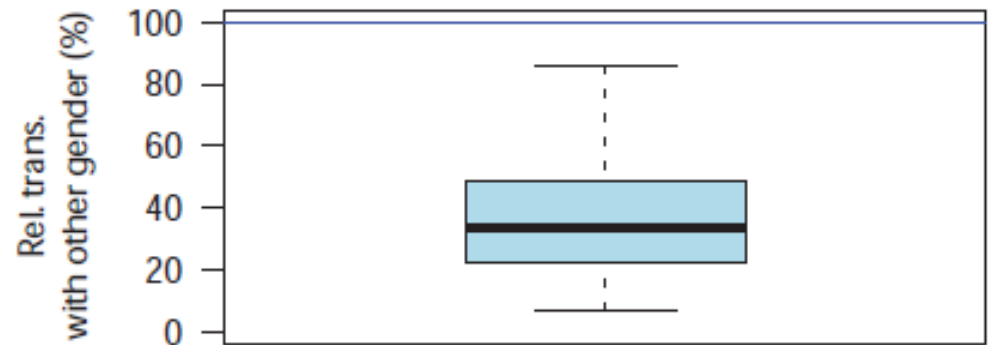
Evidence that this aspect of mixing patterns affected the transmission dynamics

i) Boys had onset before girls! (p=0.023)
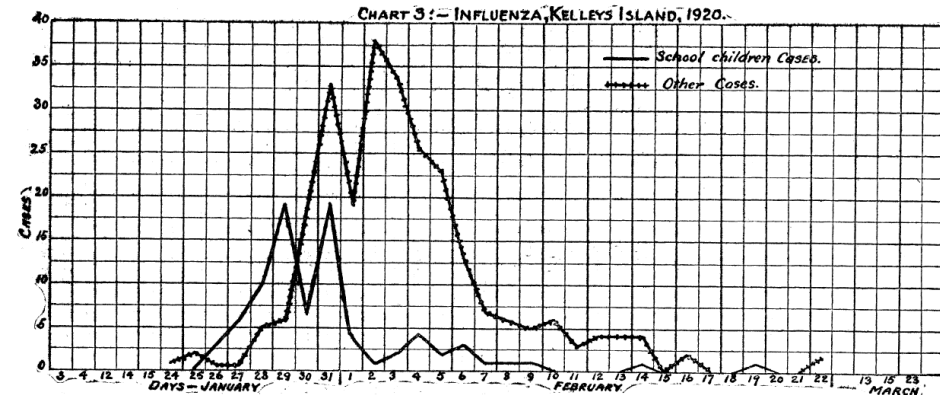
ii) Bayes Factor for model with gender-effect: 8.0
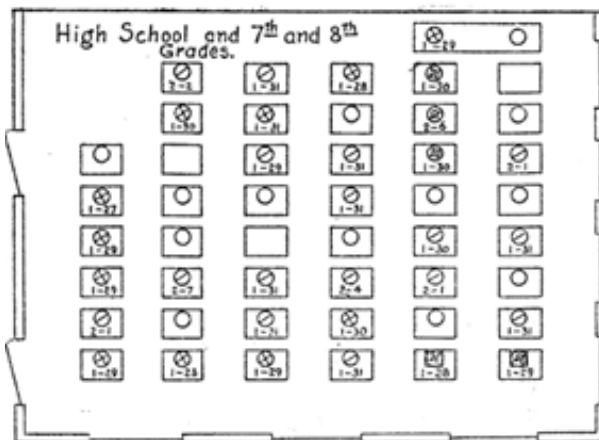
Was influenza transmission different 100 years ago? Pandemic flu in Kelly Island, 1920 (C. Armstrong)

- Isolated island, 689 inhabitants;
- Epidemic period: 24 Jan – 16 Feb 1920.
- Investigation begins 19 Feb 1920.
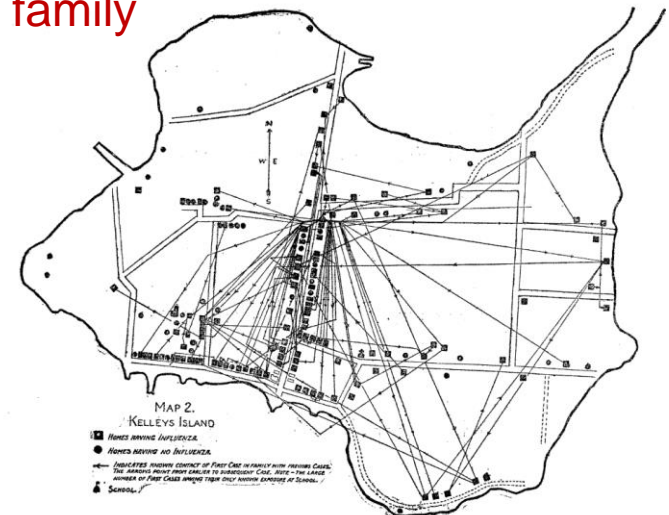- Demographic, clinical and contact information collected on the 689 inhabitants of the island.

Epidemic curves

Seating charts

Known contact for first case in family

- No individual level records available, even after Armstrong's daughter examined his personal effects.
- However, some person specific, and a lot of marginal data available in the paper
  - Epidemic curves for different types of cases e.g. students/others, first case in HH…
  - Whether first case in each HH is a student,
  - Know contact for first case in family,
  - Age and sex distribution of cases,
  - Generation times in HH,
  - General census info on HH structure…
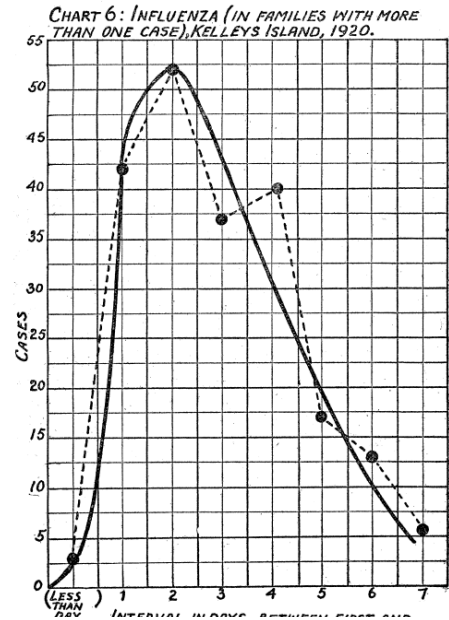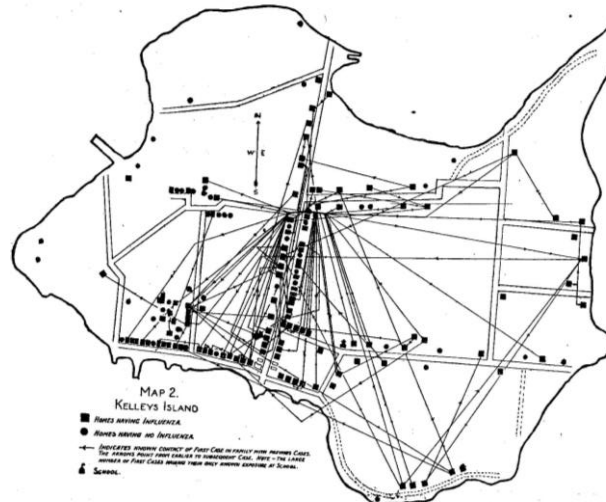
CHART 6: INFLUENZA (IN FAMILIES WITH MORE THAN ONE CASE), KELLEYS ISLAND, 1920.

TABLE XIII.—Summary of 27 known contacts followed by influenza, Kelleys Island, 1920.

MAP 2. KELLEYS ISLAND

MAP 1. EARLY CASES, KELLEYS ISLAND

Additional difficulty : Individual records have been destroyed (2)

Solution: Multiple imputation of assignments of reported cases to HHs that fulfill marginal constraints

Method 1: Exhaustive deterministic search of assignment space
- Unknown if there are contradictory constraints (full search needed to be sure)
- Even with efficient constraint propagation, no valid assignment of even 1/3 of households after >400,000,000 assignments considered (>4 days compute time),

Method 2: Probabilistic search
- Ensuring good match of marginal distributions.
- MCMC and data augmentation.
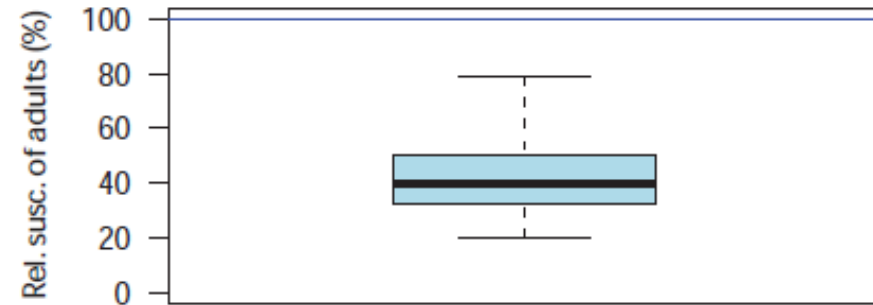- Convergence in <5 minutes (~5000 iterations).

Attack rates by age group

Attack rate by age group
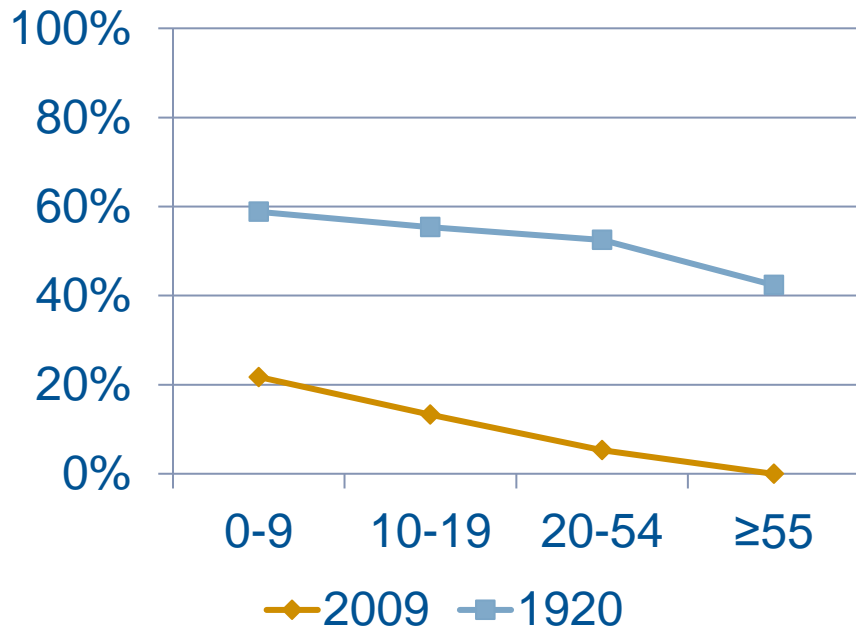
2009

Age dependent susceptibility

Institut Pasteur

Probability of transmission in the class

Same probability of transmission…
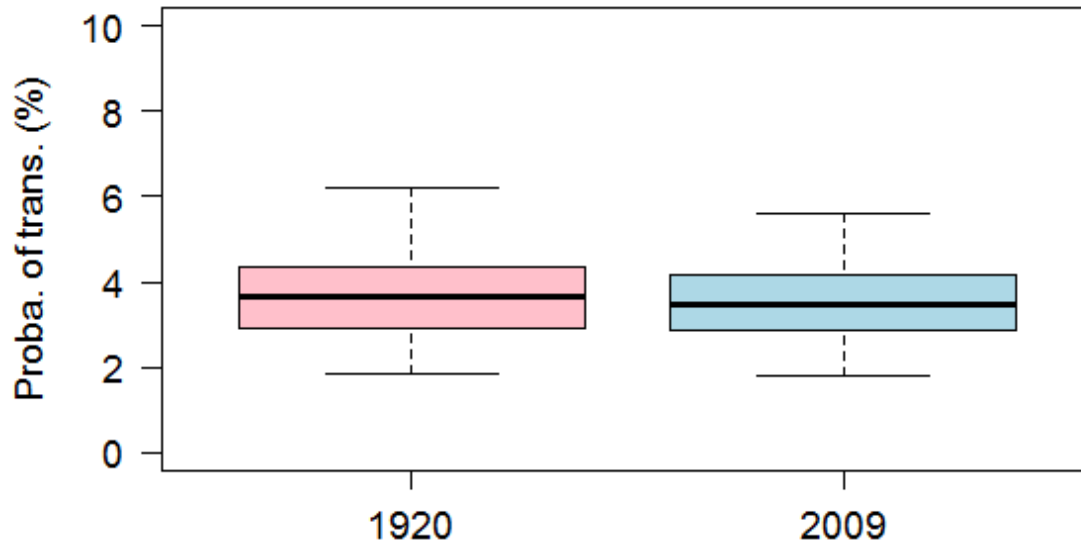
… but more children per class
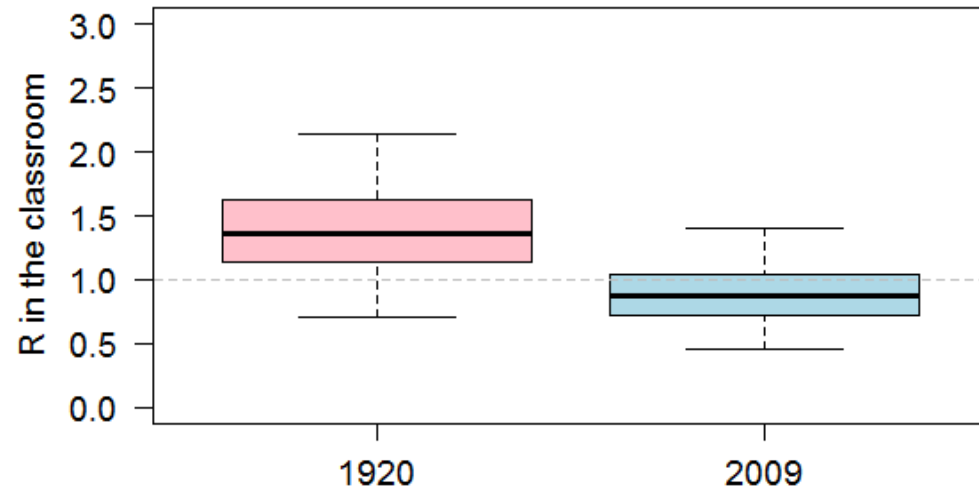
	Average number of children per class
1920	39
2009	25

Probability of transmission in the household
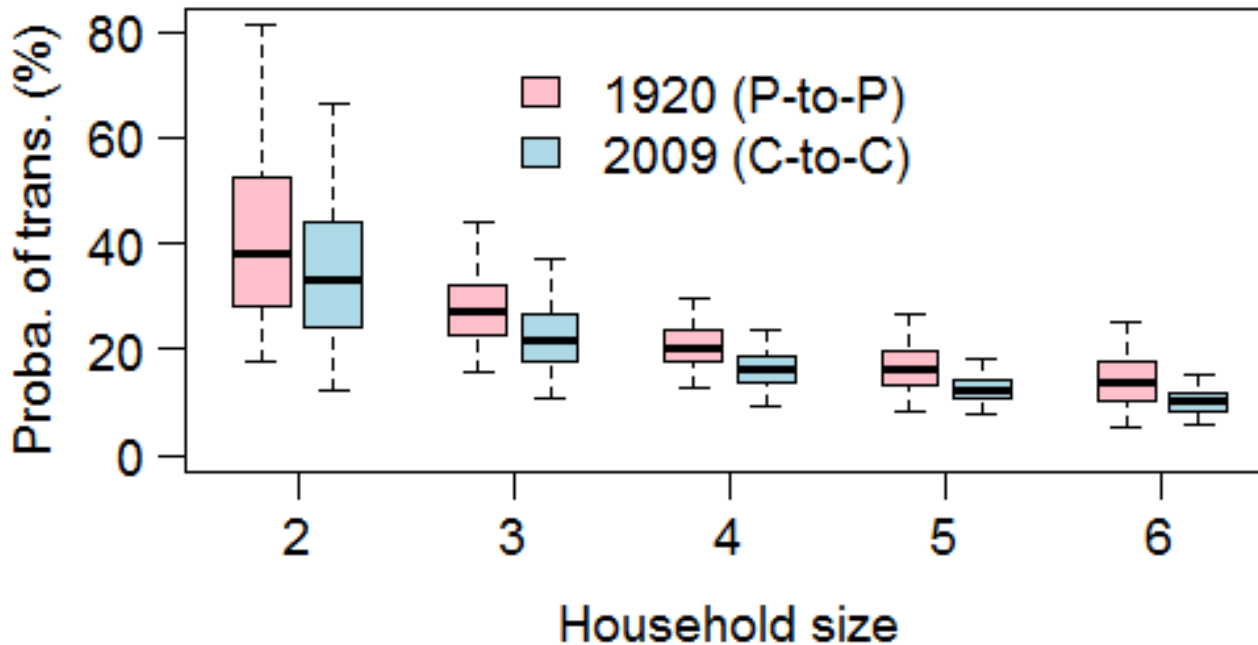
Sensitivity analysis: reconstruction of individuals records

Person-to-person transmission in the school
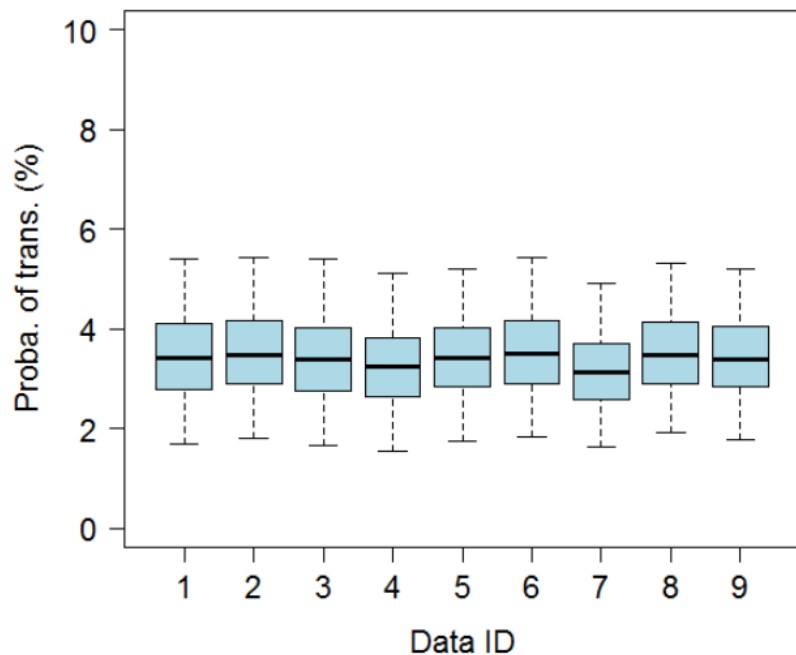
Person-to-person transmission in a household of size 3
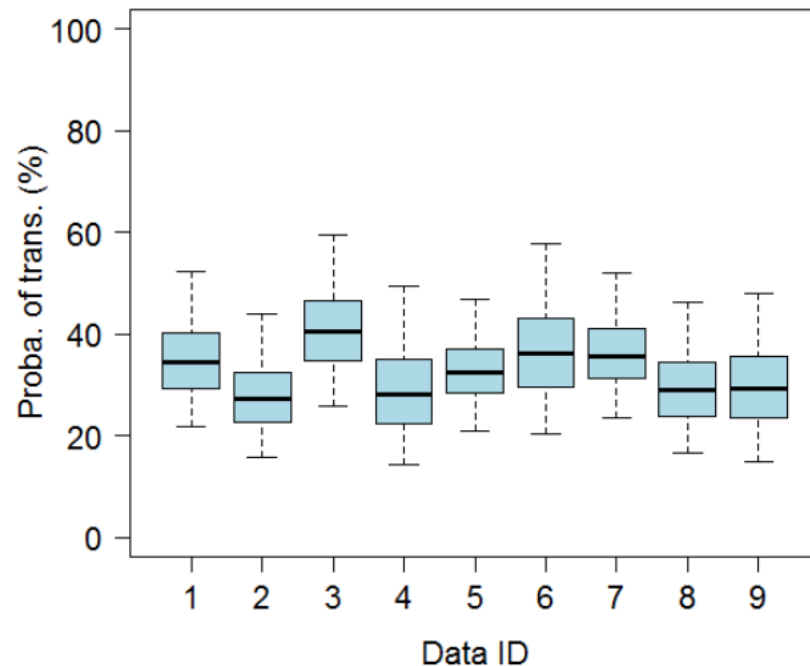
Spread of chikungunya in a village in Bangladesh
[Salje et al, PNAS, 2016]

β_{100m}

β_{20m}

β_{Hous}

Institut Pasteur
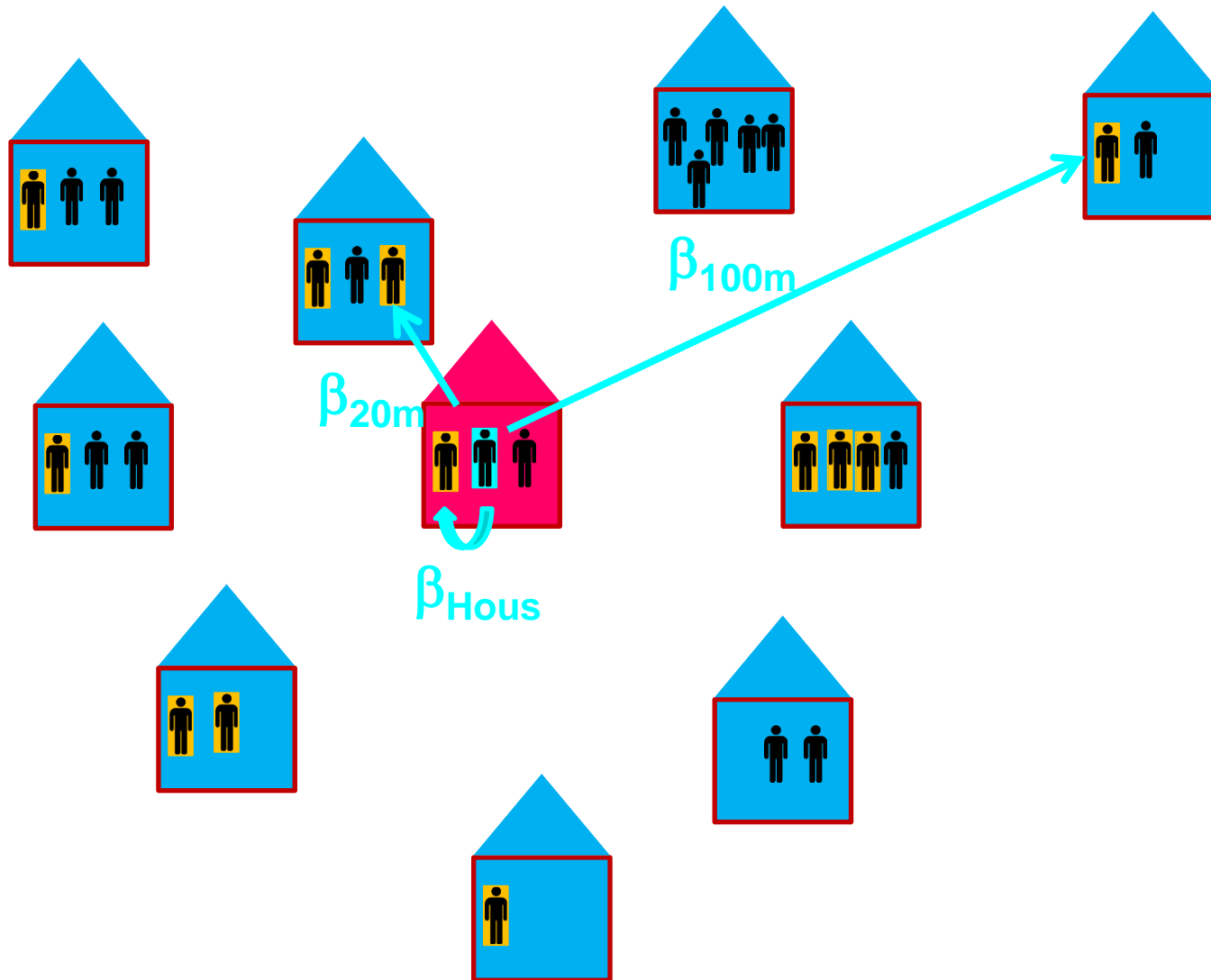
Transmission rates in households and as a function of distance

Household transmission rate
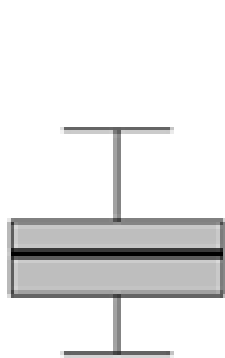
Transmission rate as a function of distance

Where does transmission occur?

Average distance at transmission
107m (89m-126m)

Institut Pasteur
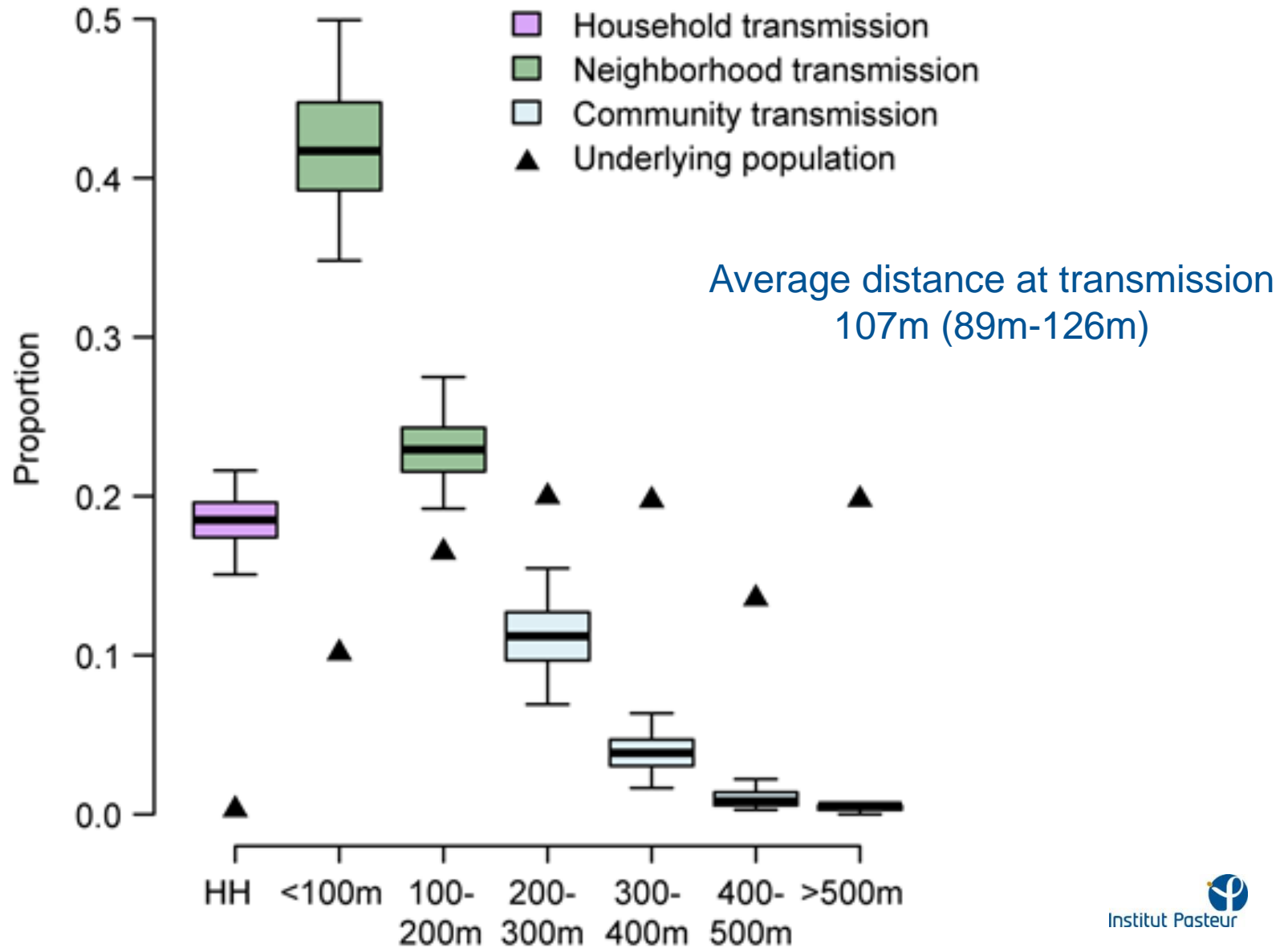
Characterizing infection burden and antibody dynamics from repeated measurements [Salje et al, Nature 2018]
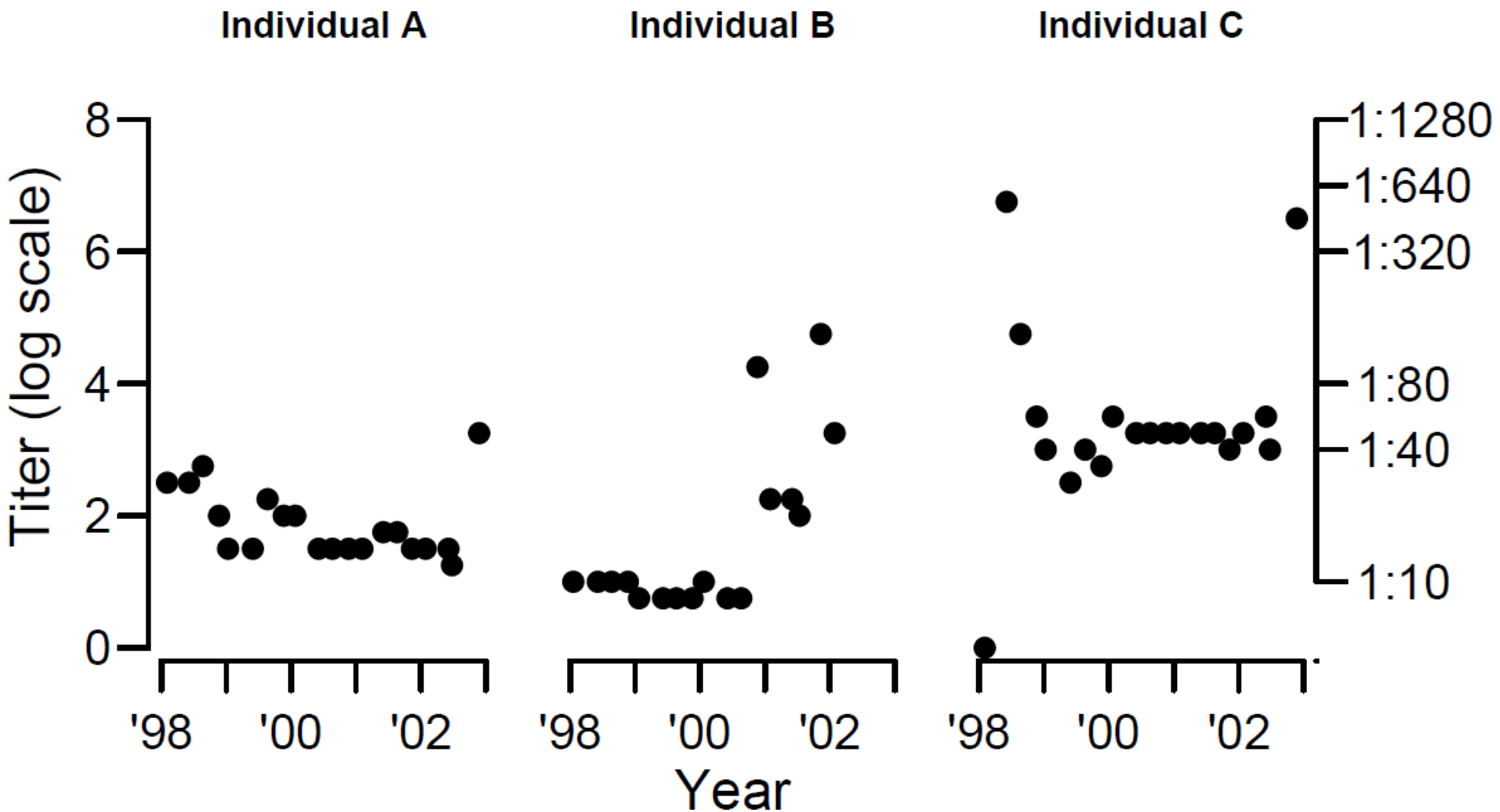
Dengue antibody titers measured every 3 months in >3000 Thai children

When were these individuals infected?

Methods

- Bayesian hierarchical model characterizing antiboby titers in individuals:
  - ➢ A model describing the history of infection.
  - ➢ A model describing antibody titer dynamics given history of infection.
  - ➢ A model describing measurement errors.
- Reversible jump MCMC used for inference.
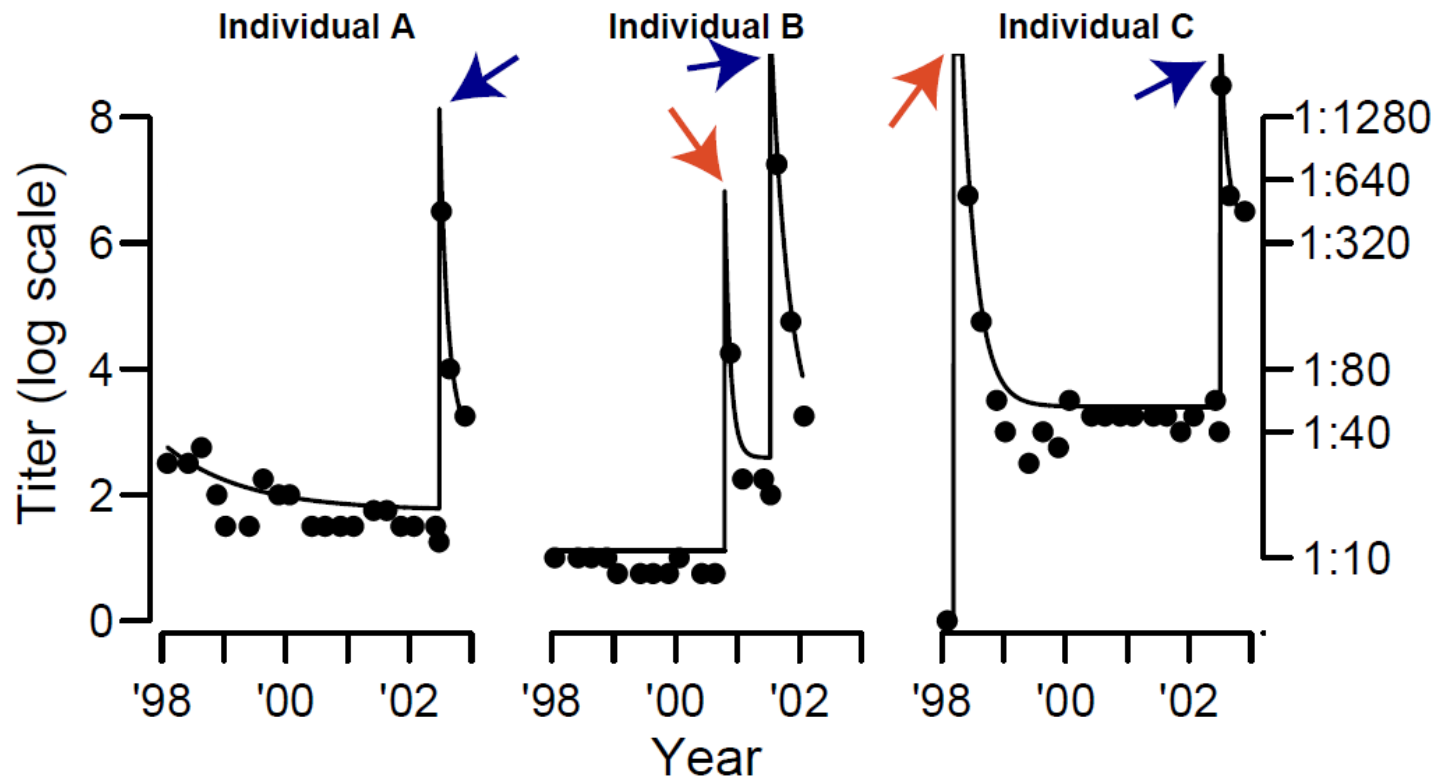
Characterizing subclinical infections

Detected infections:
- DENV1 (green)
- DENV2 (light blue)
- DENV3 (red)
- DENV4 (orange)
- Unknown (black)
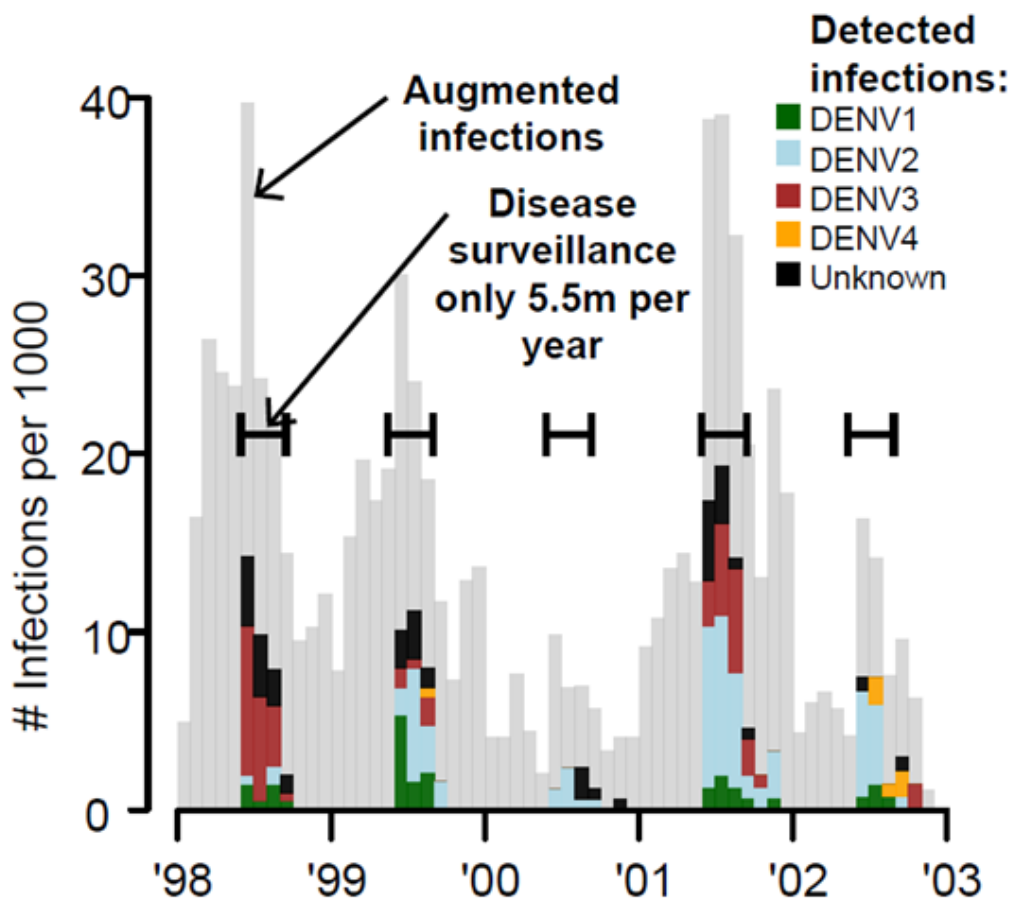
- 65% of infections subclinical.

- 34% of subclinical infections due to serotype 4 while serotype 4 only represents 3% of symptomatic infections.

- Lower risk to develop symptoms for serotype 4.

Institut Pasteur

Antibody titer dynamics
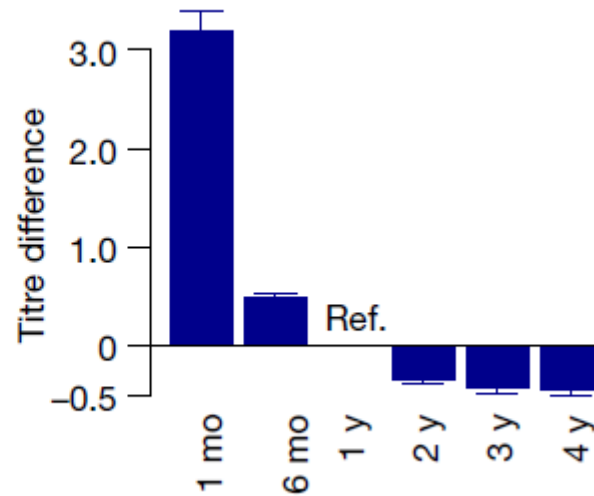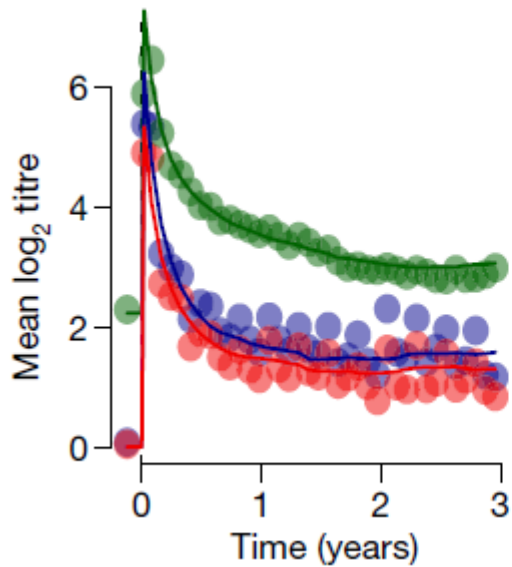
Primary infection, infecting serotype

Primary infection, non-infecting serotype

Post-primary infection
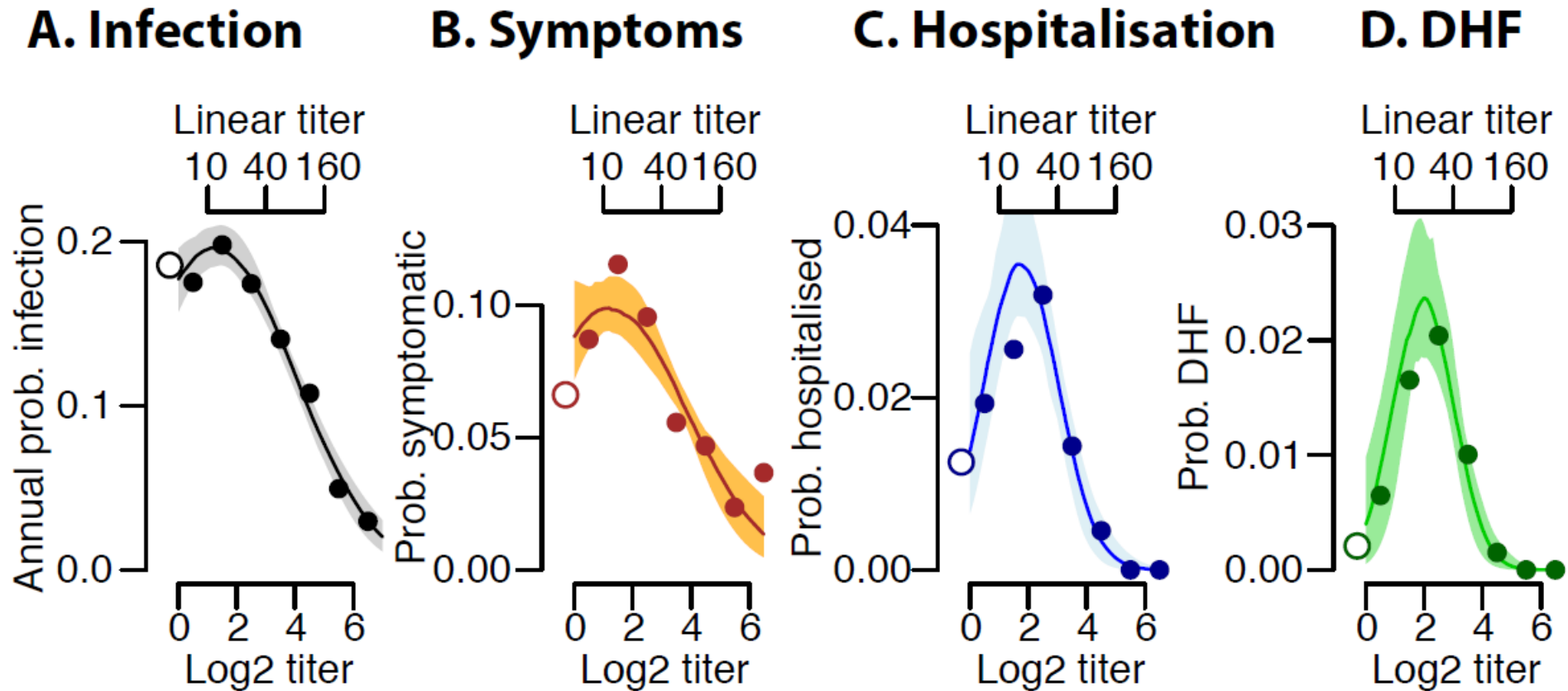
Correlates of protection and of severity

Conclusions

- Household transmission studies:
  - Within household transmission vs community transmission.
  - Many other insights…
    - Flu in France, US...
- Progressively added more features to these analyses:
  - Missing data problems
    - Lost individual records.
  - More complex social structures & space
    - School settings.
    - Community outbreaks.
  - Repeated measurements
    - Dengue cohort.

- Stricking similarities in some of the fundamental parameters that govern spread.
- But nonetheless, transmission rates are highly variable:
  - Social structures, contact networks, distance;
  - Spatial heterogeneities.
- Further developments:
  - Inclusion of contact data;
  - Model comparison.

Institut Pasteur

Acknowledgements

- **Institut Pasteur (France):** H. Salje, A. Cousien, Q. Ten Bosch, A. Andronico.
- **Imperial College London (UK):** N. Ferguson, C. Donnelly, A. Ghani.
- **Big Data Institute (UK):** C. Fraser.
- **CDC (USA):** L. Finelly, D. Swerdlow, C. Reed, A. Bhattarai.
- **Pennsylvania Department of Health (USA):** S. Ostroff
- **ICDDR,B (Bangladesh):** Emily Gurley, KK Paul.
- **Standford University (USA):** SP Luby.
- **WHO:** M. Van Kerkhove.
- **Johns Hopkins University (USA):** J. Lessler.
- **University of Florida (USA):** D. Cummings.
- **INSERM (France):** F. Carrat, P.Y. Boëlle, G. Thomas, A.J. Valleron.
- **Oxford University (UK):** P. Horby, A. Fox.

Thank you!

Institut Pasteur