Imperial College London

MRC Medical Research Council

Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza

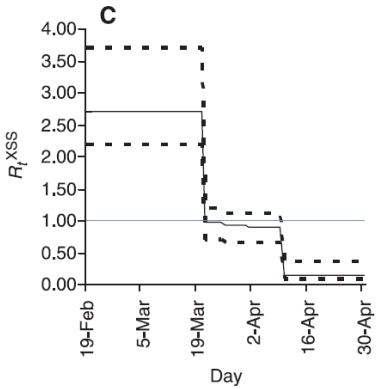
Simon Cauchemez
MRC Centre for Outbreak Analysis and Modelling
Imperial College London

Statistical motivation (1)

- Fitting parametric mechanistic models (e.g. SIR or more complex) :
  - ➢ Likelihood-based inference in the context of missing data:
    - ✓ Transmission process imperfectly observed (e.g. times of infection unknown);
    - ✓ Likelihood difficult to write down;
    - ✓ Data augmentation strategy [O'Neill et al, JRSS C, 2000; O'Neill & Roberts, JRSS A, 1999].
  - ➢ Limitations: same as fitting of parametric model in other fields - e.g. need to pre-define time intervals on which transmission rates are constant.
- Reconstruct transmission tree & derive summary stat. [Wallinga and Teunis, AJE, 2004]:
  - ➢ Less assumptions, greater flexibility, at the cost of larger variance of estimates;
  - ➢ Ignore data on uninfected individuals – can't inform on transmission risk factors.

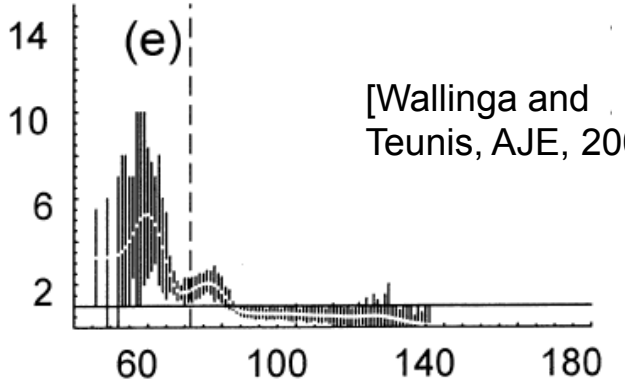
Reproduction number of SARS

Parametric approach



[Riley et al, Science, 2003]

Tree-reconstruction approach



[Wallinga and Teunis, AJE, 2004]

Imperial College London

Statistical motivation (2)

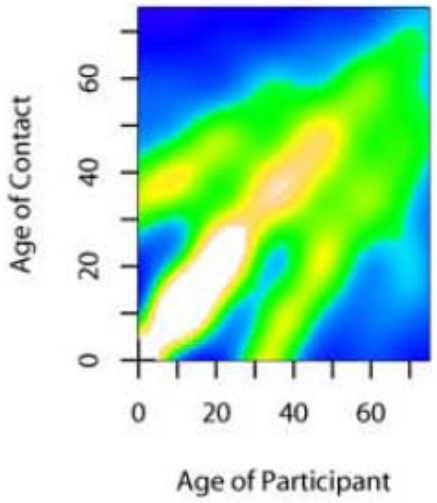
MRC Medical Research Council

- In this talk:
  - Present a single framework that integrate fitting of mechanistic model and reconstruction of transmission tree in a coherent way.
  - Estimating parameters of mechanistic model from the data is not the end of the journey – much more we can learn by mixing the 2 approaches!
  - The two approaches are complementary.

Epidemiological motivations

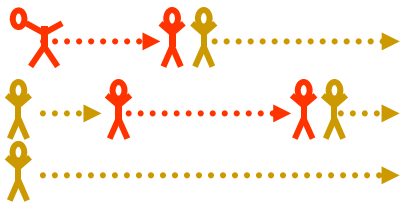
- In early May 2009, CDC received reports of a rapid-onset H1N1 outbreak in a semi-rural elementary school in Pennsylvania.
- Fast decision to investigate the outbreak.
- Very detailed investigation – rare opportunity to better understand how population structures and social networks affect influenza transmission.

Contact data



[Mossong et al, Plos Med, 2008]

Household transmission data



[Cauchemez et al, Stat Med, 2004]

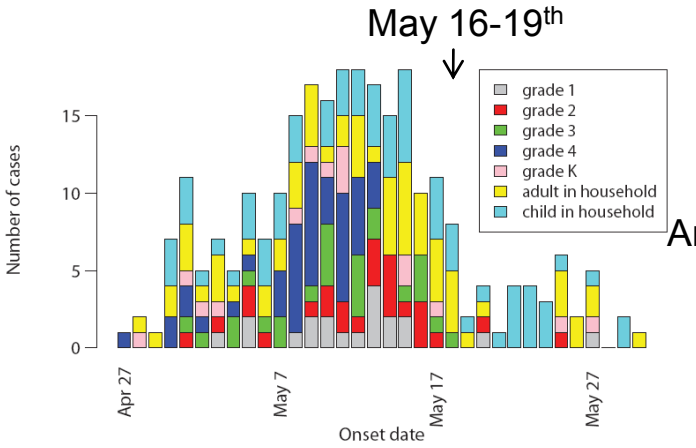
H1N1pdm community outbreak in Pennsylvania

- Demographic & clinical information collected on students and their family members – 2 phone interviews:

-141 ARI cases among 370 students (AR:38%)
- 129 ARI cases among 899 household members (AR:14%)

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

2 interviews of households



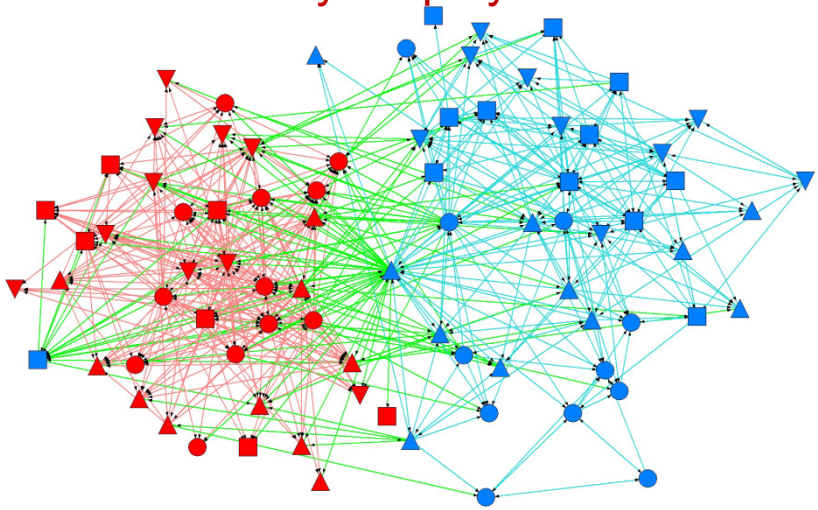
Censoring issue: 40% of subjects were interviewed on May 16-19th only.

Around June 4th

Activities

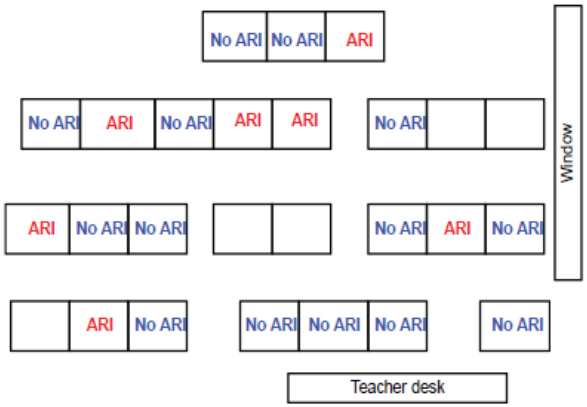
e.g. scientific demonstration with eyeball

Social networks - Who are your playmates?

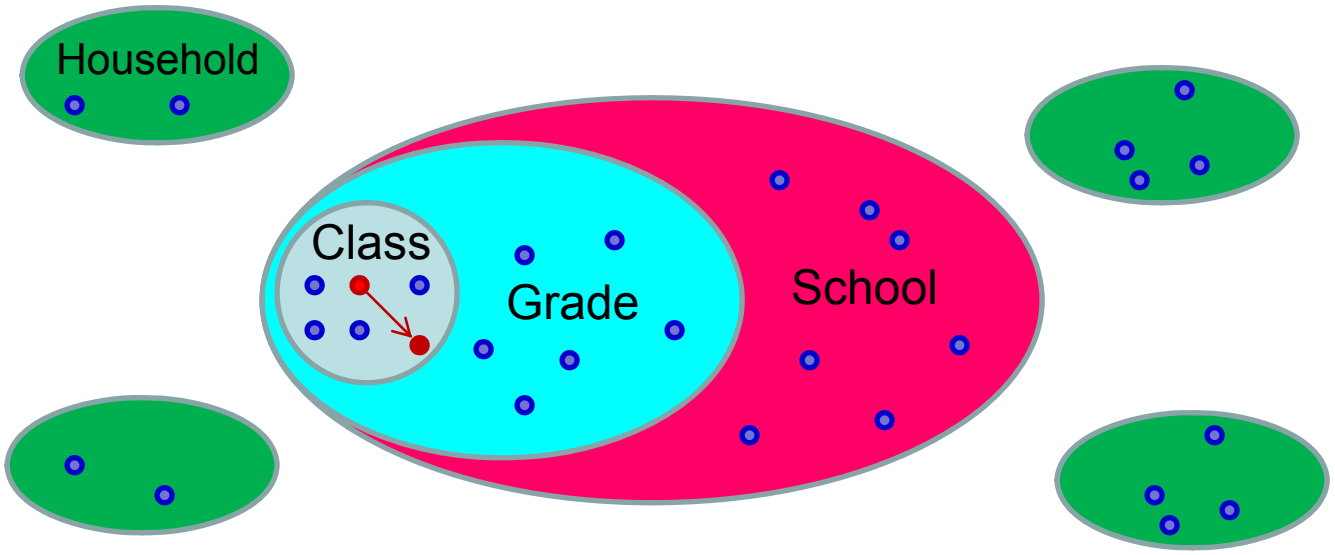


Girls Boys

Seating charts

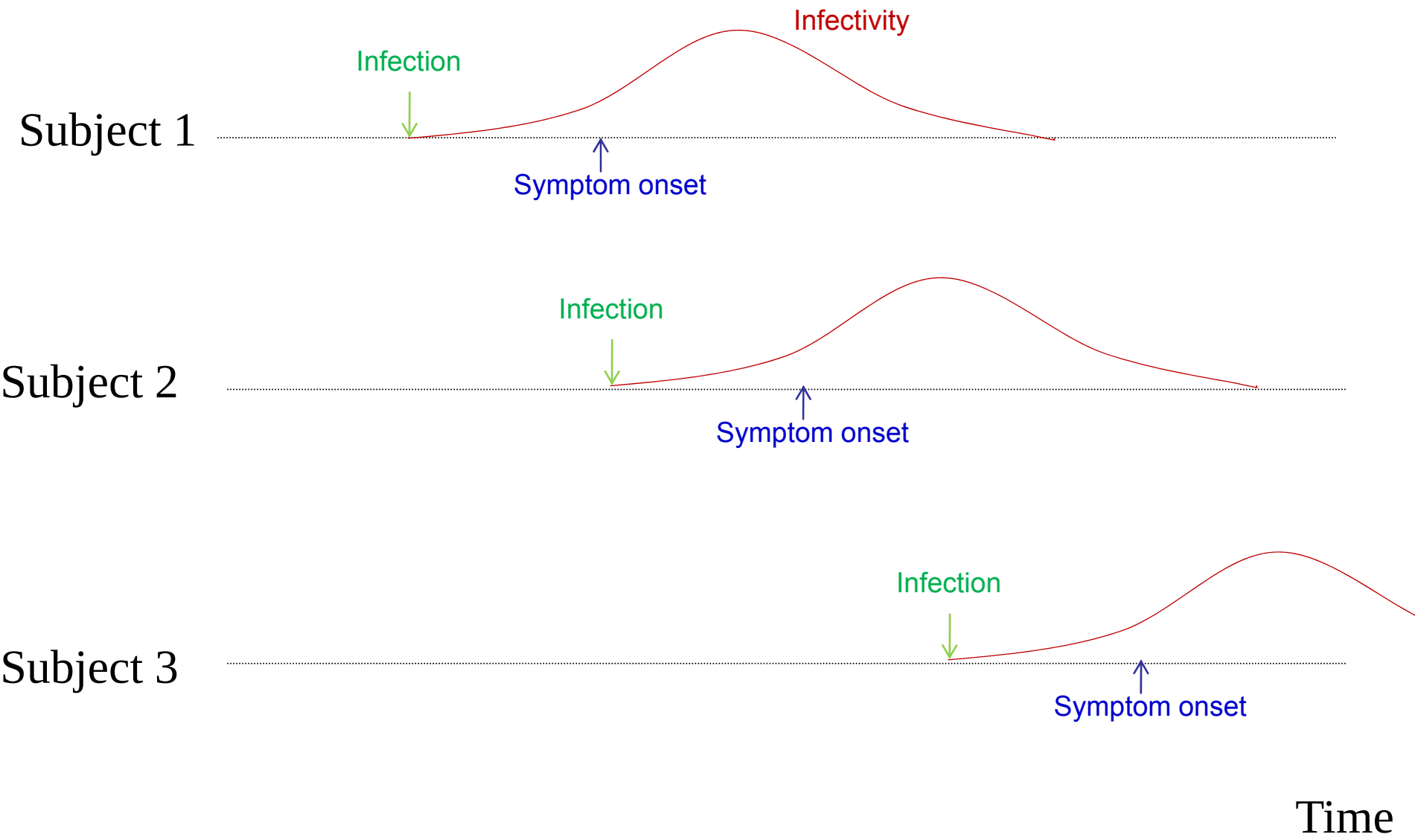


Transmission model



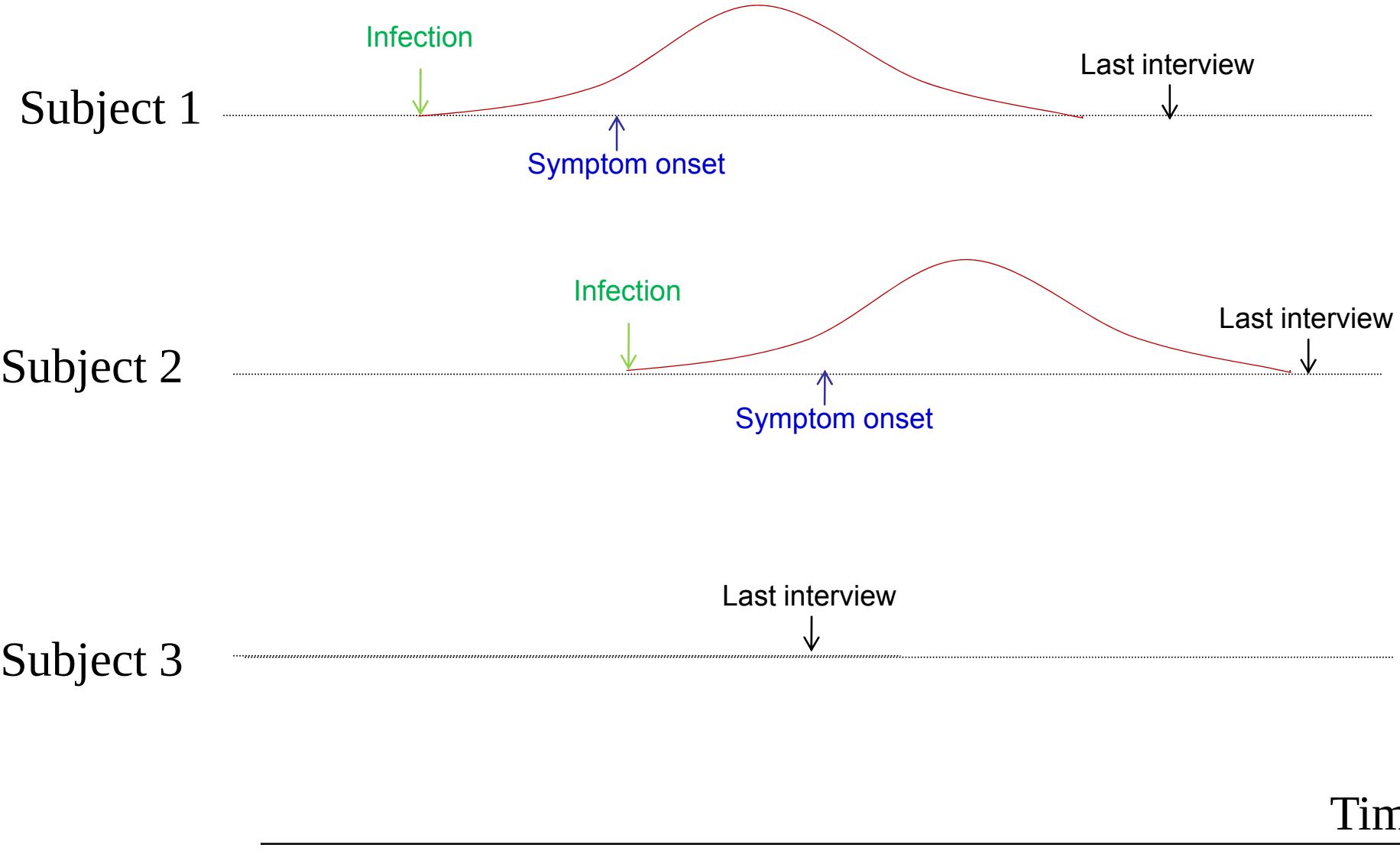
- Model **person to person transmission rates**.
  - Partition of pairs of individuals. E.g. 2 students in the same class are also in the same grade and school; But pair is classified as “classmates”.
  - Hierarchy: Household > Class > Grade > School > Community [i.e. between individuals of the dataset that are not from the same household nor from the school].
- **Infectivity profile** may be different in the school/community and in the household (for adults and for children).

Inference without missing data



Inference from the observed data

Augment the data with quantities needed to easily write down the likelihood



Imperial College London

MRC Medical Research Council

Data augmentation: general framework

- Notations:
  - Y: observed data – e.g. dates of symptom onset;
  - Z: "missing" (augmented) data – e.g. dates of infection, whether or not individual was a case, date of onset when missing;
  - θ: parameters

- Three-level hierarchical model:

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

Observation level: ensures that augmented data consistent with observed data

Transmission level: describes the latent transmission process

Prior level

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

Imperial College London

MRC Medical Research Council

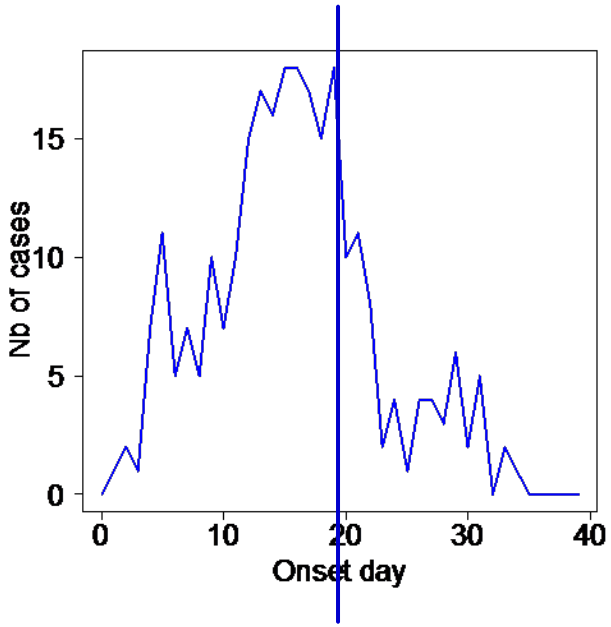
Observation level P(Y|Z) in the PA analysis

- Y=data [date of onset, household id, class id, etc]
- Z=augmented data [date of infection, whether or not individual was a case, time of onset when missing etc]

- Assumed distribution of the incubation period: mean 1.5 days & variance 0.3 days [Lessler et al, NEJM, 2009; Ghani et al, Plos Currents Influenza, 2009; Moser et al, AJE, 1979].
- Individuals with no symptom onset observed prior to last interview can have onset after last interview.
- Individuals with missing diagnosis can be infected during the outbreak.

- Analysis based on a clinical diagnoses of Acute Respiratory Illness – Potential asymptomatic cases are not accounted for.

40% of people – no information after ~ day 20



Imperial College London

MCMC output

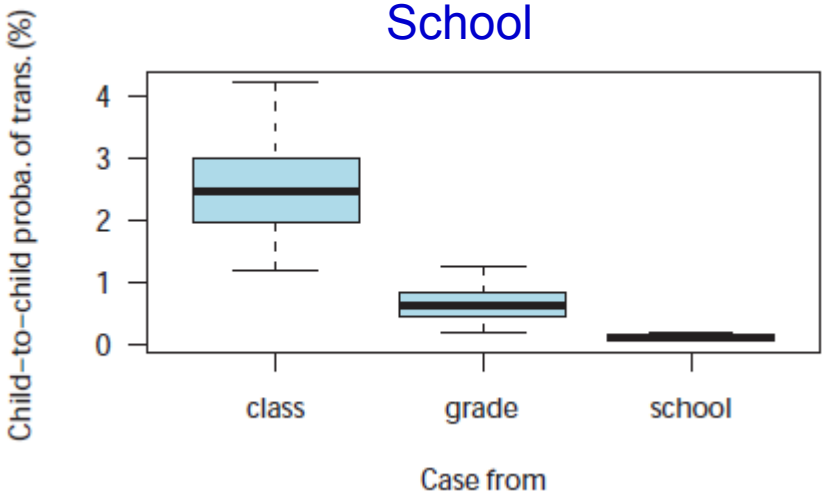
MRC Medical Research Council

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

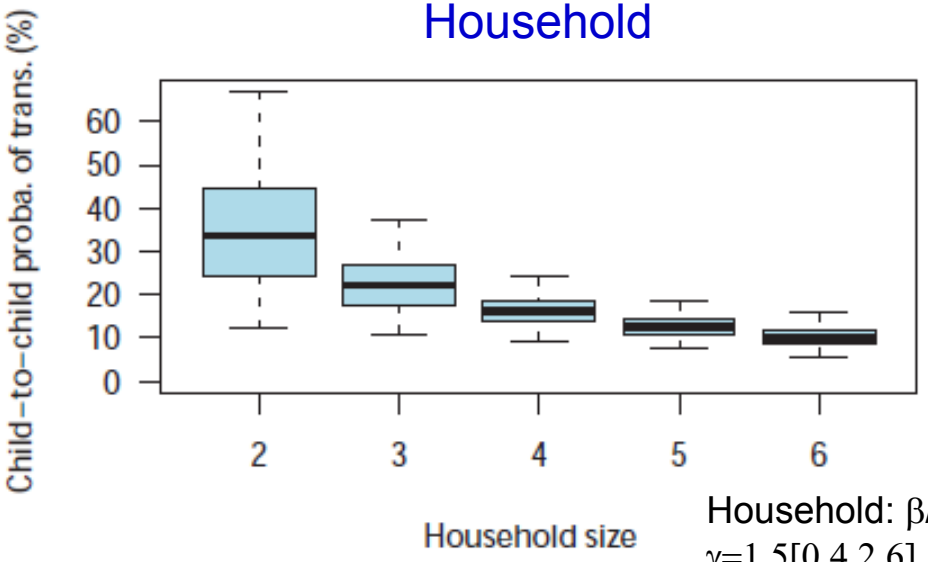
	Parameter	Augmented data for each individual m=1…M	
		Was a case	Date of infection
Iteration i=1…N	θ_1	$C_i^m = 1$	$T_i^m = 25$
	θ_i	$C_i^m = 0$	$T_i^m = NA$
	θ_N	$C_i^m = 1$	$T_i^m = 23$

Transmission rates in the school and in the household

School

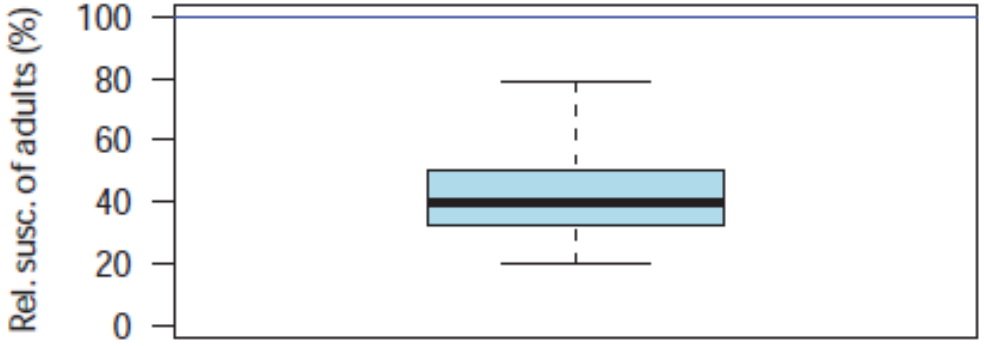


Household



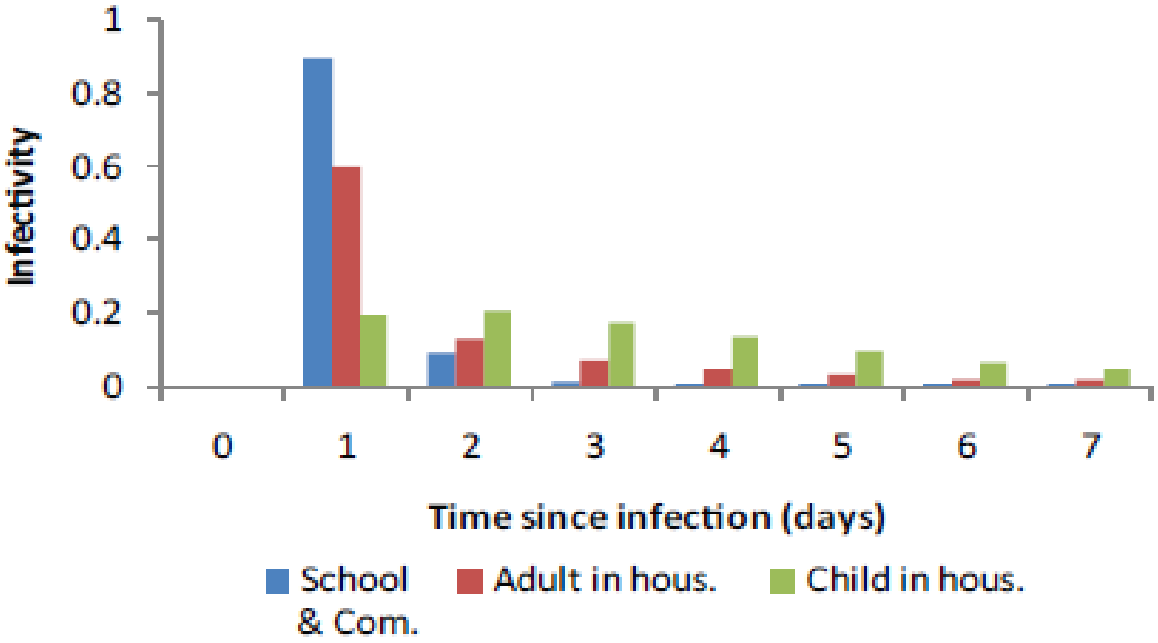
Household: β/n^{γ}
γ=1.5[0.4,2.6]

Age dependent susceptibility



No evidence of children being more infectious than adults

Infectivity profile & generation time



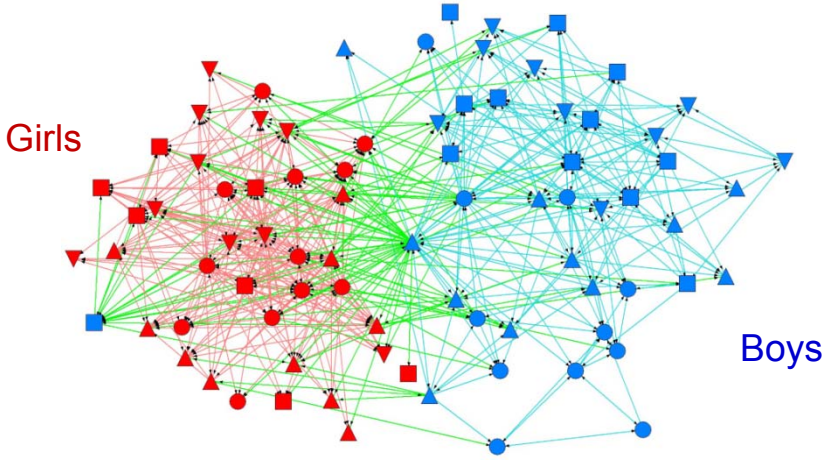
Generation time - average time between infection of a case and infection of the persons they infect:

- Household:
  - ✓ more than 10 year old: 2.3 (1.4, 3.6) days
  - ✓ 10 yr or less: 3.6 (2.2, 5.0) days

- School and community: 1.2 (1.0, 1.7) days

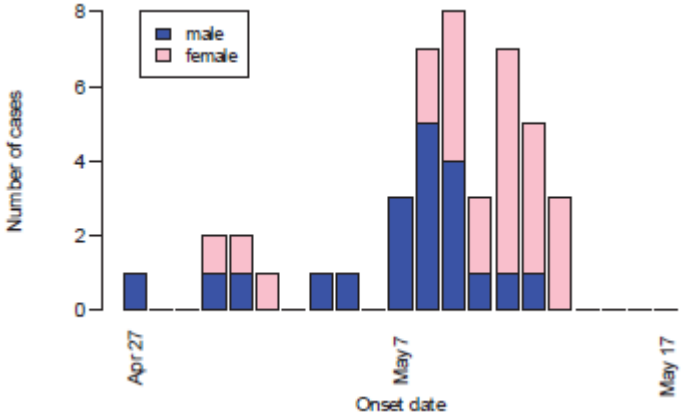
Gender-related mixing & transmission patterns

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

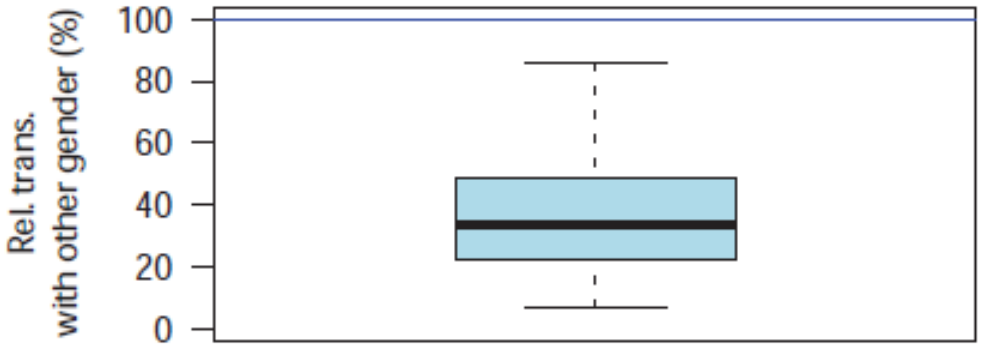


Evidence that this aspect of mixing patterns affected the transmission dynamics

i) Boys had onset before girls! (p=0.023)



ii) BF for model with gender-effect: 8.0



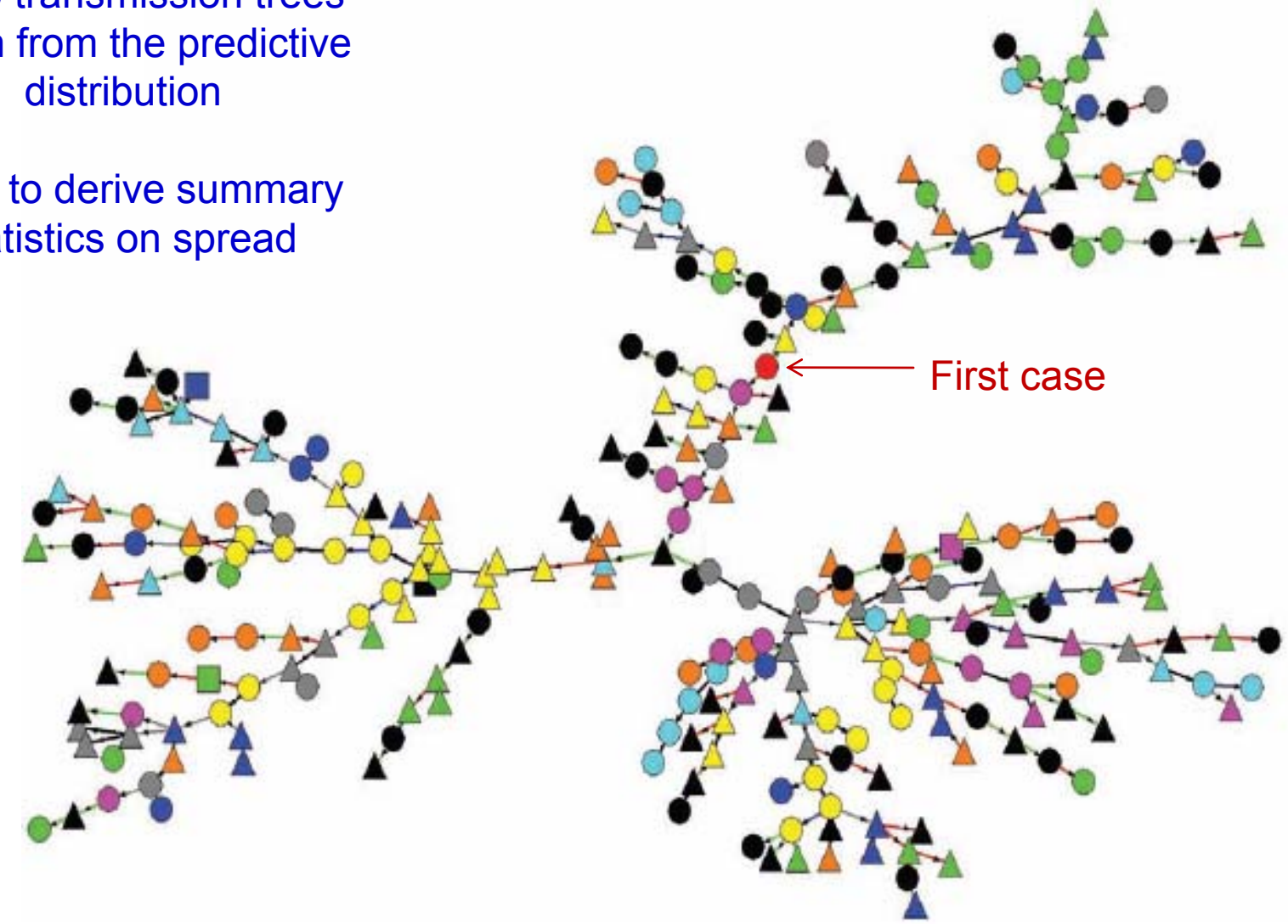
Imperial College London

MRC Medical Research Council

Reconstructing the transmission tree

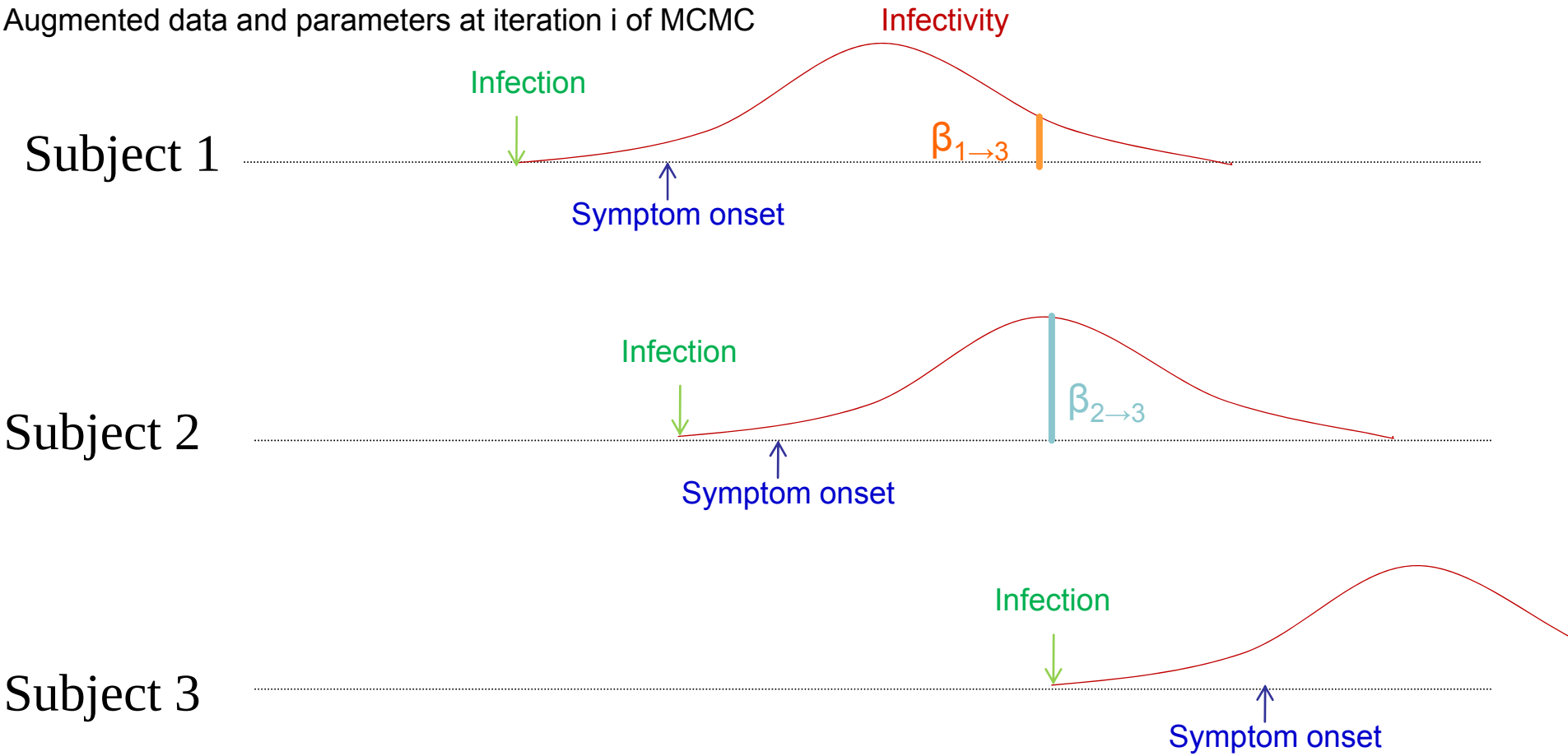
2,500 transmission trees drawn from the predictive distribution

Used to derive summary statistics on spread



Reconstructing the transmission tree from the posterior distribution

Augmented data and parameters at iteration i of MCMC



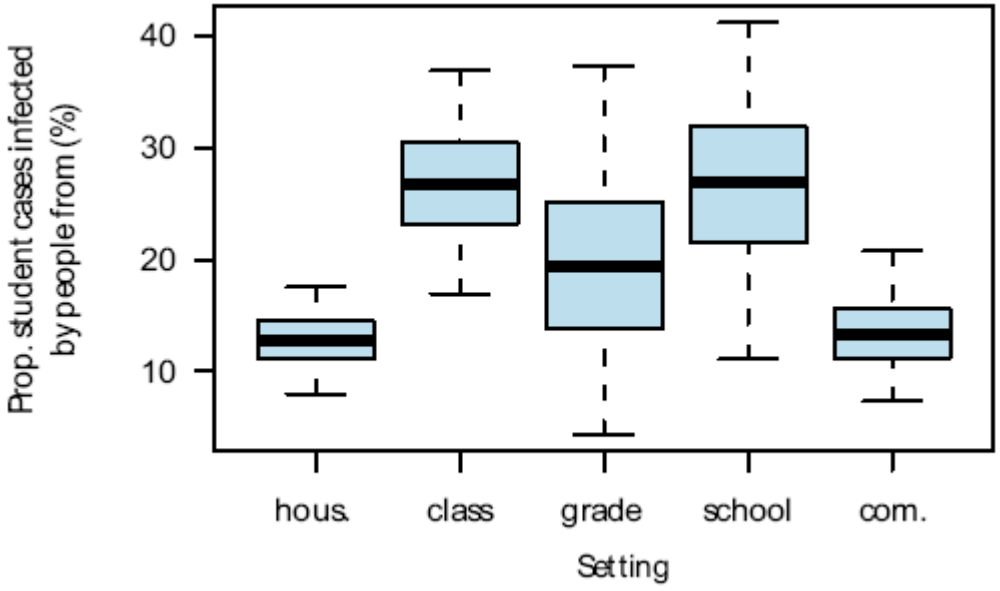
Conditional on this realisation, subject 3 was infected by subject 1 with probability:

$$\beta_{1\to3}/(\beta_{1\to3}+\beta_{2\to3})$$

Imperial College London

MRC Medical Research Council

Student cases were infected by



74% (65%, 82%) of student cases infected by other students

Imperial College London

Role of 6-18 yr old individuals in household spread

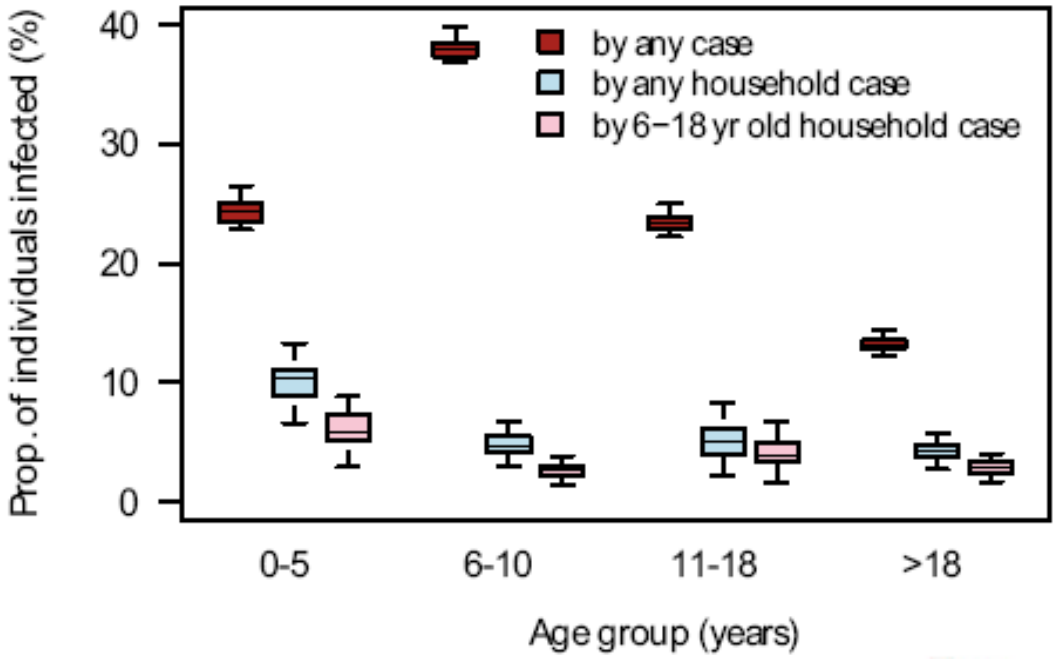
MRC Medical Research Council

- 6-18 yr old individuals facilitate the introduction of influenza in household

Relative risk that a case was the "introducer" (reference group: >18 yr):

- 0-5 yr: 1.02 (0.85, 1.23)
- 6-10 yr: 1.83 (1.60, 2.11)
- 11-18 yr: 1.20 (1.00, 1.41)

- Only 1 in 5 cases aged >18 years was infected by a 6-18 year old household member

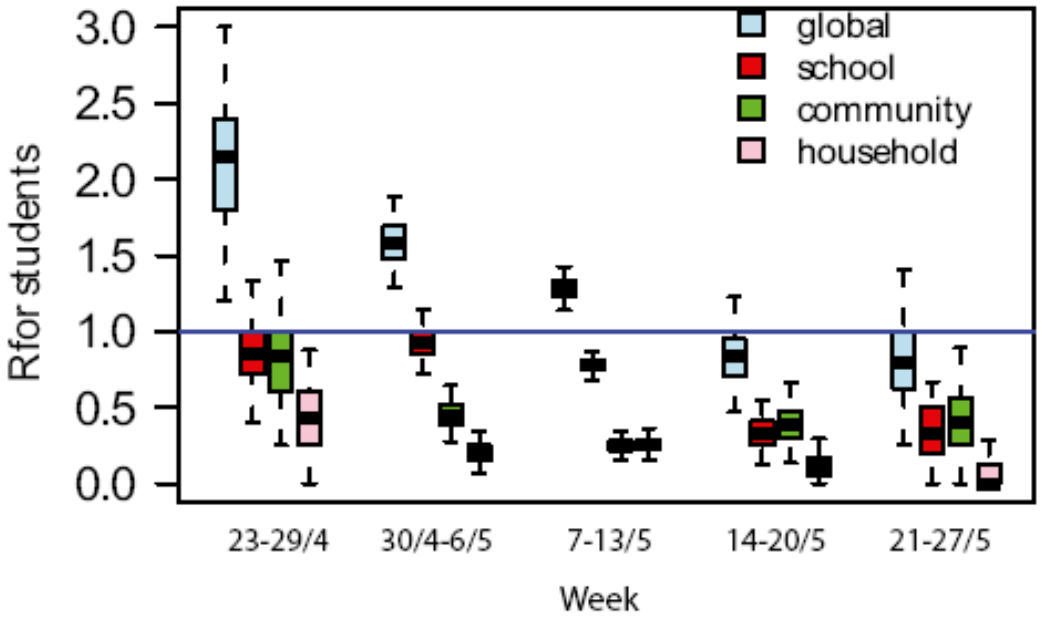


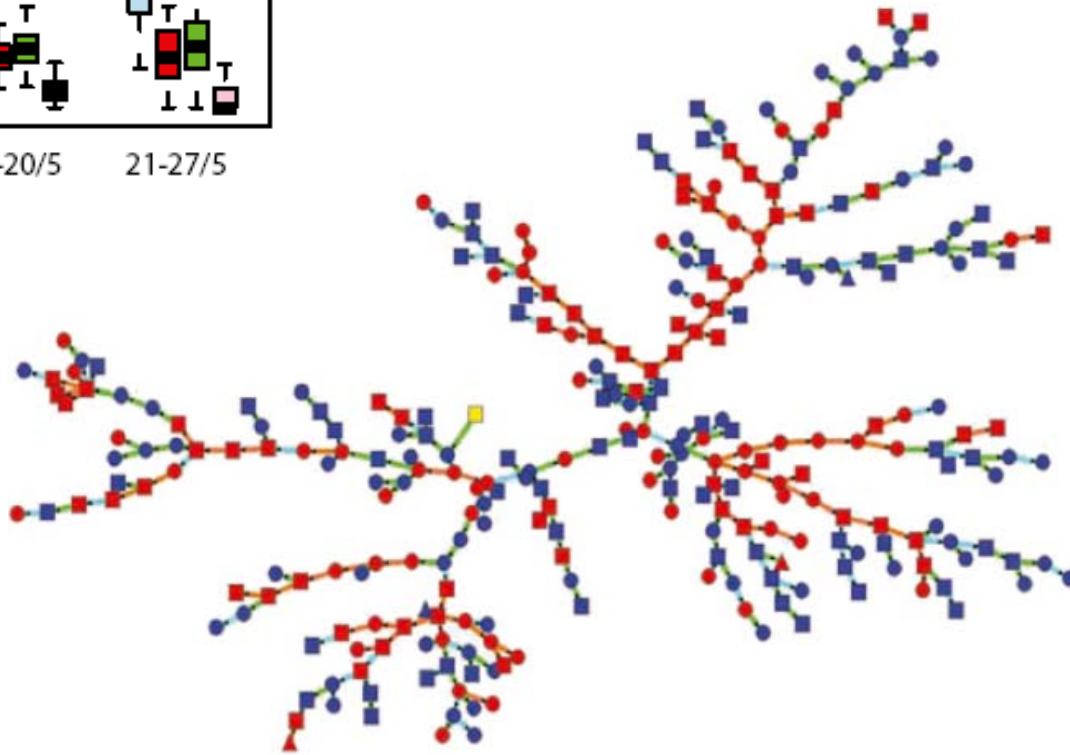
(more than 40% of cases aged >18 years were the first or only case of their household)

Imperial College London

Low estimates of global R for flu – implications for spread in places...

MRC Medical Research Council

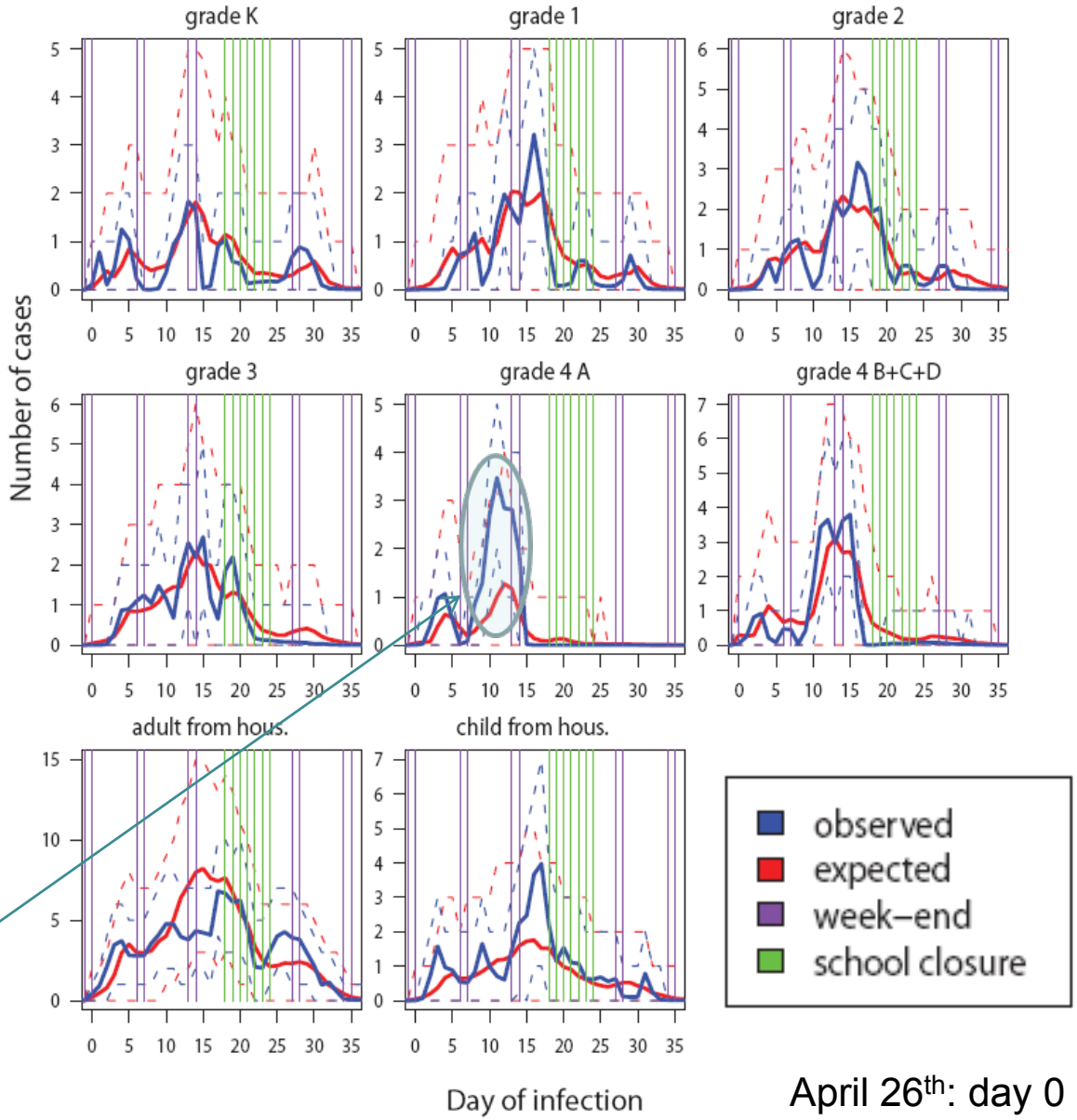




- For an outbreak to be sustained in a place, multiple introductions and/or SSE events are to be expected.
- Strong between-place interactions with back and forth waves of transmission between places.

Detecting atypical transmission events

- We consider the model where transmission rates do not change with time. Does this model satisfyingly fit the data?
- Blue - data: "reconstructed" number of infections occurring on each day and for different groups of individuals.
- Red - "next step ahead" predictions: given what has been observed up to day t-1, how many cases are predicted by the model on day t?
- Relatively good fit in general…

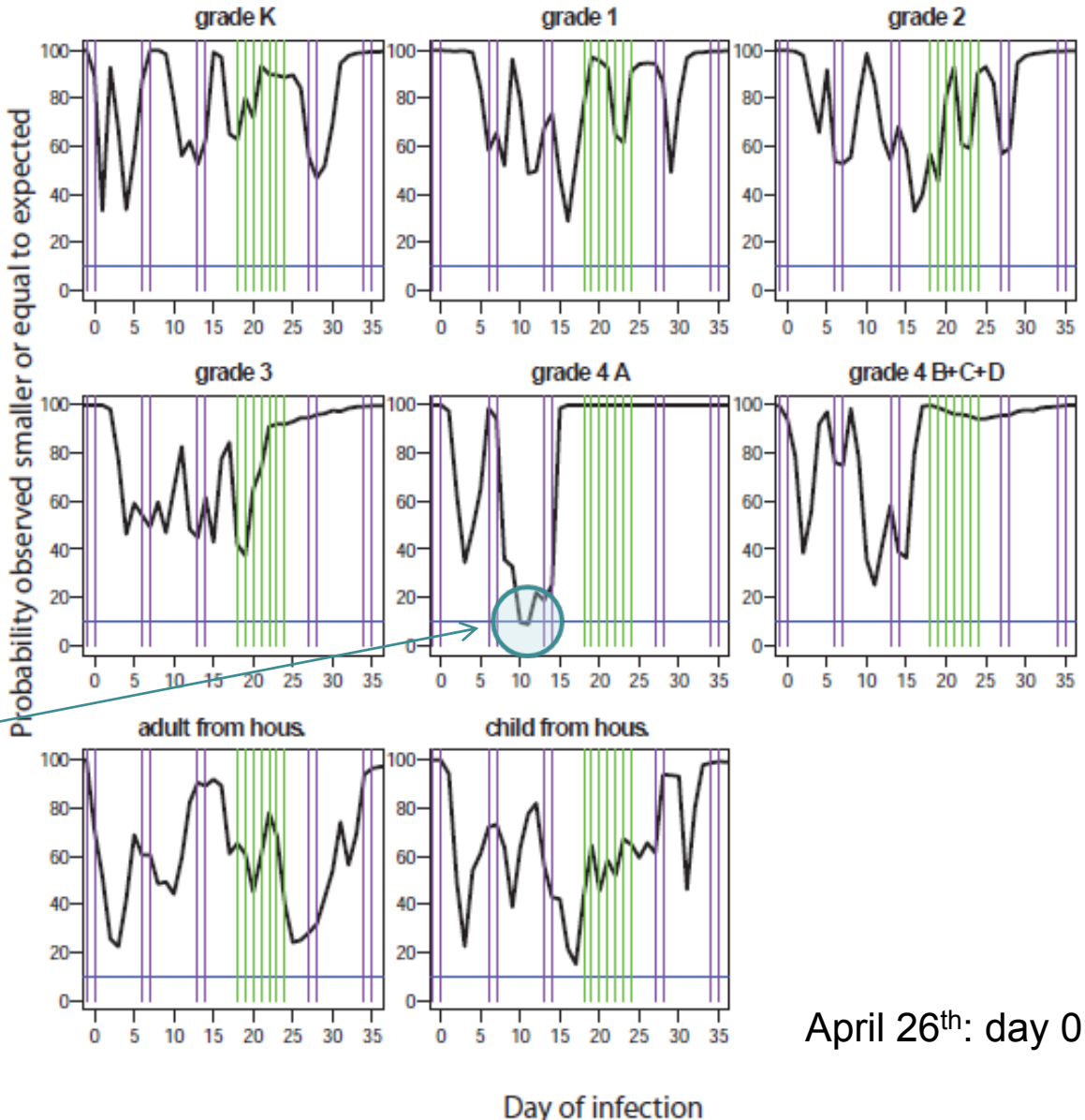


Abnormally high transmission around May 6th-7th?

April 26th: day 0

Posterior probability
Reconstructed ≤ Expected

- We can then compute the posterior probability that reconstructed numbers are smaller or equal to expected ones.



May 6th-7th Among 4th graders, classroom A: Only group/time point with a posterior probability smaller than 10%

April 26th: day 0

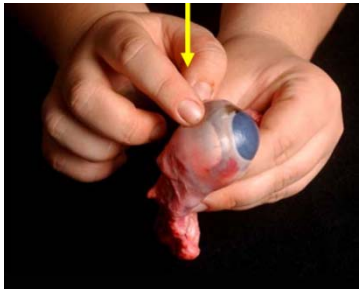
Imperial College London

May 6th-7th : activities among 4th graders… (1)

MRC Medical Research Council

Potential mechanism for super-spread

- May 6th:
  - ➢ Science demonstration with a cow eyeball in each 4th grade classroom; ← Within-class trans. in all class
- May 7th:
  - ➢4th grade class A had a physical education class; ← Within-class trans. in 1 class
  - ➢ All 4th graders congregated at a 1-day special event. ← Between-class trans.



10 models of super-spread, each with prior probability 10%

		Between-class SSE	
		no between-class SSE	between-class SSE on May 7th
Within-class SSE	no within-class SSE	10%	10%
	on May 6th, class A	10%	10%
	on May 6th, all class	10%	10%
	on May 7th, class A	10%	10%
	on May 7th, all class	10%	10%

May 6th-7th : activities among 4th graders…

Posterior probabilities estimated via Reversible Jump MCMC

		no between-class SSE	between-class SSE on May 7th
Posterior proba			
	no within-class SSE	0.4%	2.9%
	on May 6th, class A	4.1%	20.0%
Within-class SSE	on May 6th, all class	3.3%	15.6%
	on May 7th, class A	6.2%	24.9%
	on May 7th, all class	5.1%	17.6%

BF for SSE: 28 (prior:10%)

Posterior proba for between-class SSE: 81% for a prior of 50%(BF: 4.3)

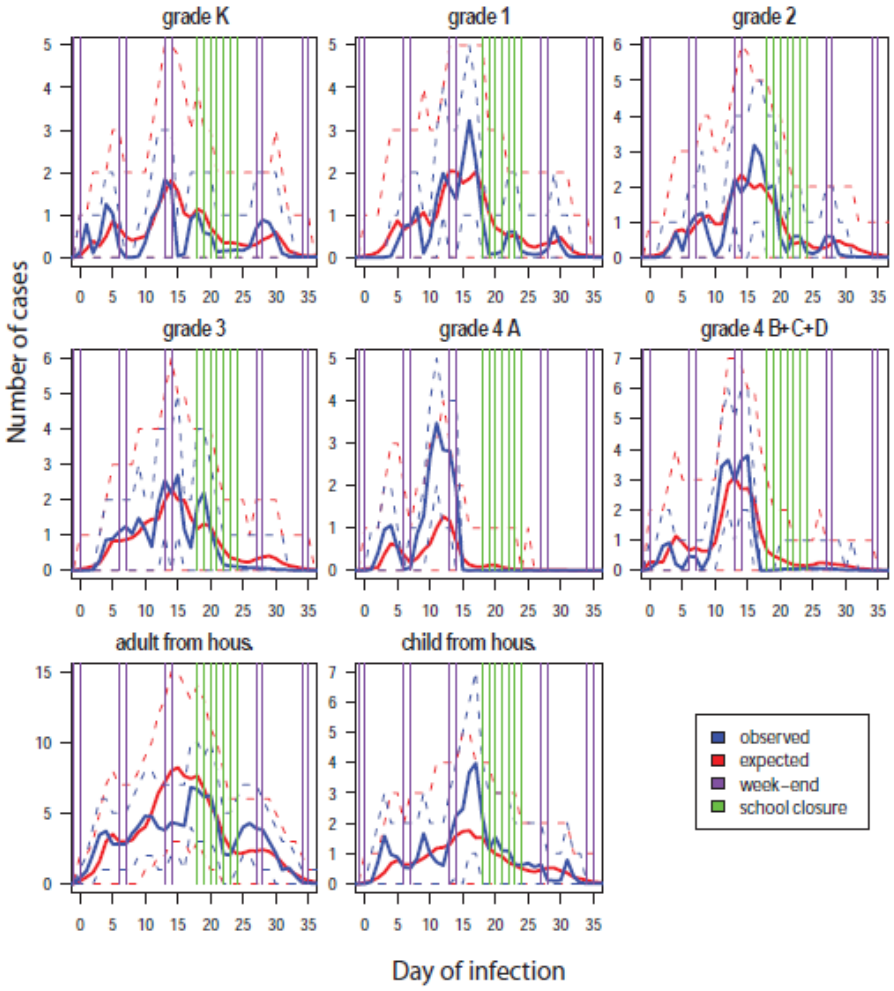
Posterior proba that within class transmission increased on May 6th or May 7th: 97% for a prior of 80%(BF: 6.3)

Imperial College London

Impact of a late 1-week school closure

MRC Medical Research Council

Late closure: 27% of students had already had symptoms at time of closure



No noticeable reduction detected

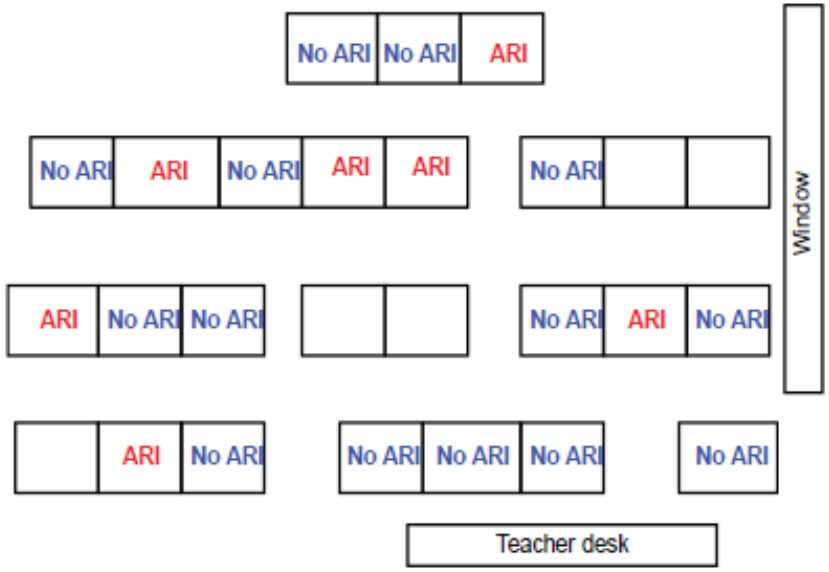
i) No “abnormally low” transmission rates detected during closure

ii) Estimate of transmission rates between students after closure relative to before closure: 0.8 (0.4,1.4)

iii) School reproduction number
- on the week of the closure: 0.3 (0.1,0.6)
- on the following week: 0.3 (0.0, 0.7)

Seating plan: higher risk of infection if neighbour is sick? (1)

- Simple permutation test:
  - Count the number of pairs of neighbours that are sick in each class room.
  - Compare with the number of pairs that would be expected if the seating plan had no impact on transmission. This expected distribution can be obtained by randomly permutating the seats of the students.
- Observed numbers of pairs are not significantly different than what would be generated by chance.



	Observed	Expected
Overall 4th graders	19	20 [16,24] (p=0.58)
Classroom A	12	13 [11,15] (p=0.62)
Classroom B	2	2 [0,4] (p=0.23)
Classroom C	1	2 [0,3] (p=0.53)
Classroom D	4	1 [4,6] (p=0.25)

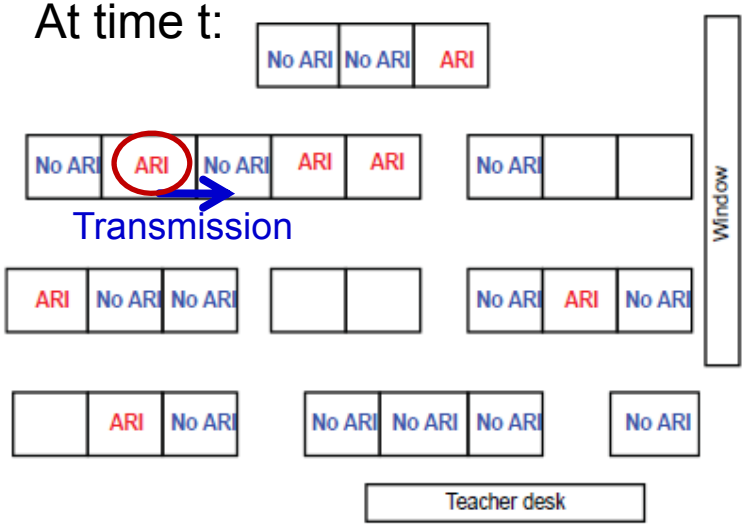
But this approach does not correct for the effect of other covariates (e.g. gender), does not use data on times of symptom onset…

Seating plan: higher risk of infection if neighbour is sick? (2)

- Why not include seating plan data in full transmission model?
  - ➢ Data only available in the 1 grade (out of 5) in which transmission was the highest.
- Alternative approach:
  - ➢ Assume seating chart has no impact on the risk of transmission & draw transmission trees from predictive distribution;
  - ➢ From tree, we can directly compute:

  Proportion of between-classmate transmission that occur between neighbours: 8% [0%,19%]
  - ➢ Is this proportion higher than expected?

At time t:



✓ 1 transmission in class;
✓ Infector known & N=14 susceptibles;
✓ If seating chart has no impact, each susceptible has the same probability of infection 1/N

✓ We generate an alternative infectee with probability 1/N

Expected proportion of between-classmate transmission that occur between neighbours: 7% [0%,19%]

Imperial College London

Conclusions on H1N1pdm transmission

MRC Medical Research Council

- Transmission risk factors
  - Structuring of school in classes and grades has strong impact on spread;
  - Seating next to a case did not significantly increase risk of transmission;
  - Gender-related mixing pattern affected spread.

- SSE and school closure
  - Suggestion of abnormally high transmission among 4th graders;
  - Late school closure had no noticeable impact on spread.

- Generation times
  - Shorter in the school than in the household;
  - Longer for those aged ≤10 years than for others.

- School-aged individuals
  - Facilitated the introduction of influenza in households;
  - But only 1 in 5 cases aged >18 yrs was infected by a school-aged household member.

Imperial College London

MRC Medical Research Council

Conclusion on methodological aspects

- Data augmentation:
  - ➢ Gold standard method to deal with missing data in detailed yet relatively small epidemic datasets (households, outbreak investigations).
  - ➢ But methodological developments are needed on model comparison. What are the alternative to Reversible Jump between models?
  - ➢ Need alternative methods for large datasets, time series etc.

- Integrating parametric modelling and tree reconstruction:
  - ➢ Fitting mechanistic model seems to be the only way to:
    - ✓ Account for depletion of susceptibles;
    - ✓ Interpret data on non-cases and estimate transmission risk factors;
    - ✓ Control variance of the estimates.
  - ➢ On the top of that, tree reconstruction methods can provide:
    - ✓ Further insights on what effectively happened during the outbreak;
    - ✓ Summary statistics on who was infected by whom etc;
    - ✓ A framework to detect abnormal features in the data that are not accounted for in the model.

Imperial College London

MRC Medical Research Council

Acknowledgements

Achuyt Bhattarai[2], Tiffany L. Marchbanks[3], Ryan Fagan[2], Stephen Ostroff[3], Neil Ferguson[1], David L. Swerdlow[2]
And the PA H1N1 working group (Samir Sodha[2], Mària E.[3], Frederick J. Angulo[2], Rakhee Palekar[2], W. Roodly Archer[2], Lyn Finelli[2])

[1]: Imperial College London
[2]: CDC
[3]: Pennsylvania Department of Health

Reference

Cauchemez et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *PNAS*; 2011.