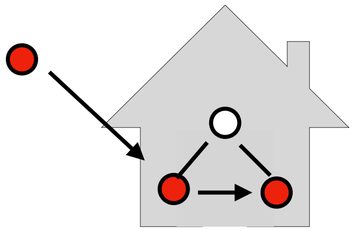
Estimating Household Transmission of SARS-CoV-2



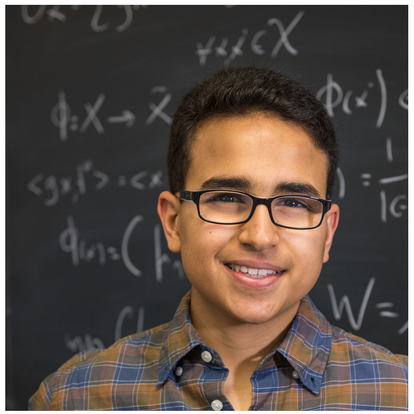
Owain Evans
University of Oxford

"If my housemate/family gets infected, is it inevitable I get infected?"

"How much of all transmission happens at home?"

"What is the risk of infection for essential workers vs. everyone else?"

Collaborators



Andrew Ilyas*
MIT

Mihaela Curmei*
UC Berkeley

Jacob Steinhardt
UC Berkeley

My co-authors on the paper

*Joint first authors

Overview

Assuming there is lockdown/social distancing, how can spread be further reduced? We address:

1. How much transmission takes place in households?
2. How much does household transmission contribute to overall spread?
3. Should policy target essential workers, some other group, or everyone?

Quantifying household transmission: R_h

R = Effective reproduction number (at time t)

= Mean infections due to infected person i

$$R = R_c + R_h$$

$R_c =$ Mean infections due to infected person i outside i's household ("community")

$R_h =$ Mean infections due to infected person i inside i's household

Quantifying household transmission: secondary attack rate

$R_h =$ Mean infections due to infected person i inside i's household

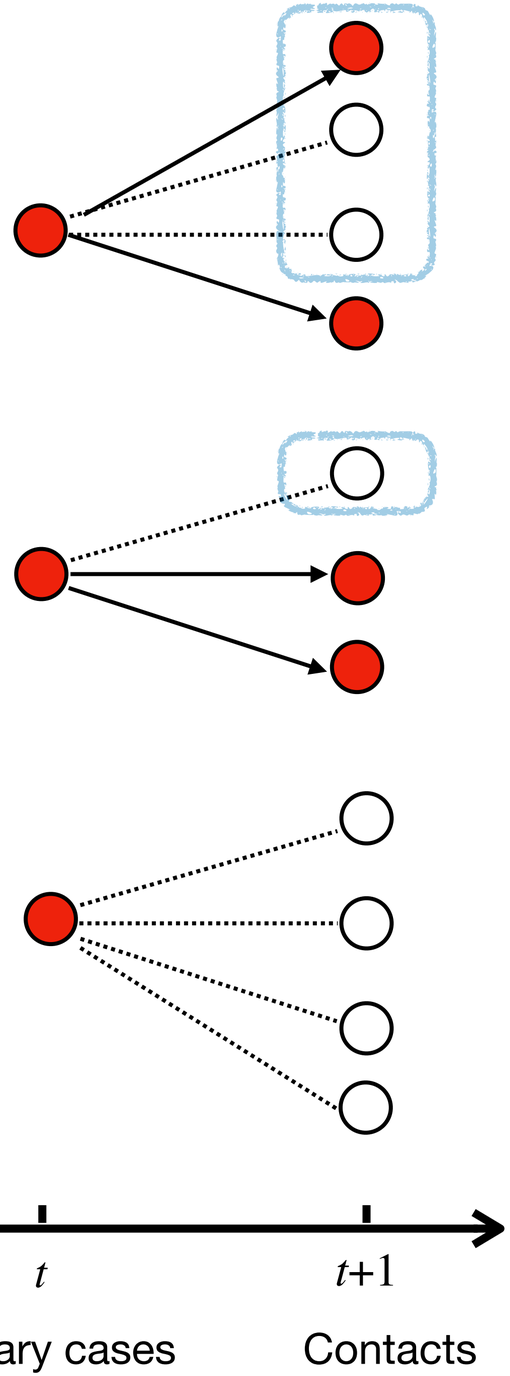
Let i and j be in same household.

SAR = household secondary attack rate

= probability i infects j, given j susceptible

= $P(\ i \rightarrow j\ \mid\ i$ infected, j susceptible)

Blue box shows contacts in same hh as primary case

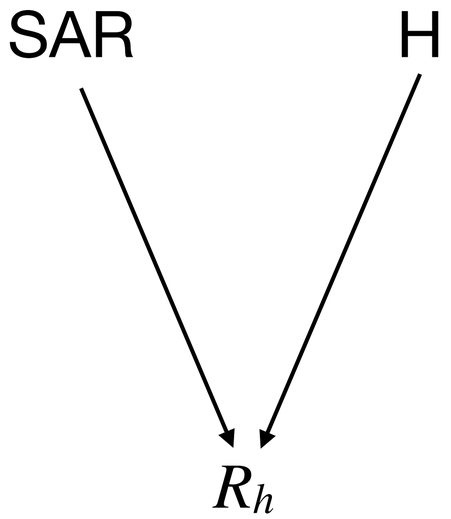


R = # positive contacts / # primary

R_h = # positive hh contacts / # primary

SAR = # positive hh contacts / # hh contacts

Functional relationships



S/N

R_c

H: mean household size

S/N: prevalence in population

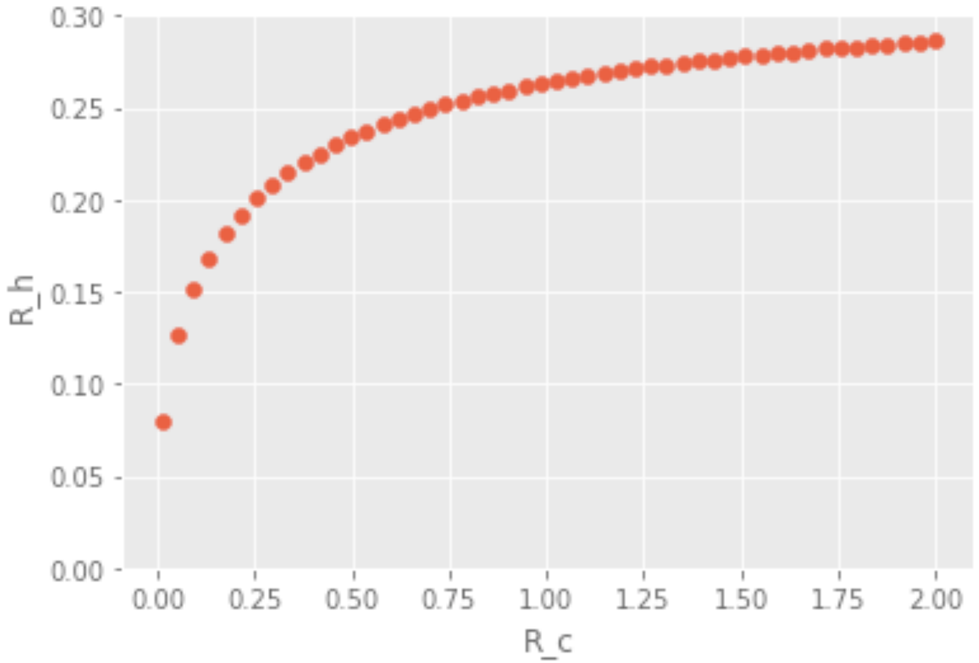
Functional relationships

H = mean household size = 2.5. Let s = SAR.

1. If i infected outside hh, i infects $s(H-1)=1.5s$ people.
2. If i infected inside hh, there's at most $H-2=0.5$ people left to infect!

SAR=0.2	
R_c	R_h
0.4	0.22
0.7	0.25

$$R_h \approx P(inf_c)s(H-1) + P(inf_h)s(1-s)(H-2)$$
$$= \frac{R_c}{R}s(H-1) + \frac{R_h}{R}s(1-s)(H-2)$$
$$\approx 1.2s$$



Conditional risk of infection

$R_h =$ Mean infections due to infected person i inside i's household

Let i and j be in same household.

SAR = $P(\ i \to j\ \mid\ i$ infected, j susceptible)

CRI = conditional risk of infection

= $P($ j infected $\mid$ i infected $)$

CRI allows for $i \to j$ and $j \to i$.

Estimating SAR from data

We found 9 studies of household SAR from China (4), Korea (2), Taiwan, US, and Germany.

Procedure:

- Identify **primary** cases (symptoms/travel + PCR test)
- Check households of primary cases for **secondary cases** (symptoms + PCR test)
- Calculate:

$$\text{SAR} = \#\text{ positive hh contacts} / \#\text{ hh contacts}$$

$$R_h = \#\text{ positive hh contacts} / \#\text{ primary cases}$$

Problems with SAR estimates

Problems with nearly all studies, which we'll correct for:

1. Biased (unrepresentative) sample of primary cases
   - e.g. <10% asymptomatic vs >20% in general
   - under-sample children
2. Failure to detect positive secondary cases
   - PCR test only for symptomatic contacts (some studies)
   - PCR test has 10-50% false-negative rate
3. Household could be infected from outside
   - Bias is probably small

Asymptomatic Infection

Asymptomatic rate (AR): 10-43%

Asymptomatic infectivity:
10-90% of symptomatic infectivity?

Study	AR
Vo', Italy	43%
Gangelt, Germany	22%
Spain, national	25%
Cambridge HCW	28%

Upshot

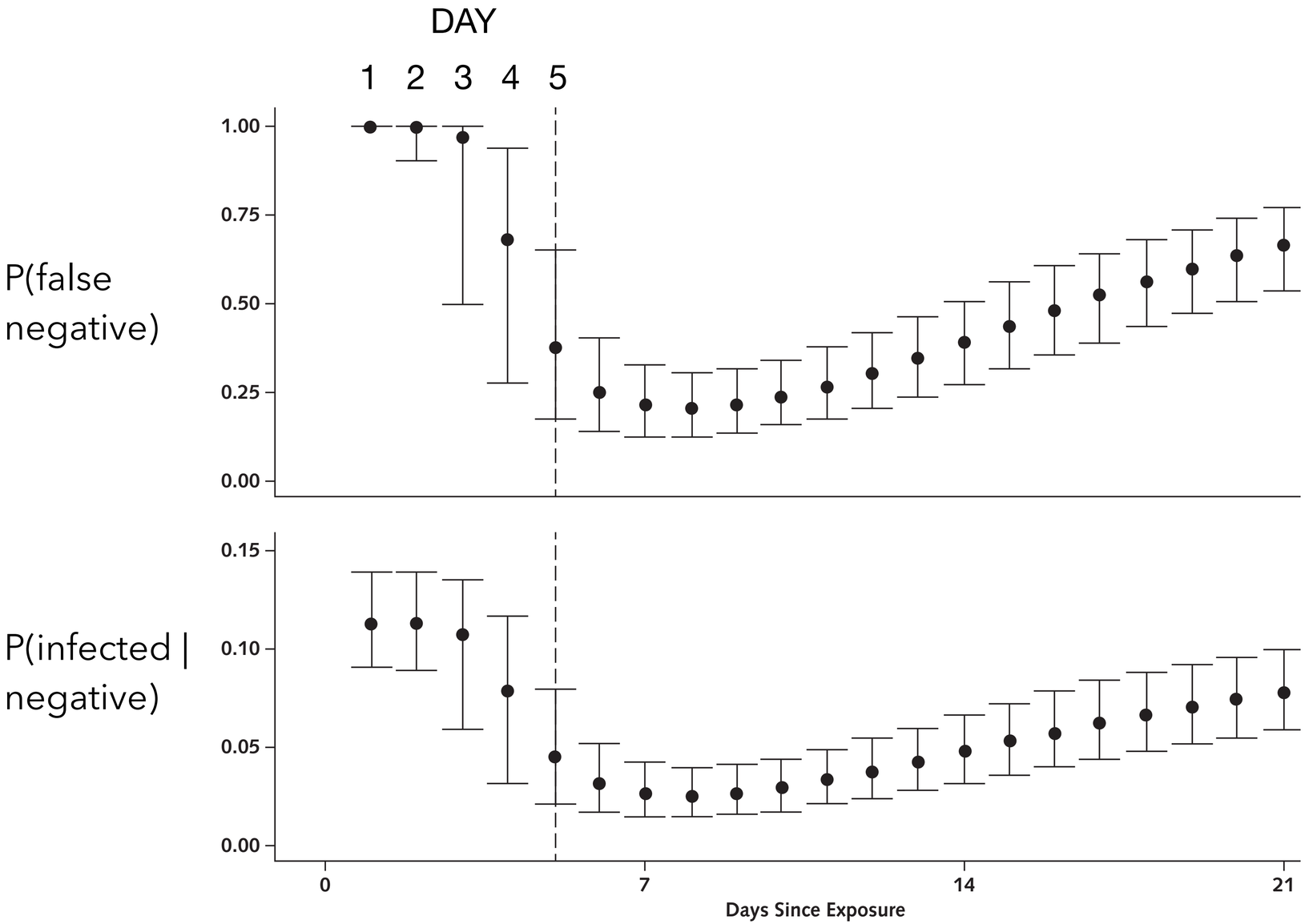
1. Lack of asymptomatics among primary cases

→ overestimate SAR if infectivity lower

2. Lack of asymptomatics among secondary cases

→ underestimate SAR

PCR false negative rate

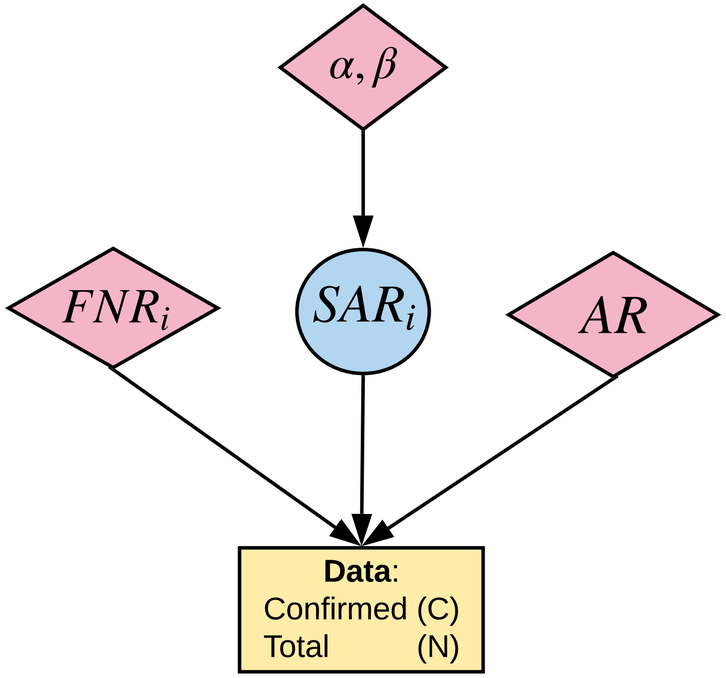


Time period	False-negative rate
Day 4	67%
Day 8	20%
Days 5-15	17-30%

Accuracy varies between swab method, lab, time since infection

Bayesian meta-analysis of SAR

- Goal: pool results from SAR studies to estimate mean SAR and heterogeneity.
- Hierarchical Bayesian random effects model (Bayesian meta-analysis).



Data

$N_i -$ number of household contacts considered in each study

$C_i -$ number of confirmed cases

$\mathbf{1}_{AR_i} -$ indicator; 0 - the study tested asymptomatics, 1 - otherwise

$\mathbf{1}_{FNR_i} -$ indicator; 0 - the study corrected for false negatives, 1 - otherwise

Priors

$FNR_i \sim Uniform(0.15, 0.35)$

$AR \sim Uniform(0.18, 0.43)$

$SAR_i \sim Beta(\alpha, \beta)$

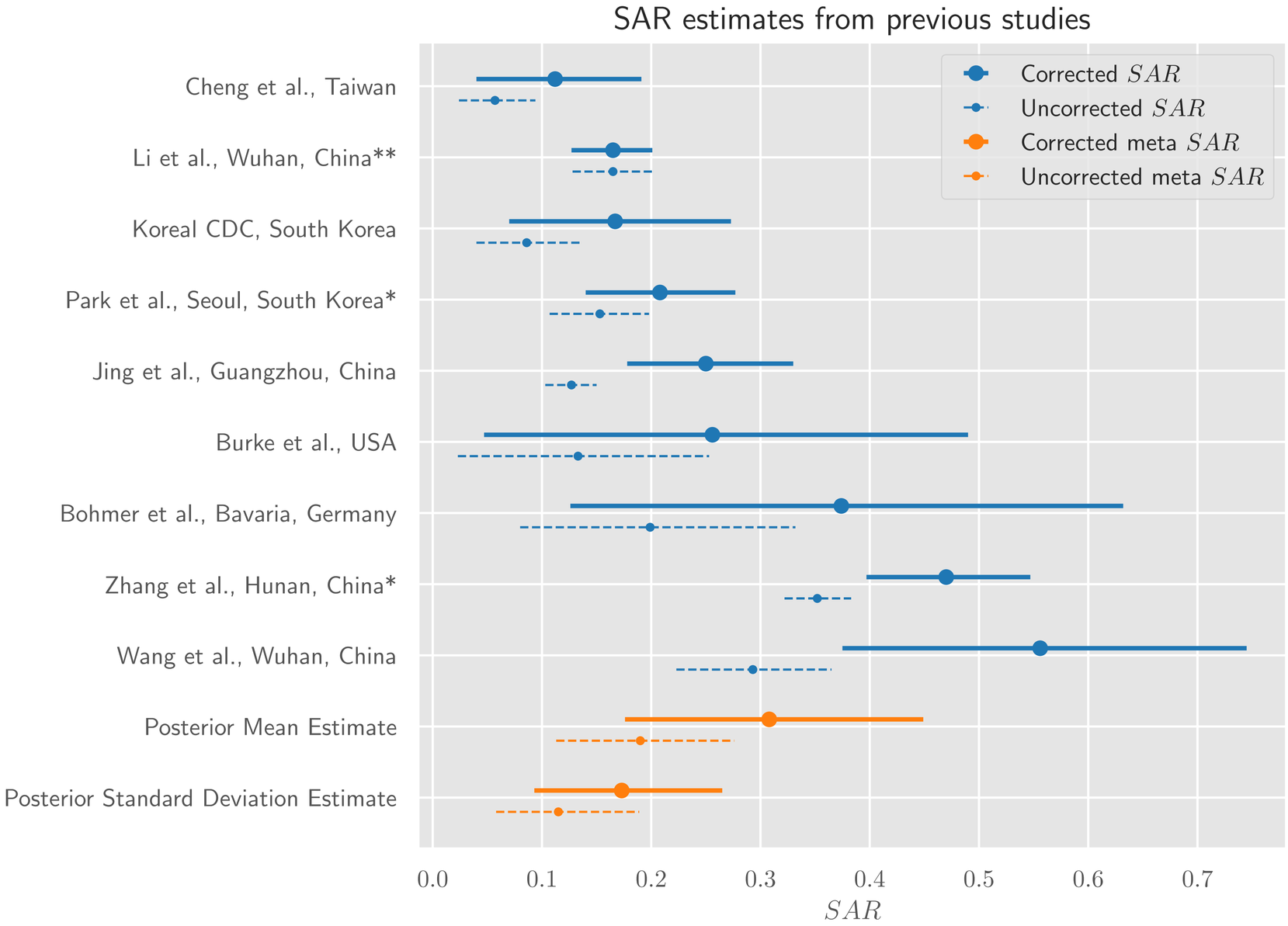
$\alpha, \beta \sim HalfFlat()$

Likelihood:

$p_i := SAR_i(1 - AR \cdot \mathbf{1}_{AR_i})(1 - FNR_i \cdot \mathbf{1}_{FNR_i})$

$\ell(C_i | SAR_i, FNR_i, AR) \propto p_i^{C_i}(1 - p_i)^{N_i - C_i}$

SAR meta-analysis results



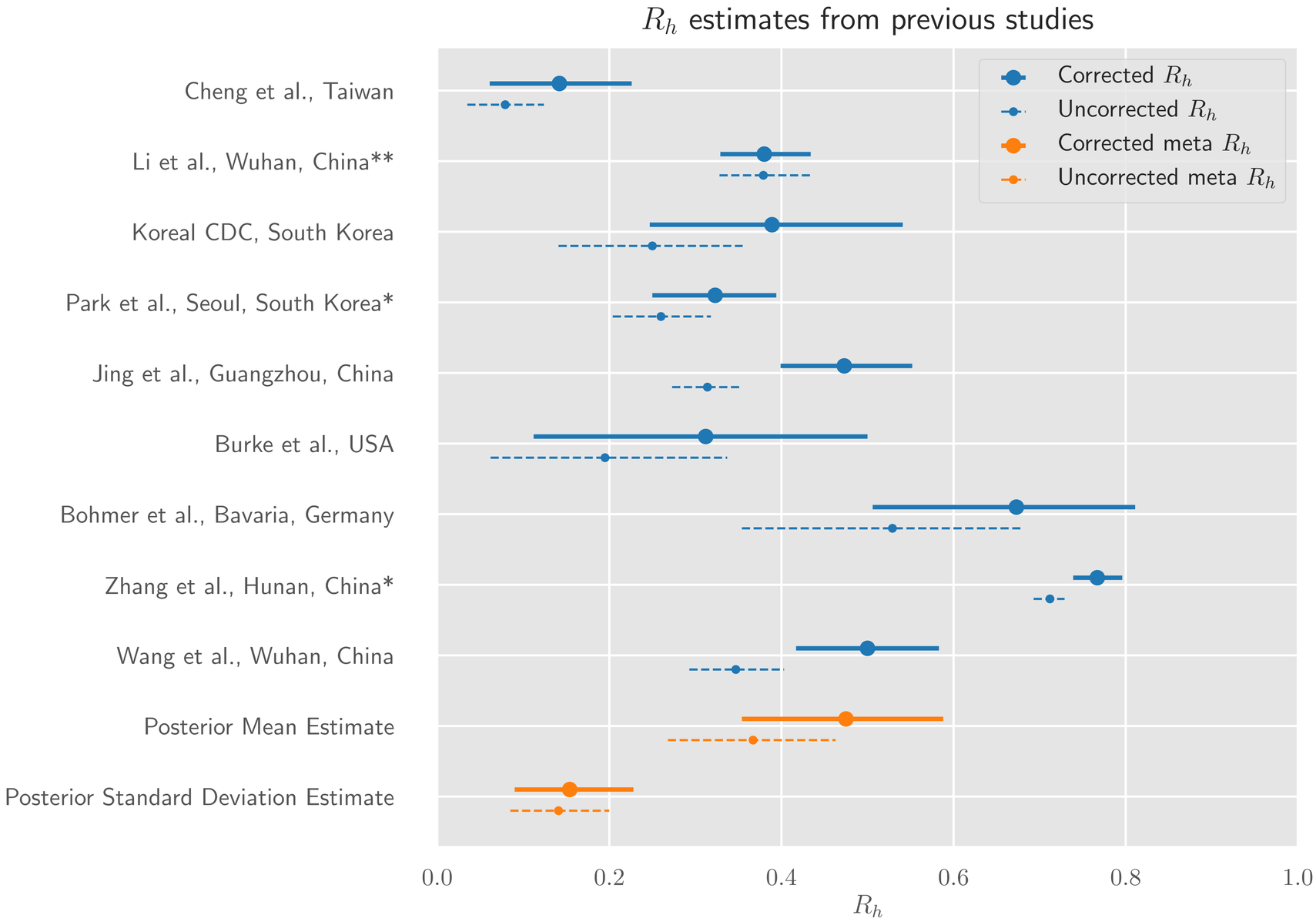
Mean and 95% Bayesian credible intervals for SAR for each study (blue).

In orange, the pooled estimate for the mean and SD for the distribution that generates the SAR. Our central pooled estimate is mean=30% and SD=15%.

SAR meta-analysis results

- Posterior mean for SAR is 30% and SD is 15%, which shows heterogeneity across studies.
- Our estimate would **increase** if FNR above 15-35%.
- Our estimate would **decrease** if asymptomatic rate (AR) below 20-40%.
- Our estimate would **decrease** if asymptomatics are less infectious. E.g. If AR=25% and relative infectiousness 60%, then SAR=30% is adjusted to 27%.
  = 0.75*0.3 + 0.25*0.6*0.3

R_h meta-analysis results



Mean and 95% Bayesian credible intervals for R_h for each study (blue).

In orange, the pooled estimate for the mean and SD for the distribution that generates R_h. Our central pooled estimate is mean=0.47 and SD= 0.15.

Vo’, Italy
Population: ~3000
Lavezzo et al.

Gangelt, Germany
Population: ~12000
Streeck et al.

Results from population sampling

Random population testing captures asymptomatics (in primary and secondary cases).

Source	Quantity	Adjusted estimate
Meta-analysis of 9 studies	$\overline{SAR}$	0.30 (0.18-0.43)
	$\overline{R_h}$	0.51 (0.40-0.62)
Meta-analysis of 9 studies	$sd(SAR)$	0.17 (0.09-0.27)
	$sd(R_h)$	0.15 (0.09-0.23)
Estimates derived from (Streeck et al., 2020), Gangelt, Germany	CRI	0.31
Our estimate from Vo', Italy data	CRI	0.50
	R_h	0.37 (0.34-0.40)
Our estimates from Singapore tracing data	R_h	0.19-0.34
Calculated from SAR= 0.3	CRI	0.41

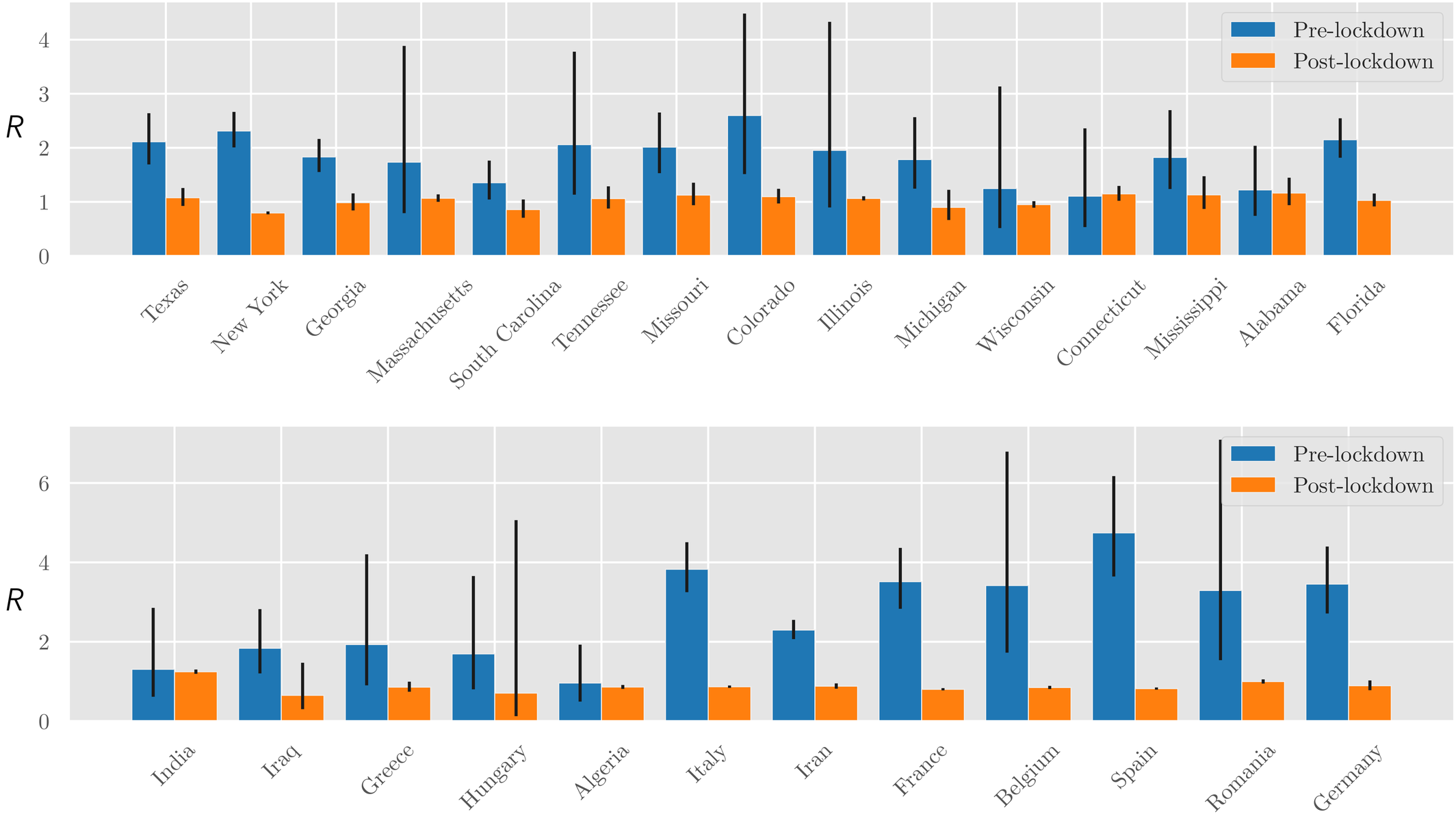
CRI in Gangelt/Vo is consistent with our SAR estimate.

This suggests our model and the SAR studies (w/ non-random testing) are reasonable way to estimate SAR.

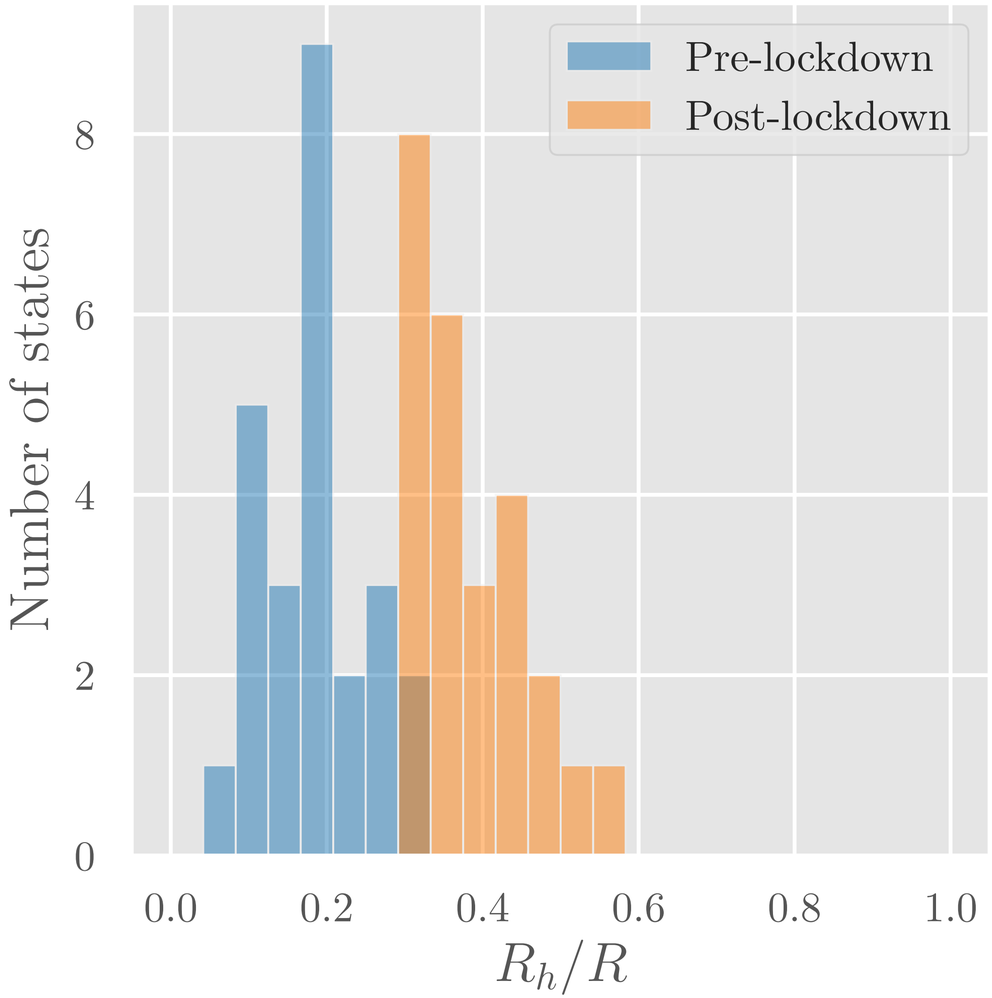
Other diseases

Disease	**SAR**	R_0
SARS-2	30%	1.4-3.9
SARS-1	8%	0.2-1.1
H1N1 Flu	15%	1.4-1.6
Colds	30-60%	2-3
Measles	70-90%	12-18

R estimates pre/post-lockdown



Household vs. total spread

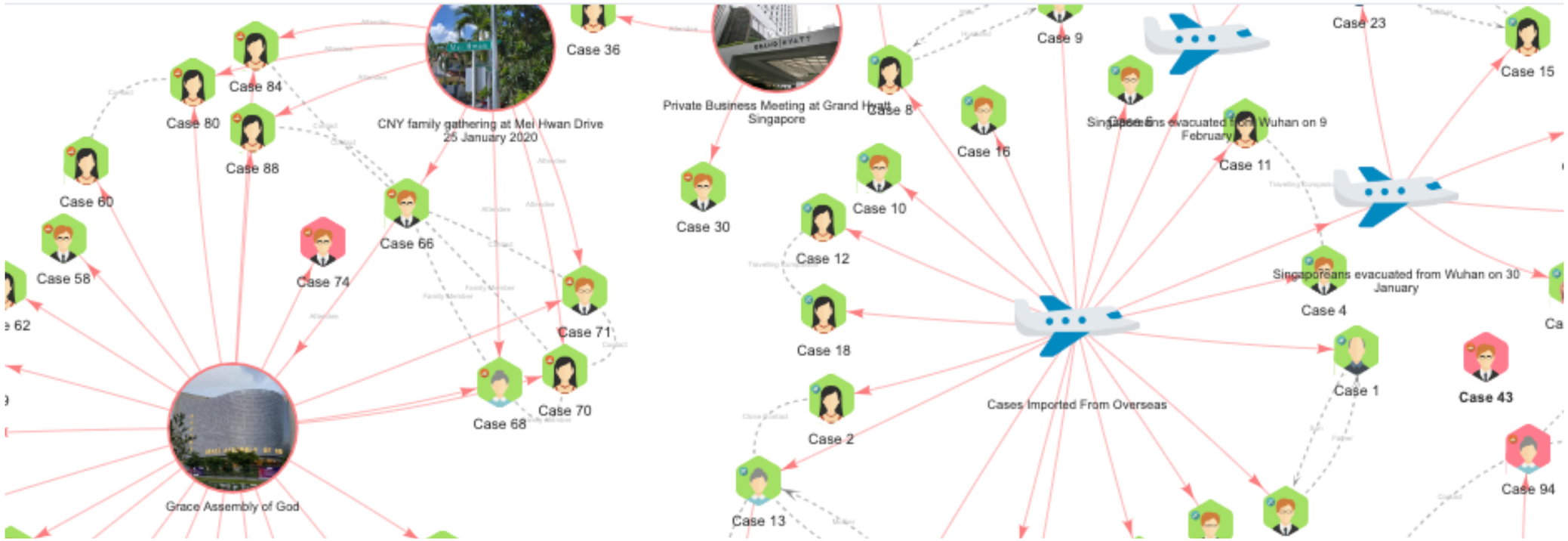


- If R_h ~ 0.3 pre-lockdown, then R_h/R was 0-25% across US states.
- After lockdown, R_h/R was 25-60%.
- Conclusion: Under social distancing, reducing household transmission is high impact.

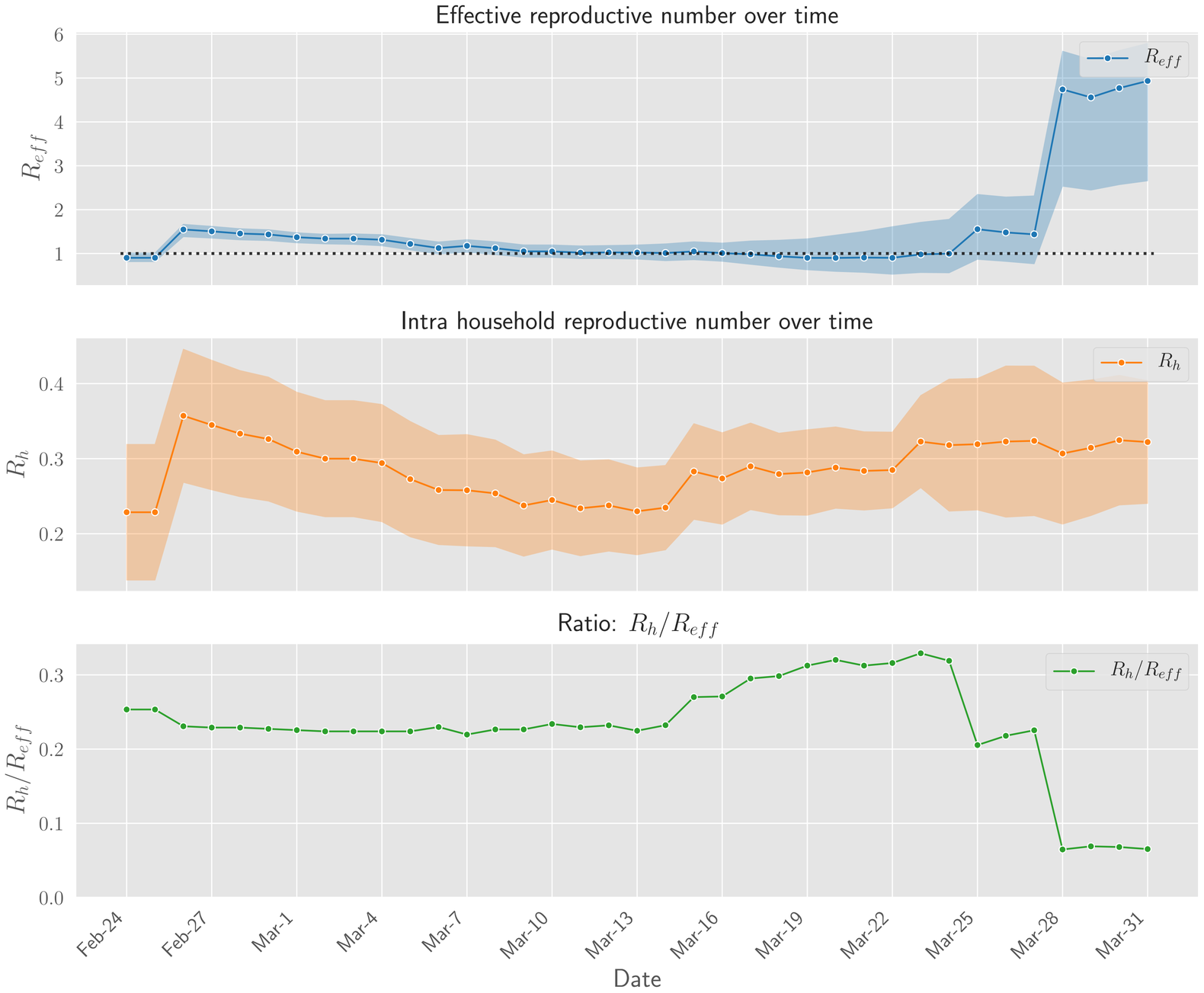
Singapore dataset

Singapore published comprehensive contact tracing with some links annotated as “family” (proxy for household).

We turned this into a dataset for inferring R_h.



R and R_h over time in Singapore



Implications: reducing SAR

Ask public, “How infectious is SARS-CoV 2?”

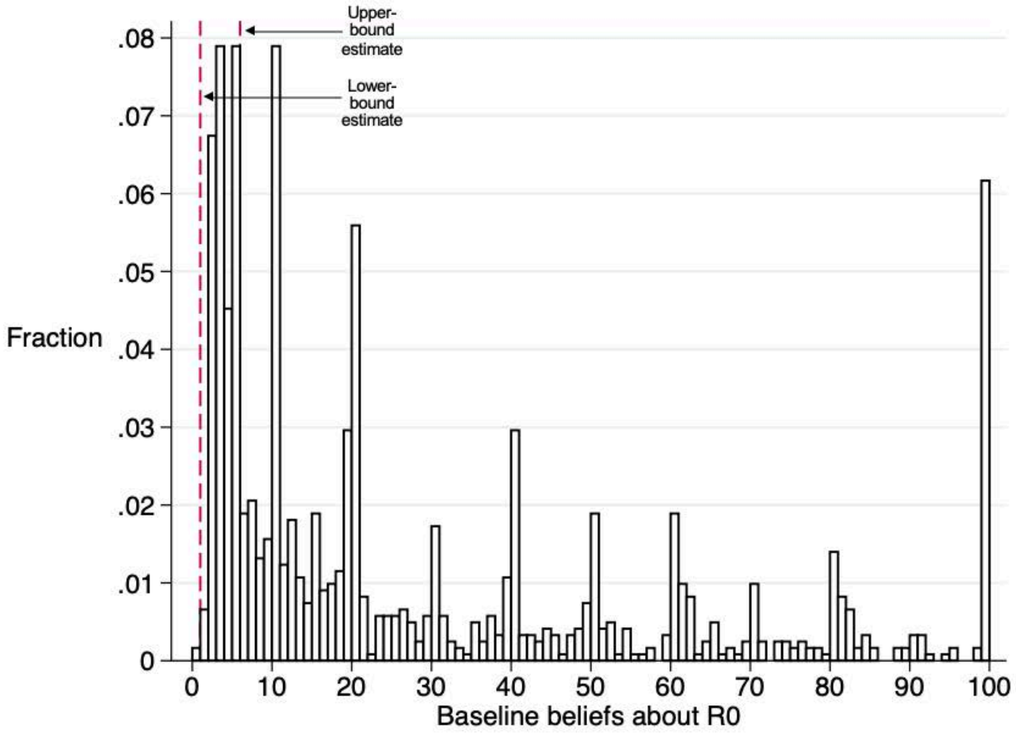
Answer:

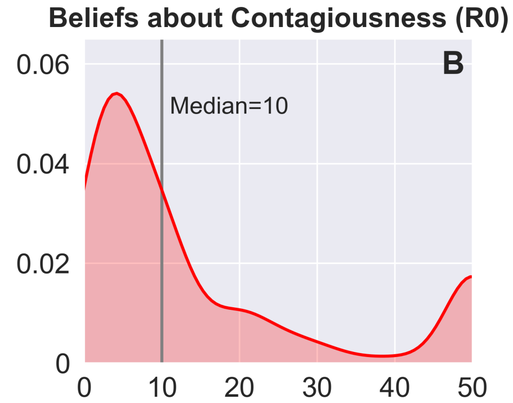
Mean R_0=28 in Akesson et al, see right.

Median R_0=10 in Fetzer et al., see bottom right.

It’s likely that people also massively overestimate household SAR.

New York Times article on transmission in Italy quotes a doctor saying household transmission was inevitable.





Can SAR be reduced?

Our meta-analysis suggest SAR <30% with some NPIs. How much can NPIs help reduce SAR?

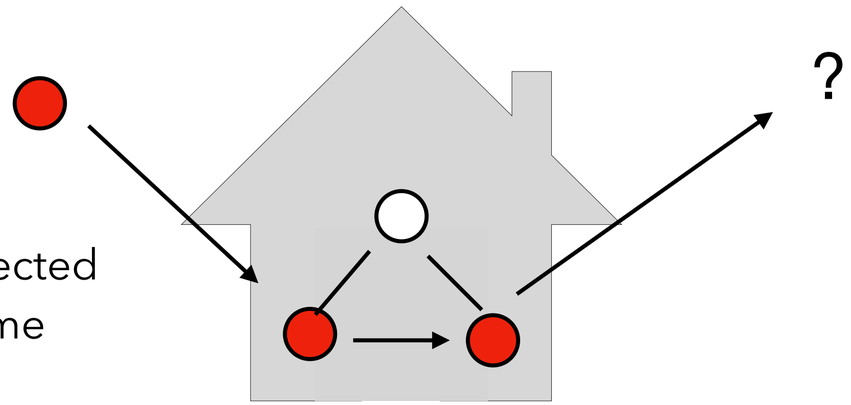
1. Li et al. find SAR drops to 0% if primary case is strictly isolated at home from symptom onset. (*n* = 14).

2. Wang et al. looked at different NPIs:

- Regular contact with primary case: 18x higher infection risk, CI = (4,85).
- Family members wearing mask *before* onset: 5x lower risk, CI = (1.25, 17)
- Disinfectant house cleaning daily: 5x lower risk, CI = (1.18, 14).

Implications: containment

Are household transmissions less bad because they stay **contained**?



Person i infected outside home

Does j infect anyone outside home?

Then i → j inside home.

Implications: containment

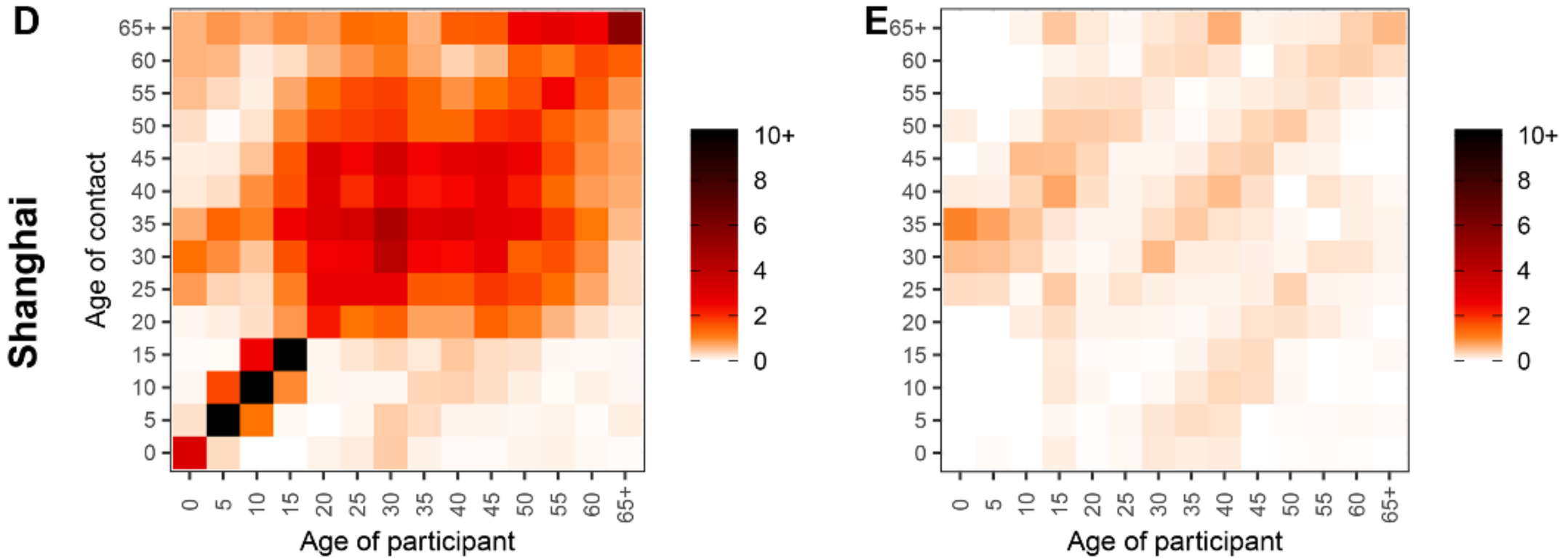
Are household transmissions less bad because they stay **contained**?

- Formally: $R_{c|h} > R_{c|c}$

  # community infections for people infected at home vs. in community.

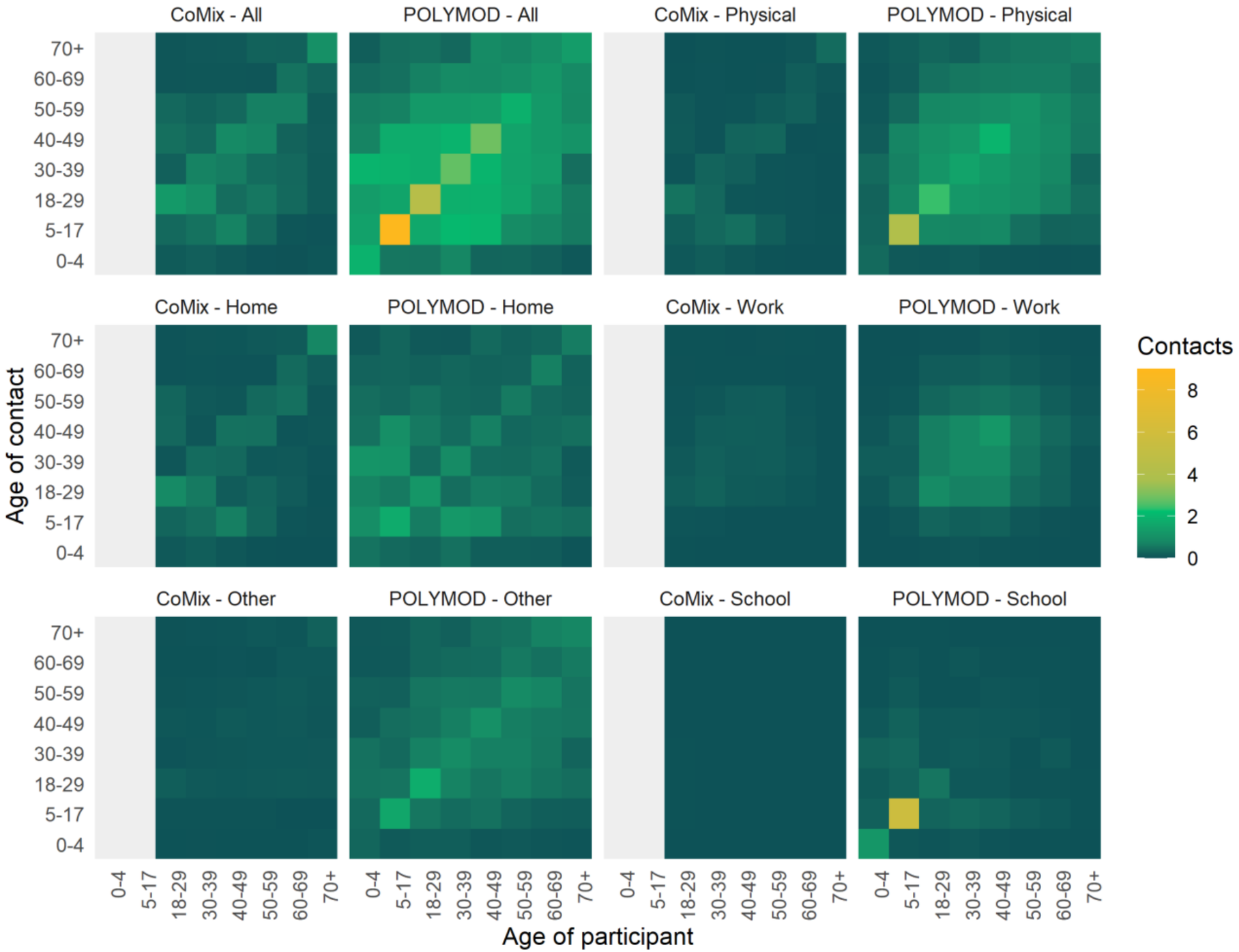
- Being infected in home is like perfect contact tracing.
- If contact tracing is weak and compliance with quarantine is high, then containment theory is probably true.
- Need better contact tracing datasets!

Lockdown contact patterns



Age-Age contact matrices for Shanghai before (left) and after (right) strict lockdown from Zhang et al. 2020

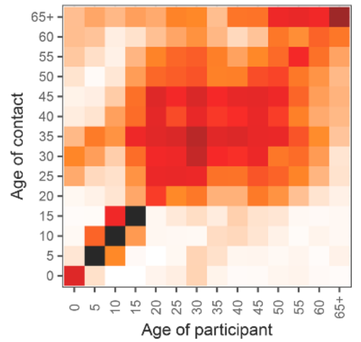
Estimated contact patterns for the UK before (POLYMOD) and after (CoMix) lockdown from Jarvis et al 2020.



Lockdown contact patterns

1. Assume that secondary attack rate constant across groups (not true for households vs work contacts!)
2. Then entry C_{ij} is proportional to mean infections in group i caused by person in group j, which is reproductive number for j restricted to i.
3. How do we "sum over" C_{ij} to get overall reproductive number R ?

   A: Find dominant eigenvalue of C_{ij}

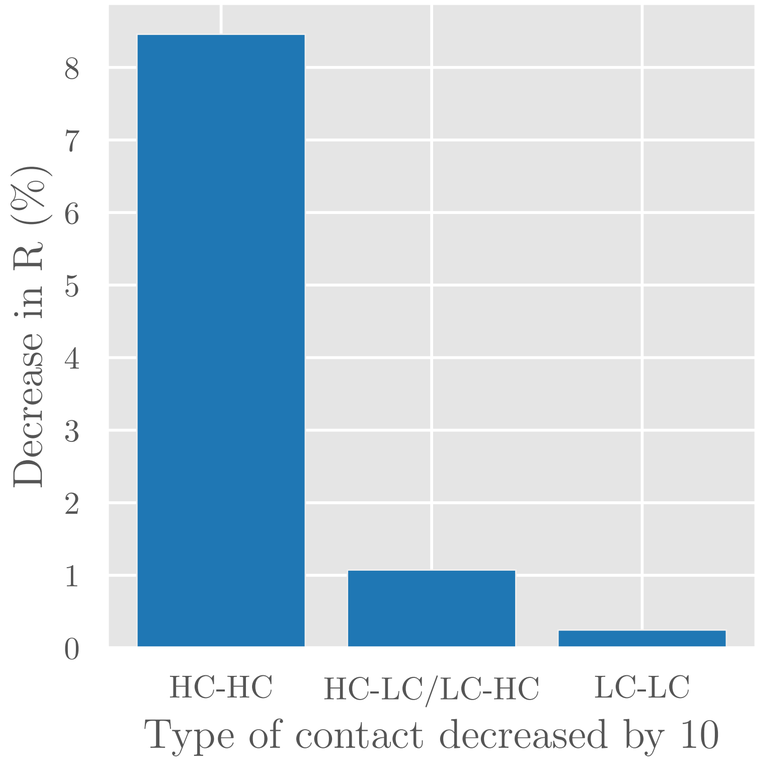


Lockdown contact patterns

- We used survey data from US to estimate 2x2 contact matrix for essential workers (high contact) and everyone else (low contact).
- What is the effect of reducing contact between i and j by 10%?

	HC	**LC**
HC	9.3	1.0
LC	4.6	3.0

Figure 5: Contact matrix estimates for the United States using data from (Rothwell, 2020).



Conclusions

- SAR has mean=30% and SD=15%. There is high heterogeneity.
- Average person infects ~0.47 household members.
- Household is small proportion of transmission pre-lockdown but large (25-60%) under lockdown.
- There’s evidence that SAR can be reduced with NPIs
- Household infections probably not “contained” but are less bad than community infections.
- If there are identifiable groups with much higher contact (e.g. essential workers), then focus interventions on them.

Bonus: Open questions outside household transmission

- How does spread work in practice?
  – kind of contact; droplets vs. fomites
  – indoors vs outdoors, duration of contact.
  – family house vs. apartments vs. dormitory.
  – superspreaders and overdispersion, can we predict who is a superspreader?
  – NPIs: masks and other PPE, distance, hygiene.
  – how do public's beliefs influence spread?
  – consider using data from Singapore, Korea.
  – need more data from Western countries. E.g. tracing, CCTV, cellphone.

- Will the virus mutate into worse or better strain? How should we update prior on lack of major mutation so far? Even if mutation is unlikely (<4%), impact would be large.

- Better analyze the overall impact of new Covid-19 tech:
  – sewage testing or other rapid prevalence testing
  – better symptomatic detection (e.g. use ML or home sensors)
  – better genetic prediction of infectiousness (e.g. superspreader risk) and severity of infection
  – treatment that reduces IFR