

# Challenges of collecting social connectivity data: an epidemiologists' perspective

John Edmunds

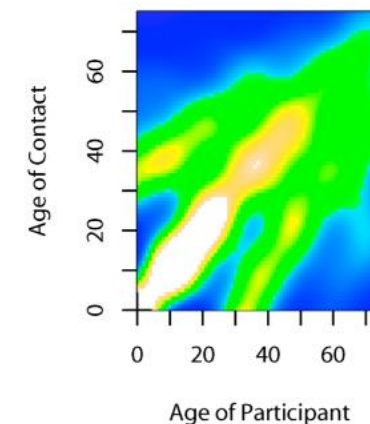
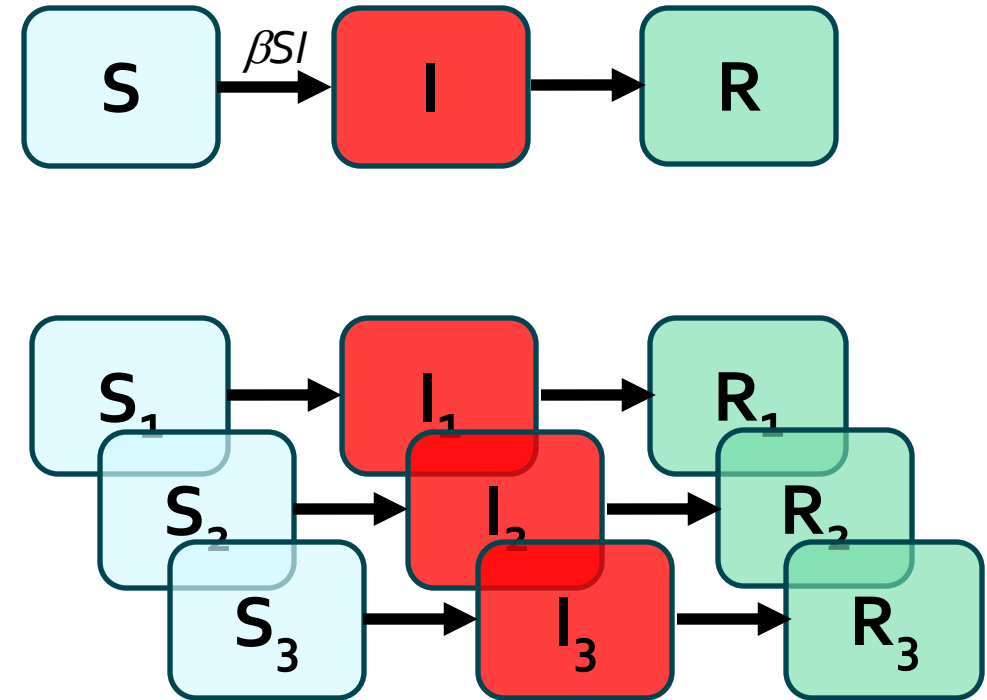
[John.Edmunds@lshtm.ac.uk](mailto:John.Edmunds@lshtm.ac.uk)

LONDON  
SCHOOL of  
HYGIENE  
& TROPICAL  
MEDICINE



# Mathematical models and social connectivity

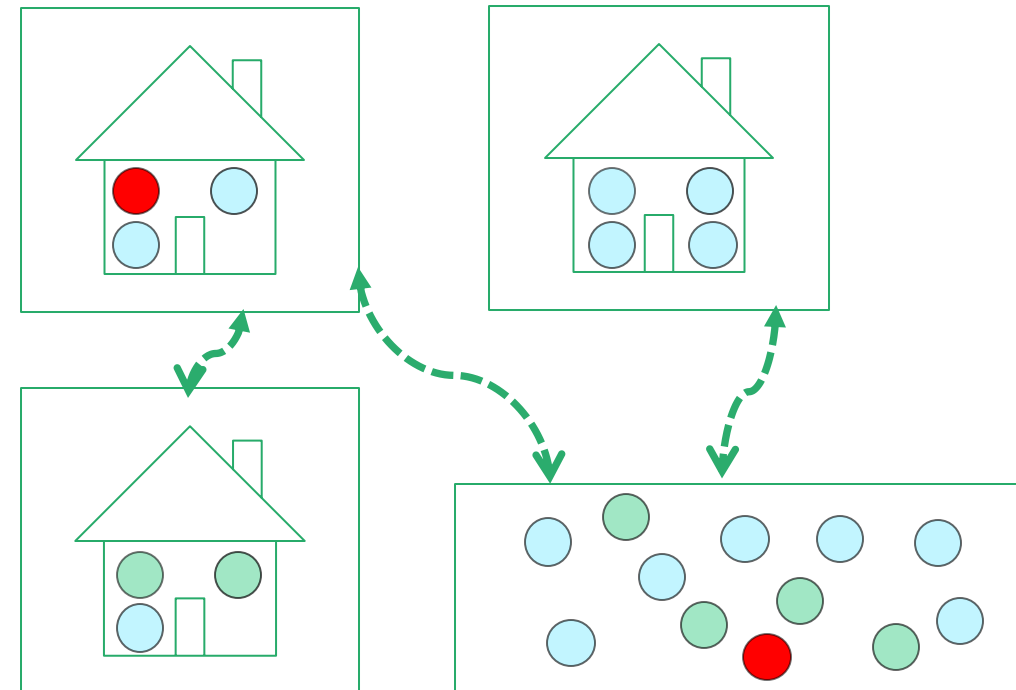
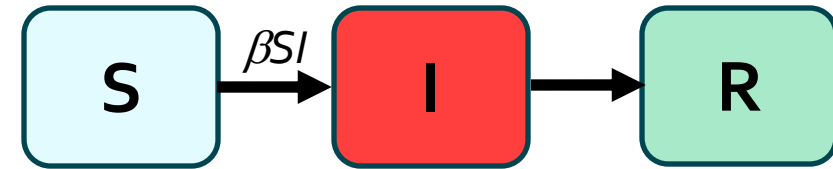
- Simple transmission models assume “mass action”
- More realistic models allow for heterogeneity in contact
  - Compartmental models
    - Contact matrix
  - Network or agent-based models



Mossong *et al.*  
PLoS Med (2008)

# Mathematical models and social connectivity

- Simple transmission models assume “mass action”
- More realistic models allow for heterogeneity in contact
  - Compartmental models
    - Contact matrix
  - Network or agent-based models



*Epidemiol. Infect.* (2012), **140**, 2117–2130. © Cambridge University Press 2012  
doi:10.1017/S0950268812000842

---

## REVIEW ARTICLE

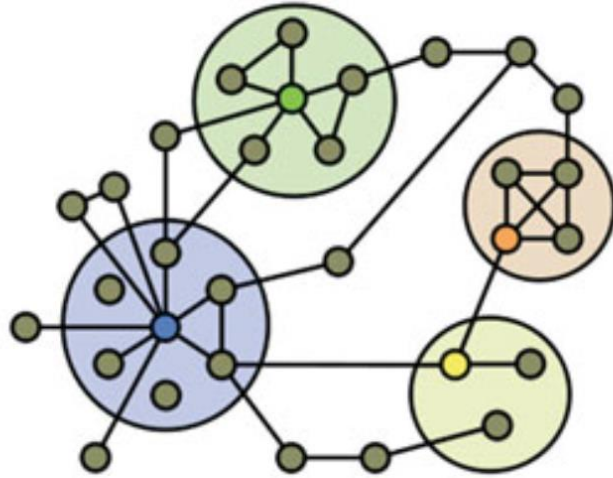
# Close encounters of the infectious kind: methods to measure social mixing behaviour

---

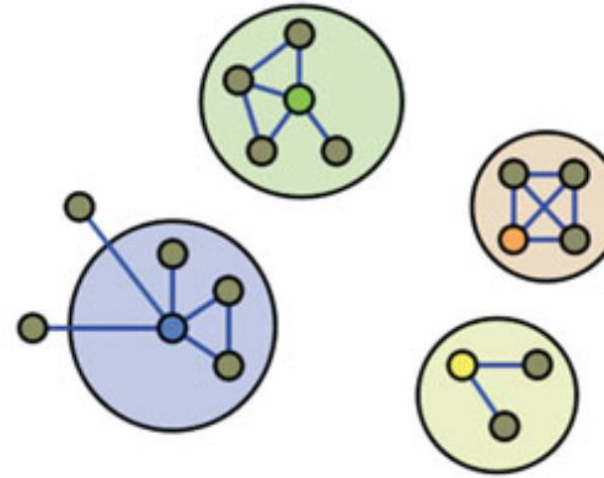
J. M. READ<sup>1\*</sup>, W. J. EDMUNDS<sup>2</sup>, S. RILEY<sup>3</sup>, J. LESSLER<sup>4</sup> AND  
D. A. T. CUMMINGS<sup>4</sup>

# Methods for measuring contacts

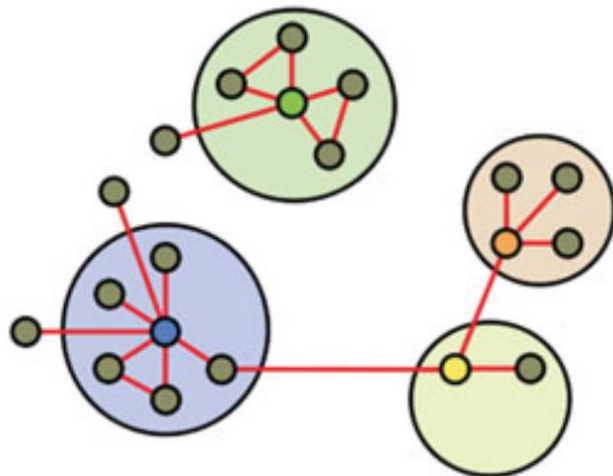
(a) True



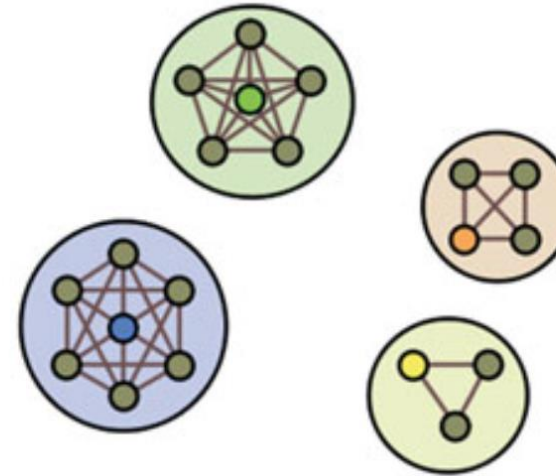
(b) Diary / recall



(c) Proximity

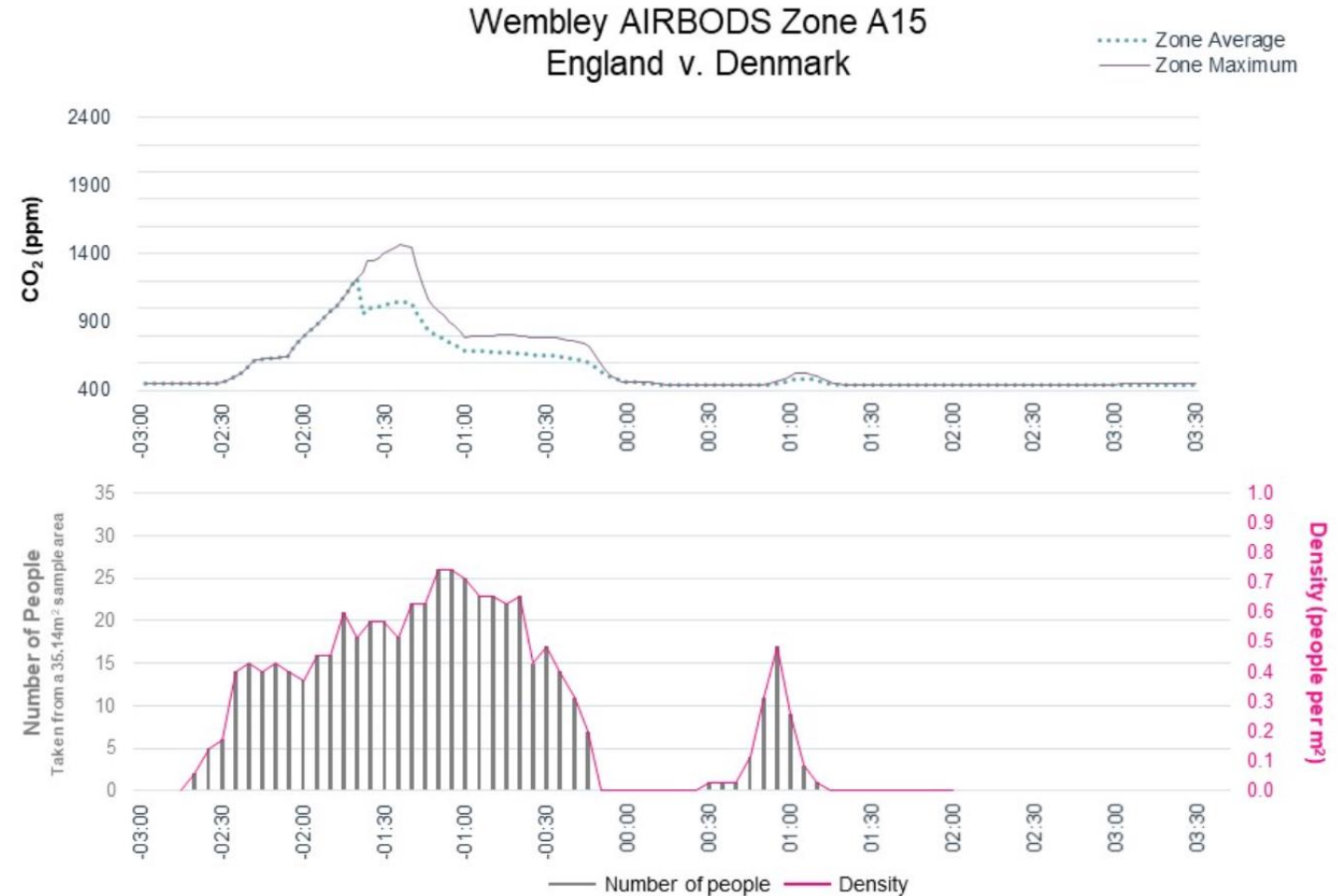


(d) Location



# Methods for measuring contacts: Direct observation

- Video recording with (e.g.) facial recognition software
- Co-location usually recorded
  - Direction (face to face)
  - Measures (e.g. mask-wearing)
- Ease of use for participants
- Limited coverage (in practice indoor settings)
  - Multiple angles often required
- Intrusive (ethics)
- Expensive

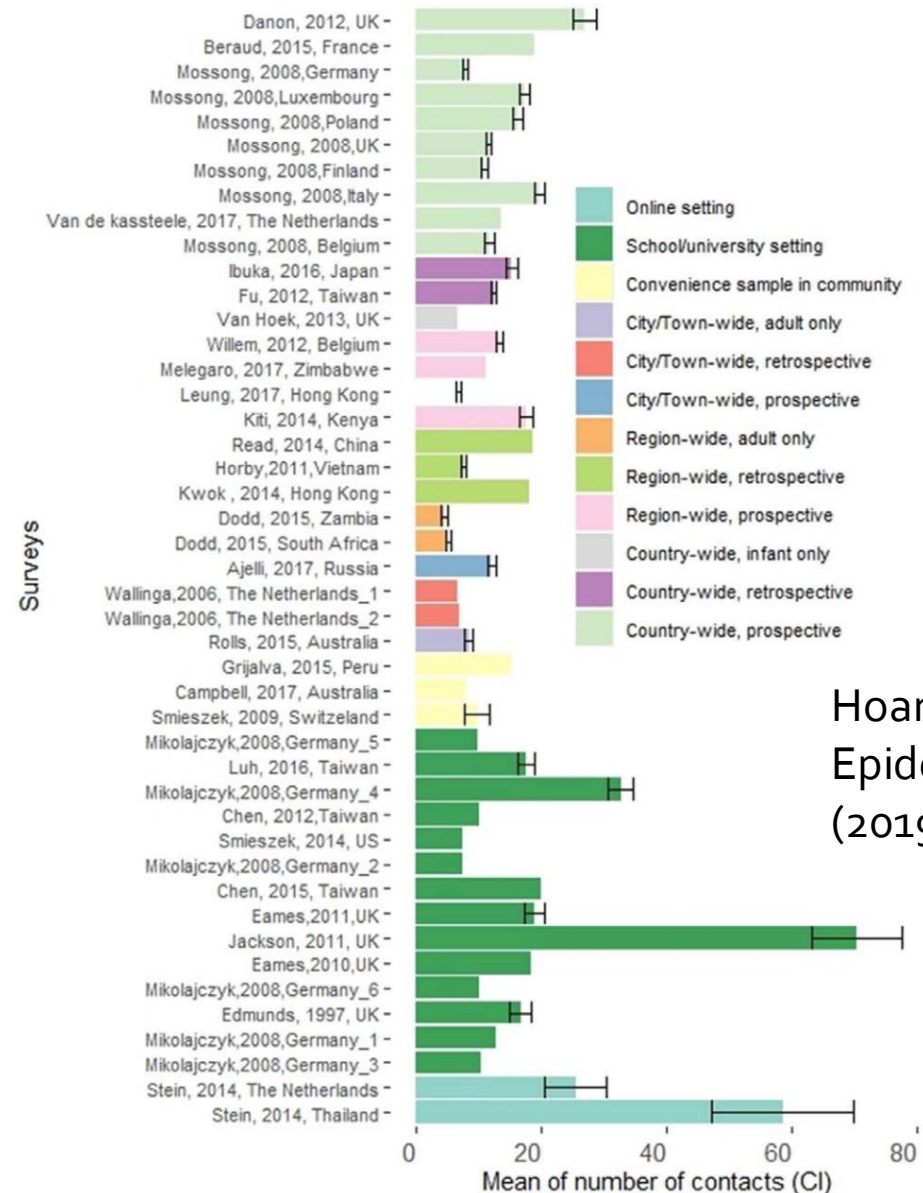


<https://www.gov.uk/government/publications/events-research-programme-phase-ii-and-iii-findings>



# Methods for measuring contacts: contact diaries

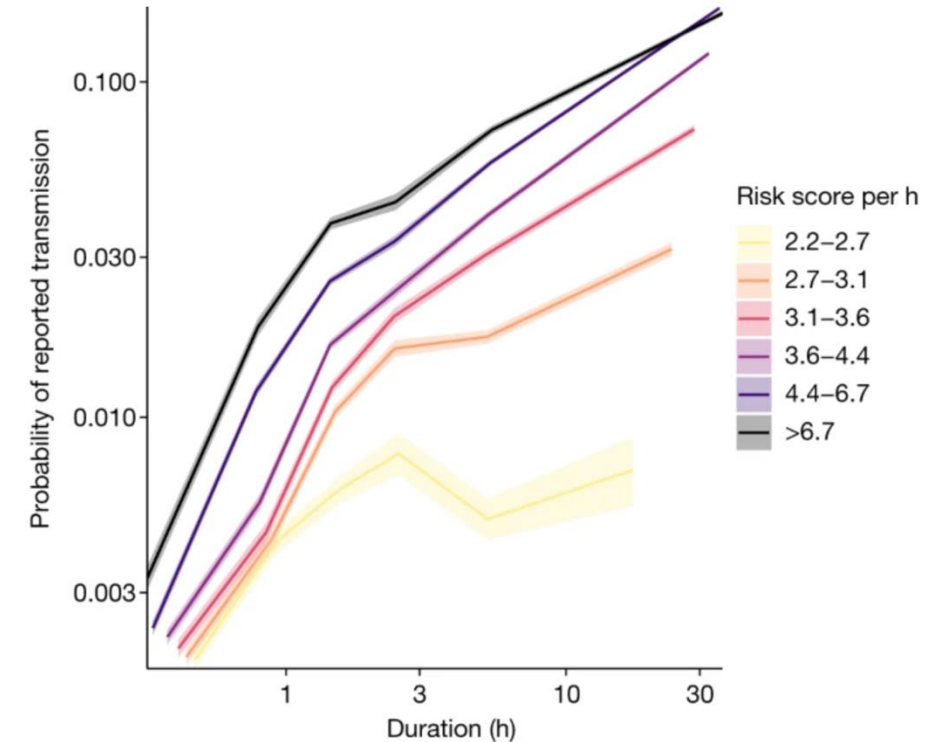
- Diaries (or similar) record conversations/touch
- Can obtain basic data on who contacted (matrices), duration, setting, frequency etc.
- Egocentric data (typically), with limited network information
- High burden on participants
  - Recall bias (forgotten contacts)
    - Particularly retrospective & non-touch, short contacts (McCaw et al. 2010)
  - Proxies used for young/old
  - High degree individuals
  - Repeat samples
- Variable results
- Small sample sizes (zero recorded contacts)
- Limited to certain types of contact (respiratory)



# Methods for measuring contacts: proximity sensors

- Phones (e.g. contact tracing app) or devices (RFID)
- Minimal effort for participant
- Can obtain detailed network data, including duration and frequency
  - Location can also be recorded
- App (linked to test results) can give insight on risk
- Best suited to specific settings/institutions
- Only record contacts between participants
  - No information from non-participants
  - Non-linear fall in contacts recorded with fall in coverage
- Difficult to tune signal to risk (sensitivity / specificity)

Fig. 2: The probability of transmission is affected by both duration and proximity as captured by risk score.

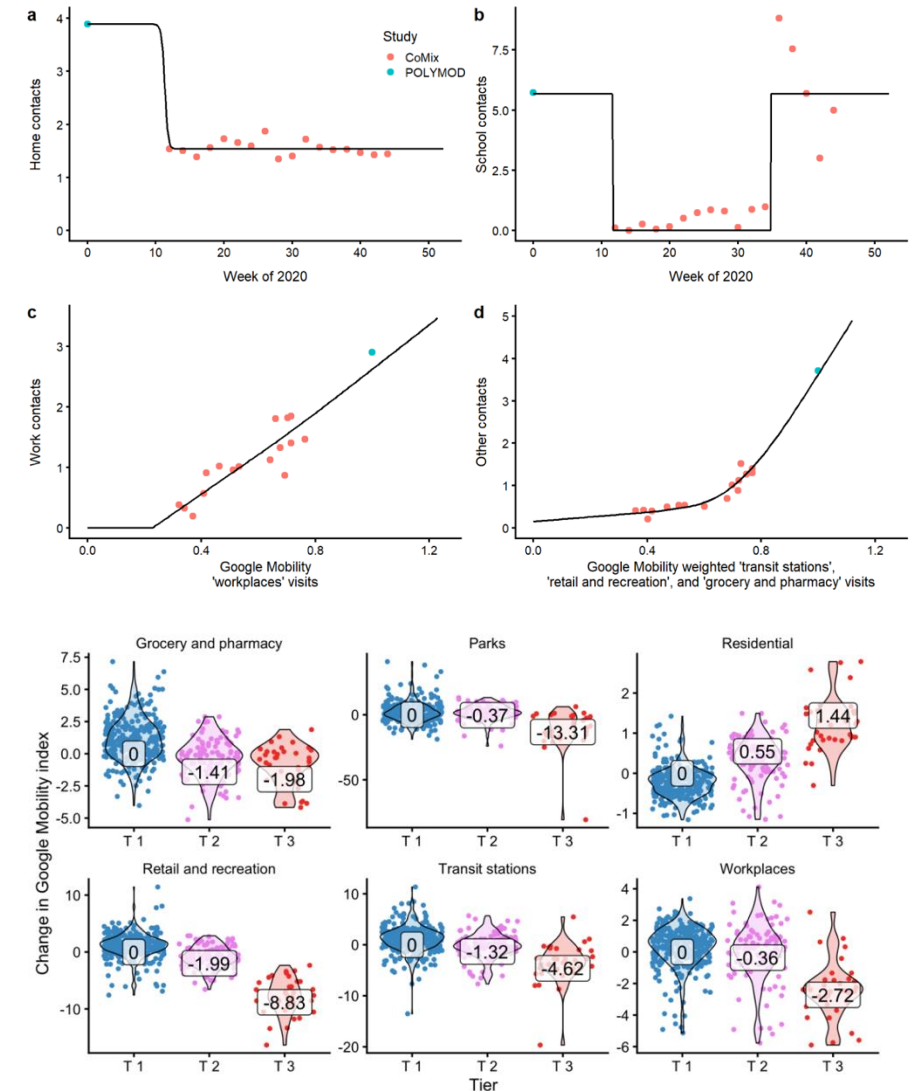


log-log plot of the probability of reported transmission—that is, the probability that the contact reported a positive test that we attributed to the transmission event traced, as a function of the binned duration of exposure and mean risk score per hour (that is, cumulative risk score divided by duration). Solid lines connect maximum-likelihood estimates for each bin and shading around these denotes 95% confidence intervals. Tabulated values can be found in Supplementary Table 8.



# Methods for measuring contacts: location/mobility data

- Phones, credit card usage etc
  - Aggregate data (e.g. Google Mobility)
  - Individual-level data (cell mast signals)
- Zero effort for participant
- Large data sets (& rapid during pandemic)
- Mobility vs risk
  - Some correlation with contacts
- Ethics
- Data typically held by large corporations
  - Methods not always clear
- Lack of covariate data (age, sex, socio-economic, etc)
- Children and others excluded (representativeness)



- Representativeness
  - Diaries & online panels
  - Mobile phones and representativeness
  - Representativeness of setting-specific findings
  - Children and others often excluded
- Sources of variation within (dynamics) and between individuals poorly understood
  - Temporal/seasonal changes under-researched (sample size?)
  - Drivers of changes in behaviour (prevalence, perceived risk, ability to respond, etc)
- Methodological variation and lack of a “gold standard”
  - Few comparative studies
- Contacts across different strata in society (e.g. ethnic and socio-economic)
  - Diaries/surveys have limited potential
- Non-respiratory routes of transmission
- Calibration/quantification of measured contacts and risk
  - Need large-scale individual-based data on outcomes and contacts
- Links between epidemiological and economic activity and wellbeing are important but missing