Social contact data for infectious disease modelling

as part of the "1st GENID Summer School on Infectious Disease Modeling"
Halle University

Dr. Pietro Coletti

DSI DATA SCIENCE INSTITUTE UHASSELT

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Health Economics & Modelling Infectious Diseases
Vaccine & Infectious Disease Institute
University of Antwerp

TransMID
www.simid.be

erc
European Research Council
Established by the European Commission

European Commission

Personal introduction

- Background: PhD in Theoretical physics
- Postdoctoral researcher at SIMID
- Research interests:
  - Modelling of infectious diseases
  - Network epidemiology
  - Social contacts

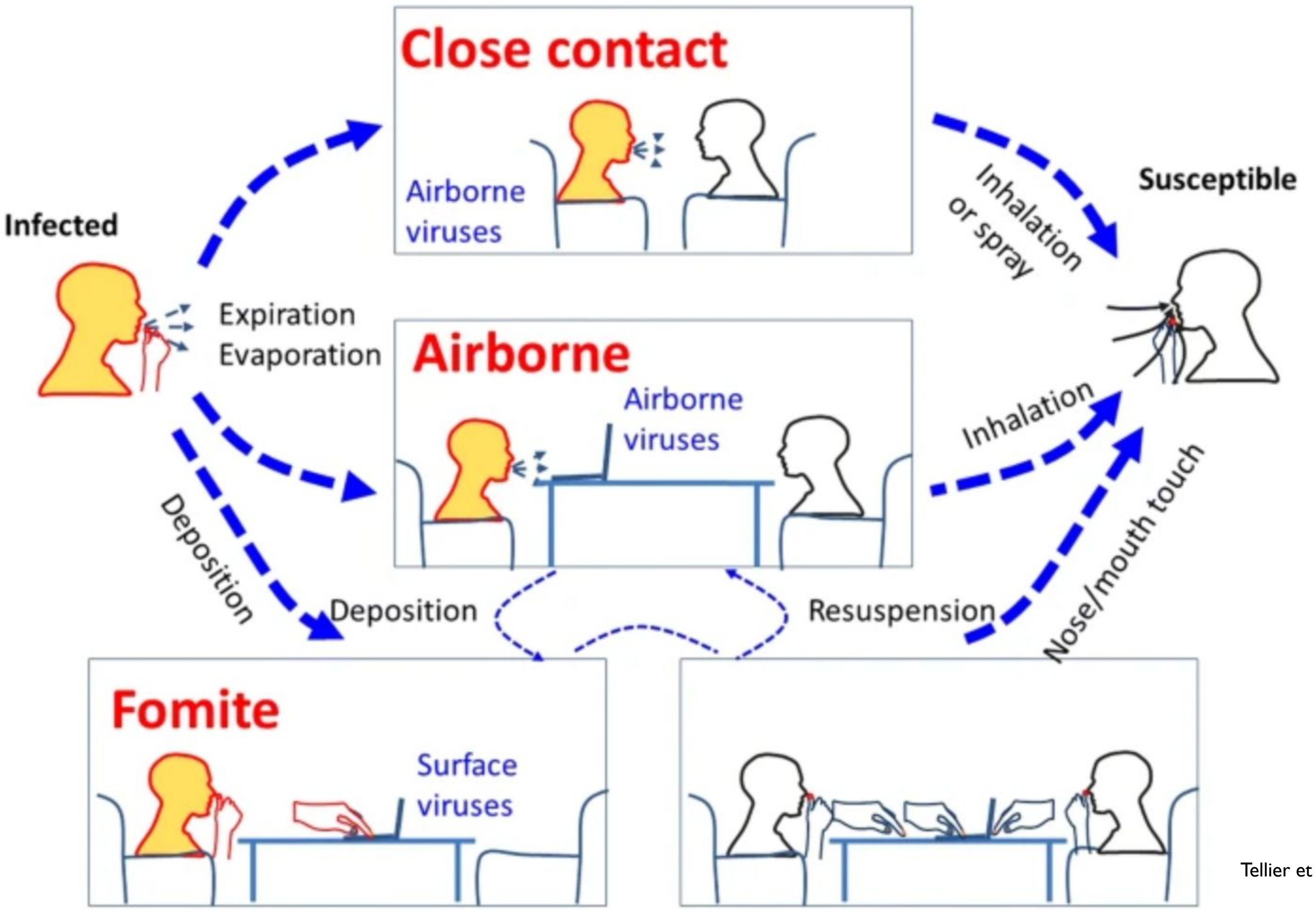
Contact: pietro.coletti@uhasselt.be

@coletti_pietro

Outline

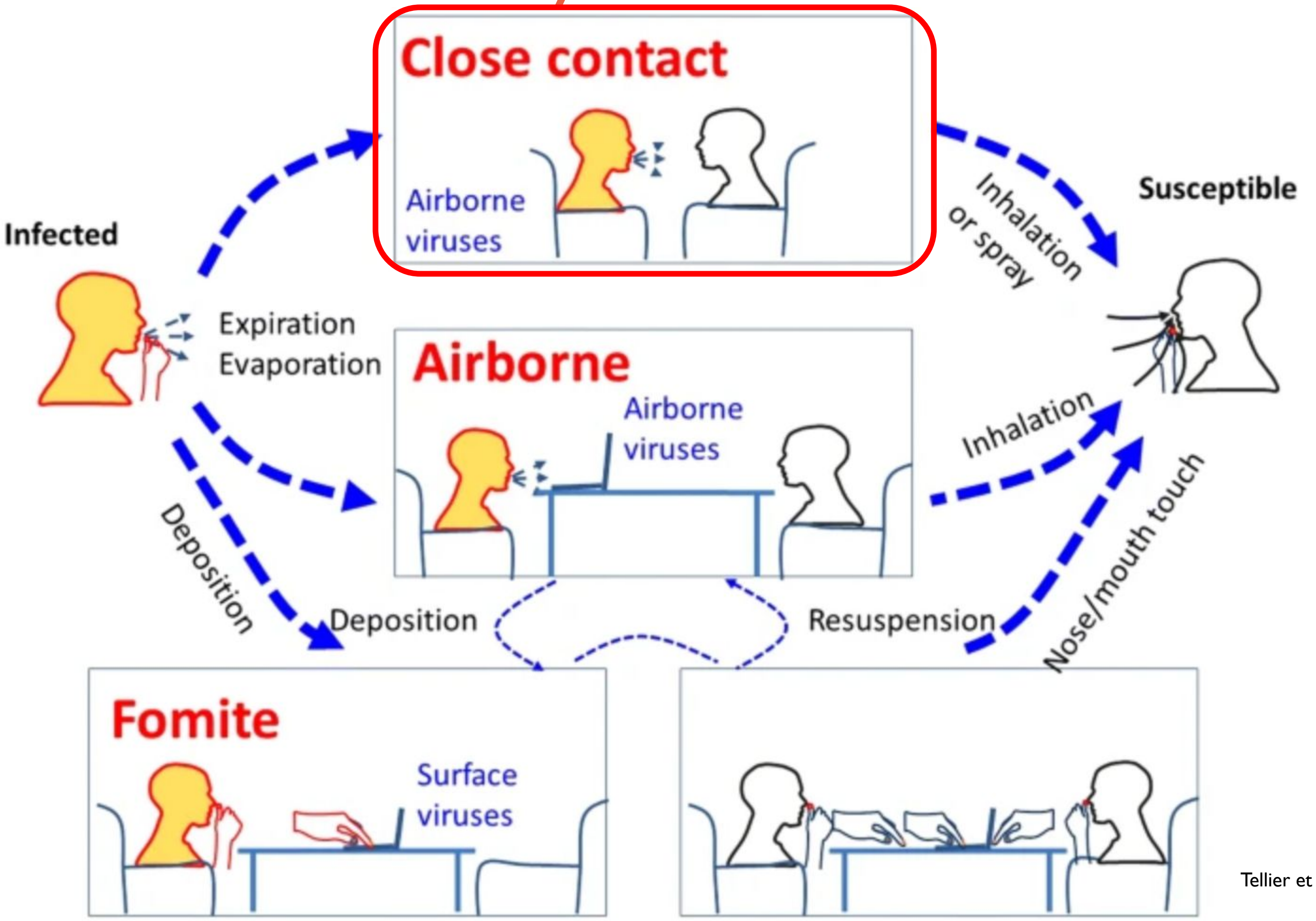
- Contact data: why and how
- Contact matrices: definition and features
- Next generation approach and R_0
- Heterogeneity in infectiousness and susceptibility
- Modelling the number of contacts
- Contacts in infectious disease models
- Other methods: Time-use survey
- Other methods: Global Positioning System data
- Conclusions

Why contacts?



Tellier et al. (2019)

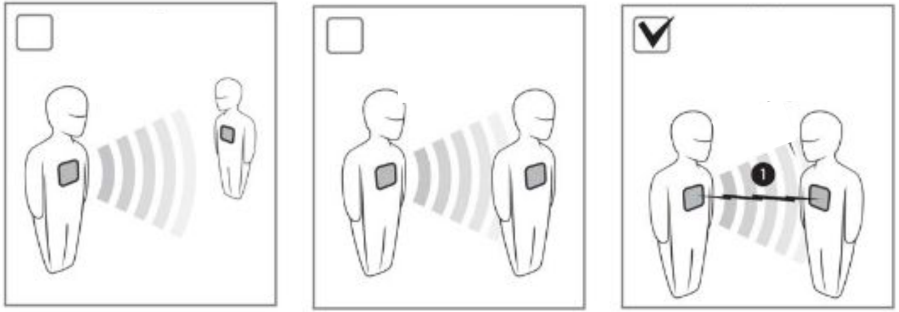
Why contacts?



Tellier et al. (2019)

Ok, but how?

- Sensors (RFID, Smartphones, …)

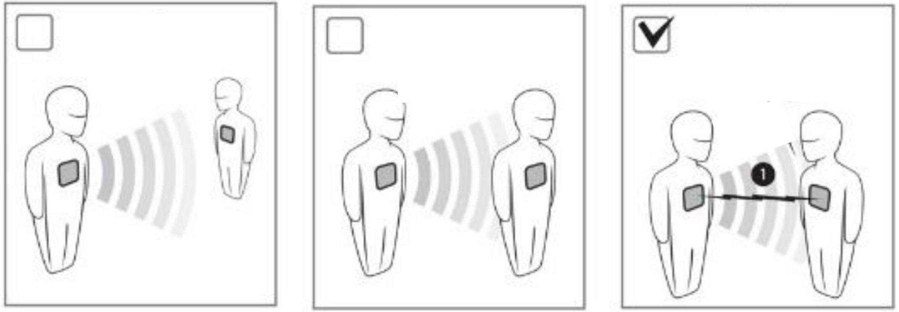


- Social contact surveys

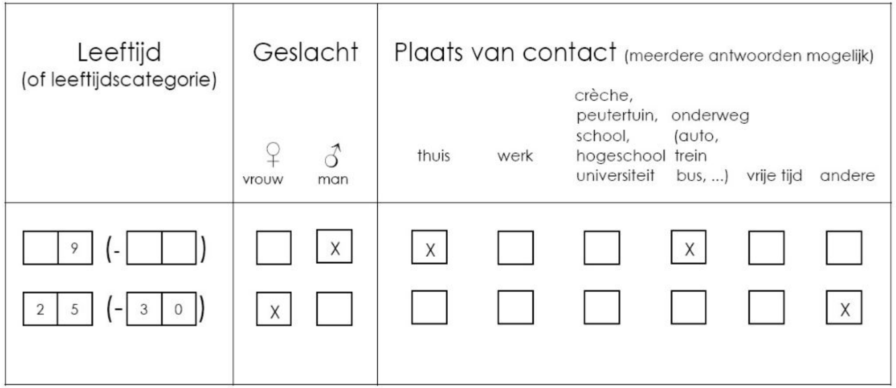
Leeftijd (of leeftijdscategorie)	Geslacht ♀ vrouw	♂ man	Plaats van contact (meerdere antwoorden mogelijk) thuis	werk	crèche, peutertuin, school, hogeschool universiteit	onderweg (auto, trein bus, ...)	vrije tijd	andere
9 (-)		X	X			X		
2 5 (- 3 0)	X							X

Ok, but how?

- Sensors (RFID, Smartphones, ...)



- Social contact surveys

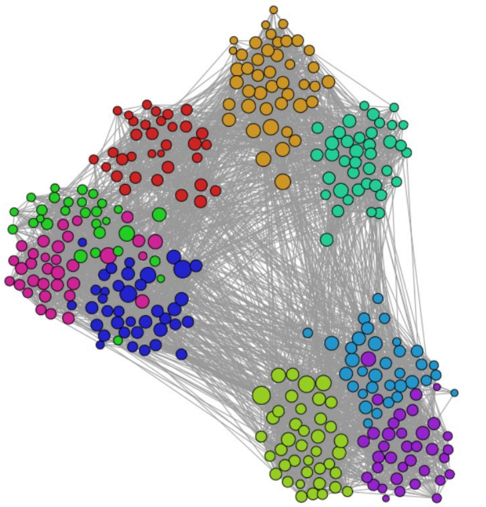


Leeftijd (of leeftijdscategorie)	Geslacht ♀ vrouw	Geslacht ♂ man	Plaats van contact (meerdere antwoorden mogelijk) thuis	werk	crèche, peutertuin, school, hogeschool universiteit	onderweg (auto, trein bus, ...)	vrije tijd	andere
9 (-)		X	X			X		
25 (- 30)	X							X

- Others

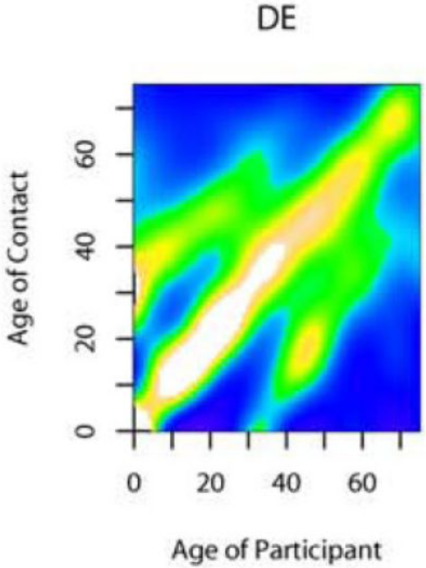
What the data looks like

Sensors



(a) Contact Network

Social contact surveys



Mastandrea et al. (2015).
Mossong et al. (2008).

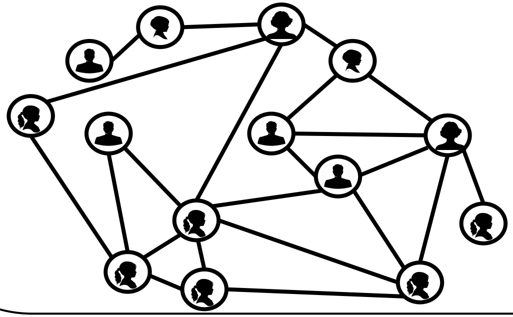
Methods at a glance

Sensors

- Collect full network information
- Difficult to collect
- Non-representative

Availability:

- Few countries/settings



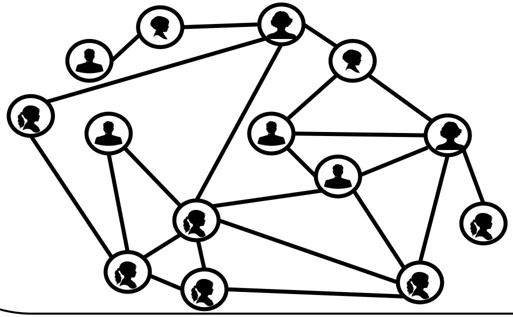
Methods at a glance

Sensors

- Collect full network information
- Difficult to collect
- Non-representative

Availability:

- Few countries/settings

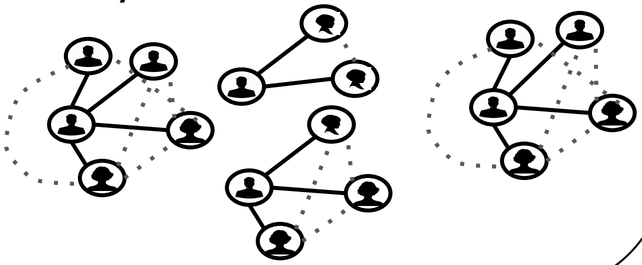


Social contact surveys

- Collect only participant information
- Easier to collect
- Generally representative

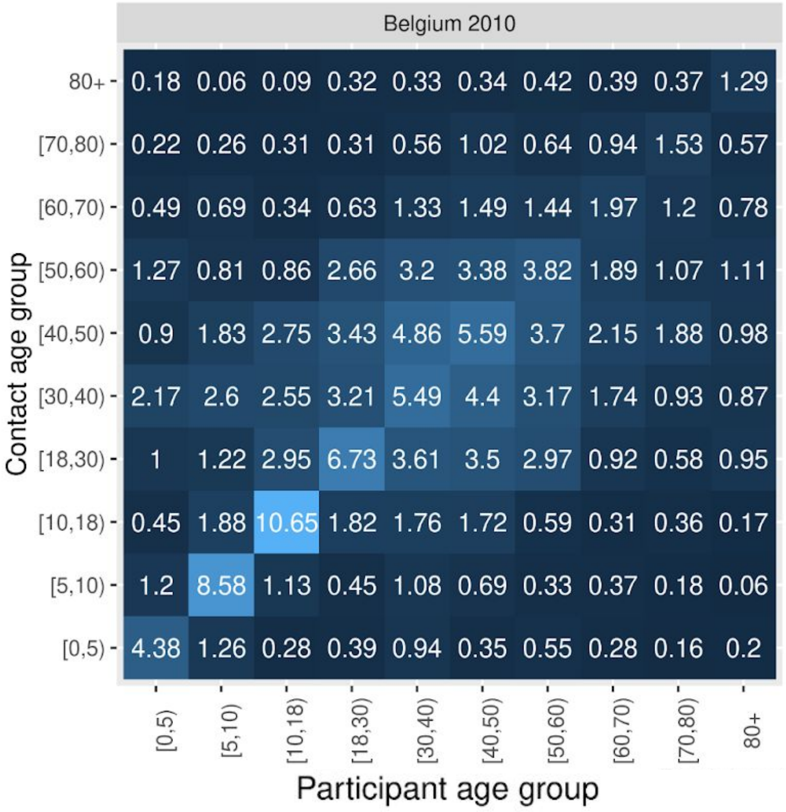
Availability:

- Many countries *



*[socialcontactdata.org]

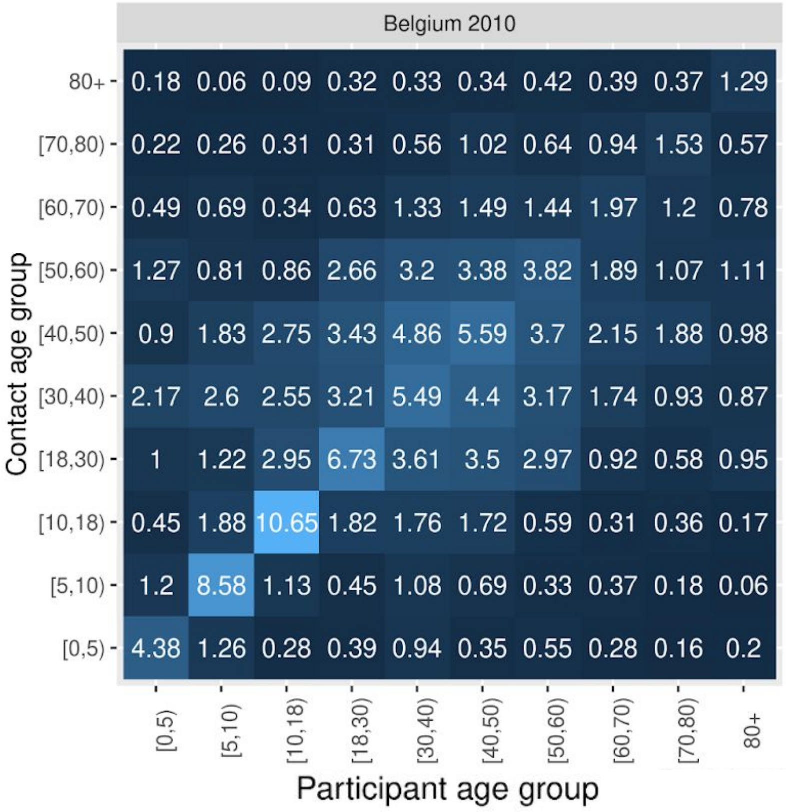
Contact matrices



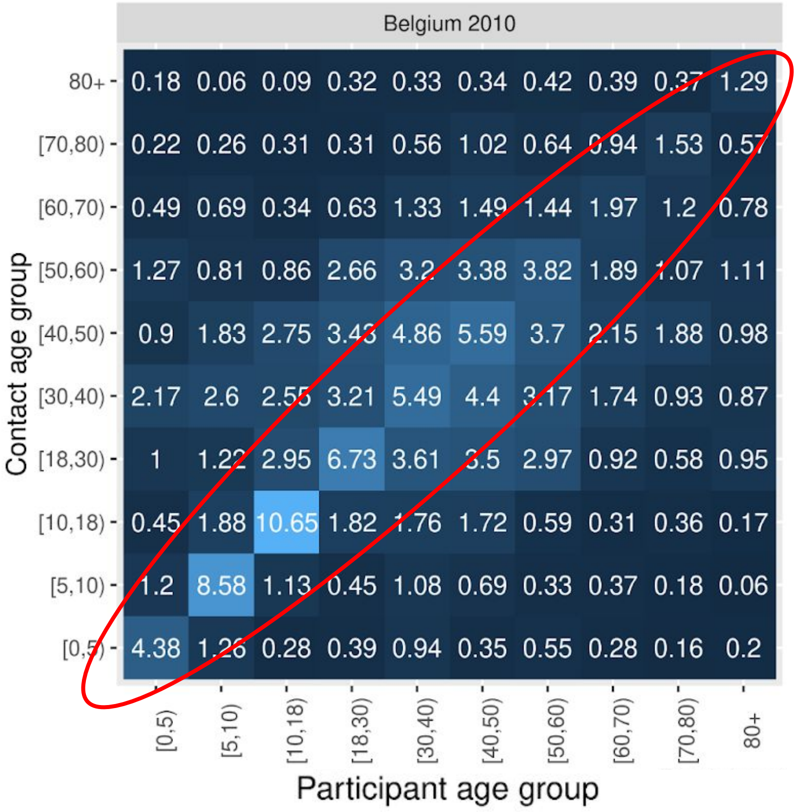
- Average number of contacts between participants in age class *i* and in age class *j*
- Usually symmetrized
- Usually location specific (e.g home, work, school,...)
- Can be defined for subset of contacts (e.g. physical contacts)

Hoang et al. (2021)

Contact matrices: pre-pandemic

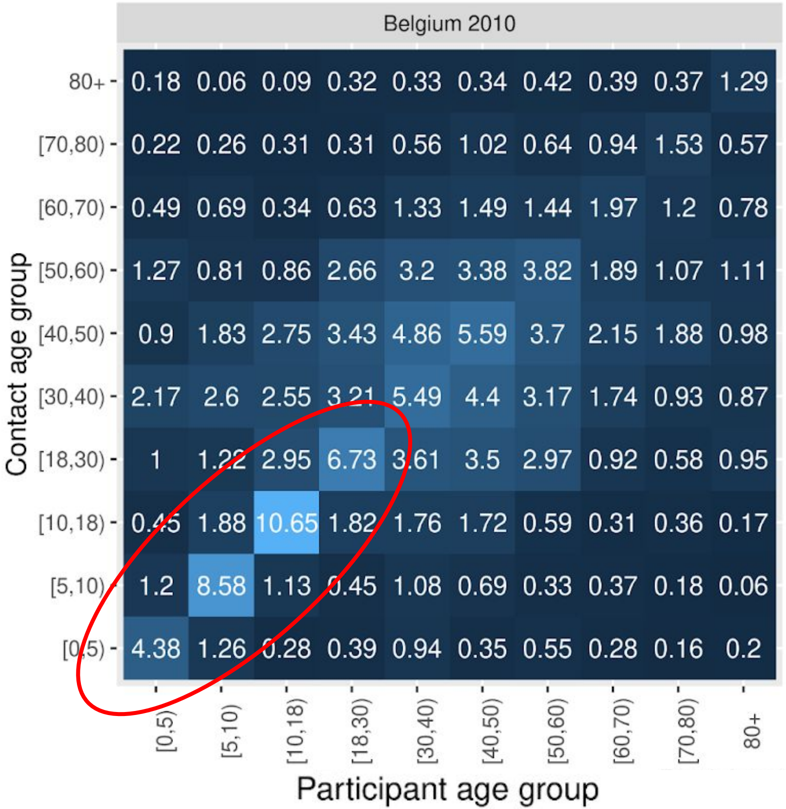


Contact matrices: pre-pandemic



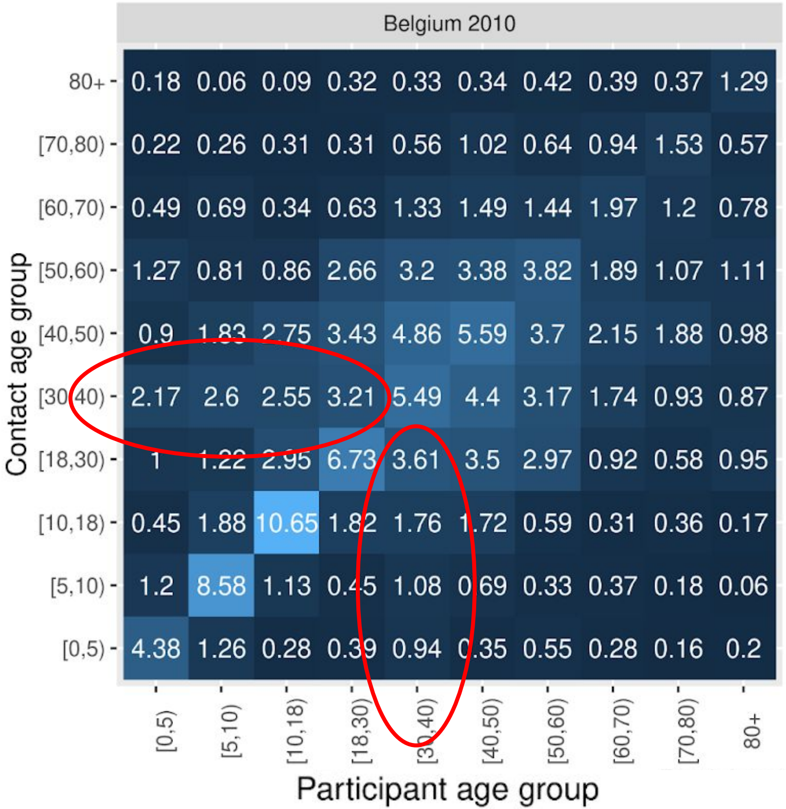
- Homophily: stronger interaction within same age category

Contact matrices: pre-pandemic



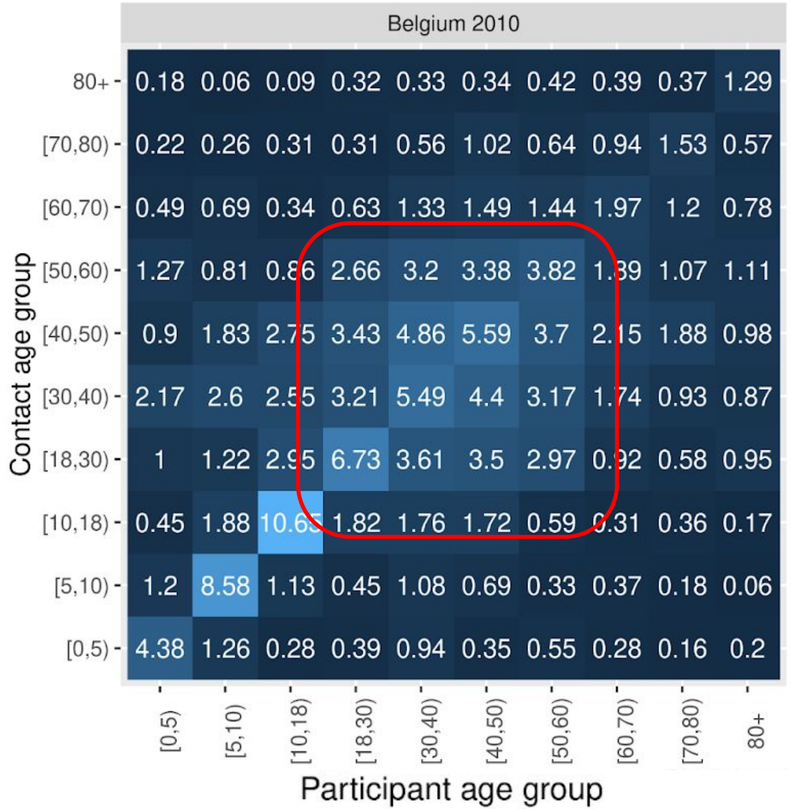
- Homophily: stronger interaction within same age category
- Children higher contact activity

Contact matrices: pre-pandemic



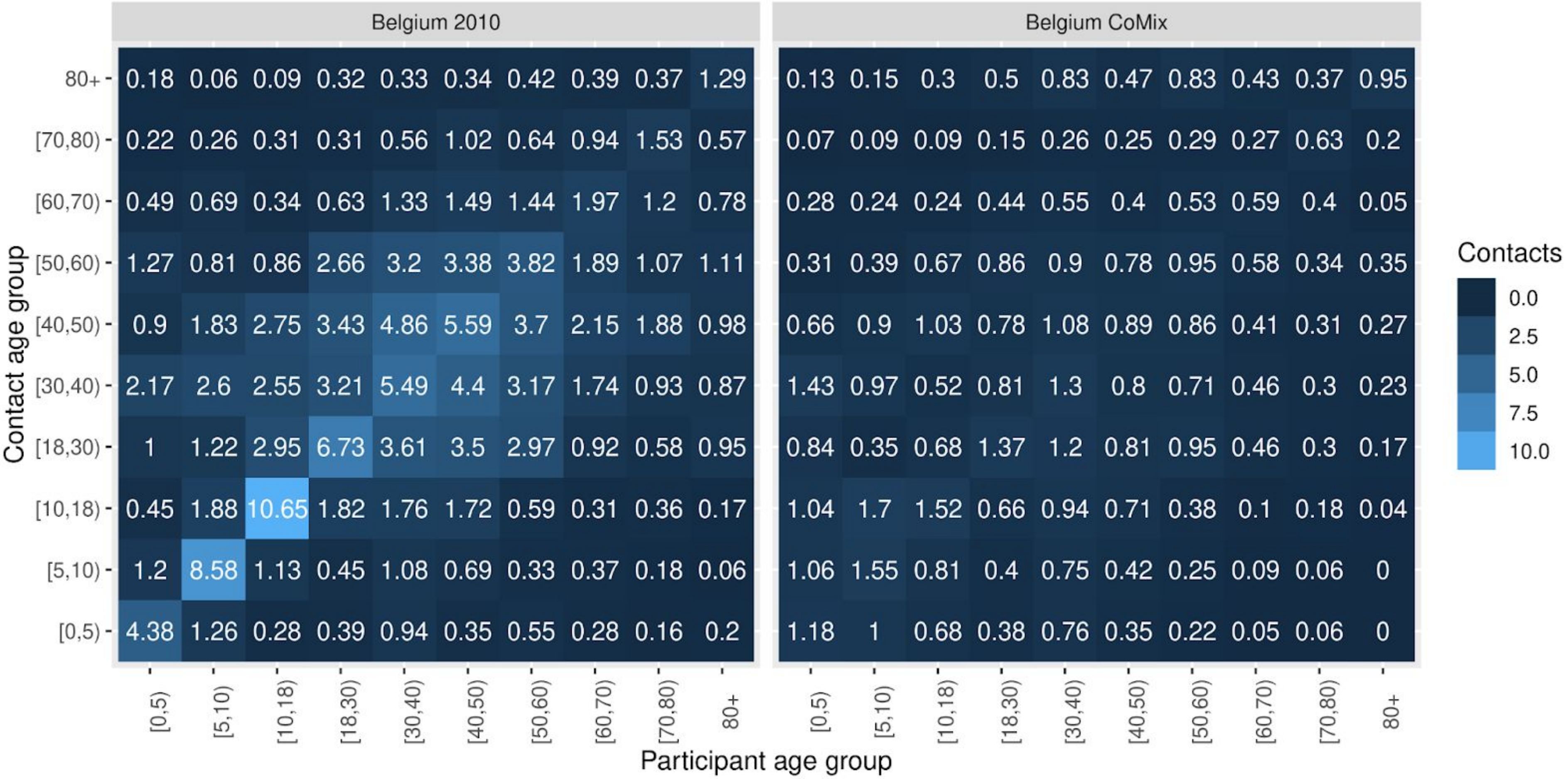
- Homophily: stronger interaction within same age category
- Children higher contact activity
- Parents-children interactions

Contact matrices: pre-pandemic

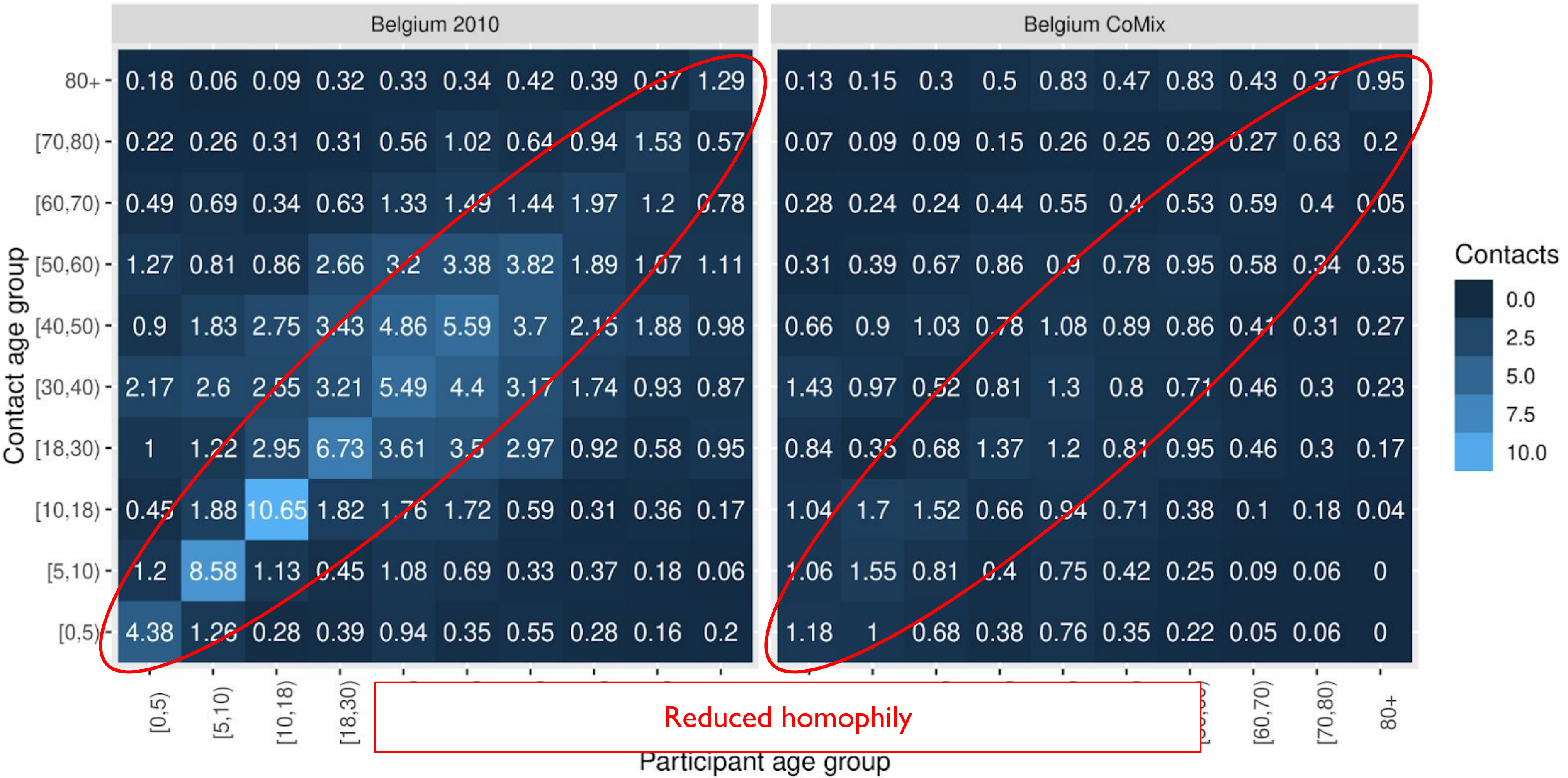


- Homophily: stronger interaction within same age category
- Children higher contact activity
- Parents-children interactions
- Working age interactions

Contact matrices: pre vs post pandemic



Contact matrices: pre vs post pandemic



Contact matrices: pre vs post pandemic

Next Generation Matrix (NGM)

$$K_{ij} =$$

Average number of secondary infections in age class *i* caused by one single infectious individual in age class *j*

Next generation matrix

Next Generation Matrix (NGM)

$$K_{ij} =$$

Average number of secondary infections in age class *i* caused by one single infectious individual in age class *j*

Next generation matrix

$$\boxed{R_0 = \mathrm{MaxEigenvalue}[K]}$$

Diekmann et al. (2010)

Next Generation Matrix and social contacts

Social contact hypothesis: Contact rates are proportional to transmission rates

Next Generation Matrix and social contacts

Social contact hypothesis: Contact rates are proportional to transmission rates

Population size

Mean duration of infectiousness

$$K_{ij} = \frac{ND}{L} \times q \times C_{ij}$$

Next generation matrix

Life expectancy

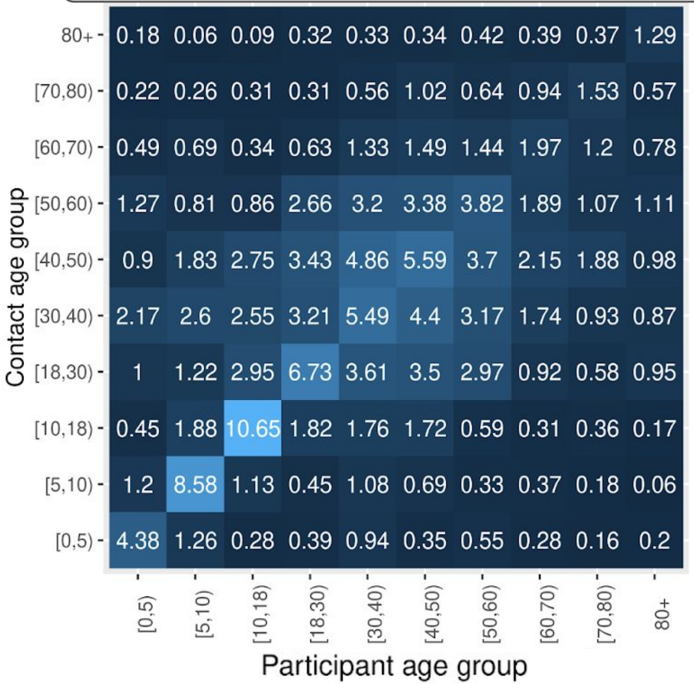
Proportionality factor

Contact matrix

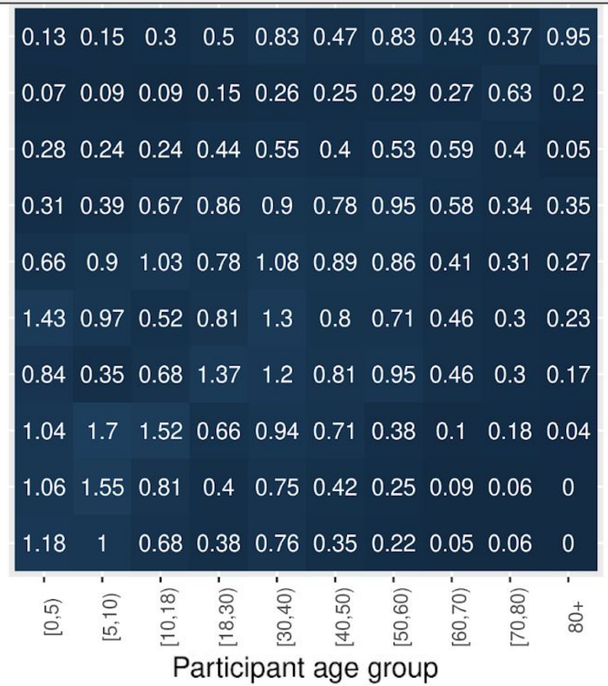
Assessing the impact of NPIs

Example: We want to assess the impact of a lockdown

Contact matrix before lockdown



Contact matrix after lockdown



Assessing the impact of NPIs

$$\frac{R_0^{\text{new}}}{R_0^{\text{old}}} =$$

Assessing the impact of NPIs

$$\frac{R_0^{\mathrm{new}}}{R_0^{\mathrm{old}}} = \frac{\mathrm{MaxEigenval}[K^{\mathrm{new}}]}{\mathrm{MaxEigenval}[K^{\mathrm{old}}]}$$

Assessing the impact of NPIs

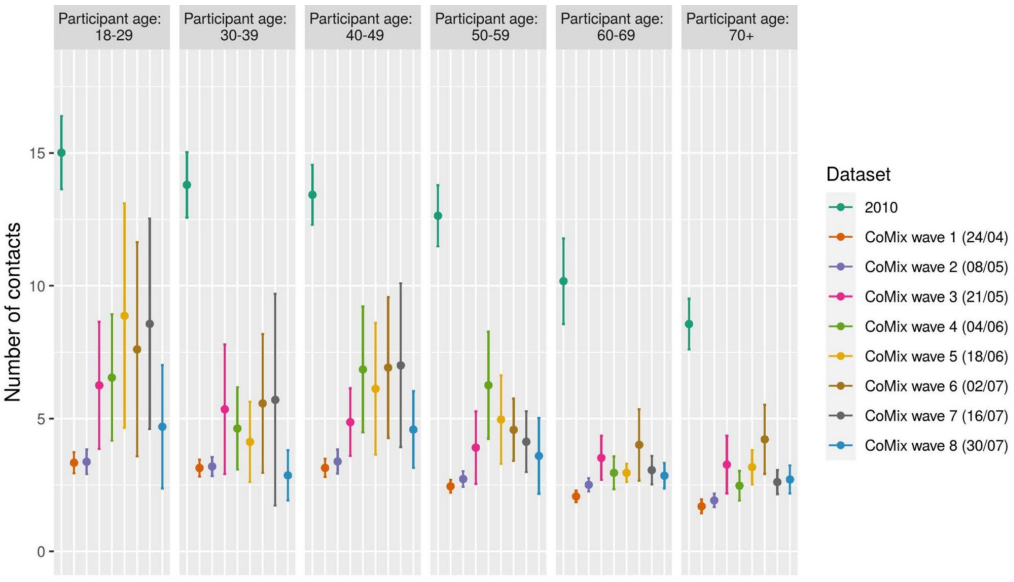
$$\frac{R_0^{\text{new}}}{R_0^{\text{old}}} = \frac{\text{MaxEigenval}[K^{\text{new}}]}{\text{MaxEigenval}[K^{\text{old}}]} = \frac{\text{MaxEigenval}[C^{\text{new}}]}{\text{MaxEigenval}[C^{\text{old}}]}$$

Social contacts:

- ►Allow the estimation of R_0 reductions
- ►Include compliance to social distancing

Longitudinal studies

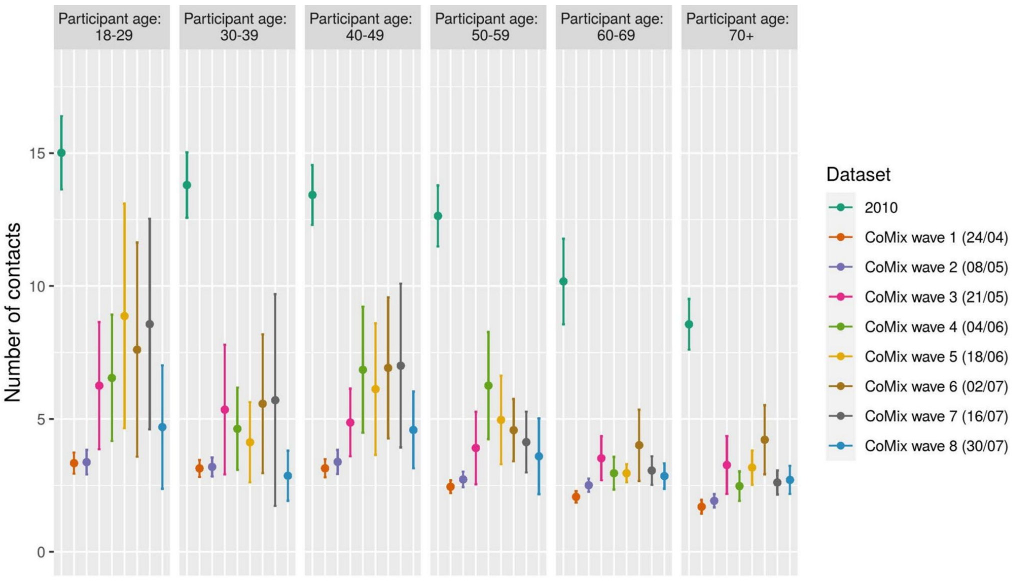
Contacts



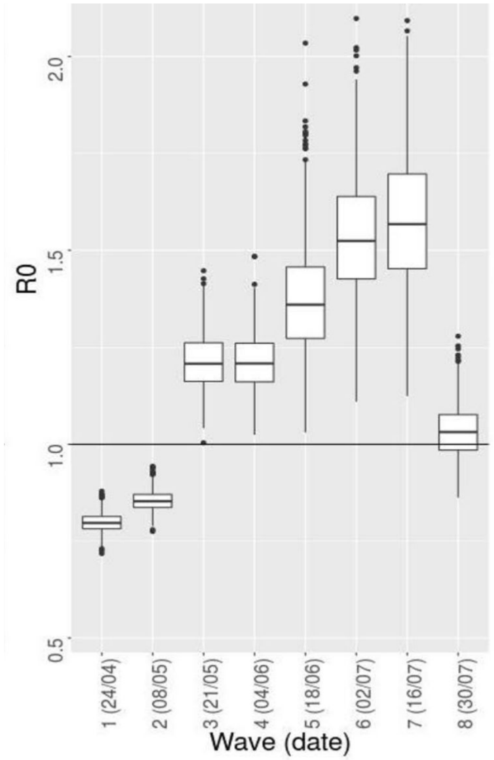
Coletti et al (2021).

Longitudinal studies

Contacts



R_0



Coletti et al (2021).

Extension: Heterogeneous q parameter

Factors related to susceptible individual
(e.g. susceptibility, mask use, …)

$$q_{ij} = \tilde{q} \cdot a_i \cdot h_j$$

Factors related to infectious individual
(e.g. infectiousness, mask use, …)

Franco et al. (2021)

Extension: Heterogeneous *q* parameter

In matrix form:

$$\tilde{q}\ \mathrm{diag}(a_i)\mathbf{C}^T\mathrm{diag}(h_j)\ \vec{w} = R\ \vec{w}$$

We can not compute R, as it depends on several unknowns (mostly q).
However, if we consider the ratio between element *i* and *j* we can get rid of q.

Extension: Heterogeneous *q* parameter

Contact matrix
(known from data)

Relative incidence
(known from testing)

$$\frac{[\mathrm{diag}(a_i)\mathbf{C}^T\mathrm{diag}(h_j)\ \vec{w}]_i}{[\mathrm{diag}(a_i)\mathbf{C}^T\mathrm{diag}(h_j)\ \vec{w}]_j} = \frac{w_i}{w_j}$$

Extension: Heterogeneous q parameter

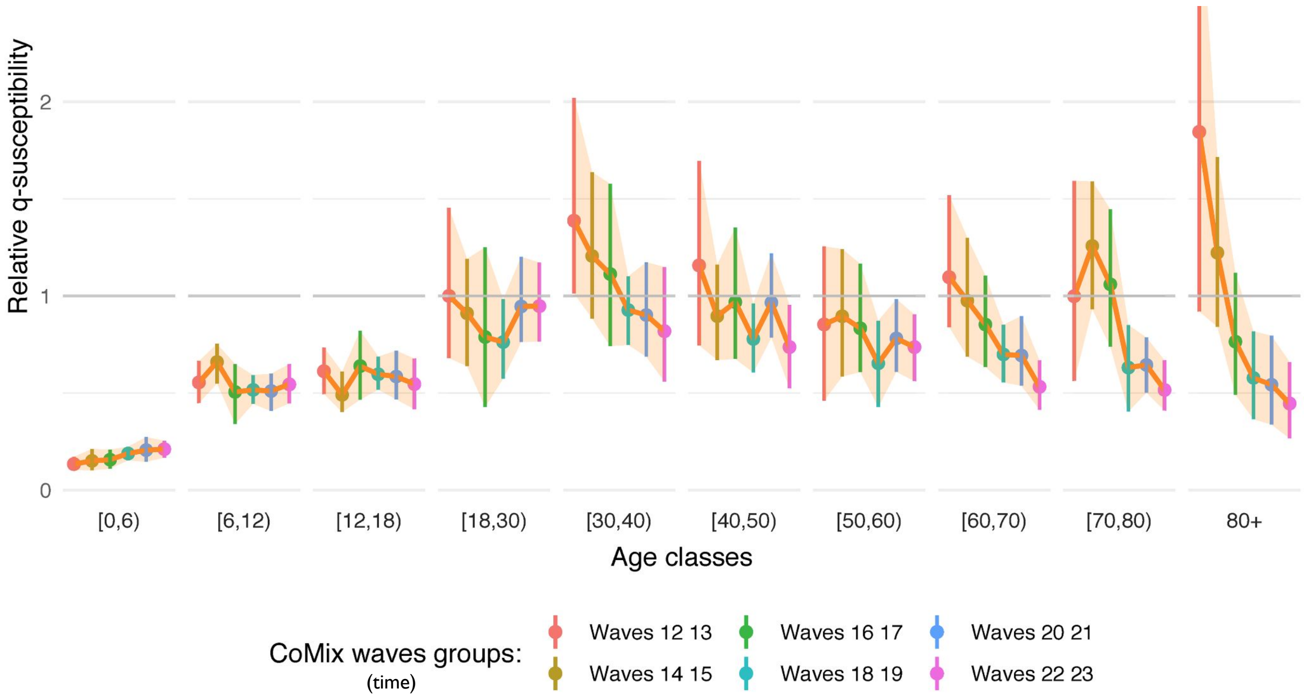
Contact matrix
(known from data)

Relative incidence
(known from testing)

$$\frac{[\mathrm{diag}(a_i)\mathbf{C}^T\mathrm{diag}(h_j)\ \vec{w}]_i}{[\mathrm{diag}(a_i)\mathbf{C}^T\mathrm{diag}(h_j)\ \vec{w}]_j} = \frac{w_i}{w_j}$$

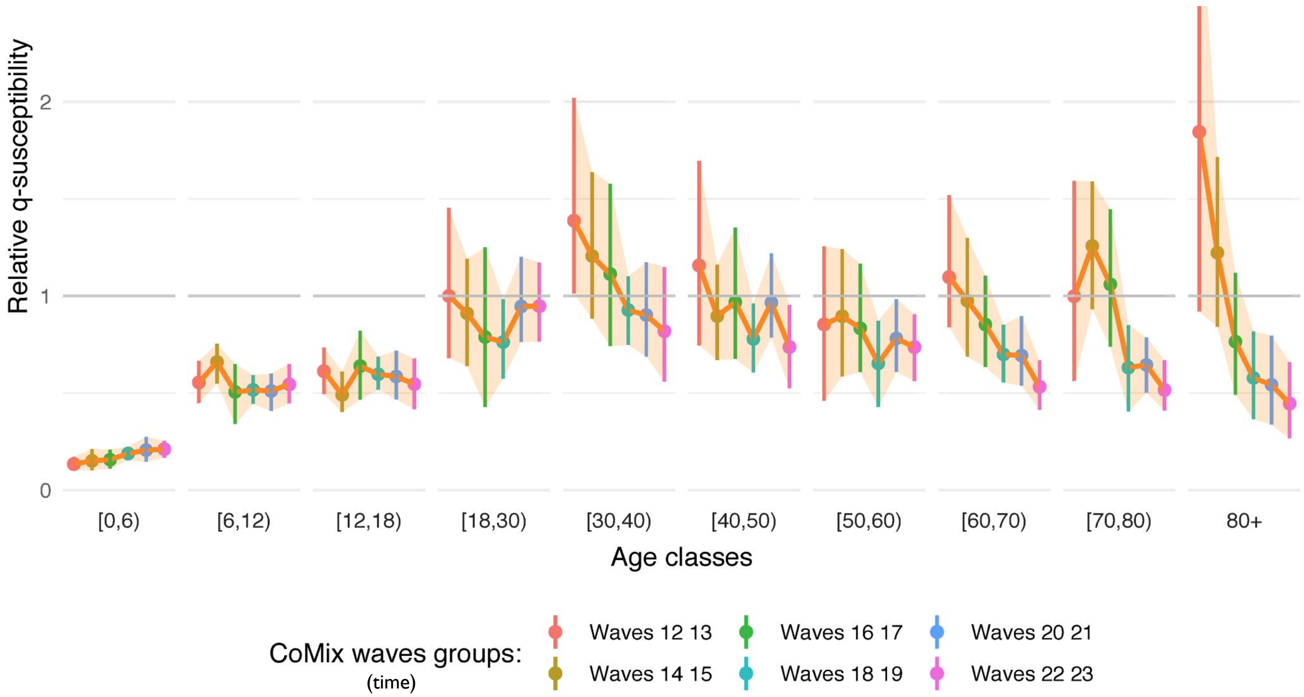
If we assume values for a_i $\Rightarrow$ we can estimate h_j

Time-varying q-susceptibility



Franco et al. (2021)

Time-varying q-susceptibility



- ► Different age-specific susceptibility of COVID-19
- ► Time varying (e.g. vaccines,)

Franco et al. (2021)

Where to get contact matrices: SOCRATES

R Library + Rshiny app to generate contact matrices, build upon SocialMixr

Takes care of:

- Country demography and age imputation
- Post stratification weights for age and day of the week
- Participants bootstrap

Generates:

- Location specific contact matrices (symmetric, per capita, …)
- Participant weights

Willem et al. (2020)
Verelst et al. (2021)

CoMix Socrates

SOCRATES CoMix

www.socialcontactdata.org/socrates

The goal of the CoMix project has been to measure social distancing during the COVID-19 pandemic. This tool is part of the SOCRATES initiative

Country

Austria 2020 CoMix (Verelst 2021)

Wave: start [panel]

All

Age breaks (comma delimited)

0,18,60

Type of day

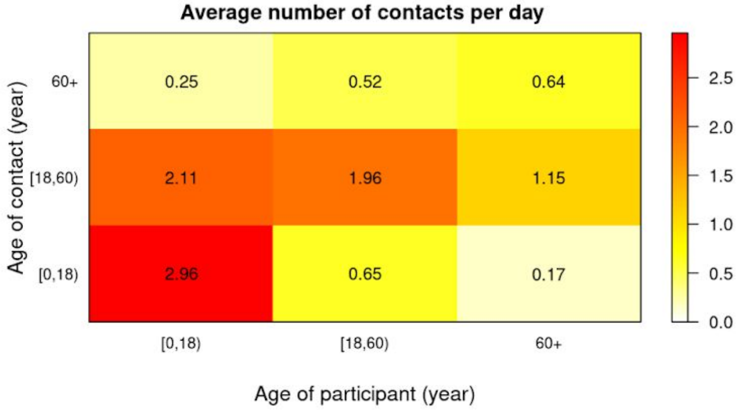
All contacts

Contact intensity

All

Gender

All



All results | Matrix per capita | Contact rates | Participants | Weights | Data sets | About CoMix | Updates

```
$matrix
      contact.age.group
          [0,18)  [18,60)       60+
  [1,] 2.9585329 2.112832 0.2523997
  [2,] 0.6470172 1.957597 0.5166117
  [3,] 0.1713386 1.145196 0.6379937

$relative_incidence
   [0,18)   [18,60)       60+
0.6284914 0.2490059 0.1225027
```

Verelst et al. (2021)

Modelling the number of contacts

In social contact survey we know several information on the participants (e.g. age, household size, vaccination status, …).

We can use this to model the number of contacts:

Modelling the number of contacts

In social contact survey we know several information on the participants (e.g. age, household size, vaccination status, …).

We can use this to model the number of contacts:

- Similar (in spirit) to linear regression
- Assess the impact of a covariate on the number of contacts (e.g. change in vaccination status)
- Include properties of contact patterns (e.g. *overdispersion*)
- (Include longitudinal nature of the survey, to account for individual variability)

Modelling the number of contacts

In social contact survey we know several information on the participants (e.g. age, household size, vaccination status, …).

We can use this to model the number of contacts:

- Similar (in spirit) to linear regression
- Assess the impact of a covariate on the number of contacts (e.g. change in vaccination status)
- Include properties of contact patterns (e.g. *overdispersion*)
- (Include longitudinal nature of the survey, to account for individual variability)

We can then *predict* the number of contacts for synthetic populations (e.g. with increased vaccination levels)

Modelling the number of contacts

In social contact survey we know several information on the participants (e.g. age, household size, vaccination status, …).

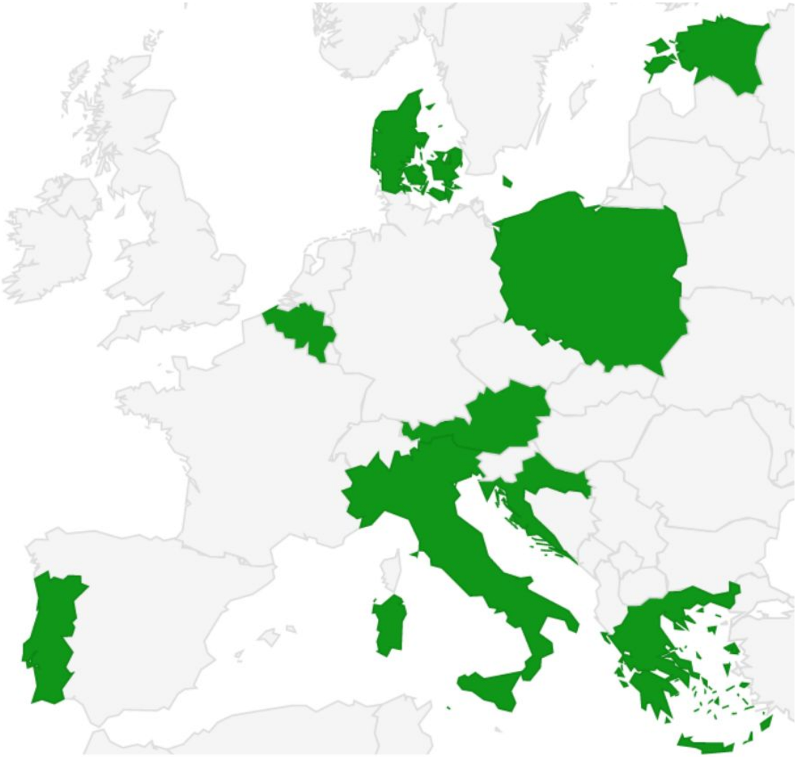
We can use this to model the number of contacts:

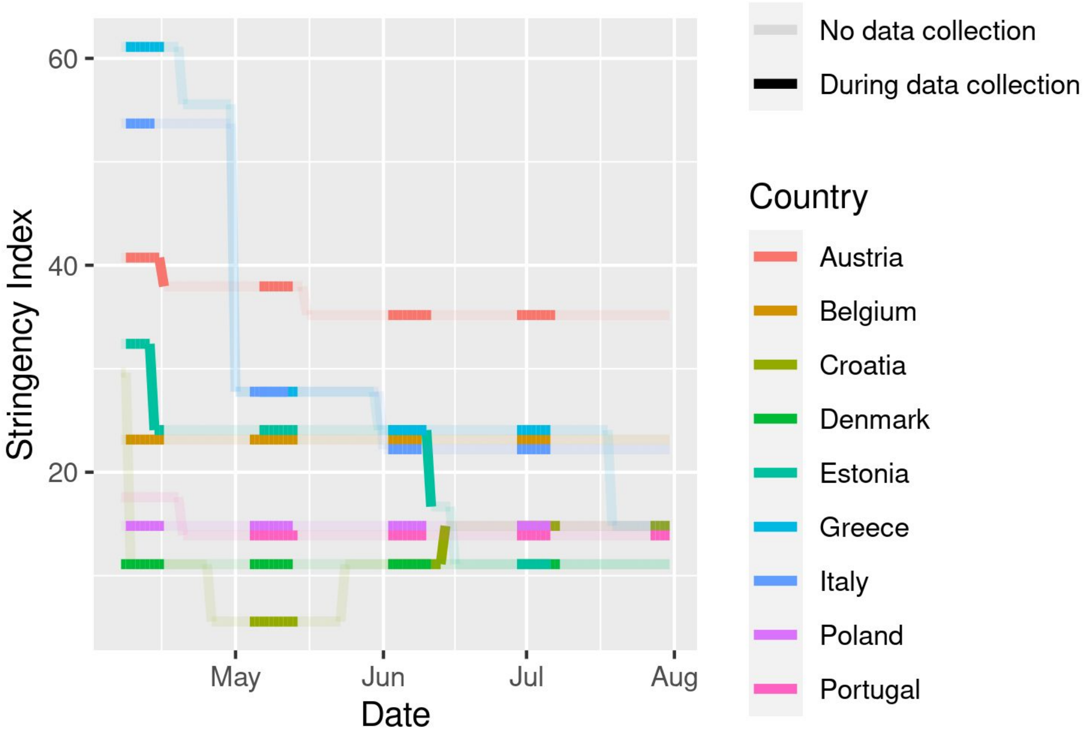
- Similar (in spirit) to linear regression
- Assess the impact of a covariate on the number of contacts (e.g. change in vaccination status)
- Include properties of contact patterns (e.g. *overdispersion*)
- (Include longitudinal nature of the survey, to account for individual variability)

We can then *predict* the number of contacts for synthetic populations
(e.g. with increased vaccination levels)

[implicit assumption: 'nothing will change' wrt parameters]

Example: CoMix 2.0





Coletti et al. in preparation

Modelling the number of contacts

Generalized linear mixed effects model with participants as random effects and accounting for overdispersion. Variables included:

- Household size, age group, day of the week (week day versus weekend), high risk status, country
- Vaccination status (any of the first dose of the vaccines)
- Perceived severity ("*COVID-19 would be a serious illness for me*")

Modelling the number of contacts

Generalized linear mixed effects model with participants as random effects and accounting for overdispersion. Variables included:

- Household size, age group, day of the week (week day versus weekend), high risk status, country
- Vaccination status (any of the first dose of the vaccines)
- Perceived severity (“*COVID-19 would be a serious illness for me*”)

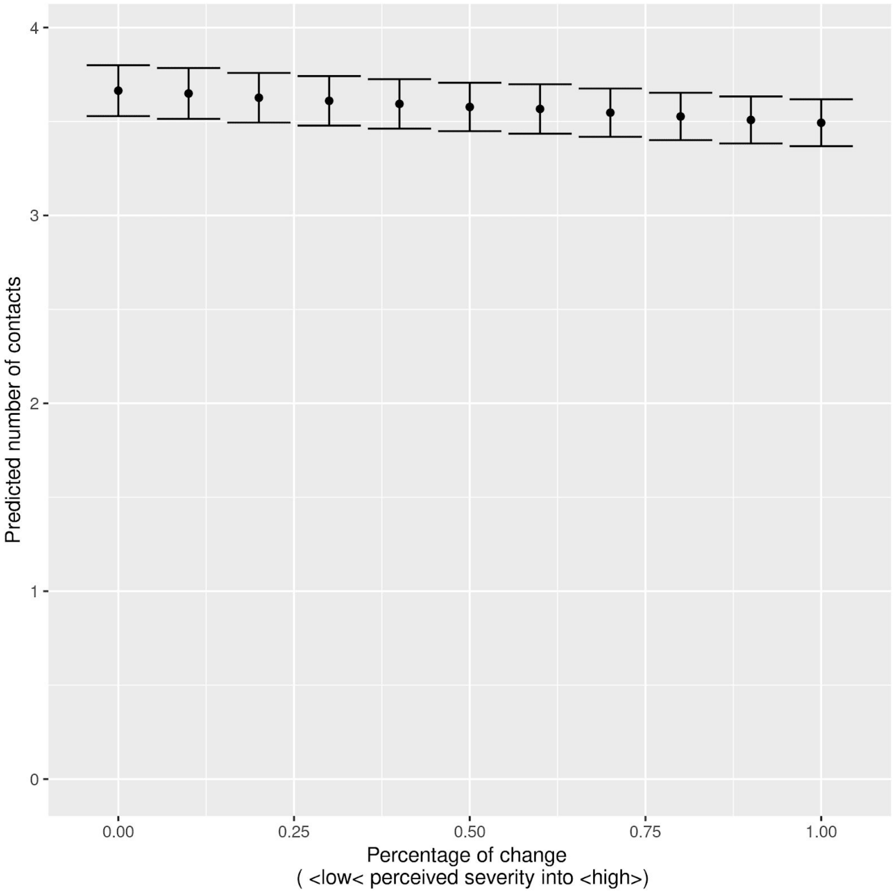
Perceived severity:

- Low perceived severity made **1.28 (95% CI 1.13 – 1.43)** times more contacts wrt high perceived severity

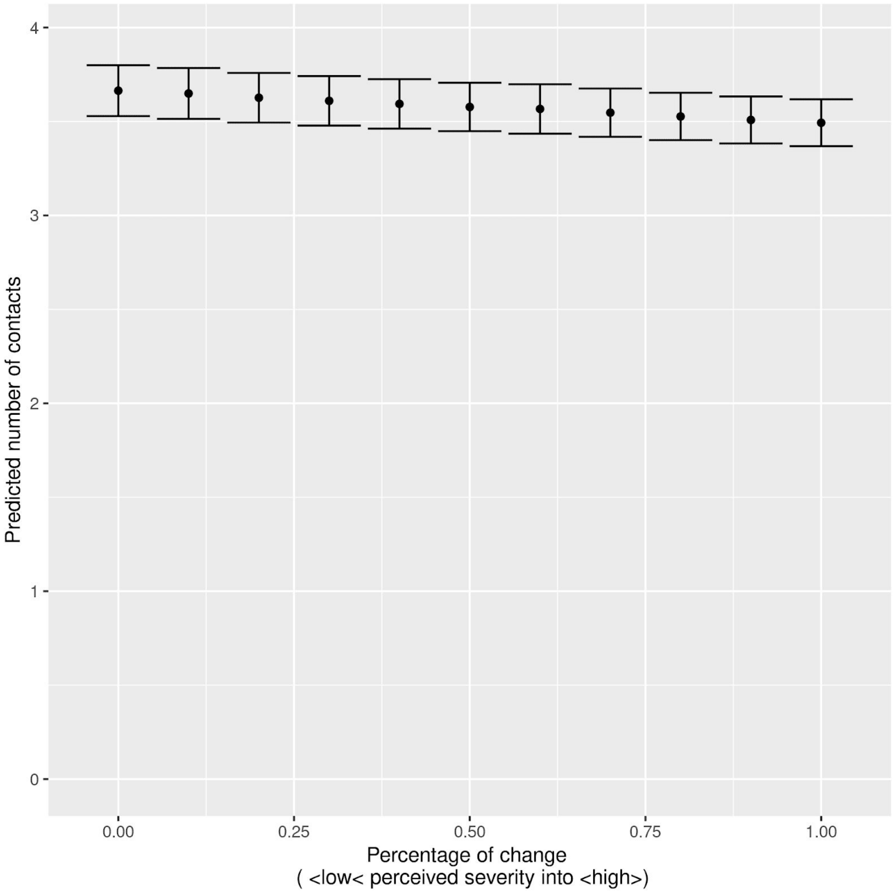
Vaccination status:

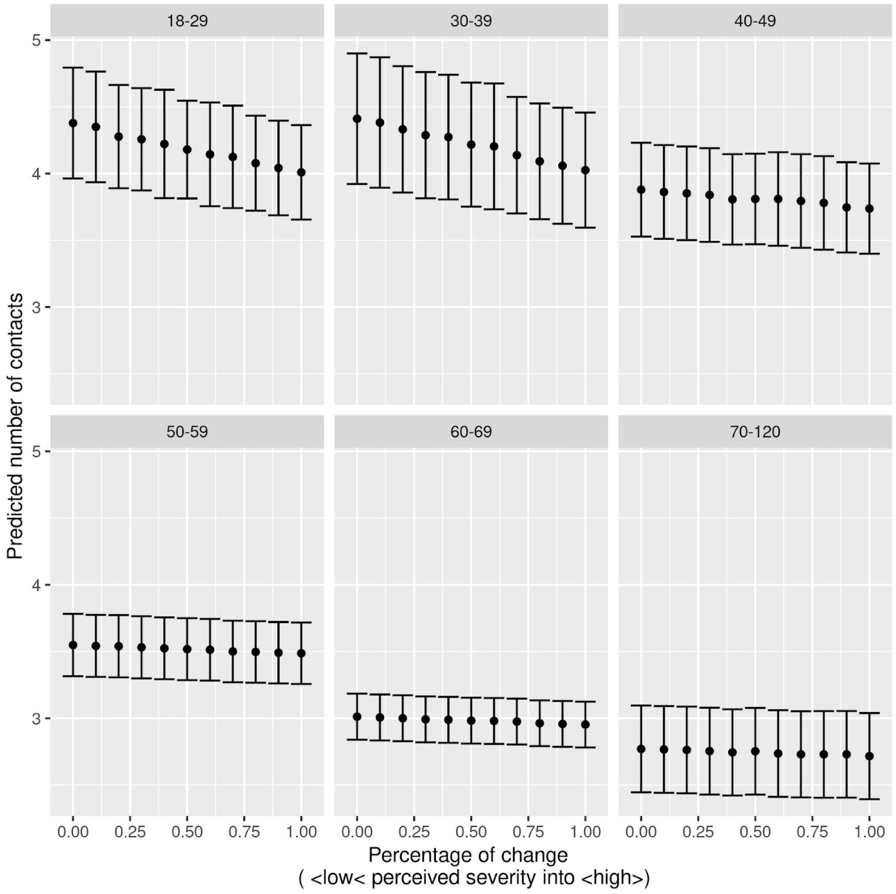
- Vaccinated made **1.15 (95% CI 1.05 – 1.24)** times more contacts wrt non-vaccinated

Scenario: increase in perceived severity



Scenario: increase in perceived severity





Including social contacts in infectious disease modelling

Recall: SIR model

$$\frac{dS}{dt} = -\beta S \frac{I}{N}$$
$$\frac{dI}{dt} = \beta S \frac{I}{N} - \mu I$$
$$\frac{dR}{dt} = \mu I$$

$$R_0 = \frac{\beta}{\mu} \begin{cases} > 1 & \text{epidemic spreads} \\ \leq 1 & \text{epidemic goes extinct} \end{cases}$$

Recall: SIR model

$$\frac{dS}{dt} = -\beta S \frac{I}{N}$$

$$\frac{dI}{dt} = \beta S \frac{I}{N} - \mu I$$

$$\frac{dR}{dt} = \mu I$$

$$R_0 = \frac{\beta}{\mu} \begin{cases} > 1 & \text{epidemic spreads} \\ \leq 1 & \text{epidemic goes extinct} \end{cases}$$

Also called "Force Of Infection" (FOI)

Denoted by λ

Force Of Infection for two age classes

Two age classes:

Children

Adults

Force Of Infection for two age classes

Children

Adults

Two age classes:

$$\lambda_i = \beta \sum_j C_{ij} P_j$$

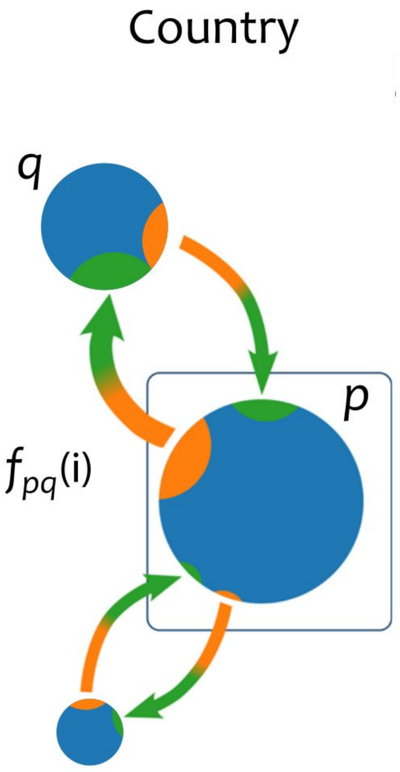
Force of infection for age class i

Per contact Infection probability

contact matrix

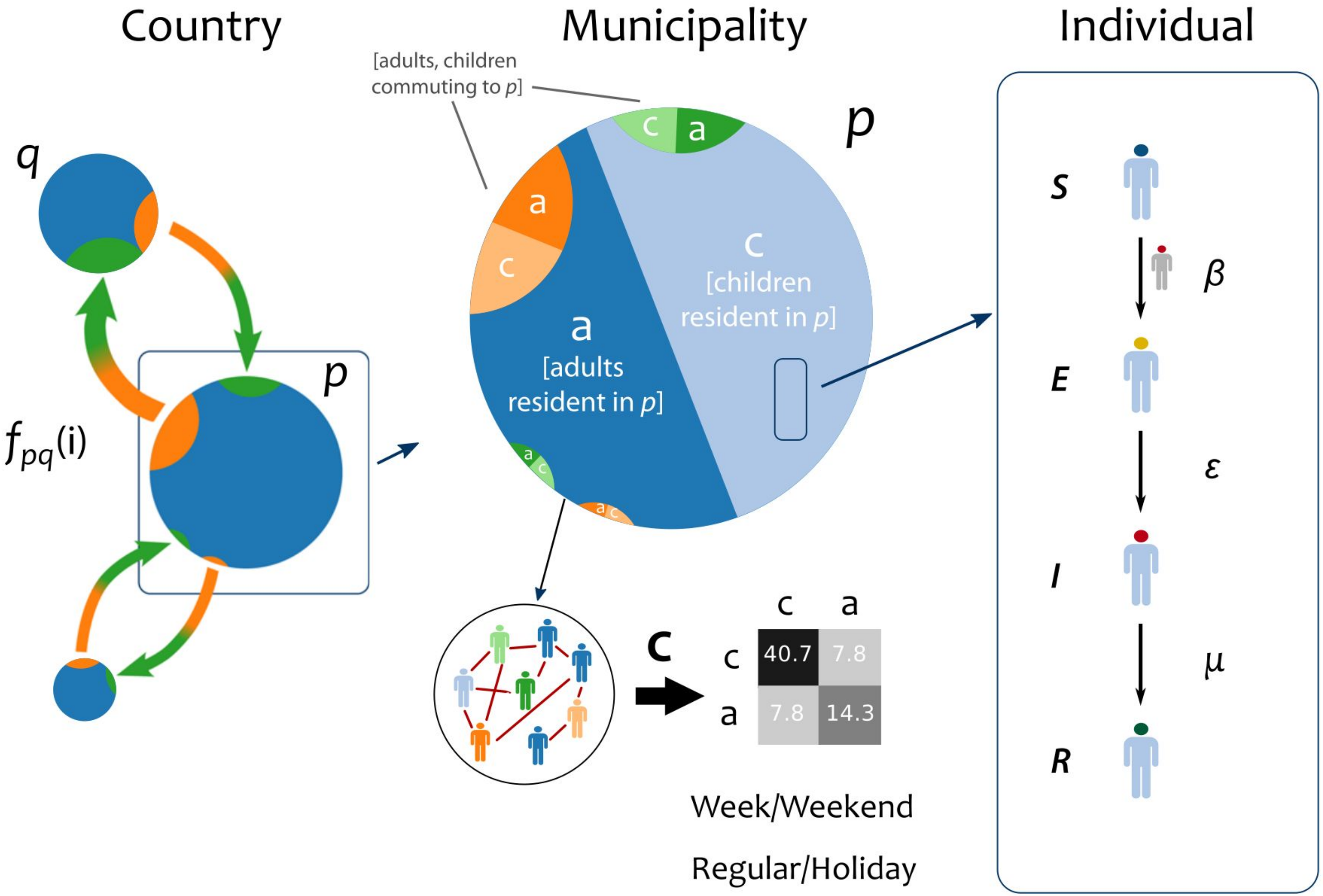
proportion of infected individuals in age class j

Contacts in space: metapopulation models

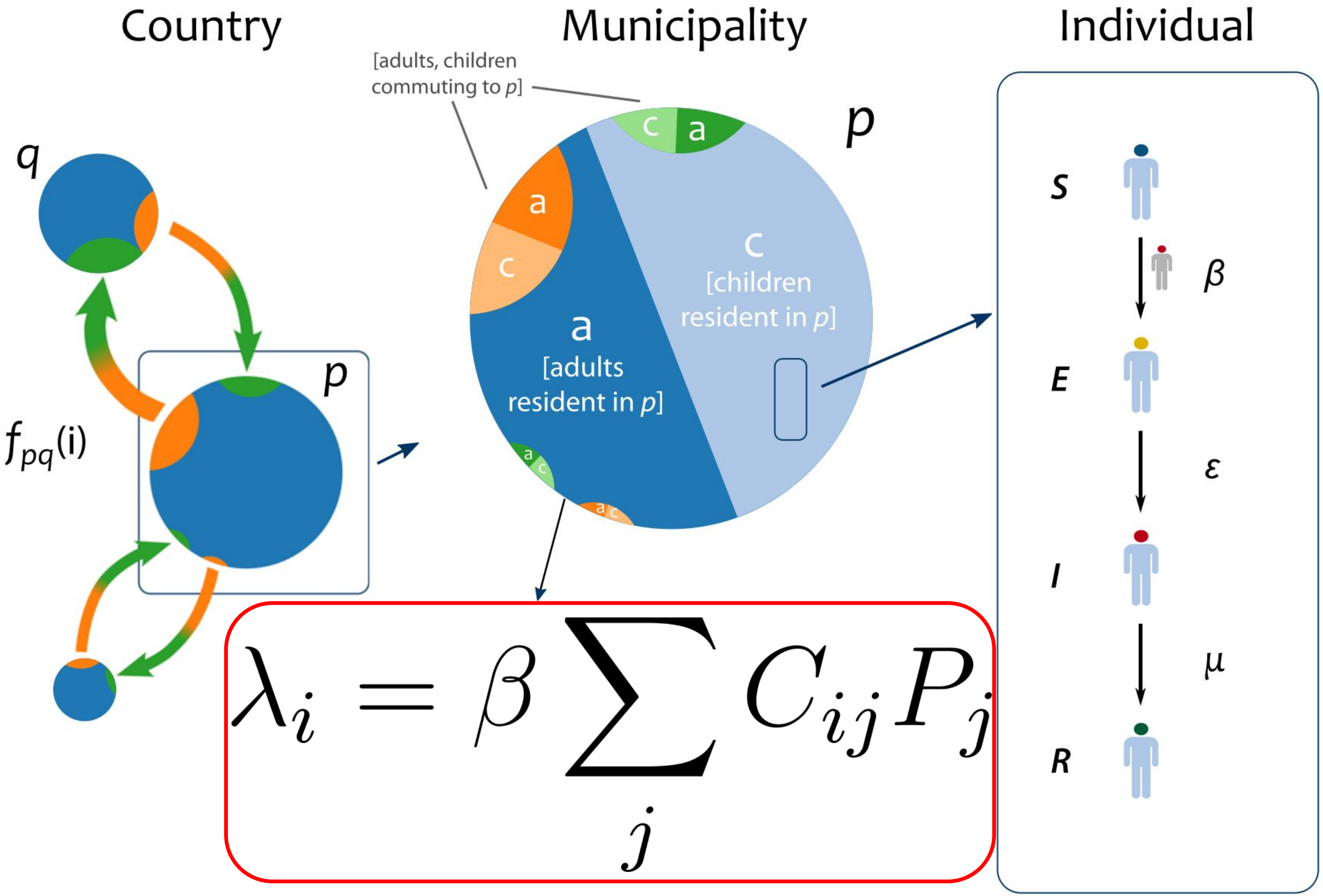


De Luca et al. (2018)

Contacts in space: metapopulation models



Contacts in space: metapopulation models



$$\lambda_i = \beta \sum_j C_{ij} P_j$$

Scenarios defined via contact matrix & mobility

$$C_{\text{asympt}} = C_{\text{home}} + p_{\text{w}} \cdot C_{\text{work}} + p_{\text{s}} \cdot C_{\text{school}} + p_{\text{o}} \cdot C_{\text{leisure}} + p_{\text{w}} \cdot C_{\text{transport}} + p_{\text{o}} \cdot C_{\text{other}} \tag{3}$$

$$C_{\text{sympt}} = C_{\text{home}} + p_{\text{w}} \cdot 0.09 \cdot C_{\text{work}} + p_{\text{o}} \cdot 0.06 \cdot C_{\text{leisure}} + p_{\text{w}} \cdot 0.13 \cdot C_{\text{transport}} + p_{\text{o}} \cdot 0.25 \cdot C_{\text{other}} \tag{4}$$

Scenarios defined via contact matrix & mobility

$$C_{\text{asympt}} = C_{\text{home}} + p_{\text{w}} \cdot C_{\text{work}} + p_{\text{s}} \cdot C_{\text{school}} + p_{\text{o}} \cdot C_{\text{leisure}} + p_{\text{w}} \cdot C_{\text{transport}} + p_{\text{o}} \cdot C_{\text{other}} \tag{3}$$

$$C_{\text{sympt}} = C_{\text{home}} + p_{\text{w}} \cdot 0.09 \cdot C_{\text{work}} + p_{\text{o}} \cdot 0.06 \cdot C_{\text{leisure}} + p_{\text{w}} \cdot 0.13 \cdot C_{\text{transport}} + p_{\text{o}} \cdot 0.25 \cdot C_{\text{other}} \tag{4}$$

+

p_w also modulates working commuters mobility

p_s also modulates school commuters mobility

Scenarios defined via contact matrix & mobility

$$C_{\text{asympt}} = C_{\text{home}} + p_{\text{w}} \cdot C_{\text{work}} + p_{\text{s}} \cdot C_{\text{school}} + p_{\text{o}} \cdot C_{\text{leisure}} + p_{\text{w}} \cdot C_{\text{transport}} + p_{\text{o}} \cdot C_{\text{other}} \quad (3)$$

$$C_{\text{sympt}} = C_{\text{home}} + p_{\text{w}} \cdot 0.09 \cdot C_{\text{work}} + p_{\text{o}} \cdot 0.06 \cdot C_{\text{leisure}} + p_{\text{w}} \cdot 0.13 \cdot C_{\text{transport}} + p_{\text{o}} \cdot 0.25 \cdot C_{\text{other}} \quad (4)$$

+

p_w also modulates working commuters mobility

p_s also modulates school commuters mobility

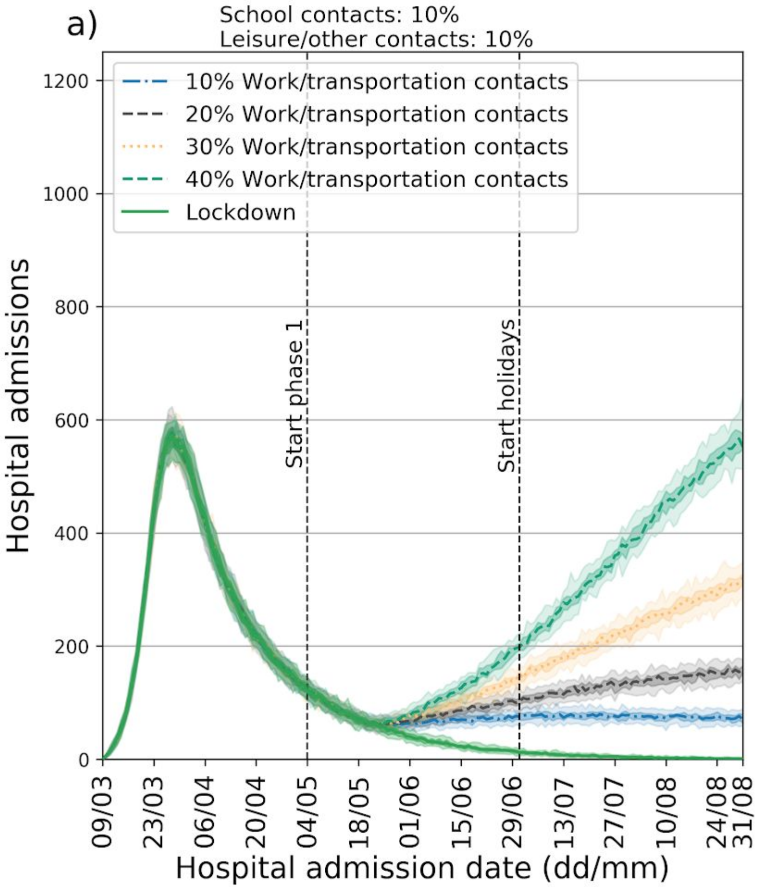
Table 2 Timing and concepts of lockdown relief

	Timing start/end	Work & transportation contacts (%) (p_w)	School contacts (%) (p_s)	Mobility adults (%)(m_a)	Mobility children(%) (m_c)	Leisure & other contacts (%)(p_o)
Phase 1 (work)	04-05/17-05	**20 [10-40]**	0	**20 [10-40]**	0	10
Phase 2 (school)	18-05/07-06	20 [10-40]	**20 [10-40]**	20 [10-40]	**20 [10-40]**	10
Phase 3 (leisure)	08-06/30-06	20 [10-40]	20 [10-40]	20 [10-40]	20 [10-40]	**20 [10-40]**
Summer holidays	01-07/31-08	20 [10-40]	**0**	20 [10-40]	**0**	20 [10-40]

Each phase is implemented incrementally with respect to the previous ones. Bold values highlight the changes with respect to the previous phase (i.e. the previous row). Intermediate parameter values are reported, with the full range between squared brackets. See sections 'Population mixing' and 'Population mobility' for parameter definitions

Impact of different phases: phase 1 (telework)

Phase 1: Work & transportation



- Strong effect on epidemic profile
- Three weeks to see difference between implementations
- School holidays have minor to no effect (10% school contacts)

Coletti et al. 2021b

Estimating contact data from other sources

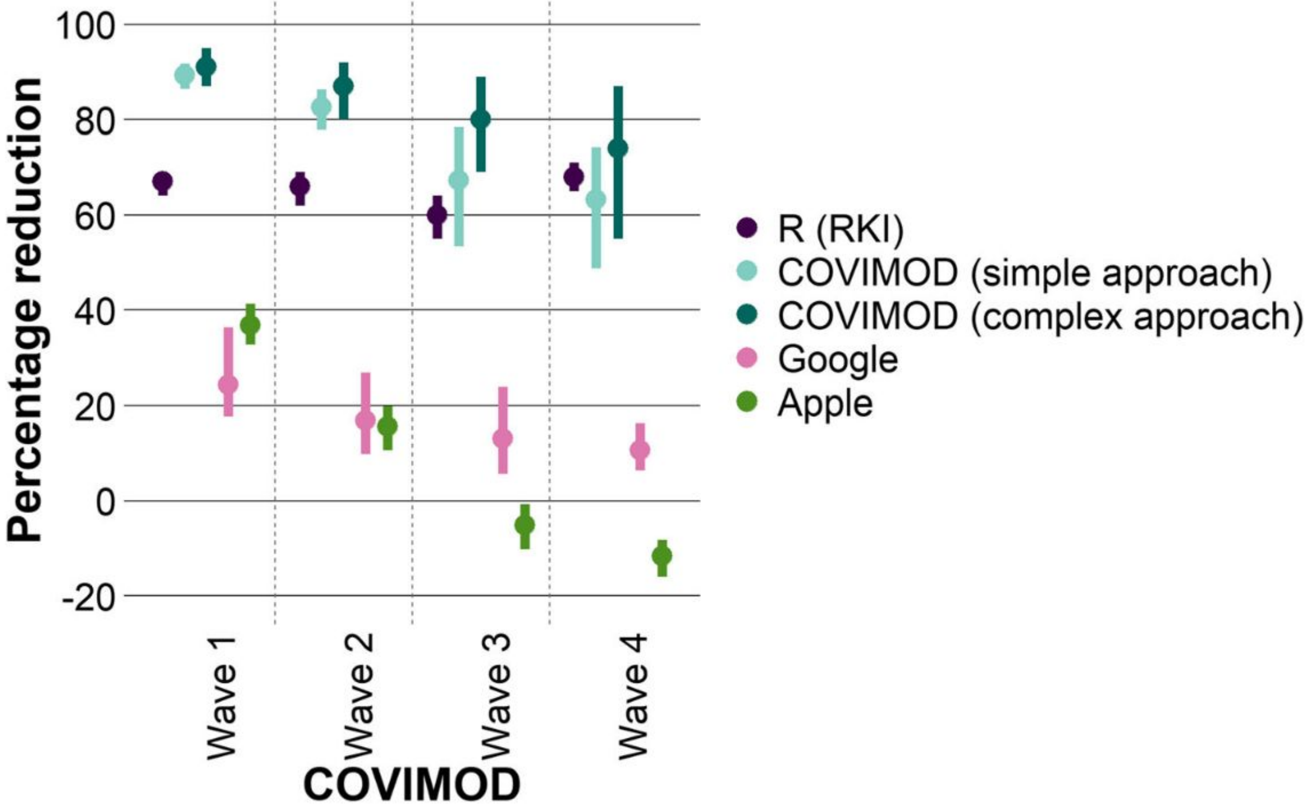
Contact reduction from Google mobility

If contact data is not available, one can use baseline data, modulated by other data sources

Google/Apple mobility data:

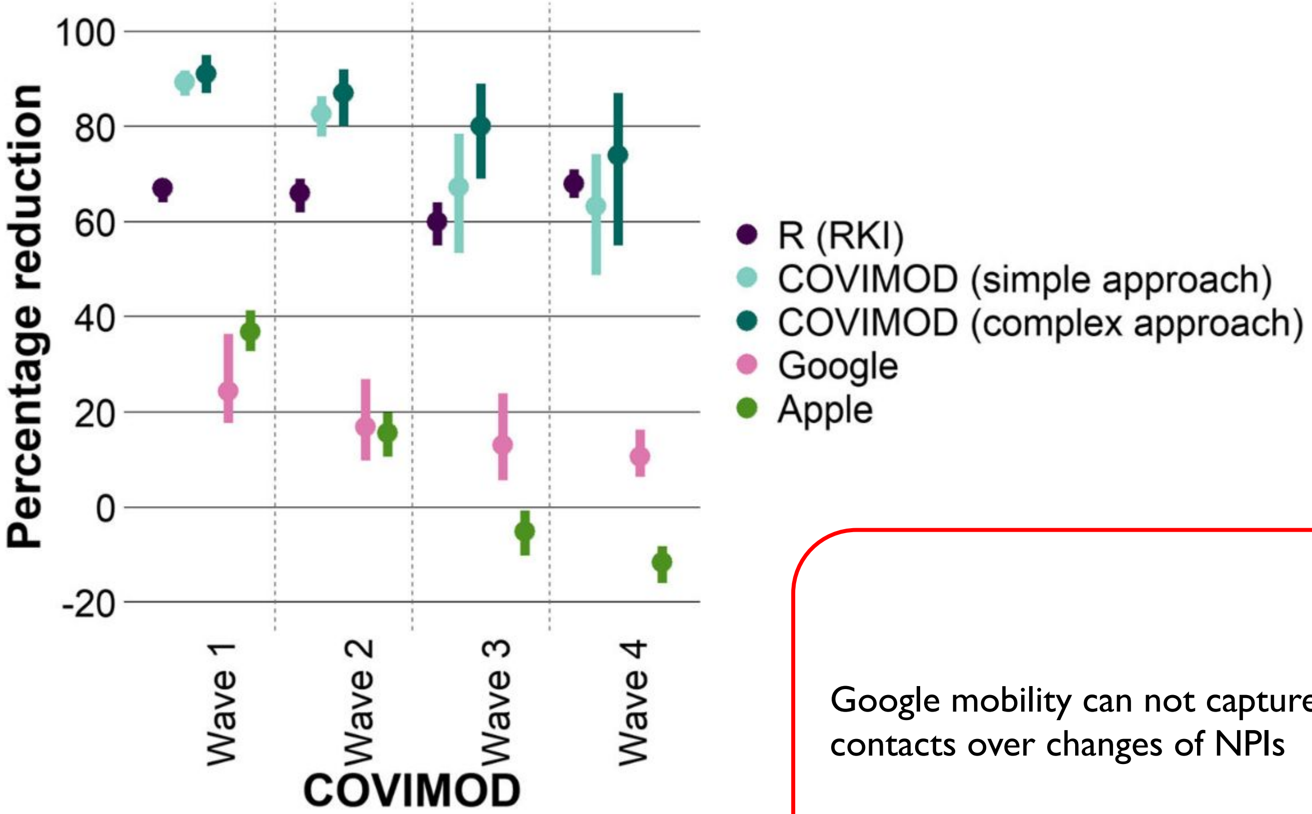
- **Time use** data
- Aggregated from individual data
- Locations are grouped
- Reports reduction/increase wrt baseline

Contact reduction from Google mobility



Tomori et al. (2021)

Contact reduction from Google mobility



Google mobility can not capture changes in contacts over changes of NPIs

Time-use surveys

From time spent at different location, one can estimate the exposure via the *proportionate time mixing* assumption:

Time-use surveys

From time spent at different location, one can estimate the exposure via the *proportionate time mixing* assumption:

$$e_{ij} = \underbrace{t_{ij}}_{\text{exposure outside home}} + \underbrace{h_{ij}}_{\text{exposure at home}}.$$

Zagheni et al. (2008)

Time-use surveys

From time spent at different location, one can estimate the exposure via the *proportionate time mixing* assumption:

$$e_{ij} = t_{ij} + h_{ij}.$$

$$t_{ij} = \sum_{s=1}^{S} \sum_{l=1}^{L} k_i^{ls} \frac{k_j^{ls}}{k_\cdot^{ls}} d^s$$

number of people in location *l* at time slot *s* of age class *j*

duration of time slot *s*

total number of people in location *l* at time slot *s*

Zagheni et al. (2008)

Time-use surveys

From time spent at different location, one can estimate the exposure via the *proportionate time mixing* assumption:

$$e_{ij} = t_{ij} + h_{ij}.$$

number of people in location *l* at time slot *s* of age class *j*

$$t_{ij} = \sum_{s=1}^{S} \sum_{l=1}^{L} k_i^{ls} \frac{k_j^{ls}}{k_\cdot^{ls}} d^s$$

duration of time slot *s*

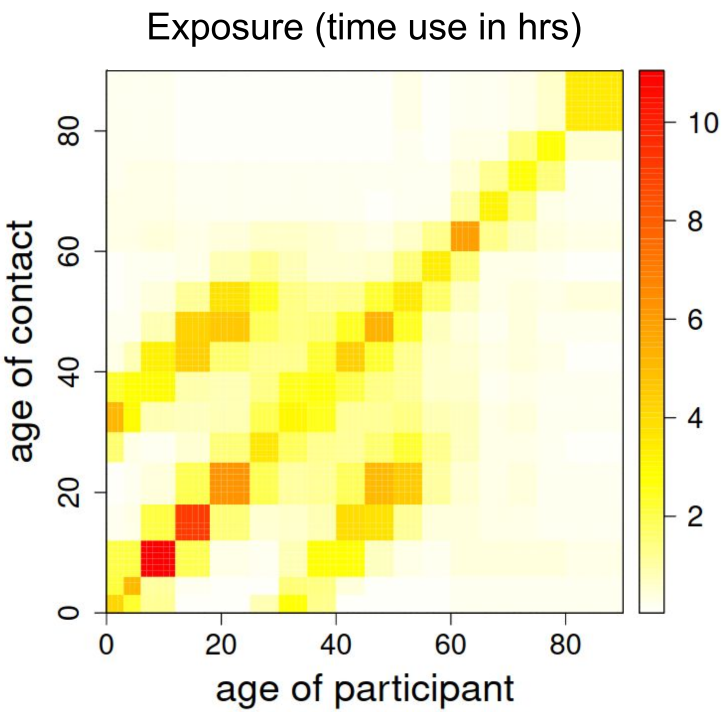
various time slots

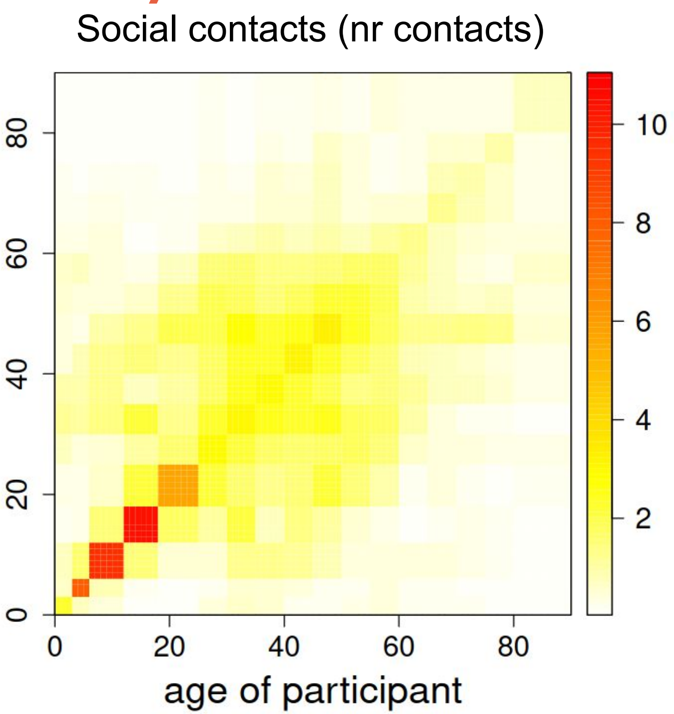
various locations

total number of people in location *l* at time slot *s*

Zagheni et al. (2008)

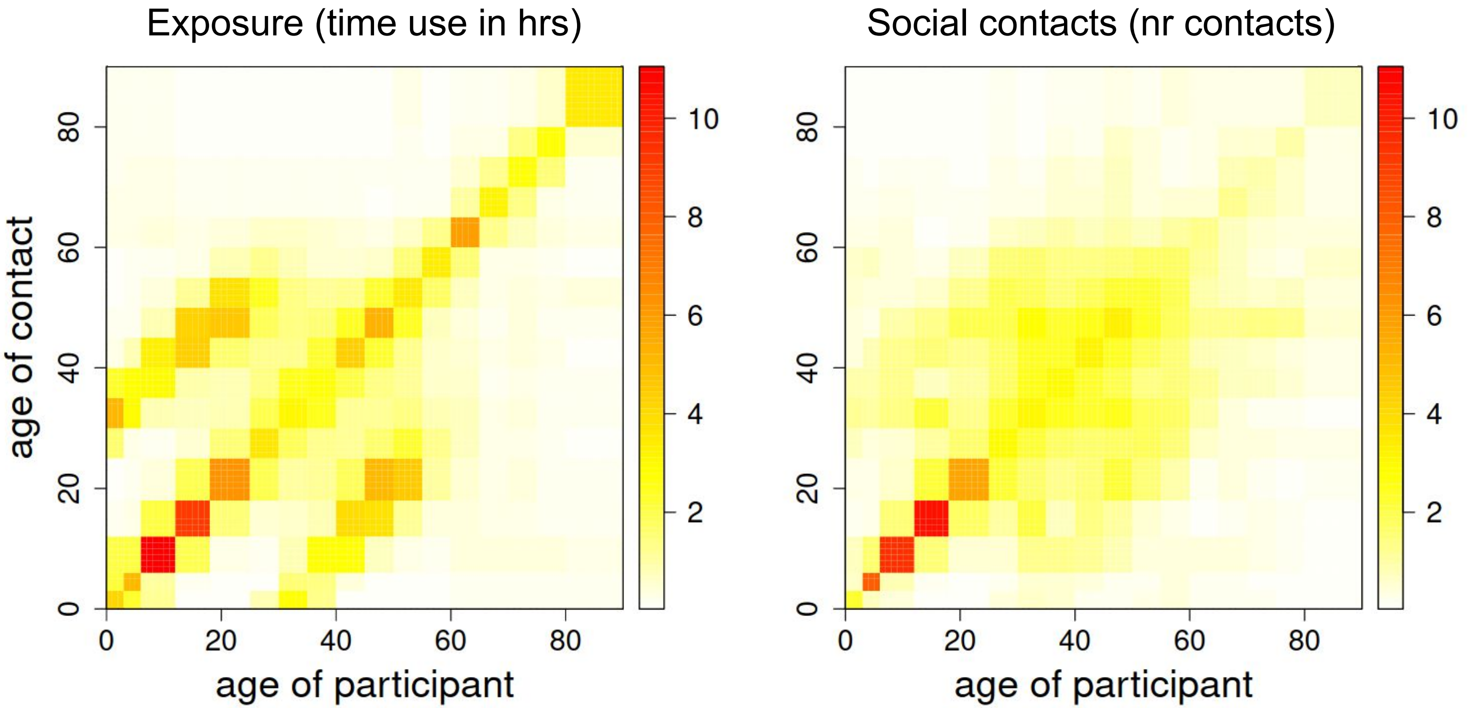
Time-use surveys





Hoang et al. (2022).

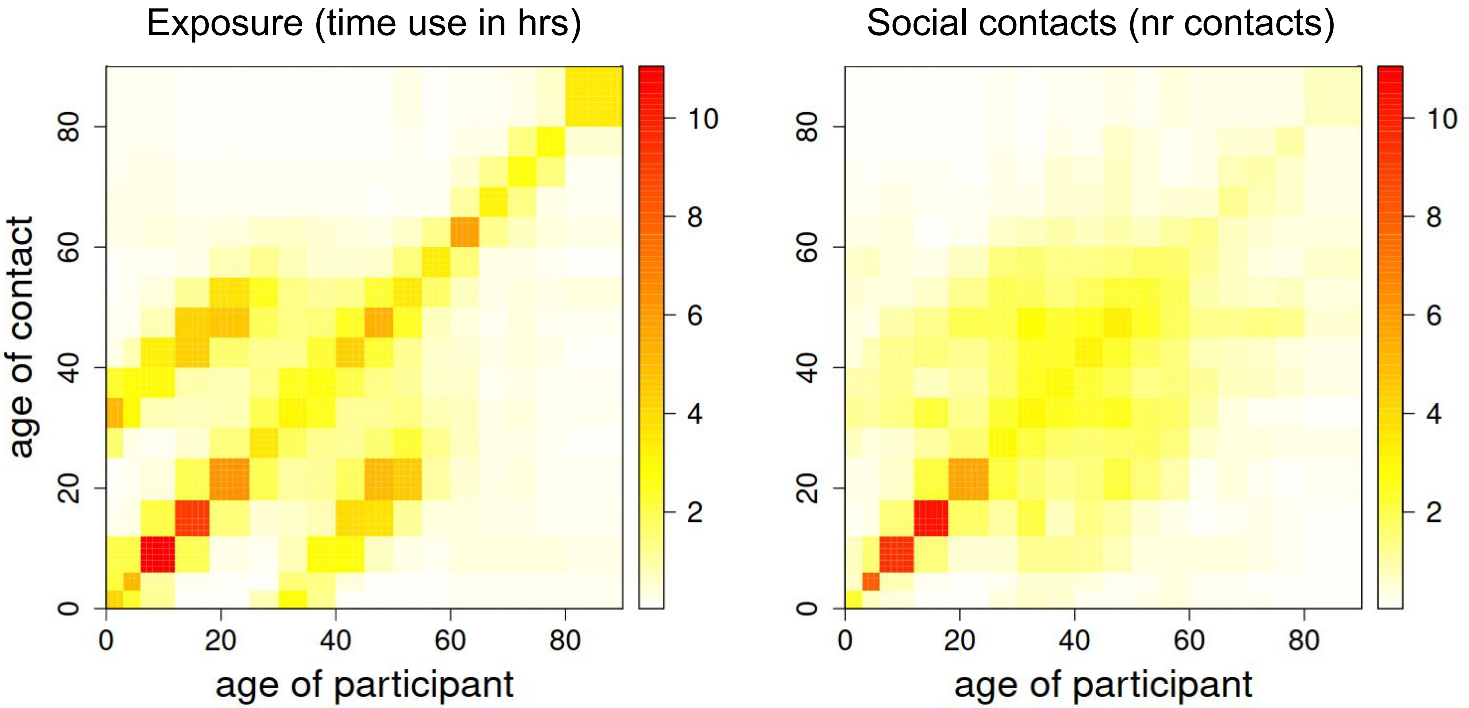
Time-use surveys



- ► Similar patterns, overall
- ► Exposure dominated by time at home

Hoang et al. (2022).

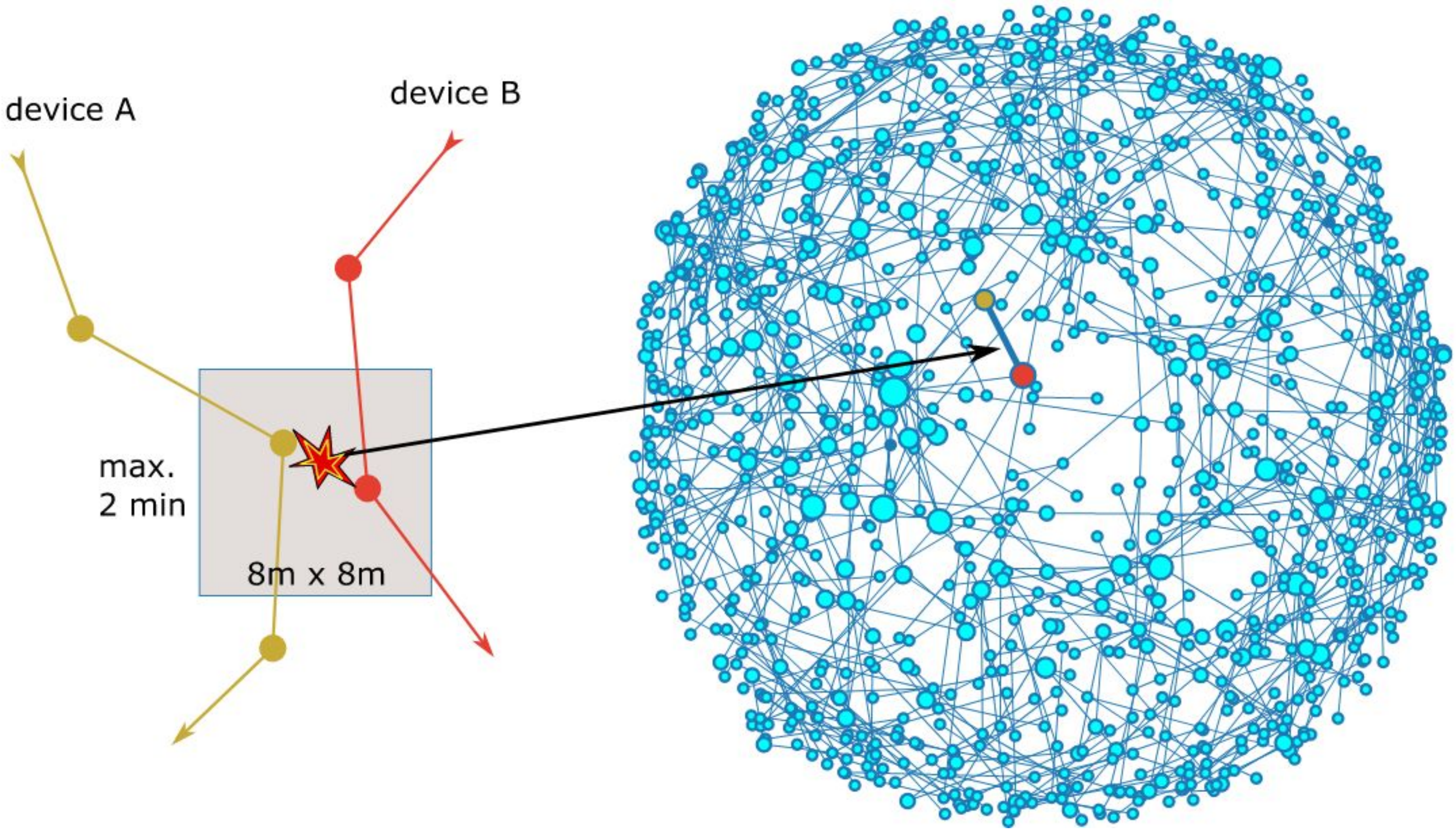
Time-use surveys



- ► Similar patterns, overall
- ► Exposure dominated by time at home
- ► (note: different relation between time/contacts and infection)

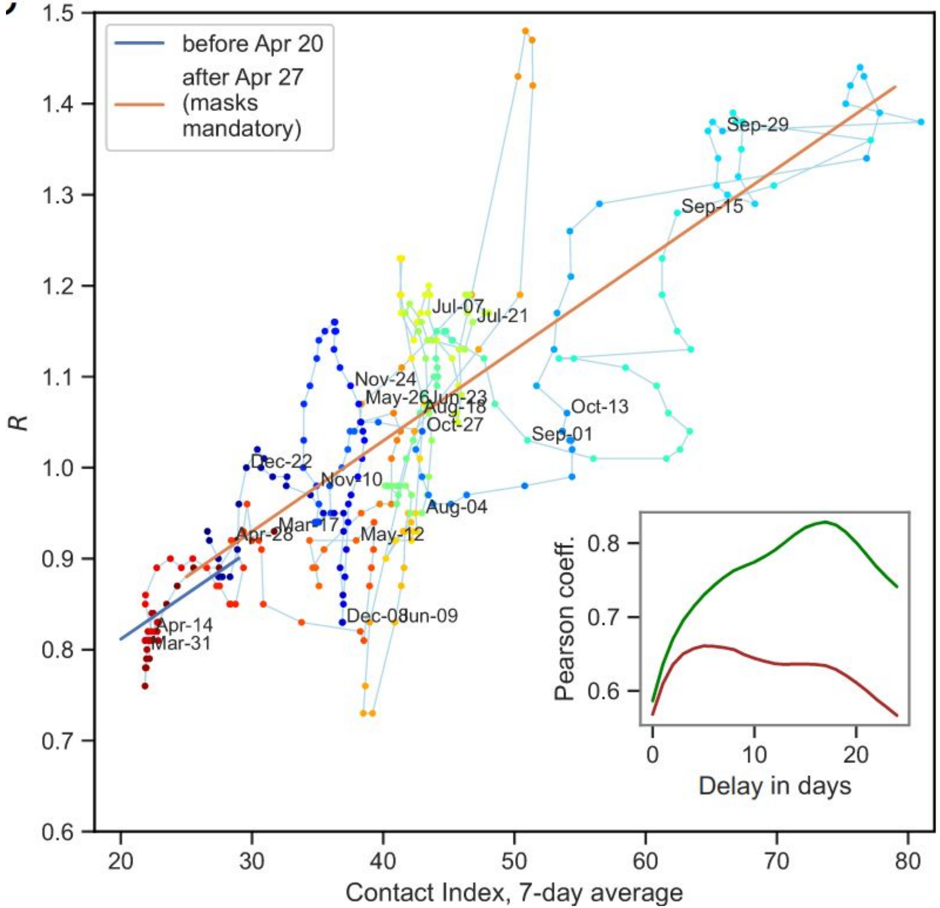
Hoang et al. (2022).

GPS data



Rüdiger et al. (2021)

GPS data

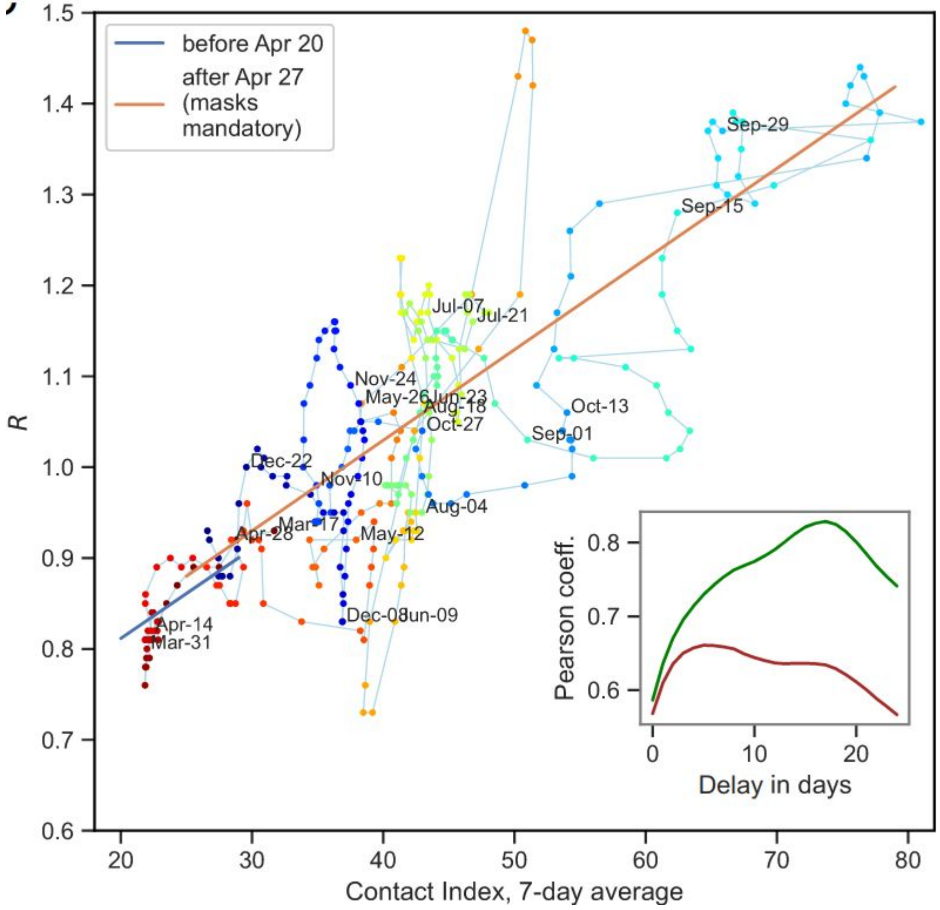


The contact index can be used to inform models, relying on millions of individuals' data.

- The *contact index* correlates with R_0
- Max correlation for delay of 17 days

Rüdiger et al. (2021)

GPS data



The contact index can be used to inform models, relying on **millions** of individuals' data.

- The *contact index* correlates with R_0
- Max correlation for delay of 17 days

However, wrt contact data, has some limitations:

- Representativity
- Not age-specific
- Easily connected to NPIs
- Difficult to connect to individual covariates

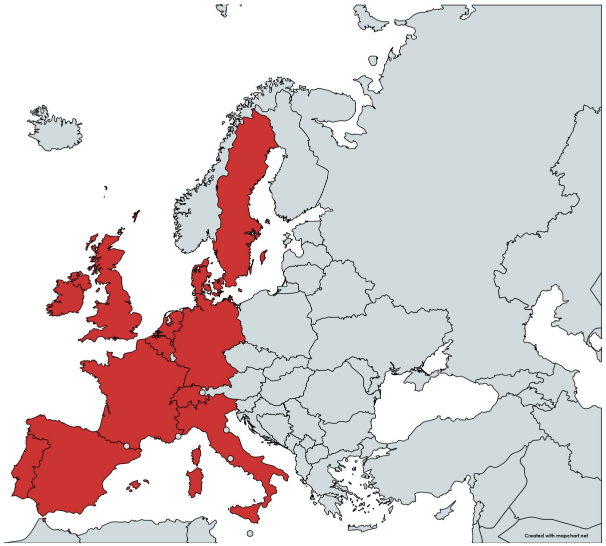
Rüdiger et al. (2021)

Influenzanet

- Europe-wide network to monitor the activity of infectious diseases
- Obtaining data directly from the population (anonymous)
- Questionnaires : background information and weekly symptoms
- Direct comparison between countries
    - Reporting to ECDC

- influenzanet.info

INFECTIERADAR.be



Conclusions

- Contact data allow to inform models of infectious diseases
- Crucial in case of behavioral changes
- Contact matrix are directly related to infectious quantities such as R_0
- Social contact data modelling allows for the *prediction* of contacts
- Contact matrix can be seamlessly included in models of infectious diseases
- Other (and future) methods to measure social contacts

Thank you for your attention

Contact: pietro.coletti@uhasselt.be

@coletti_pietro

http://www.socialcontactdata.org/

Thank y  tention

Contact: pietro.coletti@uhasselt.be

@coletti_pietro

http://www.socialcontactdata.org/

References (TBD)

Coletti et al.: Scientific Reports 10:21885 (2021).
Coletti et al.: BMC Infectious Diseases 21:503 (2021b).
De Luca et al.: BMC Infectious Diseases volume 18, Article number: 29 (2018).
Dikman et al.: J. R. Soc. Interface 7, 873–885 (2010).
Franco et al.: PLOS Computational Biology 18(3): e1009965 (2022).
Hoang et al.: medRxiv 2022.01.25.22269385 (2022).
Hoang et al.: BMC Infectious Diseases volume 21, Article number: 274 (2021).
Mastandrea et al.: PLOS ONE 10(9): e0136497 (2015).
Mossong et al.: PLOS Medicine 5(3): e74 (2008).
Rüdiger et al.: PNAS 118 (31) e2026731118 (2021).
Tellier et al.: BMC Infectious Diseases 19, Article number: 101 (2019).
Tomori et al.: BMC Medicine volume 19, Article number: 271 (2021).
Verelst et al.: BMC Medicine volume 19, Article number: 254 (2021).
Willem et al.: BMC Research Notes volume 13, Article number: 293 (2020).
Zagheni et al.: American Journal of Epidemiology, Volume 168, Issue 9, Pages 1082–1090 (2008).