Lessons Learnt from NPI Effectiveness Modelling throughout the COVID-19 Pandemic

Mrinank Sharma
@MrinankSharma

UNIVERSITY OF OXFORD

Research happens in teams. Many collaborators :)

And more who I couldn't find photos of

Why care about the effects of different interventions?

Governments across the world implemented a suite of non-pharmaceutical interventions to control the COVID-19 pandemic.

This article is more than **1 year old**

Geneva motor show cancelled as Switzerland bans large events

Ruling to prevent coronavirus spread covers events likely to attract more than 1,000 people

Coronavirus - latest updates

▲ Covered cars are pictured at the Palexpo exhibition centre, which had been due to stage the Geneva motor show from 2-15 March. Photograph: Pierre Albouy/Reuters

The Geneva International Motor Show has been cancelled after the Swiss government banned large events of more than 1,000 people as a measure to help combat the spread of coronavirus.

Why care about the effects of different interventions?

Governments across the world implemented a suite of non-pharmaceutical interventions to control the COVID-19 pandemic.

UK schools to be closed indefinitely and exams cancelled

Schools will remain open only for key workers' children and 'the most vulnerable'

- Coronavirus - latest updates
- See all our coronavirus coverage

▲ Coronavirus: UK schools to close indefinitely, says Boris Johnson – video

Why care about the effects of different interventions?

Governments across the world implemented a suite of non-pharmaceutical interventions to control the COVID-19 pandemic.

Italy set to quarantine whole of Lombardy due to coronavirus

Government's draft decree would impose fines on anyone caught entering or leaving northern region

▲ The deserted Piazza Duomo in Milan. Photograph: Piero Cruciatti/AFP via Getty Images

Why care about the effects of different interventions?

Governments across the world implemented a suite of non-pharmaceutical interventions to control the COVID-19 pandemic.

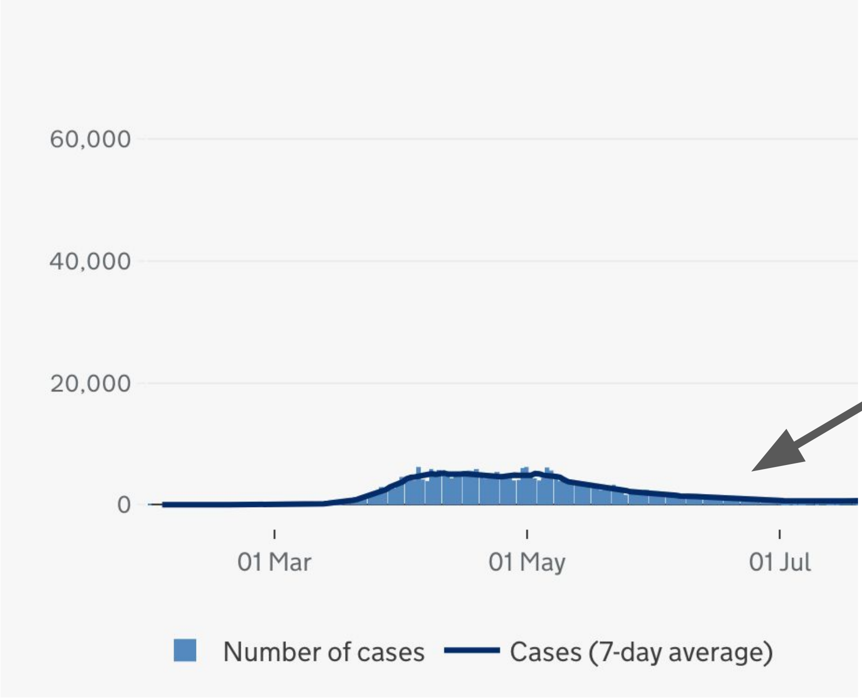
As we know, **interventions have costs—socially, economically, …**

Italy set to quarantine whole of Lombardy due to coronavirus

Government's draft decree would impose fines on anyone caught entering or leaving northern region

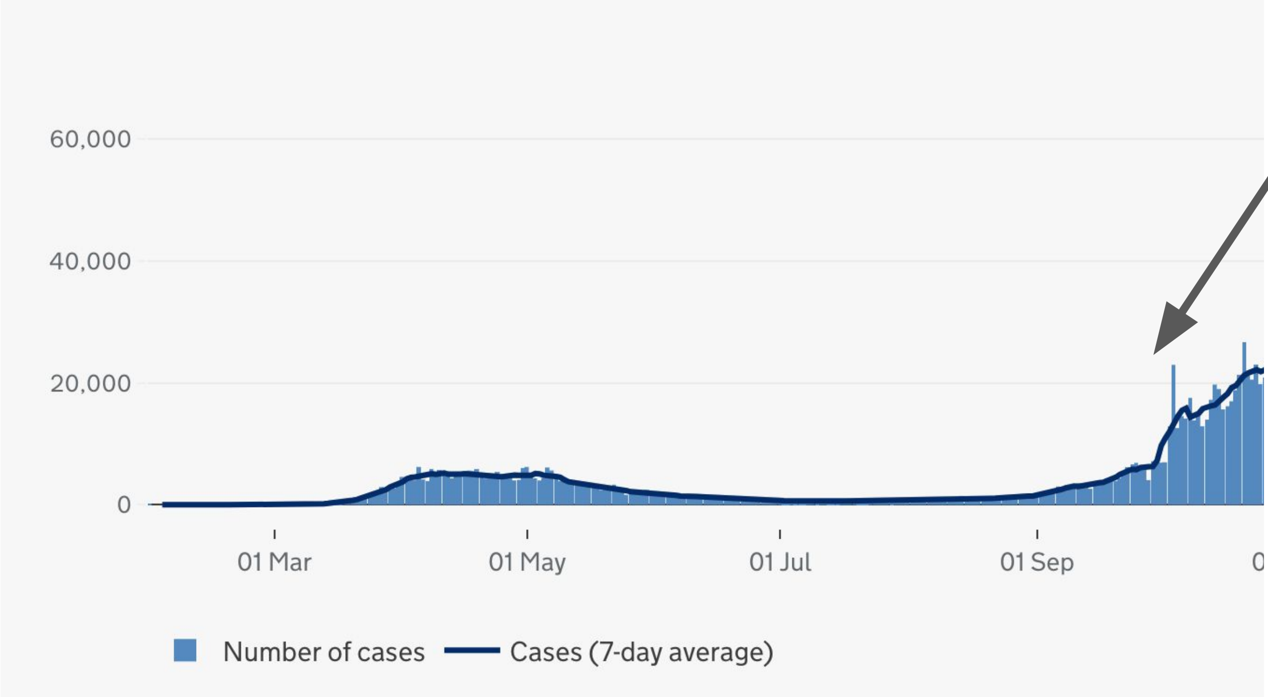
▲ The deserted Piazza Duomo in Milan. Photograph: Piero Cruciatti/AFP via Getty Images

COVID-19 in the UK: The Policymaker's Choice



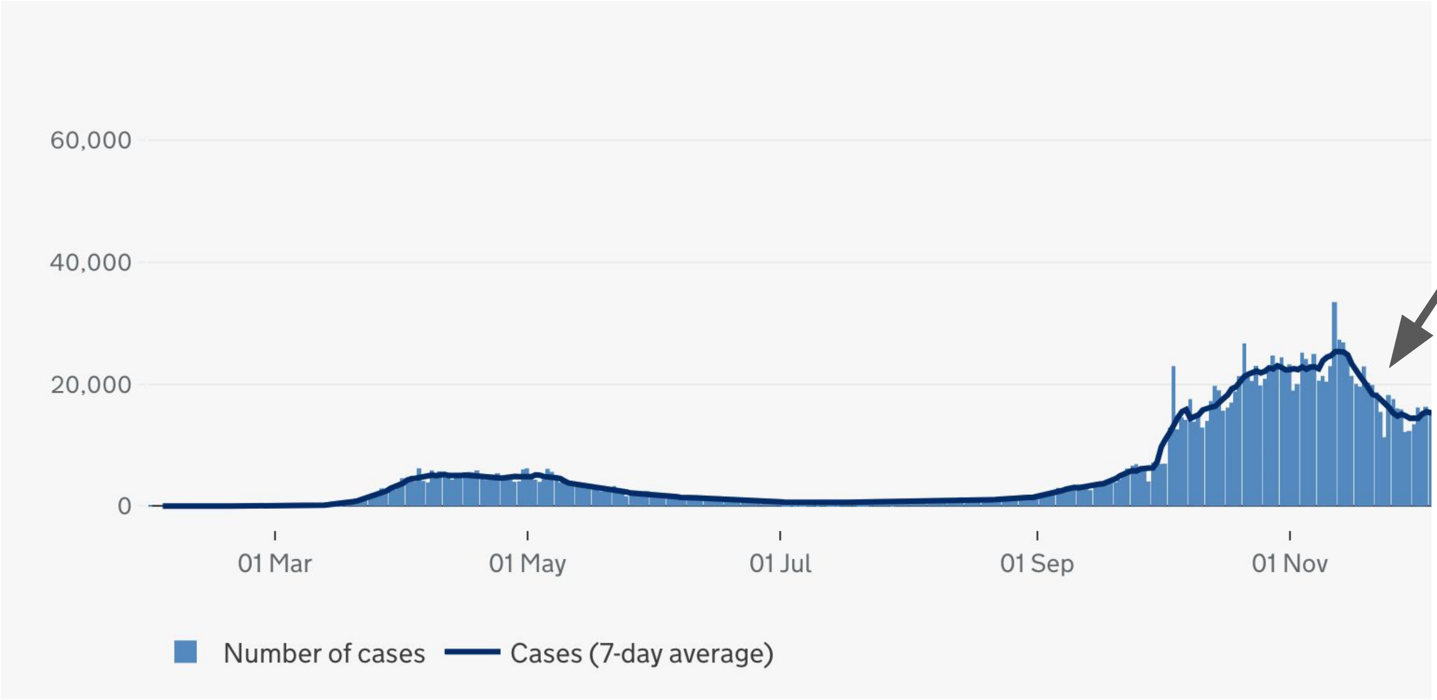
What restrictions should we ease?

COVID-19 in the UK: The Policymaker's Choice



How can we control the resurgence?

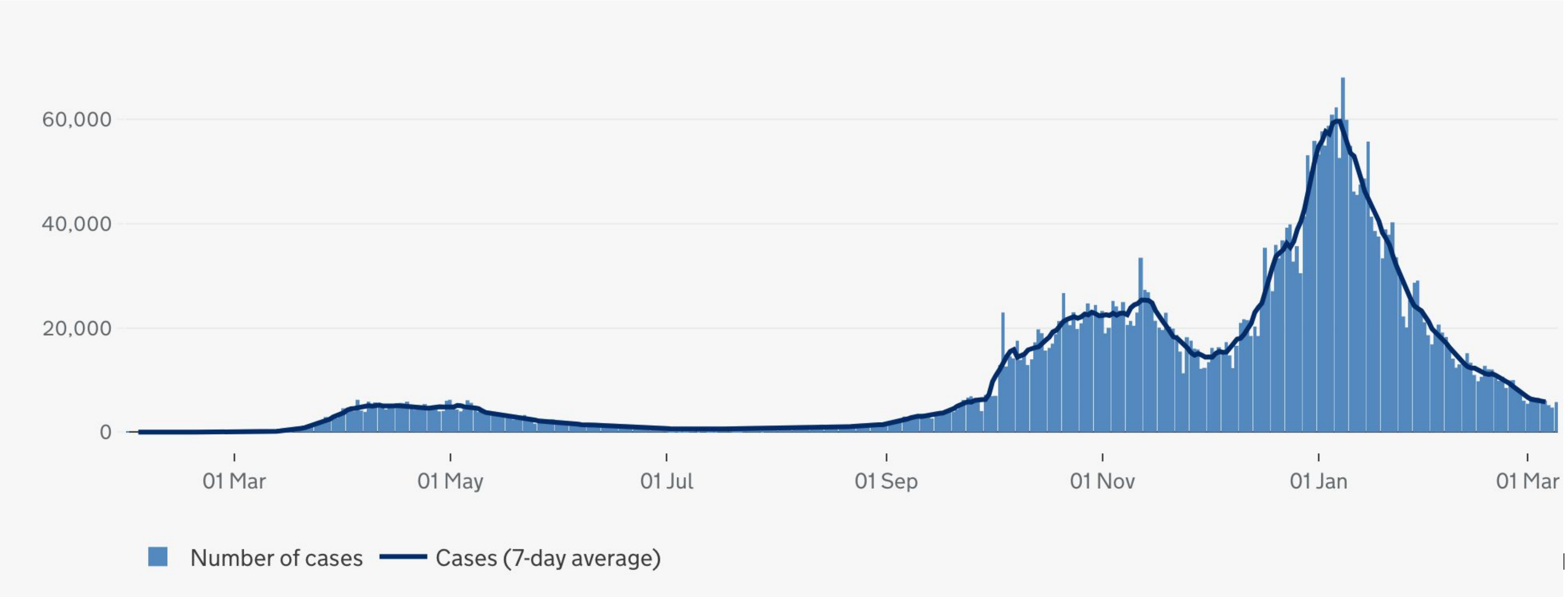
COVID-19 in the UK: The Policymaker’s Choice



Are we safe to ease restrictions?

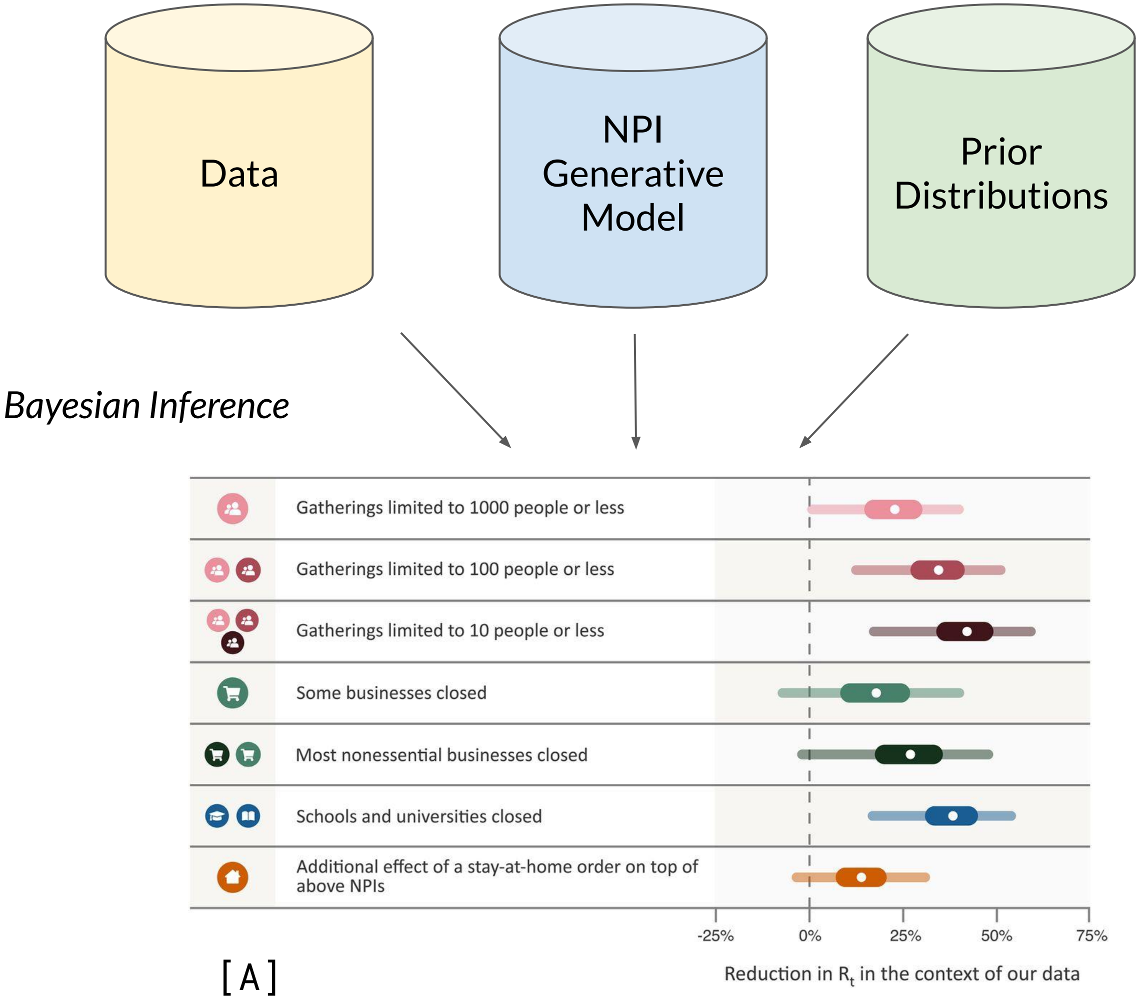
Policymaker:

How do we balance the social costs of interventions with control of COVID?



General Approach

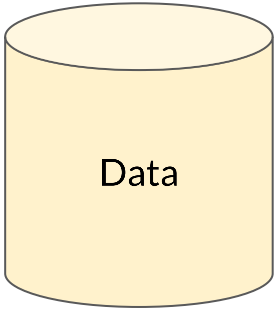
Data-driven NPI effectiveness modelling



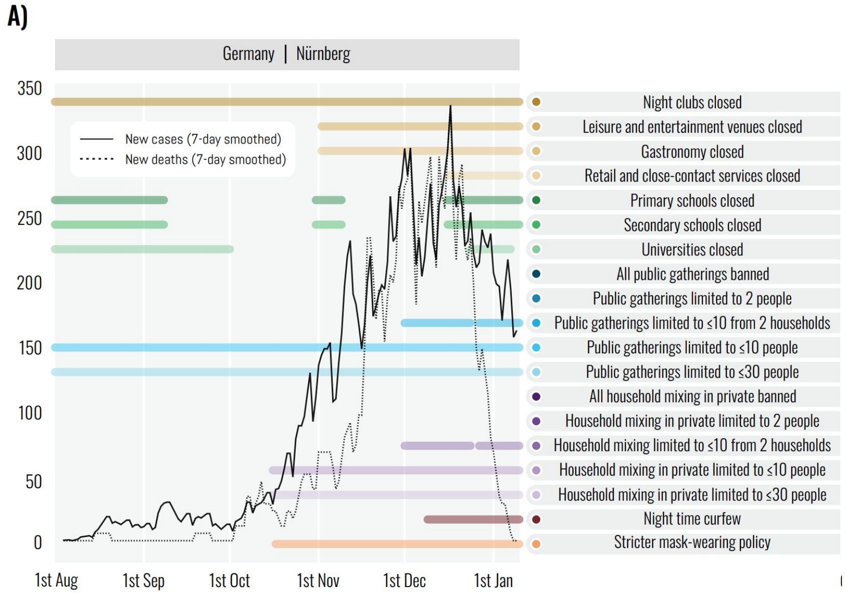
[A]

General Approach

Data-driven NPI effectiveness modelling

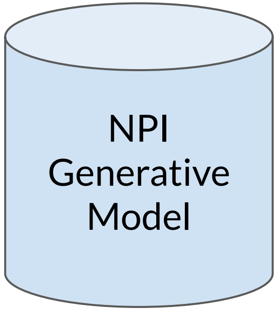


Data: timeline of different interventions across regions, number of reported cases and deaths in those regions



[C]

General Approach



Data-driven NPI effectiveness modelling.

Bayesian Generative Model:

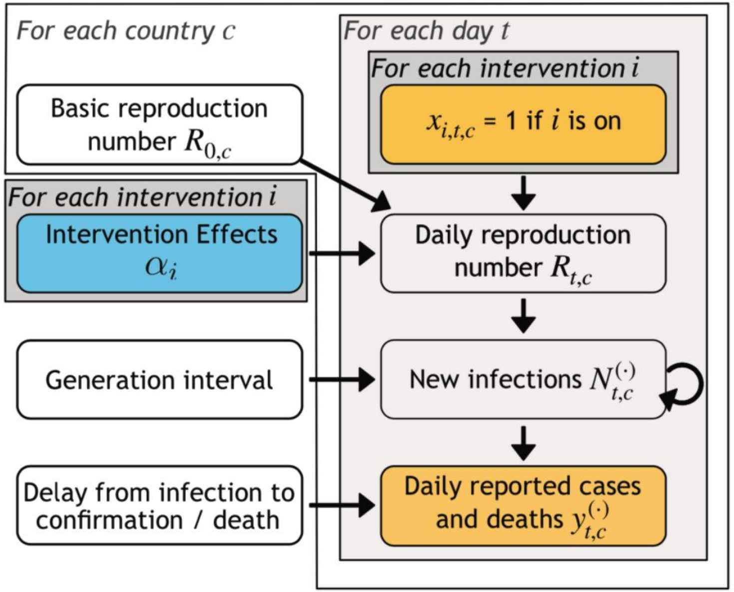
> *“What is the probability of observing data D if the intervention X has effect Y?”*

Base R

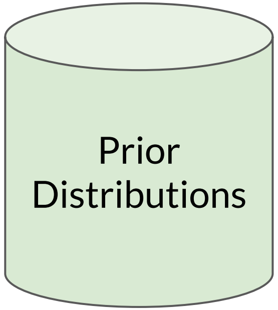
NPI effectiveness

$$R_{t,c} = R_{0,c} \prod_i \exp(-\alpha_i x_{i,t,c})$$

NPI activations



General Approach



Data-driven NPI effectiveness modelling

Prior distribution:

> *"Before observing any data, what is our belief about the effectiveness of intervention X?"*

General Approach

Data-driven NPI effectiveness modelling.

Several papers:

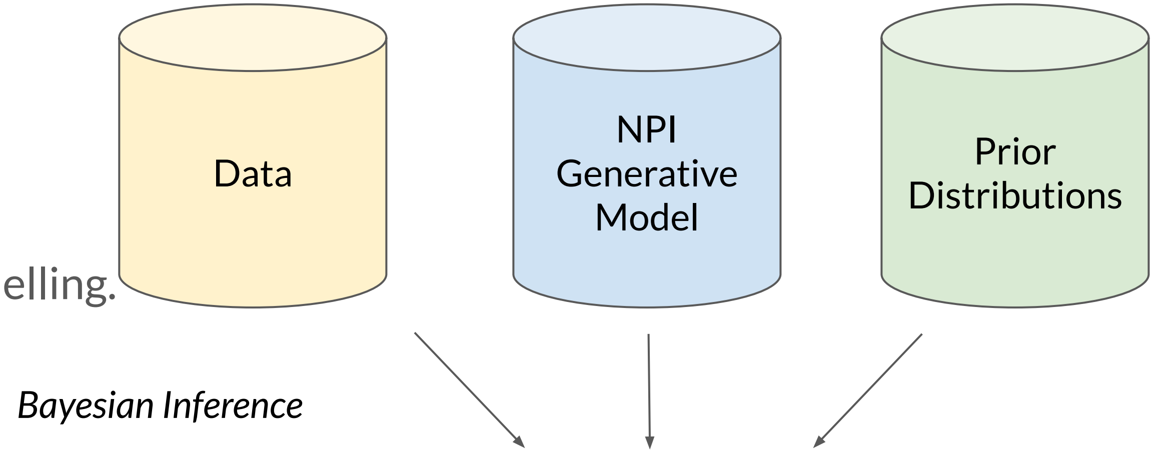
[A] "1st wave", March-July 2020

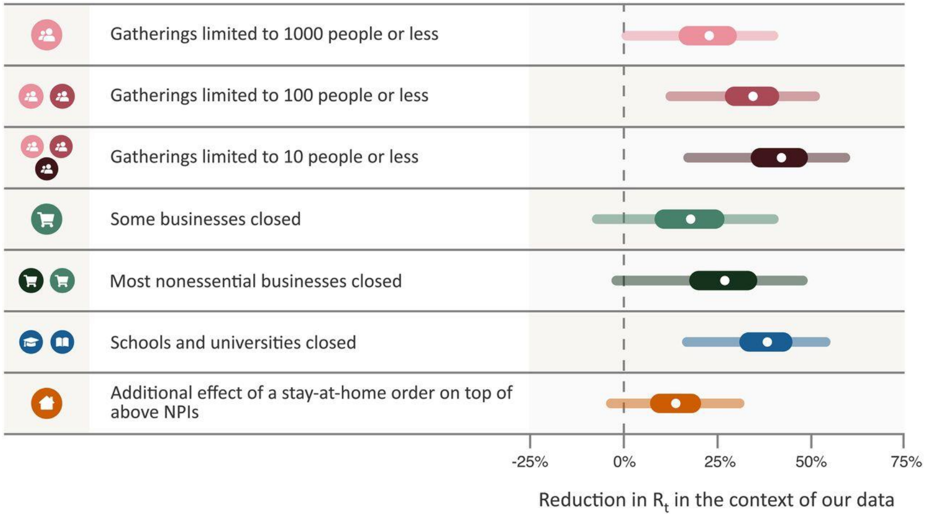
[B] "2nd wave", August-Jan 2020

[C] Mask wearing effects

[D] Seasonality

Instead of focusing on *results*, let's focus on *learnings* that mean things **today**.





Learning #1: Garbage in, Garbage Out

- Our inferences combine data with a probabilistic model. The **quality of data** critically determines the "quality" of our inferences.
- Data collection is difficult—it requires judgement calls, proper scoping, effective teamwork, …

But, contributions to data can be undervalued by academia. How can we incentivise high quality, "modelling ready" data collection?

Learning #2: Effectiveness Changes in Time.

- Our inferences are made in the *context of our data.*
  - **Midpandemic schools ≠ prepandemic schools**

- During the pandemic, organisations implemented **safety measures** and people changed their **behaviour.**
  - If people in the population no longer meet in large groups, banning large gatherings doesn't affect transmission although it previously did!

How to communicate limitations and nuance to policymakers?

Learning #3: Adapting to time sensitivity

- Exponential growth leads to **incredible time sensitivity.** A delay by a few days can *drastically change* the number of infections/cases/deaths.
    - E.g., consider investigation into transmission advantage of new variants of concern.

How can we produce **decision relevant research** in a timely fashion?

Having tools ready, and releasing high quality, documented tools so others can easily build upon them!

Alternative peer review protocols, with more dialogue and real time discussion?

Summary

- Understanding the effectiveness of different interventions is crucial for smart policy.
- We can tackle this question by combining Bayesian modelling with high quality collected data.
- What have we learnt?
  - In short, *academic structures and incentives are not well designed for performing time-sensitive research in a pandemic. How can we equip researchers with the skills to perform and communicate research in emergencies?*
    - Data, communication of limitations and nuance, changes to peer review and publicly releasing code and data, making high quality modelling data available, training for time sensitive situations, …

Our work

[A] **Inferring the effectiveness of government interventions against COVID-19.** *Jan Brauner*, Sören Mindermann*, Mrinank Sharma*, Anna B Stephenson, Tomáš Gavenčiak, David Johnston, Gavin Leech, John Salvatier, George Altman, Alexander John Norman, Joshua Teperowski Monrad, Tamay Besiroglu, Hong Ge, Vladimir Mikulik, Meghan A. Hartwick, Yee Whye Teh, Leonid Chindelevitch, Yarin Gal, Jan Kulveit* (2020). **Science**.

[B] **How robust are the estimated effects of nonpharmaceutical interventions against COVID-19?** *Mrinank Sharma*, Sören Mindermann*, Jan Markus Brauner*, Gavin Leech, Anna B. Stephenson, Tomáš Gavenčiak, Jan Kulveit, Yee Whye Teh, Leonid Chindelevitch, Yarin Gal* (2020). **NeurIPS**.

[C] **Understanding the effectiveness of government interventions against resurgence of COVID-19 in Europe.** *Mrinank Sharma*, Sören Mindermann*, Charlie Rogers-Smith, Gavin Leech, Ben Snodin, Janvi Ahuja, Jonas B. Sandbrink, Joshua Teperowski Monrad, George Altman, Gurpreet Dhaliwal, Lukas Finnveden, Alexander John Norman, Sebastian B. Oehm, Julia Fabienne Sandkühler, Thomas Mellan, Jan Kulveit, Leonid Chindelevitch, Seth Flaxman, Yarin Gal, Swapnil Mishra, Samir Bhatt, Jan Markus Brauner*. **Nature Communications**.

Our work

[D] Gavenčiak, Tomáš, Joshua Teperowski Monrad, Gavin Leech, Mrinank Sharma, Sören Mindermann, Jan Markus Brauner, Samir Bhatt, and Jan Kulveit. "Seasonal variation in SARS-CoV-2 transmission in temperate climates." *MedRxiv* (2021).

Thank you for your attention

Check out the papers for more detail, limitations, …

Why study the second wave?

Safety Measures

Coronavirus: pubs and restaurants across England to be forced to shut at 10pm

Boris Johnson to set out limited nationwide coronavirus restrictions on Tuesday

- Coronavirus - latest updates
- See all our coronavirus coverage

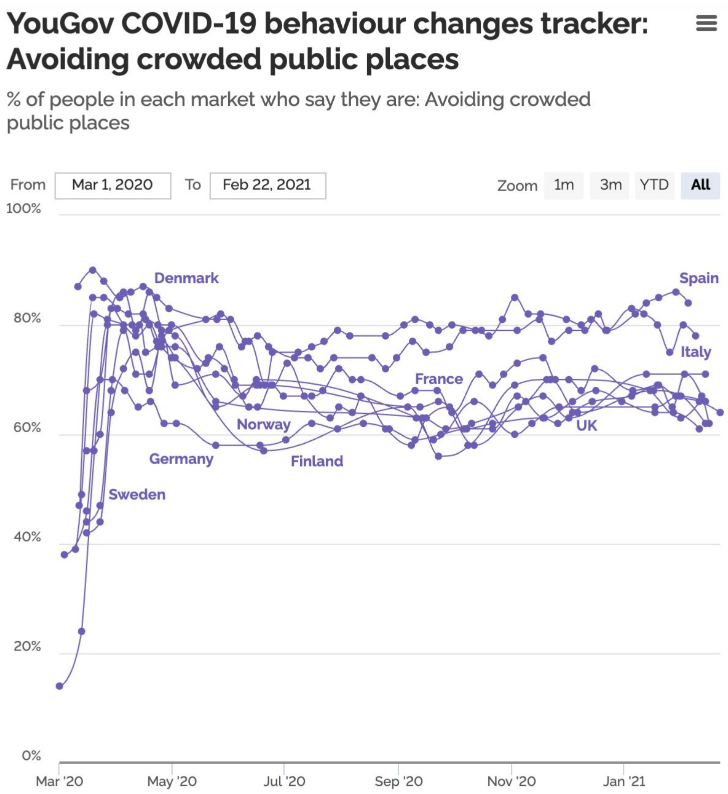
▲ Hospitality venues in England will have to close their doors at 10pm and offer table service only from Thursday. Photograph: Jonathan Brady/PA

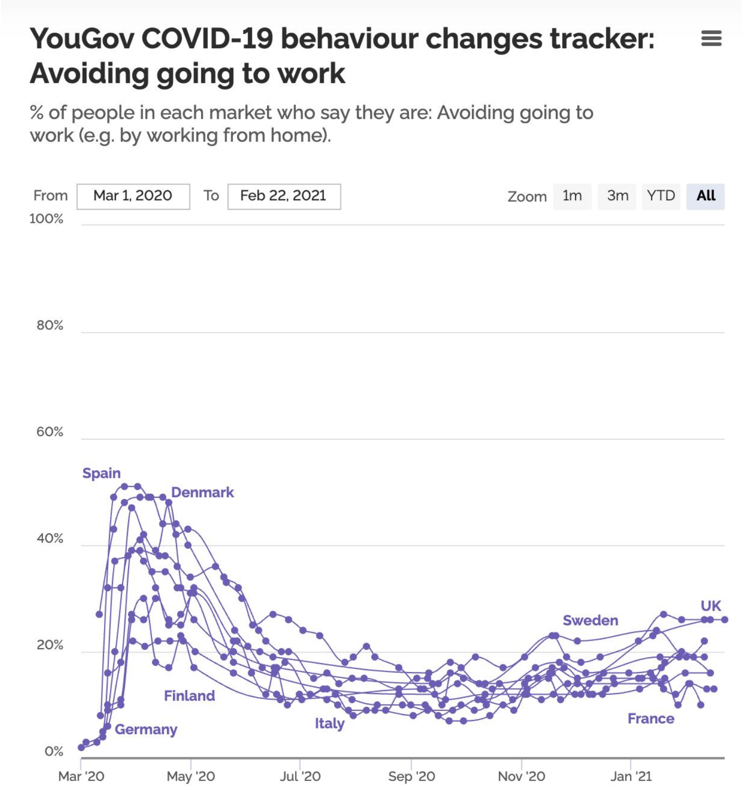
Could maximum class sizes of 15 pupils significantly improve our children's school life?

When children begin to return to school in June, they will be in classes of up to 15 pupils. How will that change British education?

By Sally Peck, FAMILY, EDUCATION AND CAREERS EDITOR
12 May 2020 • 3:17pm

Behaviour Stability





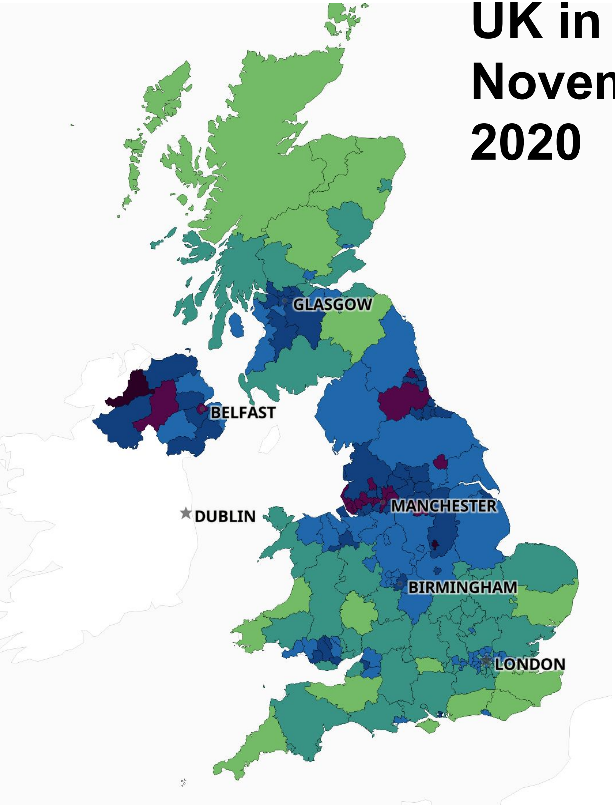
The first wave effects will not generalise to the second (and future waves)

The first wave effects will not generalise to the second (and future waves)

Also: New NPI constellations (and more data) allow new insights.

How to study the second wave?

Local epidemics



UK in November 2020

Lockdown tightened in north-east England as Covid-19 infections rise

For first time since pandemic began it will be illegal for people from different households to mix in pubs and restaurants

- Coronavirus - latest updates
- See all our coronavirus coverage

▲ Shoppers in Newcastle, one of seven north-east council areas subject to tighter restrictions after a rise in the coronavirus infection rate. Photograph: Ian Forsyth/Getty Images

New interventions

Second Wave NPI Effectiveness Estimates

Data

Data

- Problems with existing intervention datasets:
    - National level intervention data
    - Intervention definitions not suitable for second wave
    - Lack of validation procedures (low data quality)

Data

- How to start? - Proper scoping!
  - Questions
    - What NPIs mattered?
    - What level of geographic granularity?
    - Is case and death data available at that level?
    - What period of analysis?
  - Solutions
    - Exploratory data collection
    - Talking to local epidemiologists
    - Many judgement calls -> this is a job for a team

Data

- We collect **fine grained intervention data** in **114 areas from 7 European countries.**
  - We use **stratification by deaths in the first wave to ensure our estimates generalise.**
- **NPIs:**
  - **Gathering and Household Limits (Public/Private/Indoor/Outdoor)**
  - **School Closures (primary/secondary school)**
  - **University Closures**
  - **Gastronomy/Nightclubs/Leisure Venues/Retail Business Closures**
  - **Curfews**
  - **Mask Wearing (5 stringency levels)**

Data

Countries	7
Regions of analysis	114
Period	1 August 2020 - 9 January 2021
Days across all regions	19,000
NPI entries in the dataset	> 5,800
Data validation (manual)	Semi-independent double entry; interviews with local epidemiologists; validation against external sources; cross-country consistency checks
Time spent on data collection (excluding exploratory collection and design)	950 hours (9 researchers)

Data

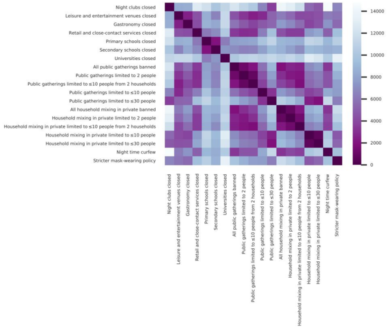
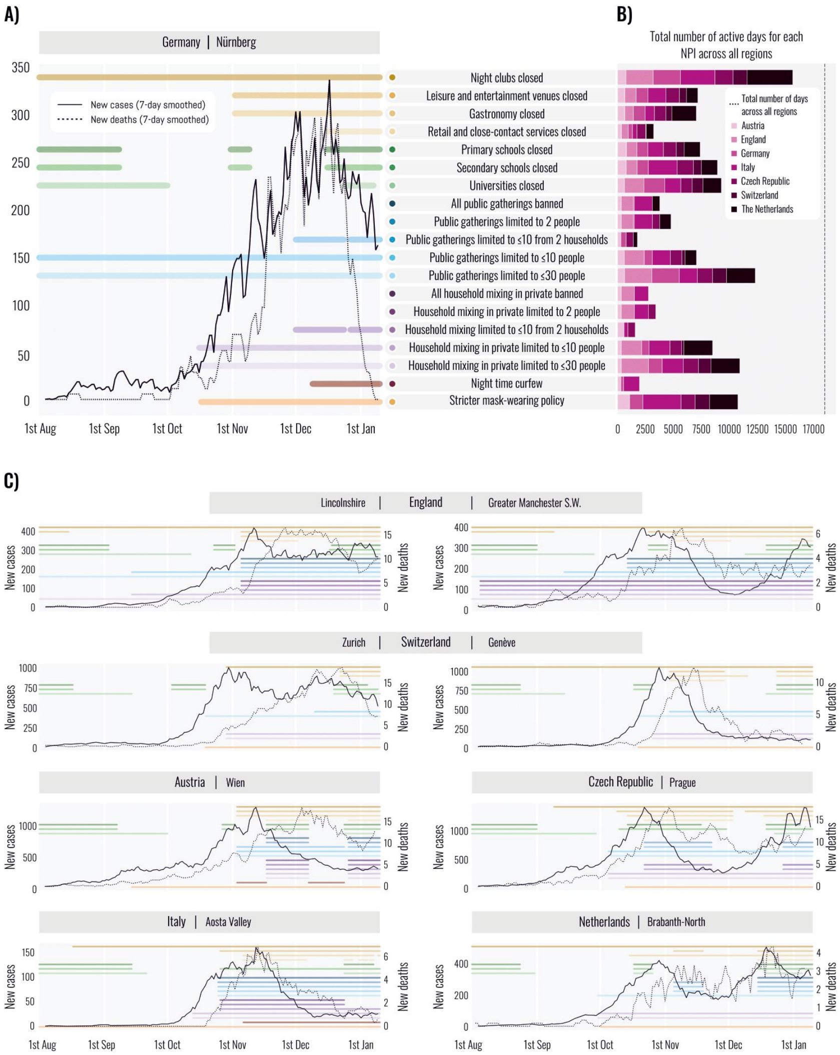


Figure S28: Number of days across all regions available to distinguish NPI effects. For every pair of NPIs (row - column), the entry shows the number of days on which exactly one of the two NPIs was active.



Data

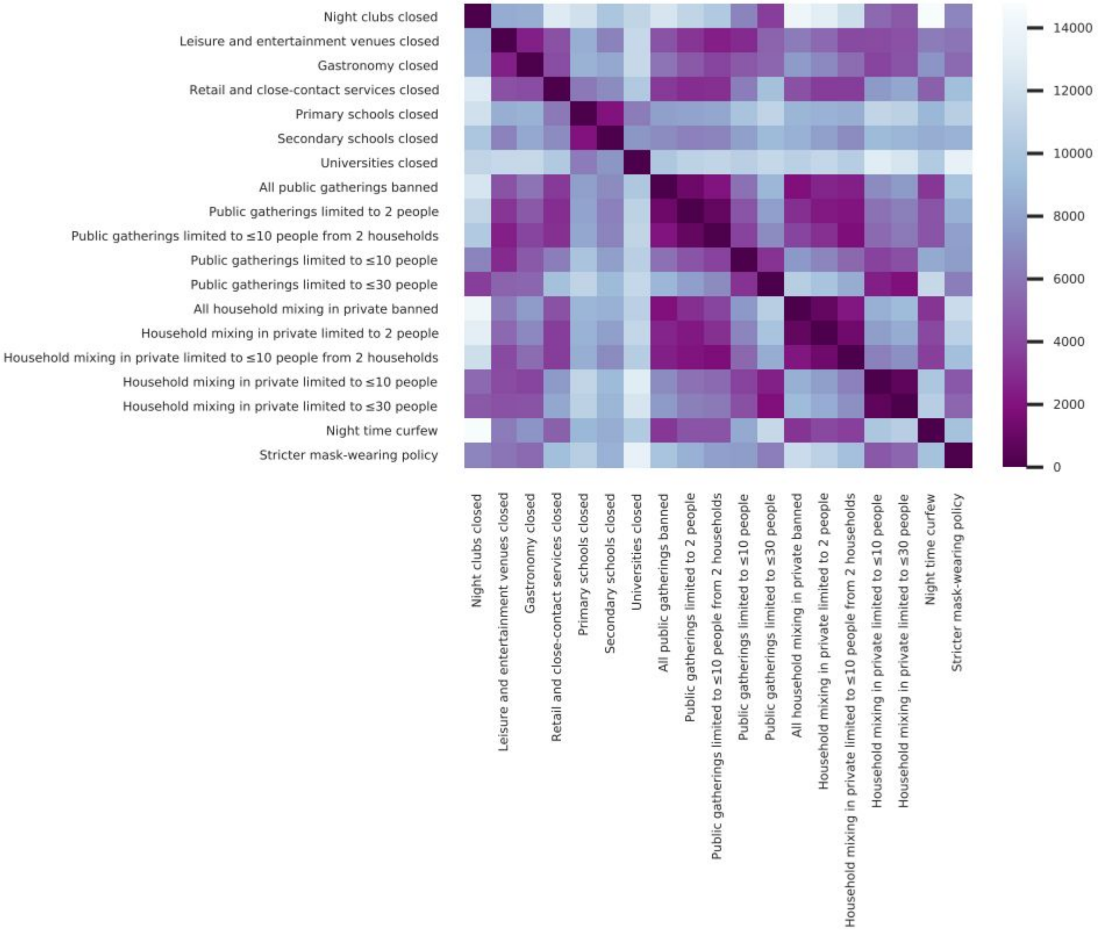


Figure S28: Number of days across all regions available to distinguish NPI effects. For every pair of NPIs (row - column), the entry shows the number of days on which exactly one of the two NPIs was active.

Modelling

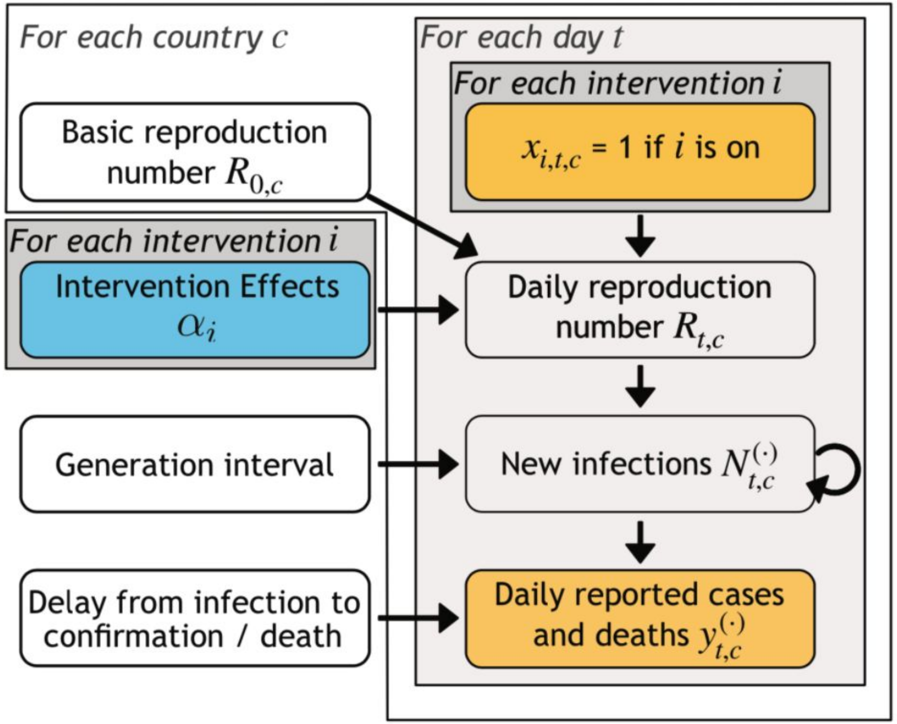
Modelling Approach - First Wave

Base R

NPI effectiveness

$$R_{t,c} = R_{0,c} \prod_i \exp(-\alpha_i x_{i,t,c})$$

NPI activations



Modelling Approach

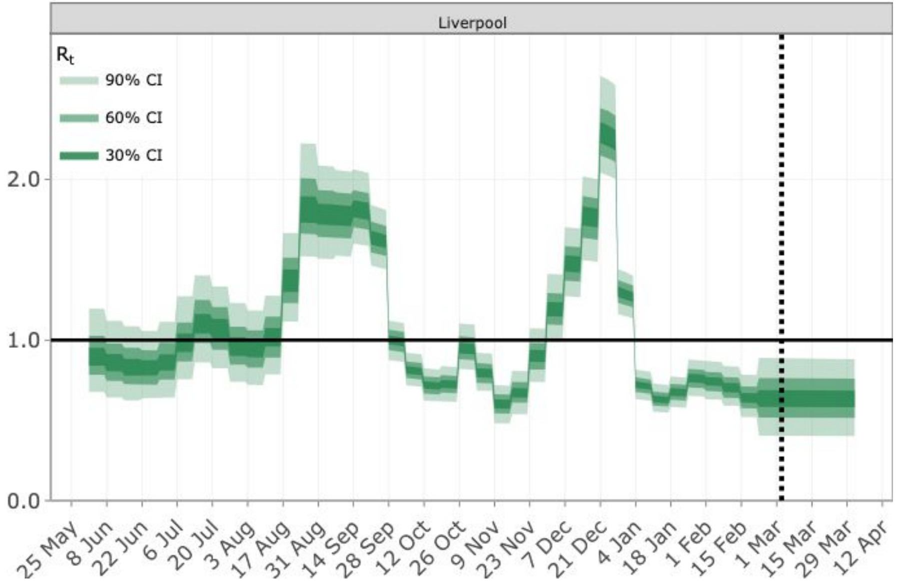
- We observe increases in transmission **unrelated to changes in NPIs.**
- Solution: **random walk term in transmission.**

R: secondary infections generated by primary infection

$$R_{t,l} = R_{0,l} \exp[\sum_i -\beta_{i,c_l} \phi_{l,i,t} + \zeta_{t,l}]$$

Weekly Random Walk Noise

Allows for smooth changes in R every week. This can explain the increases in transmission unrelated to NPIs.



Modelling Approach

- The number of infections N is determined by a renewal equation:

$$\overline{N}_{t,l} = R_{t,l} \sum_{\tau=1}^{32} (\overline{N}_{t-\tau,l} \cdot \pi_{GI}[\tau])$$

Modelling Approach

- First wave NPI effectiveness estimates use **national data** (usually).
- In the second wave, you have to go local.
  - Problem: fewer cases and deaths in each area! More difficult to estimate R.
  - **Solution:** additional noise in the modelling, reducing the influence of small case and death counts.

$$N_{t,l} = \text{softplus}[\tilde{N}_{t,l} + \epsilon_{t,l}]$$

If cases increase from 1 to 2 in a week, is this R=2?

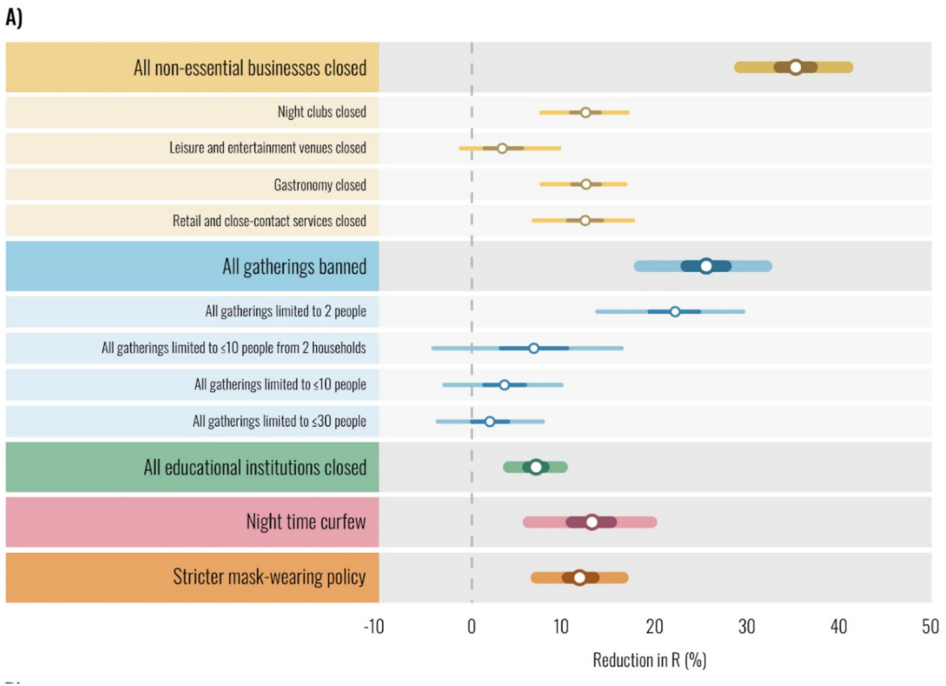
If cases increase from 1000 to 2000 in a week, is this R=2?

Modelling Approach

- R tells us the amount of infections generated by the currently infected people.
- Therefore, we can predict the number of infections that will occur in the future.
- Infections **today** show up as cases and deaths **in the future.**
    - These infections are **smoothed and delayed.**
    - Then, they are matched to the observed cases and deaths.

Now, given a probabilistic model and a dataset, **we can perform Bayesian inference using standard MCMC sampling algorithm.** *Big thanks to the Numpyro team!*

Results

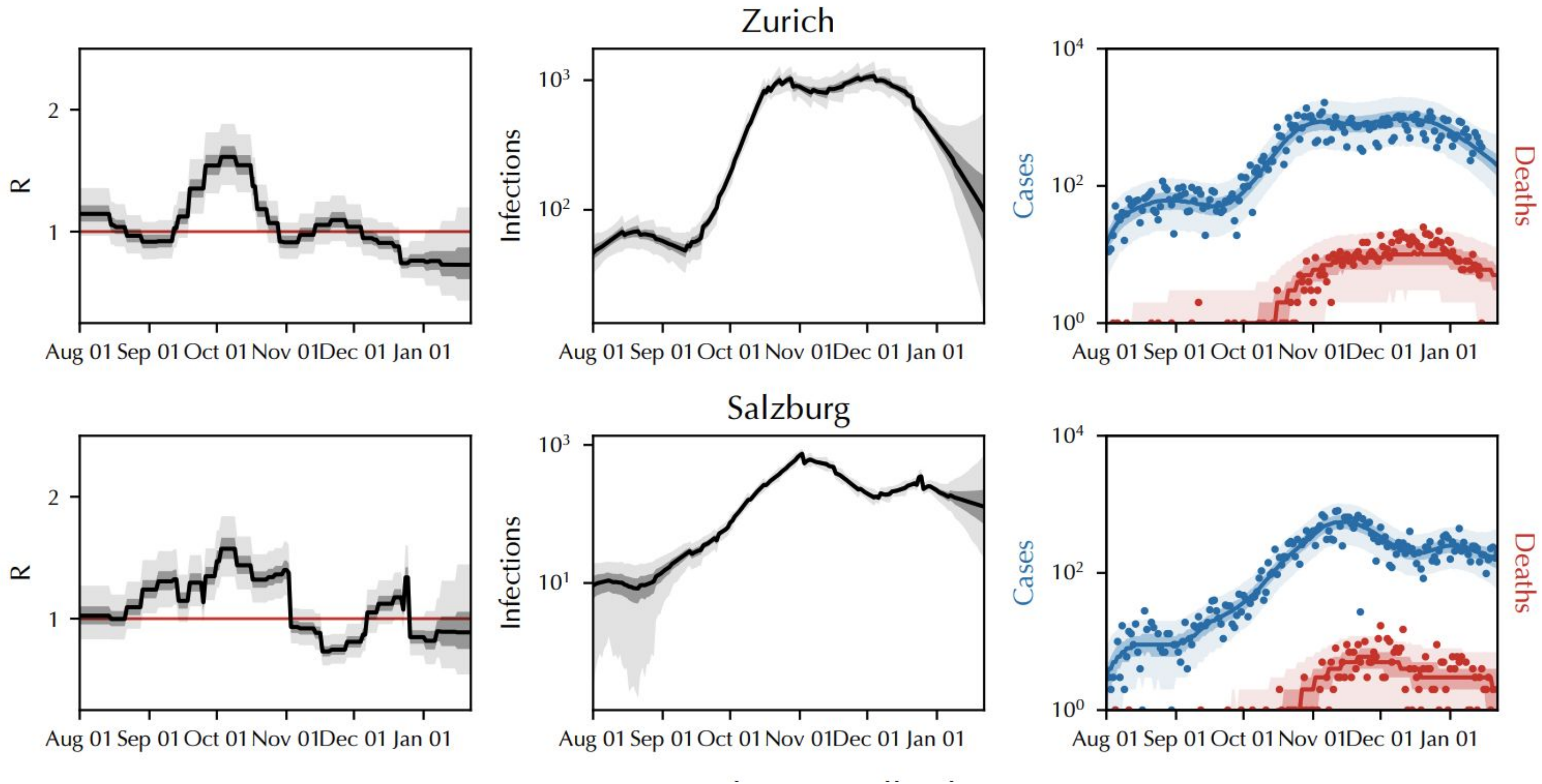


Overall, the interventions **that regions *actually used in the second wave*** were less effective.

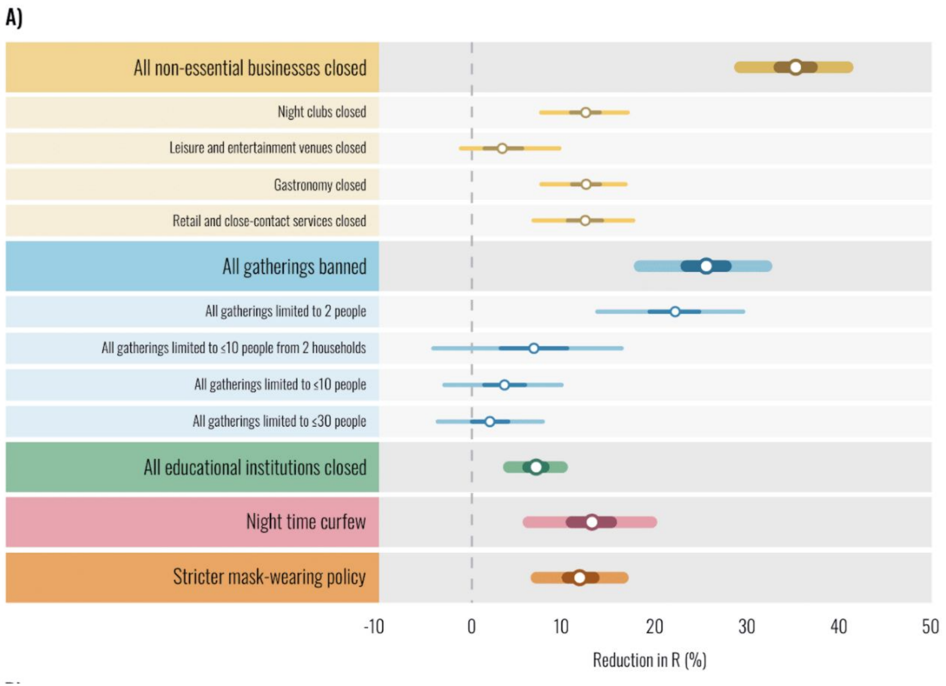
The most **stringent** set of NPIs in each region reduced Rt by average ~55%.

But, in the first wave estimates are 76%-82%.

Behaviour changes & safety measures.


Zurich
Salzburg
R
Infections
Cases
Deaths
Aug 01 Sep 01 Oct 01 Nov 01 Dec 01 Jan 01

Results

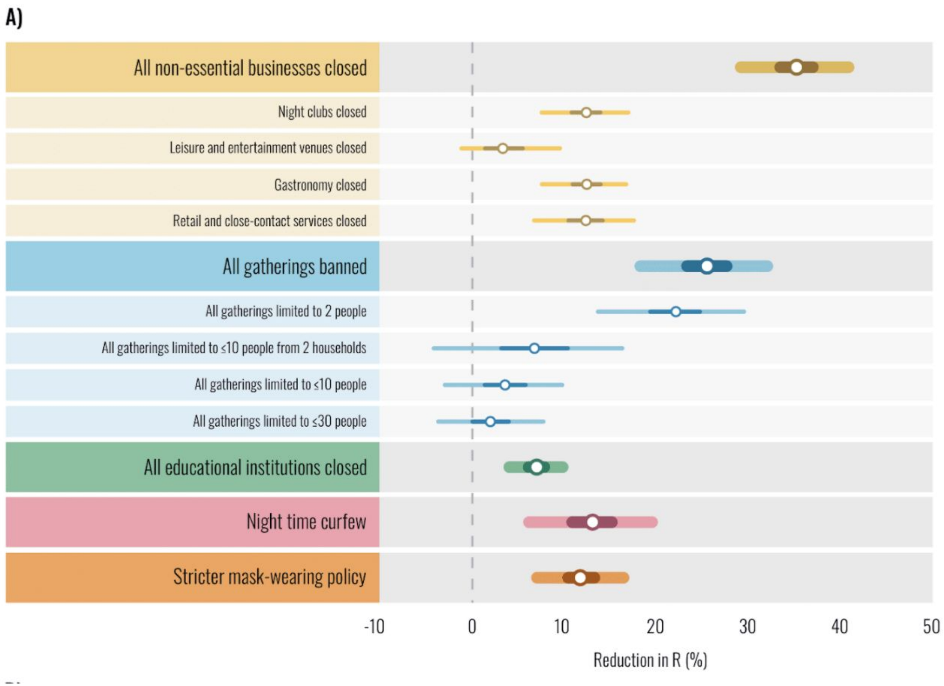


Business closures were very important!

Similar effects for nightclubs, retail businesses and gastronomy.

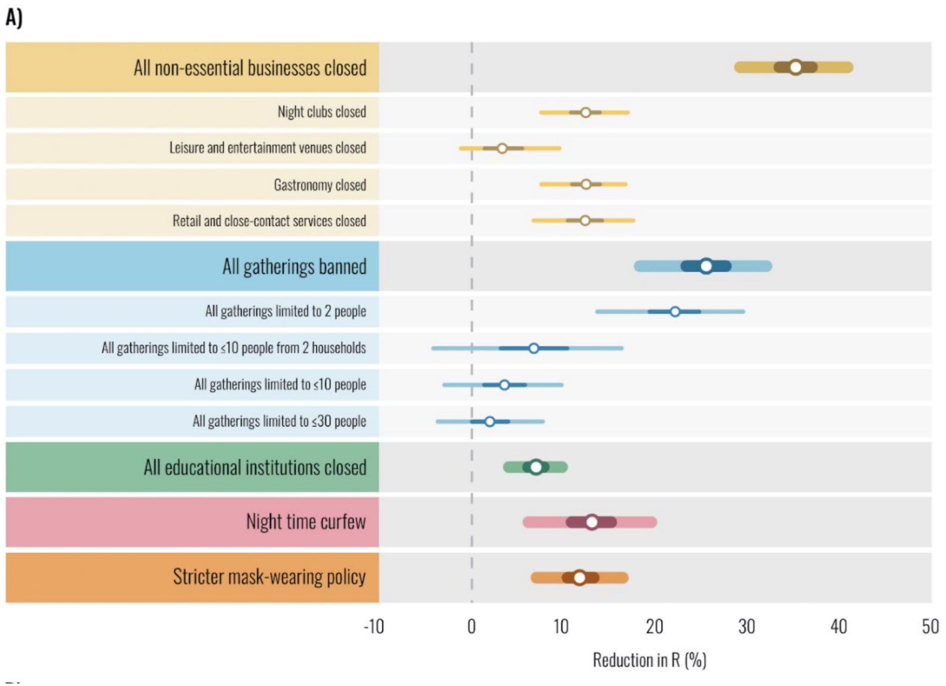
Smaller effects for Leisure Venues.

Results



Education institutions were very important in the first wave. Smaller effects in the second wave suggest successful safety measures.

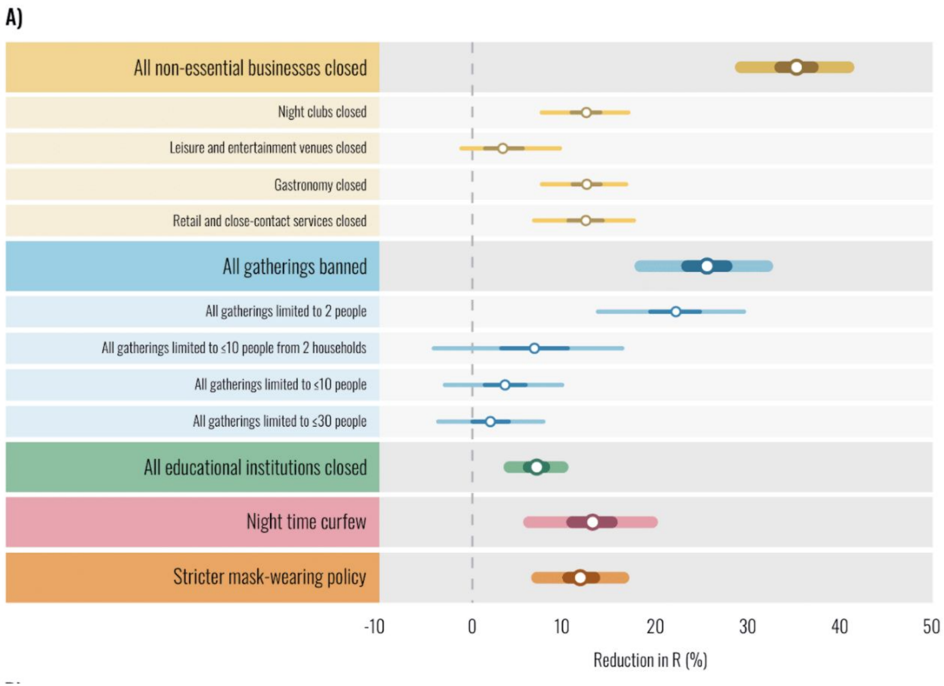
Results



Weaker gathering bans were not particularly effective.

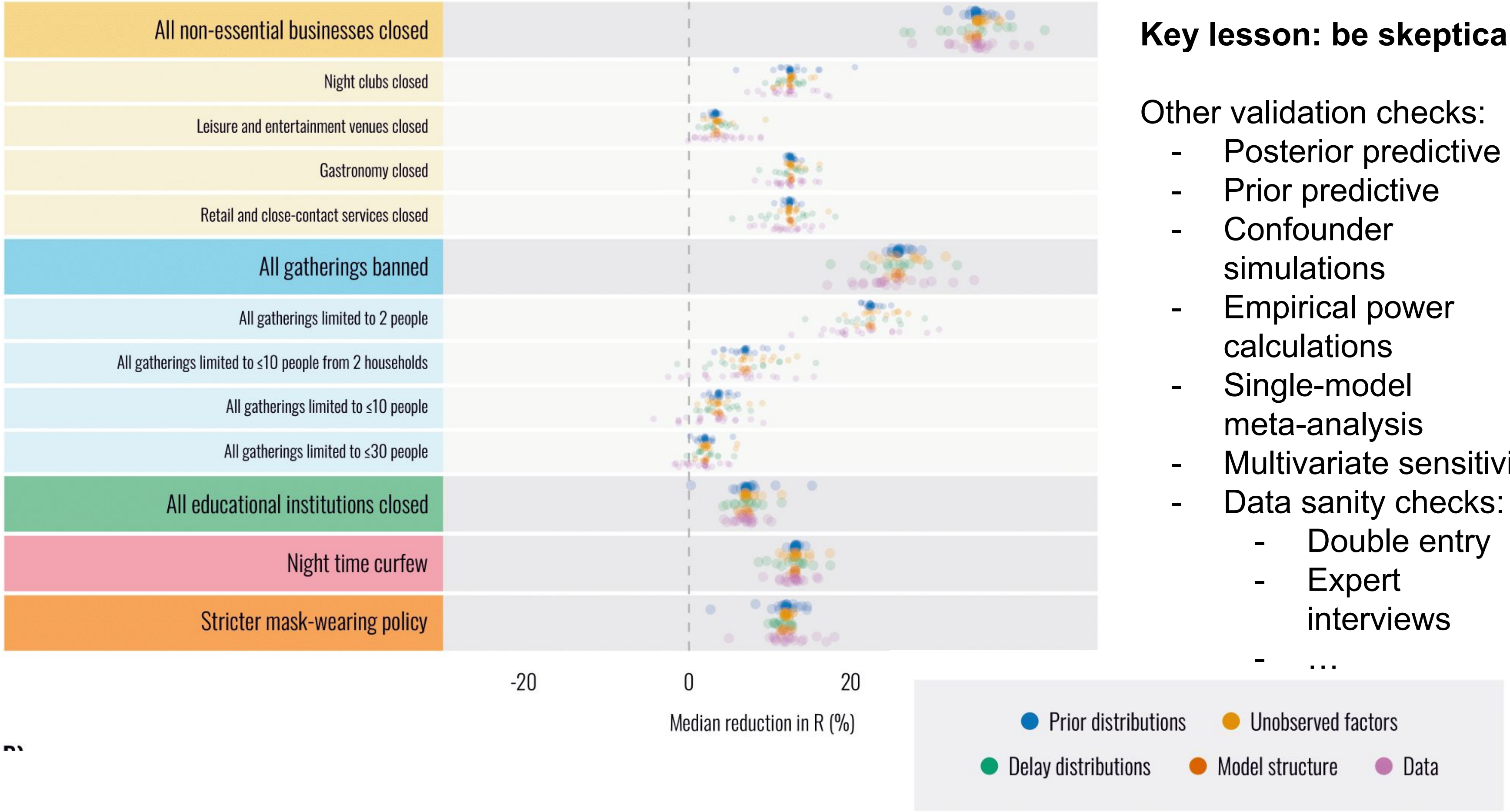
Significant reductions in transmission from the **strictest** bans, namely banning **all** gatherings, or only allowing gatherings with 1 other person.

Results



Curfews and mandatory mask wearing also helped.

Robustness / validation



Key lesson: be skeptical!

Other validation checks:

- Posterior predictive
- Prior predictive
- Confounder simulations
- Empirical power calculations
- Single-model meta-analysis
- Multivariate sensitivity
- Data sanity checks:
    - Double entry
    - Expert interviews
    - …

Robustness / validation

How to interpret results under misspecification / when model assumptions are broken?

Theorem 2. The ML solution of α_i, given $\{\alpha_j\}_{j \neq i}$, under the *Simplified Default* Model satisfies:

$$\exp(-\alpha_i) = \left(\sum_{(t,c) \in \Phi_i} \tilde{R}^{1/\nu}_{(-i),t,c} \bar{R}^{1/\nu}_{t,c} \right)^{\nu} \Bigg/ \left(\sum_{(t,c) \in \Phi_i} \tilde{R}^{1/\nu}_{(-i),t,c} \tilde{R}^{1/\nu}_{(-i),t,c} \right)^{\nu} = \frac{M^{W_i}_{1/\nu}(\{\bar{R}_{t,c}\}_{\Phi_i})}{M^{W_i}_{1/\nu}(\{\tilde{R}_{(-i),t,c}\}_{\Phi_i})} \tag{10}$$

Limitations

- Correlation analysis.
- Unobserved factors may be assigned to NPIs.
- Assume constant IFR/IAR.
- NPI effects assumed to be the same across countries.
- **How will these effects generalise e.g., to the new variants of concern?**

Take Home Story

1. We believed that second wave effects would be markedly different to the first wave...
2. … and **that's exactly what we found.**
3. **But, our estimates are historic,** and policymakers ***still*** need to balance the costs of COVID control and COVID transmission.
4. For now, a combination of **second wave effects** with **real-time monitoring and surveillance may be the best we can do.**

First wave slides

Goal

- Governments worldwide implemented nonpharmaceutical interventions (NPIs) to control the spread of COVID-19.
    - e.g., closing schools, restaurants, etc…
- We know, *in combination*, that these interventions were successful at reducing transmission significantly.
- But, *how effective was each NPI*?
- **And why do we care?**

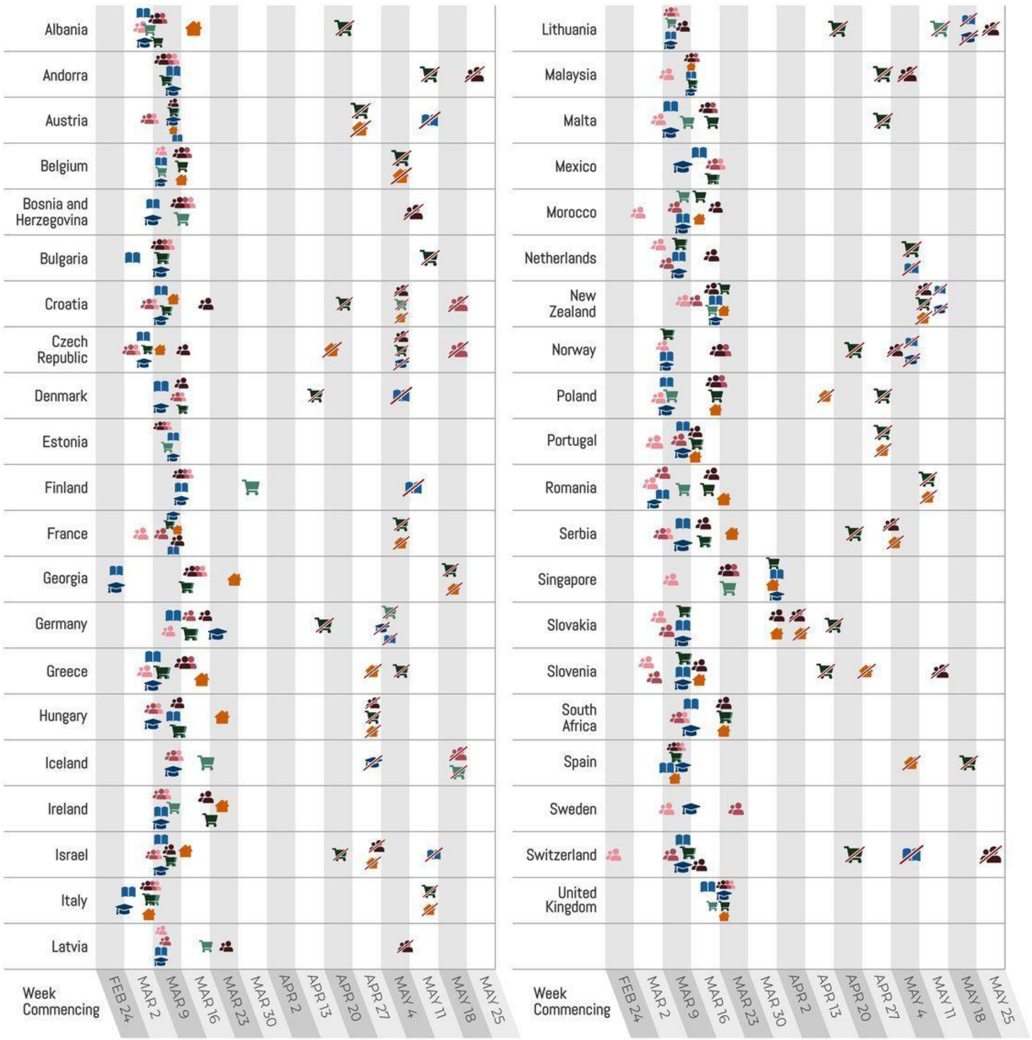
Possible approaches and challenges

- 1) Controlled trials
  - Politically and ethically challenging
- 2) Simulations
  - Assumptions lead to foregone conclusions
- 3) Cohort studies
  - Confounding
  - Only works for some interventions
- **4) Observational multi-region studies**
  - Need diverse data
  - Need high-quality intervention data
  - Need assumptions that can affect conclusions
  - Confounding
  - Past and future effectiveness may be different

Our NPI Data

41 Countries

- The European response was somewhat uncoordinated!
- Different countries implement *different* NPIs at different times!
- **Verified with independent double entry!**

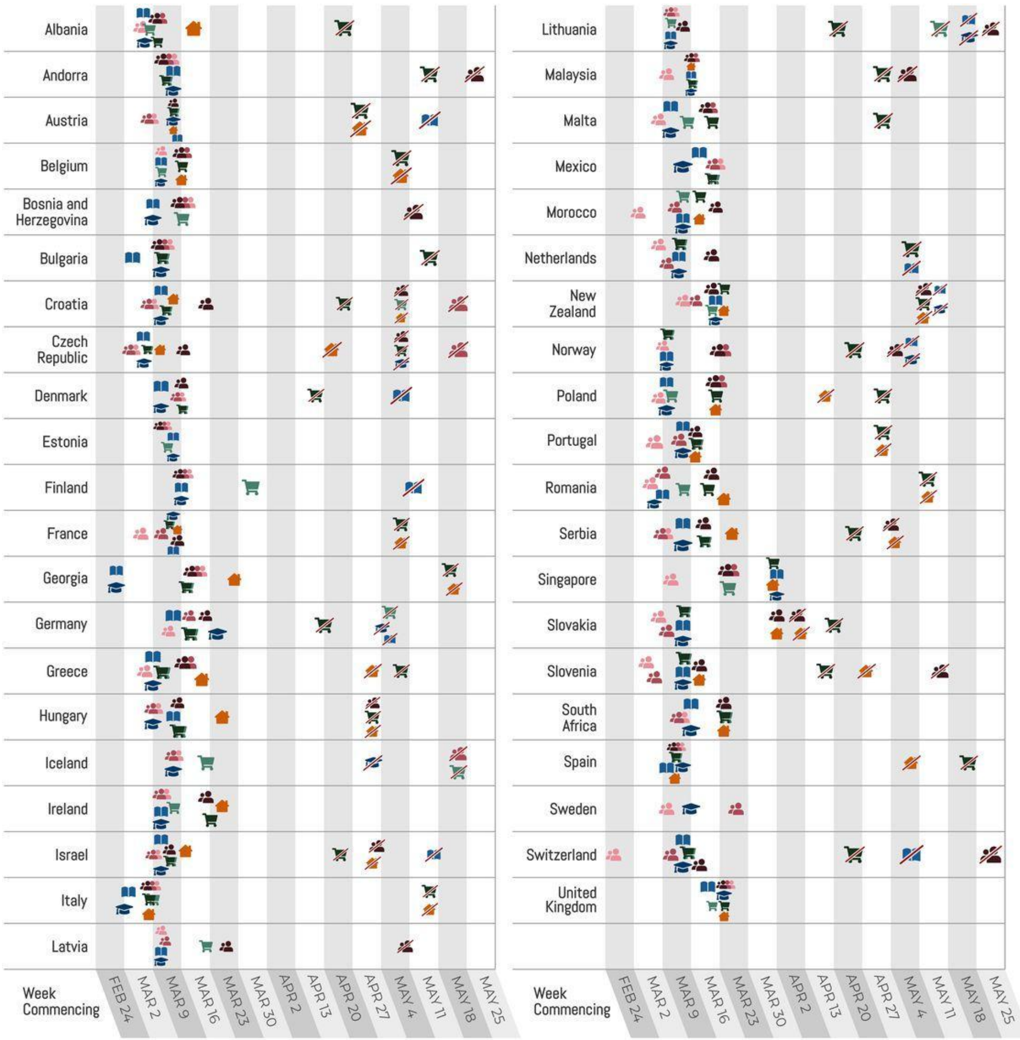


Our NPI Data

Why collecting our own data?

- Many public datasets, but the data quality is poor. (Focus on breadth rather than accuracy.)

- **Verified with independent double entry!**



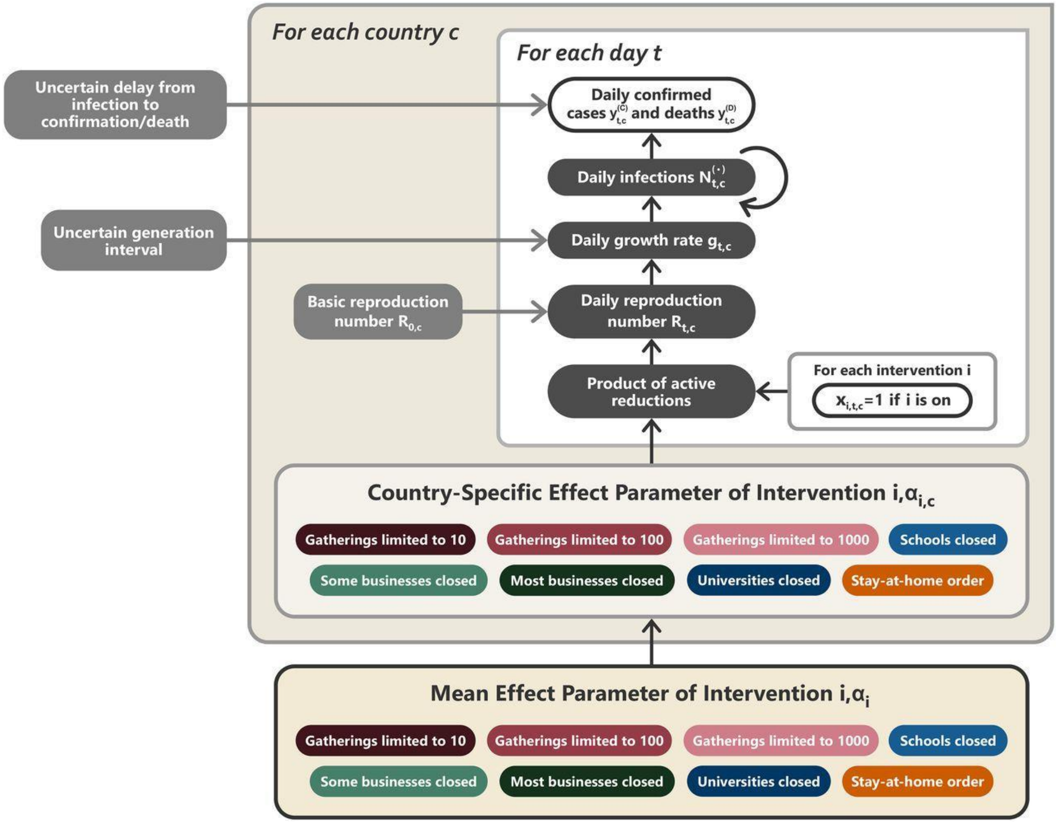
Cases and deaths

- Several possible sources: **John Hopkins University**, ECDC, WHO
- Obvious problems:
  - Changes in testing
  - Reporting is chaotic

Epidemiological parameters

- You need at a minimum:
  - Serial Interval/Generation Interval
  - Delay from infection to case confirmation
  - Delay from infection to death
- This is handled sloppily in most work.
- You want:
  - Distributions
  - Uncertainty over parameters

Model Overview



Modelling Challenges

- Constant, country-level differences in the:
  - ascertainment rate - *the proportion of infections reported*.
  - infection-fatality rate - *the proportion of infections that lead to death*.
- **Time-varying differences** in the ascertainment rate and infection-fatality rate.
- Biases in testing and reporting

Approach

Key Components

- NPI Interactions:

$$R_{t,c} = R_{0,c} \prod_{i} \exp(-\alpha_i x_{i,t,c})$$

Base R

NPI effectiveness

NPI activations

Approach

Key Components

- NPI Interactions:

Base R

NPI effectiveness

$$R_{t,c} = R_{0,c} \prod_i \exp(-\alpha_i x_{i,t,c})$$

NPI activations

- Infection Model:

$g_{\tau,c} = f(R_{t,c}, \text{GI})$

  - Cases: $N^{(C)}_{t,c} = N^{(C)}_{t-1,c} g_{t,c} \cdot \exp \varepsilon^{(C)}_{\tau,c}$

$\varepsilon^{(C)}_{\tau,c} \sim \text{Normal}(\mu = 0, \sigma = \sigma_g)$

Transmission Noise

An infection that *later* is confirmed

Infections on previous day

Approach

Key Components

- NPI Interactions:

Base R

NPI effectiveness

$$R_{t,c} = R_{0,c} \prod_i \exp(-\alpha_i x_{i,t,c})$$

NPI activations

- Infection Model:

$g_{\tau,c} = f(R_{t,c}, \mathrm{GI})$

  - Cases: $N^{(C)}_{t,c} = N^{(C)}_{0,c} \prod_{\tau=1}^{t} [g_{\tau,c} \cdot \exp \varepsilon^{(C)}_{\tau,c}]$
  - Deaths: $N^{(D)}_{t,c} = N^{(D)}_{0,c} \prod_{\tau=1}^{t} [g_{\tau,c} \cdot \exp \varepsilon^{(D)}_{\tau,c}]$

An infection that *later* dies

Latent initial sizes

$\varepsilon^{(C)}_{\tau,c} \sim \text{Normal}(\mu = 0, \sigma = \sigma_g)$

$\varepsilon^{(D)}_{\tau,c} \sim \text{Normal}(\mu = 0, \sigma = \sigma_g)$

Transmission Noise

Approach

Key Components

- NPI Interactions:

$$R_{t,c} = R_{0,c} \prod_i \exp(-\alpha_i x_{i,t,c})$$

Base R

NPI effectiveness

NPI activations

- Observation model

$$\overline{C}_{t,c} = \sum_{\tau=1}^{t} N_{t-\tau,c}^{(C)} P_C(\text{delay} = \tau)$$

$$\overline{D}_{t,c} = \sum_{\tau=1}^{t} N_{t-\tau,c}^{(D)} P_D(\text{delay} = \tau),$$

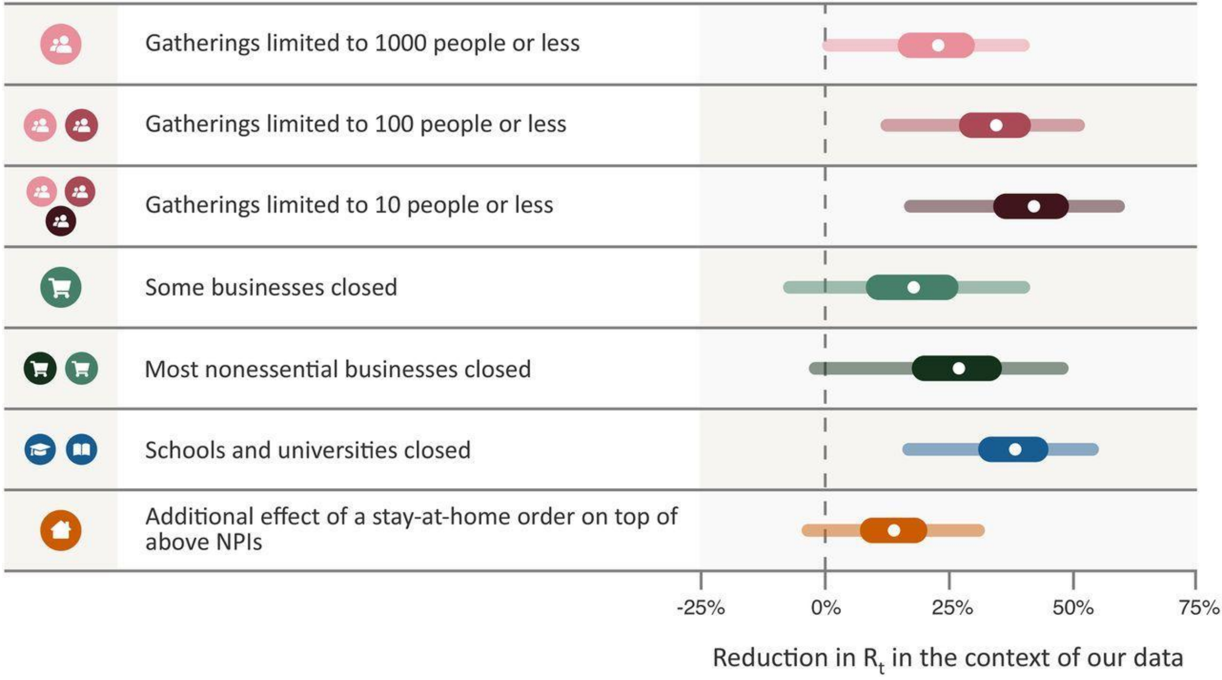
Recap: Key Model Features

- We extend the model of Flaxman, S., Mishra, S., Gandy, A. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261 (2020). https://doi.org/10.1038/s41586-020-2405-7
- Our model observes **both cases and deaths.**
- We **account for uncertainty in key epidemiological parameters**, such as the delays between infection and case/death reporting.
- We add noise to the measure of transmission i.e., we use **transmission noise**.

Main Results

Default Settings

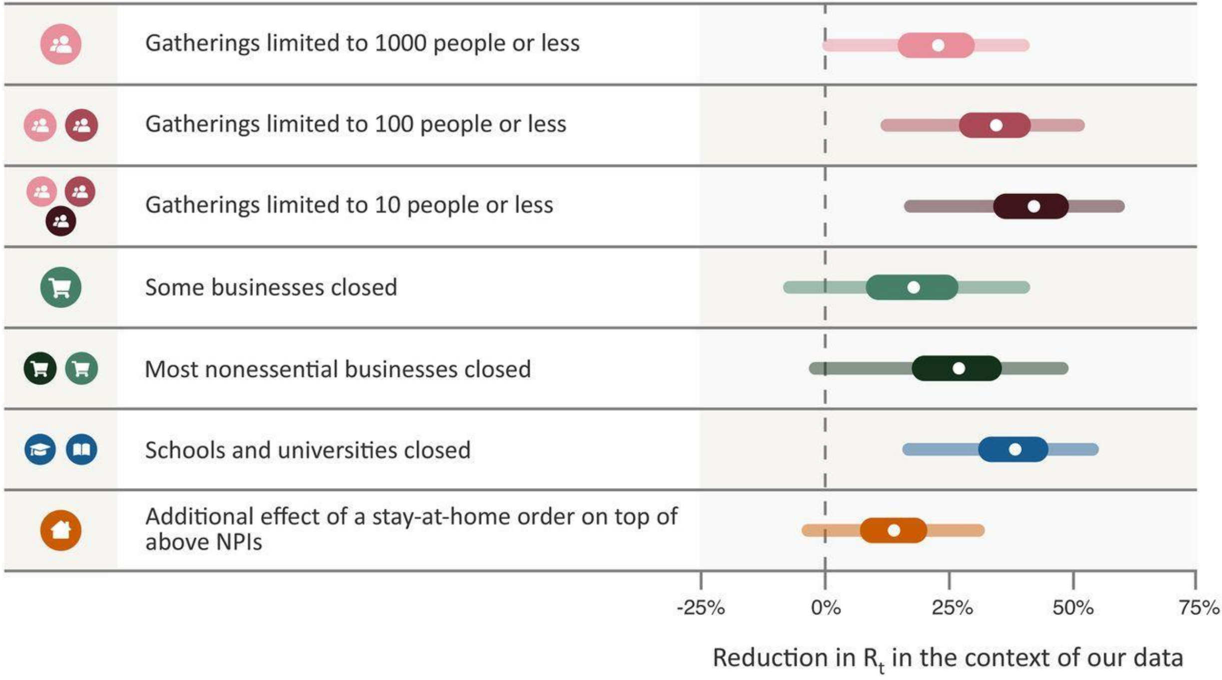
- Note: with *priors* over uncertain epidemiological parameters.
- (obtained under default settings)



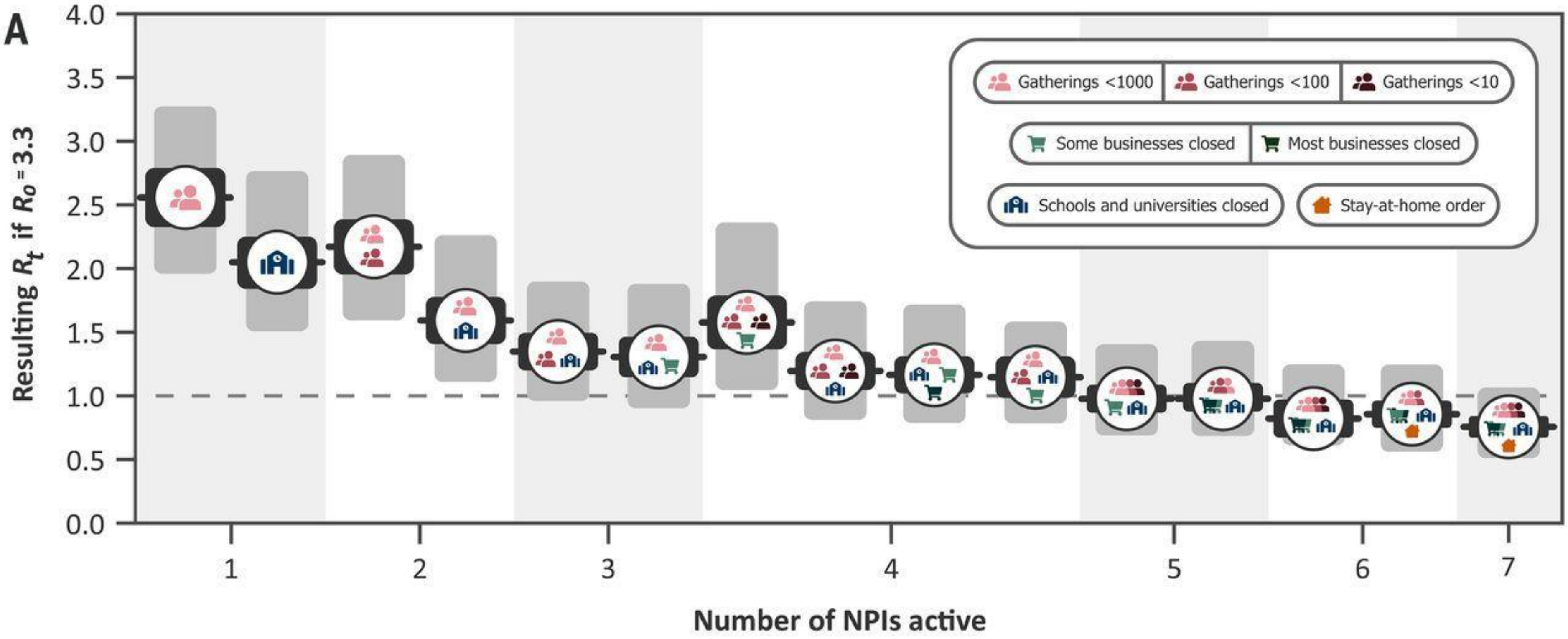
Results

Default Settings

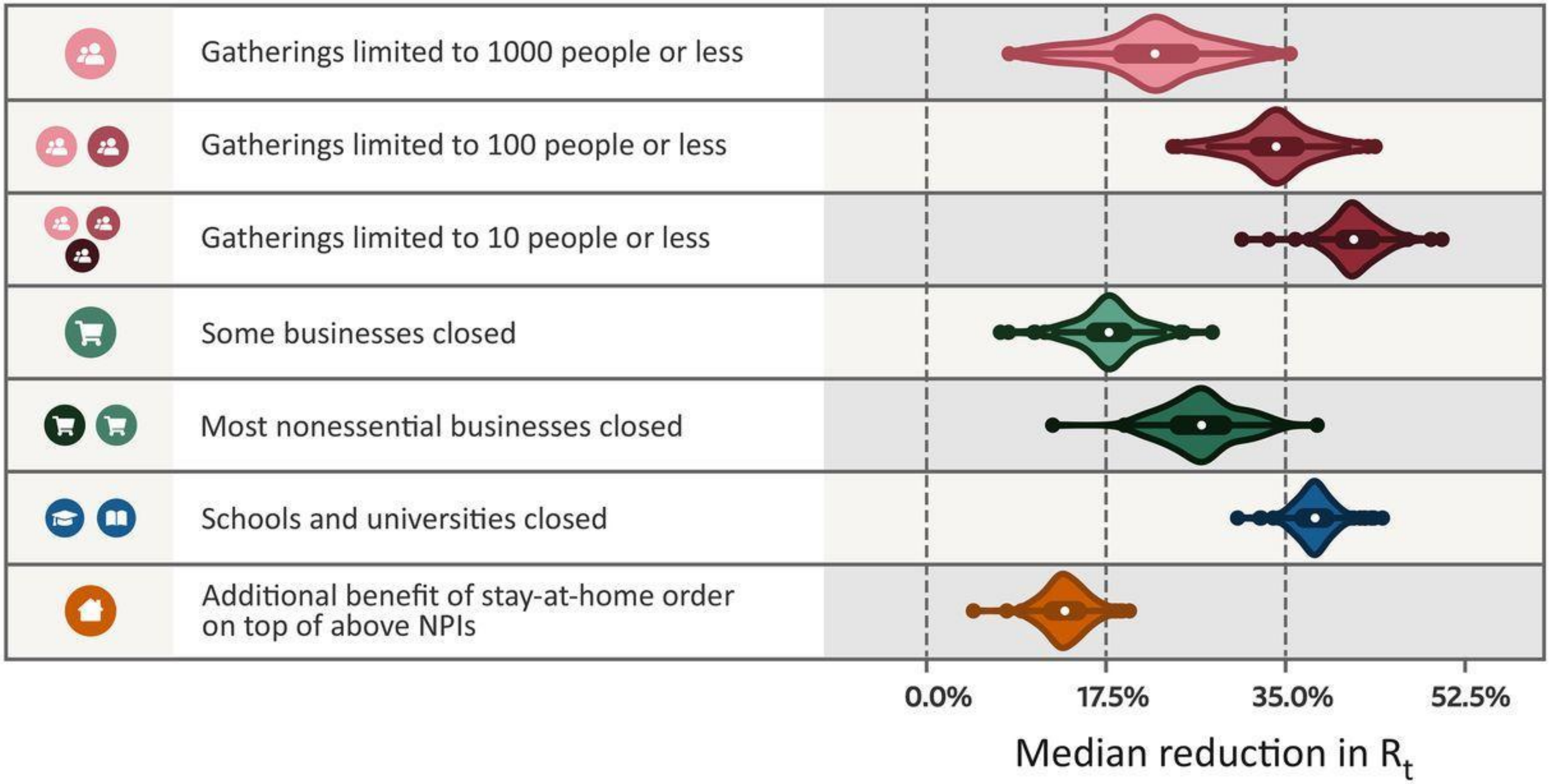
- Note: with *priors* over uncertain epidemiological parameters.
- Adjustment required for local circumstances.



NPI Combinations

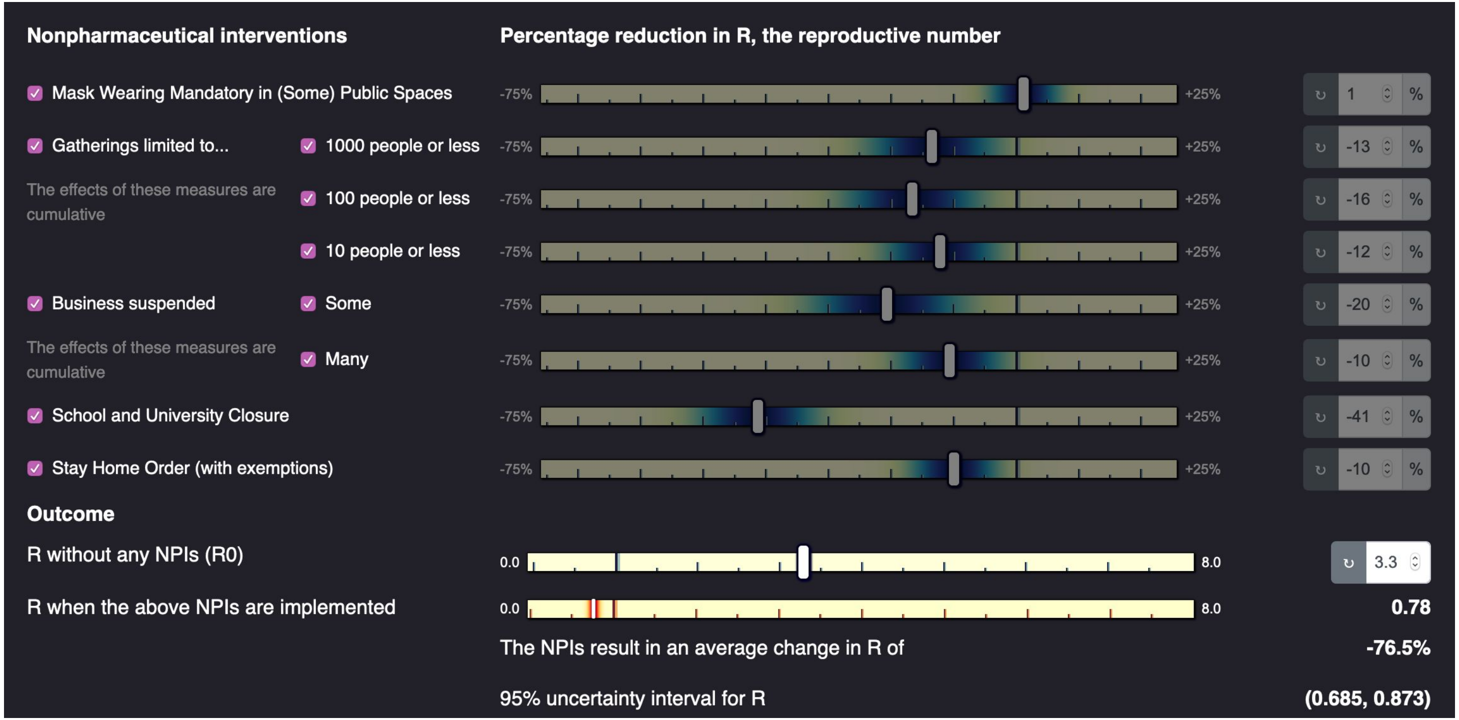


Sensitivity analyses (206 conditions)



Mitigation Calculator

http://epidemicforecasting.org/calc



Some **Limitations**

- Assumed that NPI effectiveness doesn’t vary across countries and time.
- Assumed that NPIs *don’t interact*.
- Our model doesn’t account for numbers of susceptible people *changing* over time.
- No age-stratification ...

More **Limitations**

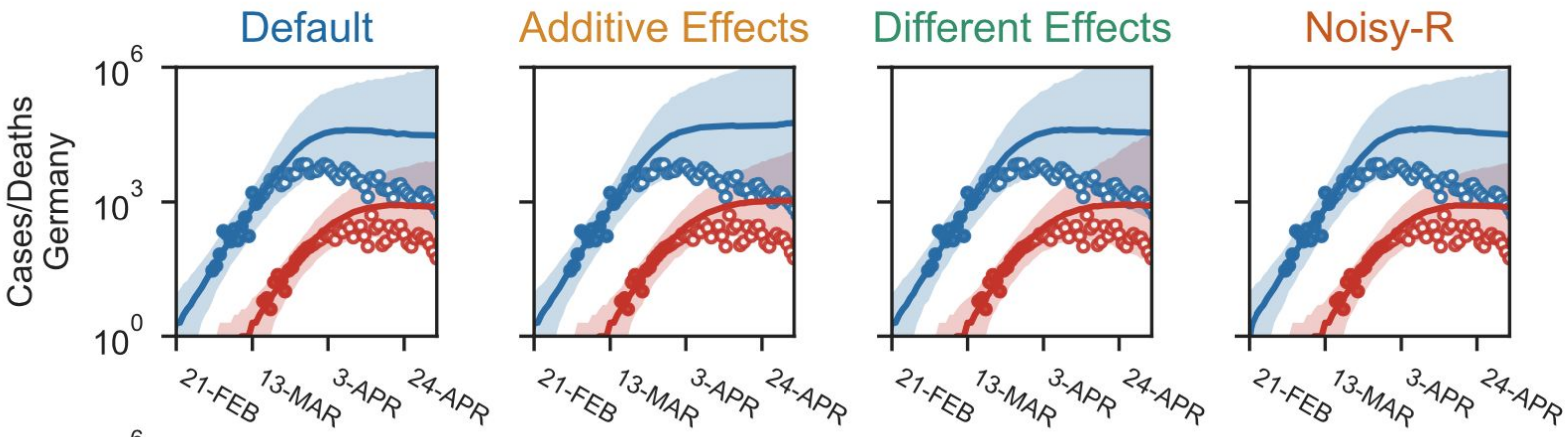
(There are many others too …)

- General comment: **we've made a lot of assumptions**.
  - e.g., NPI interactions, infection model, parameter values, …..
- How much are unobserved factors attributed to our NPIs?
  - And, we *know* we have unobserved factors! Behaviour change, unrecorded NPIs, …

Can we trust our estimates?

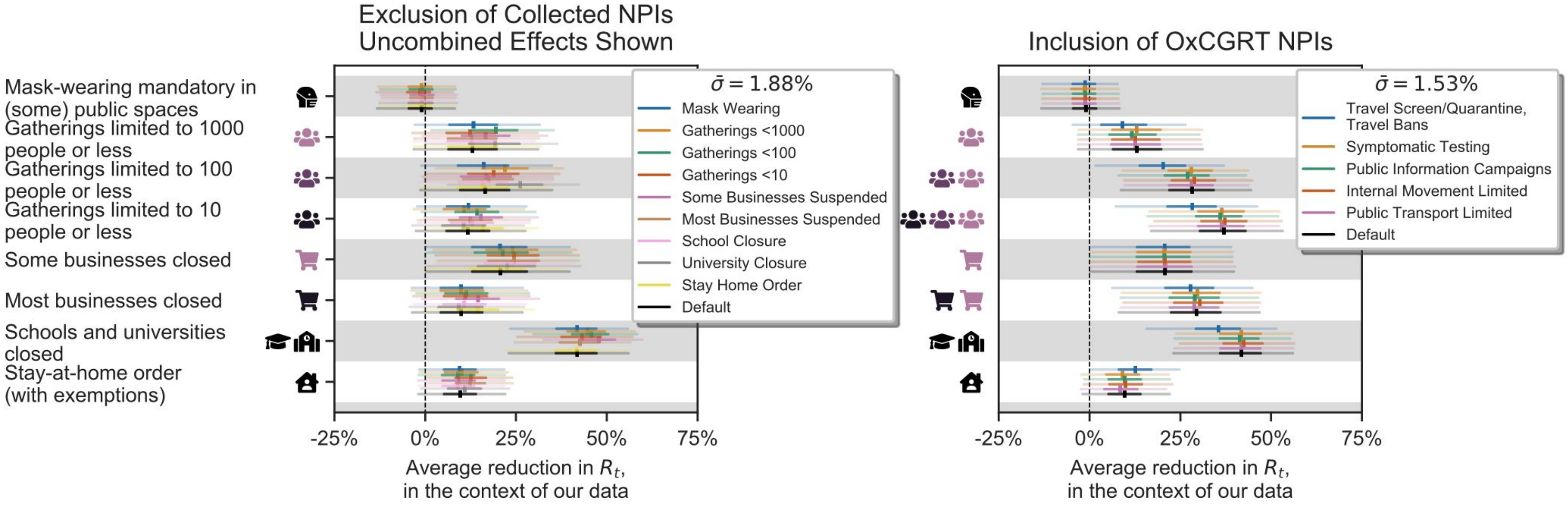
Holdout Validation

- We don't aim to forecast cases and deaths.
- But if our estimates don't *help* us to predict cases and deaths, they aren't useful!



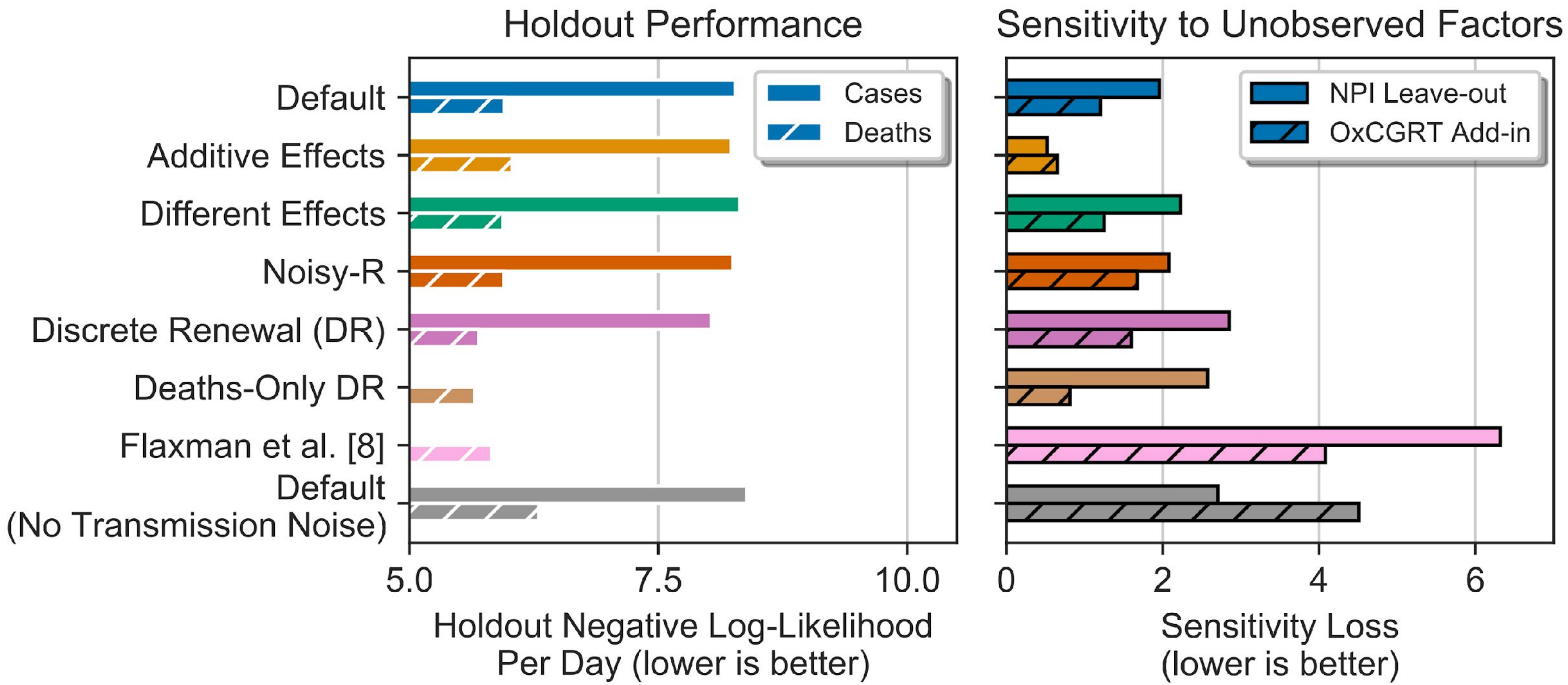
Can we trust our estimates?

How to test for unobserved factors?

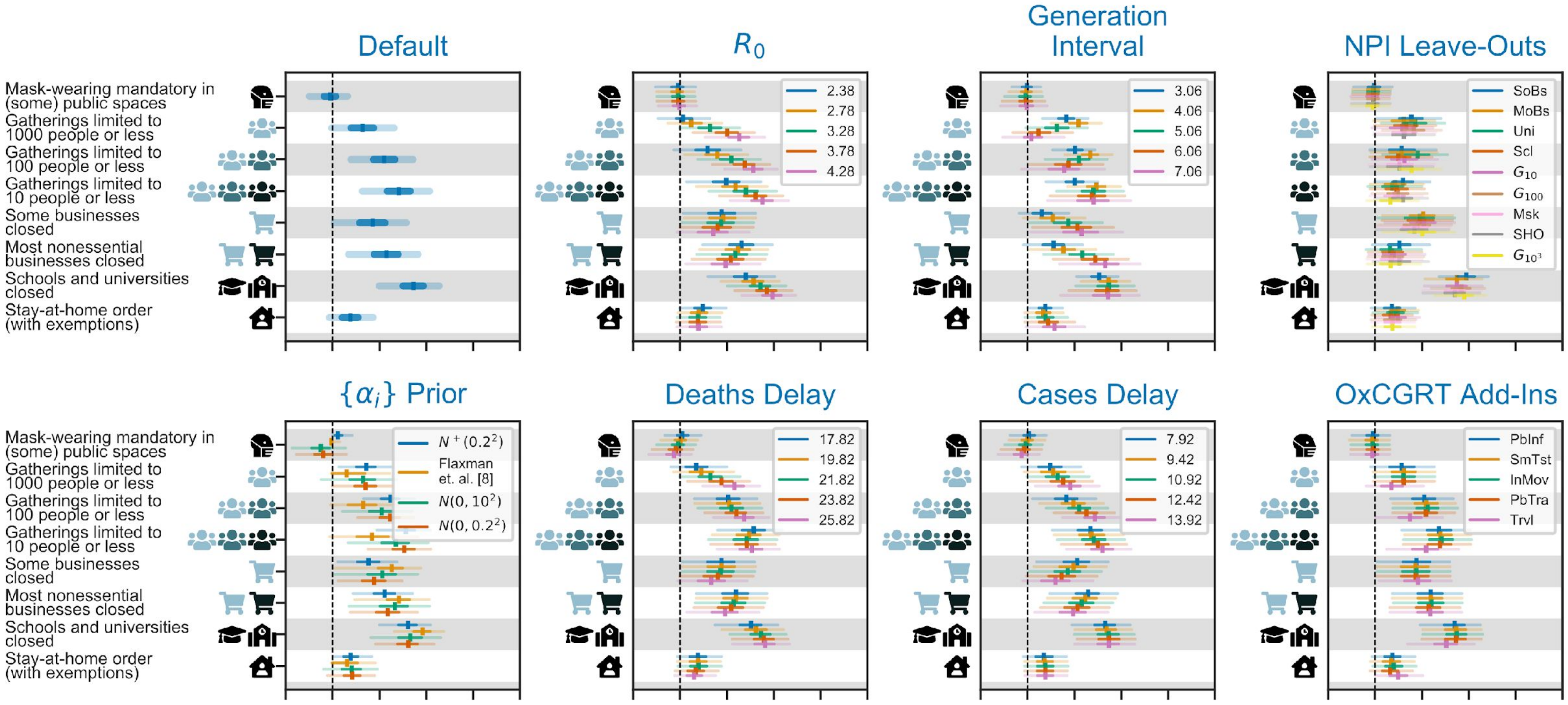


Model Comparison

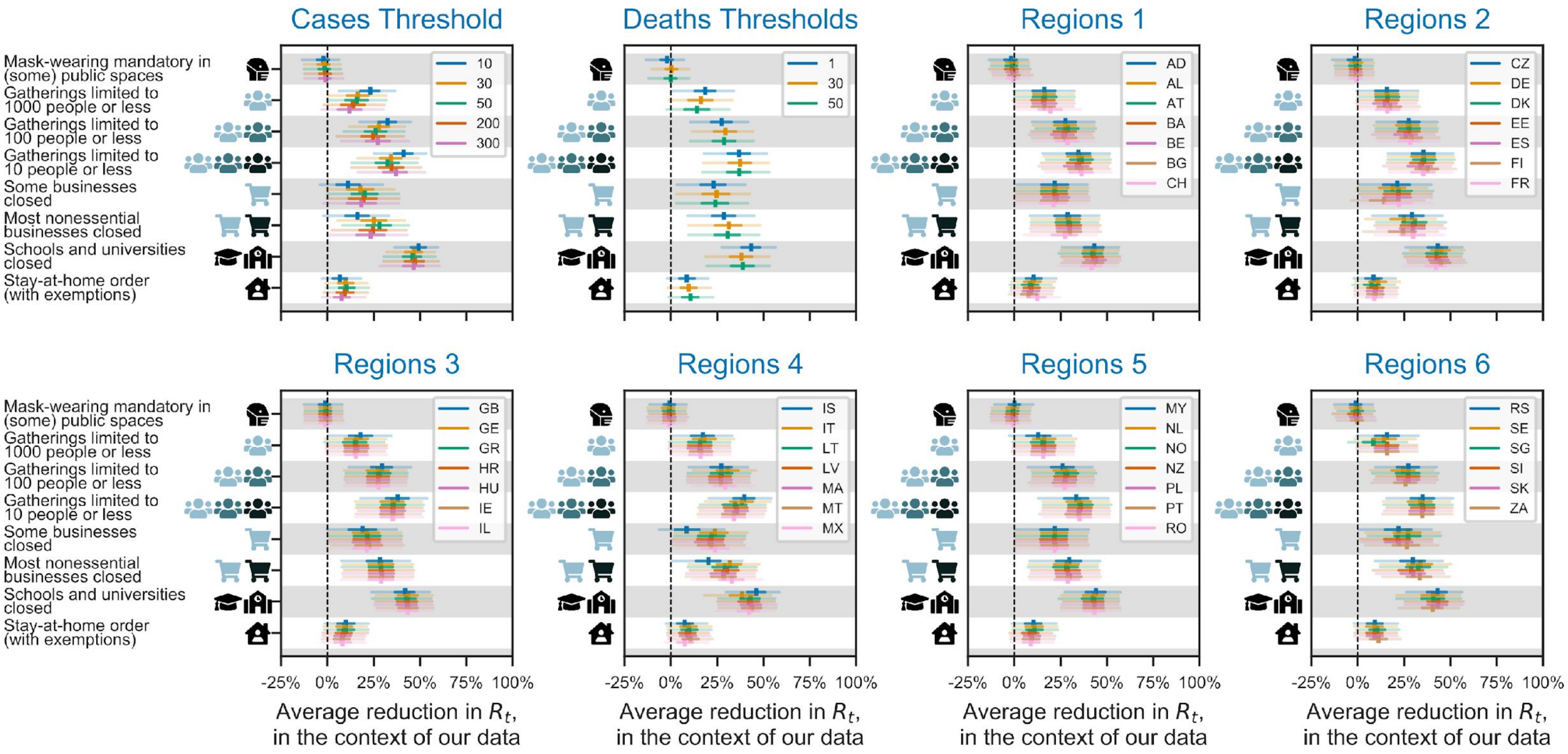
Transmission noise helps!



Sensitivity Analysis I



Sensitivity Analysis II



Structural Sensitivity