

Modelling the importance of population behaviour and seasonality on SARS-CoV-2 spread using Google community mobility reports and weather data

Nordita workshop, Stockholm

Felix Günther¹, Hilde Kjelgaard Brustad², Arnaldo Frigessi², Tom Britton¹
May 31, 2023

¹ Mathematical Statistics, Department of Mathematics, Stockholm University

² Oslo Center of Biostatistics and Epidemiology, University of Oslo



UiO • University of Oslo



NordForsk



Stockholms
universitet



NIPH
Norwegian Institute of Public Health

The NordicMathCovid collaboration

*For preparedness against Covid-19
and to create a Nordic network of
epidemic modelers.*

PIs: Tom Britton (SWE), Lasse
Leskelä (FIN), Arnaldo Frigessi
(NOR)

3 Post Docs:

Felix Günther (Stockholm University)

Mikhail Shubin (Aalto University)

Hilde Kjelgaard Brustad (University of
Oslo)

0.9 M EUR, 2021-2022



<https://nordicmathcovid.cs.aalto.fi/>

- Retrospective modelling of the SARS-CoV-2 pandemic (applied to data for Norway and Sweden)
- Main research questions:
 1. How was spread of pandemic associated with different categories of population behaviour?
(Do we find association between Google mobility reports and infection dynamics?)
 2. What role does seasonality play?

- Google COVID-19 community mobility reports
- The model
 - Data description
 - Metapopulation model
 - Transmissibility as a function of behaviour
 - Observational model - hospital admissions
- Results
 - Model applied to data for Norway and Sweden
- Summary
- Limitations/discussion

Google COVID-19 Community Mobility Reports

- Mobility reports describe the **daily activity level relative to pre-pandemic baseline** in six categories/transmission settings:

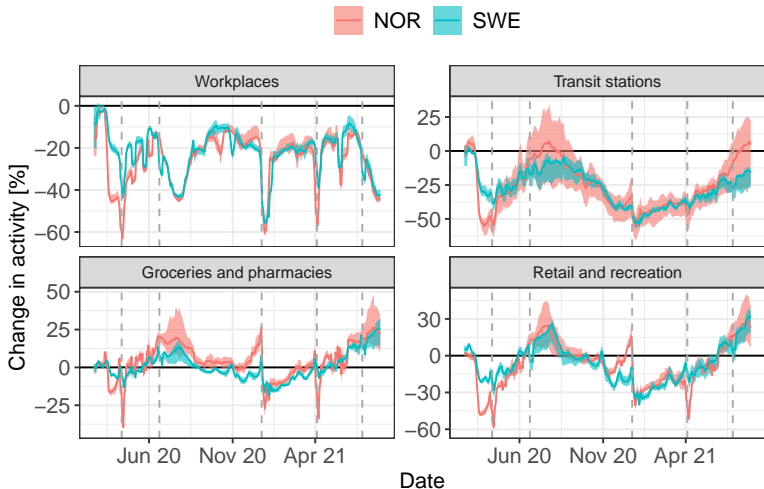
$$\frac{\text{Number of mobile devices in a category at day } t}{\text{Baseline activity level}}$$

- Baseline mobility: mean activity of 5 consecutive pre-pandemic weeks (Jan 3 - Feb 6, 2020), weekday specific
- Noise is added by Google for privacy reasons, some imputation necessary
- Reported from beginning of pandemic until Oct 15, 2022
- Available in fine grained regional resolution

COVID-19 community mobility report settings

- **Grocery & pharmacy:** grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies
- **Retail & recreation:** restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters
- **Transit stations:** public transport, subway, bus, and train stations, sea ports, taxi stands, car rental agencies
- **Workplaces:** time spent in places of work
- **Residential:** time spent in places of residence
- **Parks:** public gardens, castles, national forests. Much missing data.

COVID-19 community mobility reports



Average community mobility reports in NOR and SWE (right-aligned 7-day average and 25/75% quantiles).

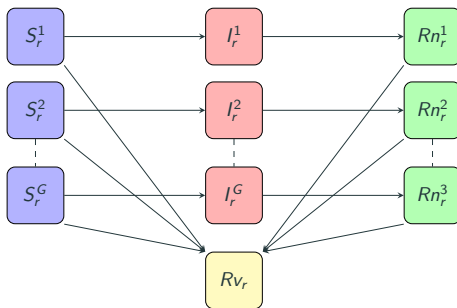
Model for spread spread

- Extended **discrete time SIR model** per region (daily)
- Use age-groups to account for **vaccination with age-prioritization**
- Vaccination: remove known number of vacc. individuals from **Susceptible** and **Recovered Naturally** to **Recovered Vaccinated**
- No further age-structure (homogeneous mixing)
- Fixed fraction of Susceptibles are **imported** infections per day
- **Transmissibility parameter** $\beta_r(t)$ will be modeled as function of region-specific covariates
- Account for VoCs (α and δ)
- **Fit model to weekly hospitalization counts** using likelihood that incorporates external information on hospitalization risk per age-group; increased by VoCs

- **Pandemic spread:** Weekly number of new hospital admissions per region
- **Covariates** (daily, per region):
 - Google COVID-19 community mobility reports
 - Temperature
 - Share of circulating virus variants (only have national data)
- **Other data:** Daily number of new vaccinated individuals (received two doses), by region and age group

Model for disease spread

Extended discrete time (daily) SIR model per region



- S_r^i - Susceptible
- I_r^i - Infectious
- Rn_r^i - Recovered naturally
- Rv_r - Recovered from vaccination.
- $i = 1, 2, \dots, G$ - Age groups,
- r - Region (fylke, län)

Model for disease spread

Transitions between compartments:

$$S_{t+1,r}^i - S_{t,r}^i = - \frac{\beta_{t,r}}{N_r} I_{t,r} S_{t,r}^i - \frac{n_{imp}}{10^5} S_{t,r}^i - p_{vac} V_{t,r}^i \frac{S_{t,r}^i}{Rn_{t,r}^i + S_{t,r}^i}$$

$$I_{t+1,r}^i - I_{t,r}^i = \frac{\beta_{t,r}}{N_r} I_{t,r} S_{t,r}^i + \frac{n_{imp}}{10^5} S_{t,r}^i - \gamma I_{t,r}^i$$

$$Rn_{t+1,r}^i - Rn_{t,r}^i = \gamma I_{t,r}^i - p_{vac} V_{t,r}^i \frac{Rn_{t,r}^i}{Rn_{t,r}^i + S_{t,r}^i}$$

$$Rv_{t+1,r} - Rv_{t,r} = p_{vac} \sum_{i=1}^G V_{t,r}^i \frac{S_{t,r}^i}{Rn_{t,r}^i + S_{t,r}^i} + p_{vac} \sum_{i=1}^G V_{t,r}^i \frac{Rn_{t,r}^i}{Rn_{t,r}^i + S_{t,r}^i}$$

- $\beta_{t,r}$ - Average number of infectious contacts/time unit
- N_r - Total population of region r
- n_{imp} - imported cases per 100k susceptibles/day (=0.5)
- $V_{t,r}^i$ - Number of new vaccinations
- p_{vac} - Vaccine efficacy (=0.9)
- γ - Inverse of infectious period (=1/6)

Principled model for $\beta_{t,r}$

$\beta_{t,r}$: average number of secondary infections produced by infect. individual in region r at day t in fully susceptible population

- We want to relate $\beta_{t,r}$ to the Google mobility reports
- Assume: secondary infections occur in K different settings, e.g., workplaces, public transport, ...
- $\beta_{t,r} = \sum_{k=1}^K \beta_{t,r,k}$
- $\beta_{t,r,k} = c_{t,r,k} \times p_{k,r}^{inf}$
 - $c_{t,r,k} > 0$: average number of contacts in setting k at day t
 - $p_{k,r}^{inf} \in [0, 1]$: probability of a contact leading to an infection (depending on setting and region; Assumption: time-constant)

Principled model for $\beta_{t,r}$ (II)

- Available Google data $x_{t,r}^k$ do not correspond to absolute number of contacts $c_{t,r,k}$
- Assume linear association between activity level and number of contacts

$$x_{t,r}^k = \frac{c_{t,r,k}}{c_{r,k}^{BL}} - 1 \Leftrightarrow c_{t,r,k} = (1 + x_{t,r}^k) \times c_{r,k}^{BL}$$

- $\beta_{t,r,k} = (1 + x_{t,r}^k) \times c_{r,k}^{BL} \times p_{r,k}^{inf} = (1 + x_{t,r}^k) \times \phi_{k,r}$
 - $\phi_{k,r} > 0$ (daily) number of secondary infections in setting k for region r at baseline activity (unknown parameter, estimated)
- Linearity assumption: when activity reduced by 50%, contacts reduced by 50% and secondary infections are $0.5 \times \phi_{k,r}$
- Basic model: $\beta_{t,r} = \sum_{k=1}^K (1 + x_{t,r}^k) \times \phi_{k,r}$

More refined model for $\beta_{t,r}$

- Linearity assumption plausible for some settings, less plausible for others (residential areas, parks, ...)
- We only use the mobility reports for **four categories: Grocery and pharmacies, Retail and recreation, Transit stations, and Workplaces**
- Transmission can also happen in settings not captured by these four above
- Include week- and region-specific iid. parameter $v_{\text{week}(t),r}$ (estimated from data) that accounts for transmission that is not captured by those categories:

$$\beta_{t,r} = \sum_{k=1}^K (1 + x_{t,r}^k) \times \phi_{k,r} + v_{\text{week}(t),r}$$

More refined model for $\beta_{t,r}$ (II)

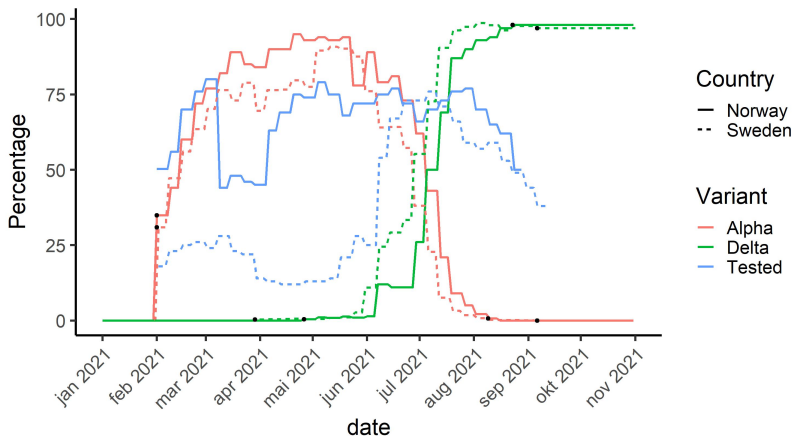
- Transmission dynamics not only related to contact frequencies but also (changes in) transmission probability
- Important drivers can be temperature and virus variants (VoCs)
- We assume that these factors act multiplicatively and specify the final model as

$$\beta_{t,r} = \left(\sum_{k=1}^K (1 + x_{t,r}^k) \times \phi_{k,r} + v_{\text{week}(t),r} \right) \times e^{\phi_{\text{temp}} x_{t,r}^{\Delta \text{temp}} + \phi_{\alpha} x_{t,r}^{\alpha} + \phi_{\delta} x_{t,r}^{\delta}},$$

- $x_{t,r}^{\Delta \text{temp}}$: temp.-diff. in region r at day t compared to day 1
- $x_{t,r}^{\alpha}, x_{t,r}^{\delta}$: share of Alpha and Delta VoC among infected
- $e^{\phi_{\text{temp}}}$: mult. change in transmission prob. for a one-degree increase in temperature
- $e^{\phi_{\alpha}}, e^{\phi_{\delta}}$: mult. change in transmission prob. for VoC compared to wildtype

Variants of Concern

Alpha (B.1.1.7) and Delta (B.1.617.2) among reported cases



Proportion of Alpha and Delta variant among sequenced/screened positive tests.

Likelihood

We use weekly hospitalization counts per region $H_{w,r}$ to fit our model and estimate the set of parameters $(\phi_{k,r}, \nu_{\text{week}(t),r}, \phi_{\text{temp}}, \phi_{\alpha}, \phi_{\delta}, \sigma_r^v)$

Define:

- $\nu_{t,r}$: number of new infected day t , region r
$$\nu_{t,r} = \sum_{i=1}^G S_{t-1,r}^i - S_{t,r}^i - p_{\text{vac}} V_{t,r}^i \frac{S_{t,r}^i}{Rn_{t,r}^i + S_{t,r}^i}$$
- $\kappa_{t,r}$: probability of going to hospital when infected at day t in region r (\rightarrow next slide)
- τ_j : probability of going to hospital j days after infection (given hospitalization, from data)

Set up the likelihood of the weekly hospitalization counts as

$$H_{w,r} \sim \text{Pois} \left(\sum_{t \in w} \sum_{j=0}^m \nu_{t-j,r} \kappa_{t-j,r} \tau_j \right)$$

Likelihood (II)

- $\kappa_{t,r}$: probability of going to hospital when infected at day t in region r
- Changes over time
- We assume that infected drawn randomly from all susceptibles at time $t - 1$
- Relative age-composition among susceptibles in region r changes over time due to age-prioritized vaccination
- VOCs increase hospitalization probability compared to wild-type
- We assume (published) increase of hospitalization risk by 1.9 for α and Δ variant compared to wildtype

Likelihood (III)

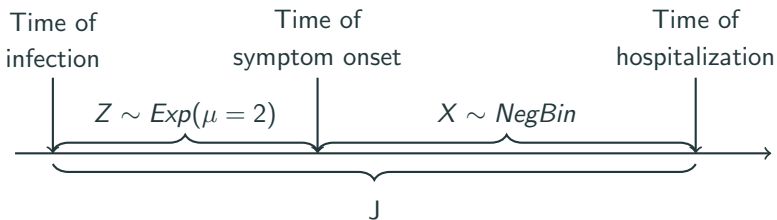
- Age-specific hospitalization risk (wildtype) from literature, adapted to Norwegian population
- For specific age-group i , hospitalization risk day t in region r :

$$\alpha_{t,r}^i = \underbrace{\alpha^{i,WT}}_{\text{risk age } i, \text{ WT}} \left(\underbrace{1 - (x_{t,r}^\alpha + x_{t,r}^\Delta)}_{\text{share of WT}} + 1.9 \cdot \underbrace{(x_{t,r}^\alpha + x_{t,r}^\Delta)}_{\text{share of } \alpha \text{ and } \Delta} \right)$$

- $\kappa_{t,r}$ is weighted sum of age-groups: $\kappa_{t,r} = \sum_{\forall i} \alpha_{t,r}^i \frac{S_{t,r}^i}{S_{t,r}}$

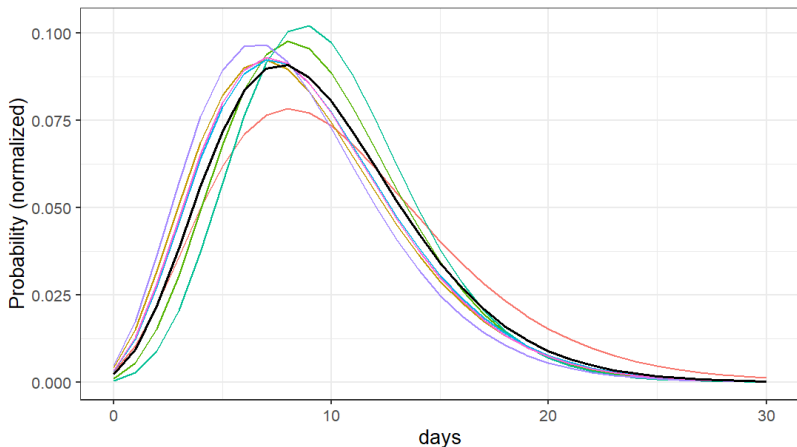
Likelihood (IV)

- τ_j : probability of going to hospital j days after infection (**given hospitalization**, from data)



$$\tau_j = P(J = j) = \sum_{s=0}^j P(X = j - s)P(Z = s)$$

Time from symptom onset to hospitalization



Distribution (normalized) of time from symptom onset to hospitalization, given that an infected individual goes to hospital. Each line corresponds to a time period. Black line gives the "average" over all time periods.

Interpretation of estimated parameters

- During model fitting we estimate large amount of parameters:
 - $\phi_{k,r}$ (4 per region)
 - $\nu_{\text{week}(t),r}$ (75 per region)
 - $\phi_{\text{temp}}, \phi_{\alpha}, \phi_{\delta}$
- Direct interpretation:
 - $e^{\phi_{\text{temp}}}$: Mult. change in transmission prob. for a one-degree increase in temperature
 - $e^{\phi_{\alpha}}, e^{\phi_{\delta}}$: Mult. change in transmission prob. for VoC compared to wildtype
 - $\phi_{k,r}$: average daily infections per infectious individual in transmission setting k at BL activity level in region r
 - $\nu_{\text{week}(t),r}$: daily infections per infectious at day t in region r outside of transmission settings captured by mobility reports
 - Note that $\phi_{k,r}$ and $\nu_{\text{week}(t),r}$ are scaled by the multiplicative effects on a given day

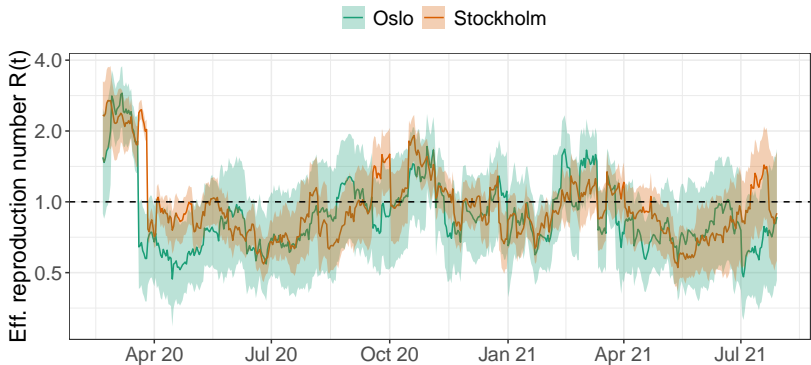
Interpretation of estimated parameters (II)

Derived quantities:

- Based on the estimated parameters $\phi_{k,r}$ and $v_{\text{week}(t),r}$ as well as the observed mobility data $x_{t,r}^k$, we can retrospectively quantify the **share of secondary infections that can be attributed to specific transmission setting**
- For given day t and setting l in region r :
$$\frac{(1+x_{t,r}^l) \times \phi_{l,r}}{\sum_{k=1}^K (1+x_{t,r}^k) \times \phi_{k,r} + v_{\text{week}(t),r}}$$
- Can be aggregated over multiple settings:
$$\frac{\sum_{l=1}^4 (1+x_{t,r}^l) \times \phi_{l,r}}{\sum_{k=1}^K (1+x_{t,r}^k) \times \phi_{k,r} + v_{\text{week}(t),r}}$$
- averaged over whole time-period (or even regions)
- Retrospective quantity that contains observed behaviour patterns (no quantification of potential for reduction of transmission!)

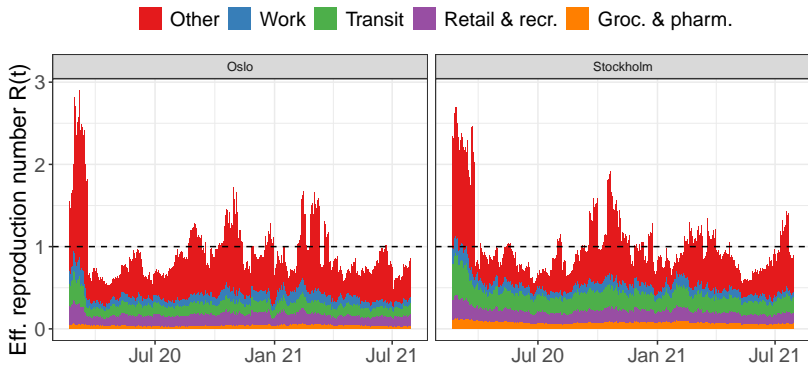
- Implementation the model in `cmdstanr`
- Parameter estimation via Hamiltonian Monte Carlo
- Fit model separately to all Norwegian and all Swedish regions
- NOR: 11 fylker, SWE: 21 län
- Feb 21, 2020 - July 31, 2021 (75 weeks)
- Initialization: 0.5 infectious per 100.000 of population in each region and age-group
- MCMC converged (4 chains, nice mixing, \hat{R} close to 1 for all parameters)

Effective reproduction number



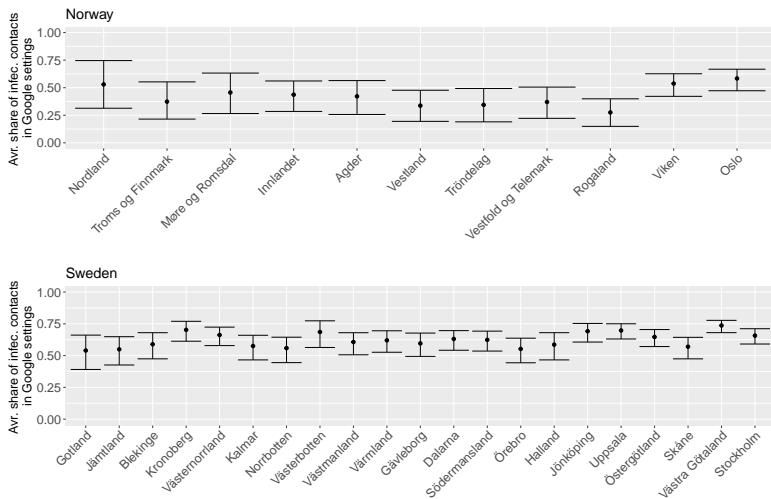
Model-based R_t (posterior median + 90% credible interval)

Effective reproduction number by Google setting

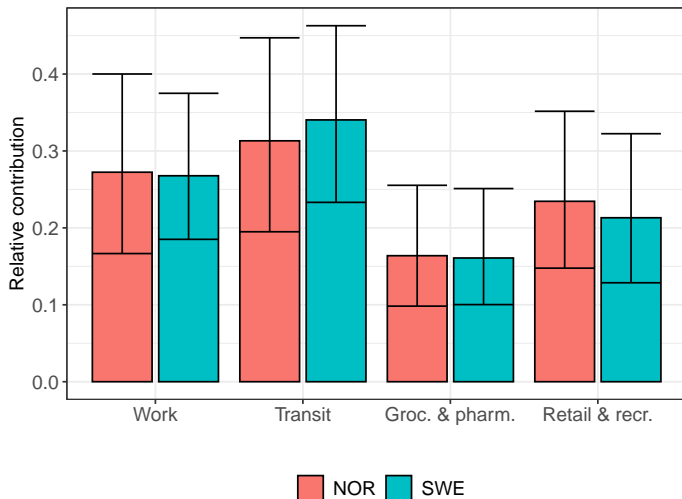


Contribution of different transmission settings to the effective reproduction number $(\beta_{t,r,k}/\gamma)$

Share of infectious contacts assigned to Google settings

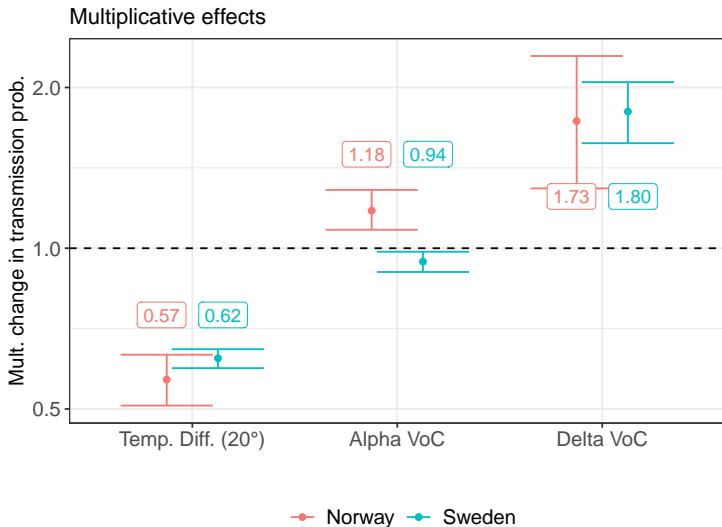


Relative contribution of Google settings at baseline



Share of secondary infections assigned to each Google mobility report setting at base line level, weighted by population density + 90% credible interval

Multiplicative effects



Estimated multiplicative effects (posterior median + 90% credible intervals)

Summary

- Developed principled model to analyse association of Google community mobility reports with infection dynamics of SARS-CoV-2 in Norwegian and Swedish regions
- Depending on the region, Google mobility data explained 25-75% of (changes) in dynamics in disease spread (linear model)
- Stronger association in SWE than in NOR
- Evidence that public transport and workplaces might play larger role for disease spread than groceries/pharmacies and retail/recreation (at baseline activity)
- Strong seasonality (accounted for using multiplicative effect of temperature): $\sim 50\%$ less transmission in summer than winter

Limitation and Discussion

- Results depend on (strong) model assumptions:
 - Linear relation between measured activity and number of contacts in four settings
 - Assumptions on SIR structure and parameters, e.g., vaccination efficacy, recovery rate, infection-hospitalization ratio, no waning of immunity in observation period, ...
- Small association between transmission setting and observed disease dynamics might be related to few infections in this setting or bad measurement quality for specific setting/region
- Google community mobility reports capture (some) changes in population behaviour that is related to dynamics of pandemic
- Interpretation of raw numbers difficult
- More direct measurement of (absolute) population activity or contact frequencies desirable

The End