# Estimating Local Protective Behavior in Denmark with dynamic MRP

HOPE Project

2021-03-16

The goal of this brief report is to introduce the HOPE project's efforts in estimating COVID-19 protective behavior at local levels in Denmark.

# The Data

We are relying on two rolling panel surveys collected by Gallup with a new wave of 500 respondents on each day. Here we rely on data from the last 3 months, since October 2. Gallup recruits participants through stratified random sampling based on a database of CPR numbers (Danish social security numbers).

**Note that due to a server error the samples between December 30 and January 2 have only between 44 and 234 respondents. Estimates for these days are imprecise, and pulled towards the mean of the 3 month study period. Our simple models do not make any assumption about time trends, even though this may improve estimates for these days.**

As of 2021-03-16 our sample size is:

- Gallup N = 54335

Variables from the Gallup survey:

- **ContactFamily1m**: Number of physical contacts with non co-residing family members.[1]
- **ContactFriends1m**: Number of physical contacts with friends.
- **ContactColleagues1m**: Number of physical contacts with colleagues.
- **ContactStrangers1m**: Number of physical contacts with stranger.
- **SumContact1m**: Total number of non-home contacts - the sum of the four contact variables (family, friends, colleagues, strangers).
- **Contact_attention**: An index on compliance with 5 protective behaviors: avoiding physical contact, distancing from elderly and sick, keeping 1-2m distance from others in general, avoiding crowds
- **Hygiene_attention**: An index on compliance with 3 protective behaviors: handwashing, sneezing and coughing in sleeve, ensuring frequent cleaning.

# The Method: MRP – Multilevel Regression with Post-stratification.

MRP has been developed as a technique to estimate attitudes on subnational level relying on data from national public opinion surveys. In recent years, it has been validated as a tool for generating reliable and valid subnational estimates of public opinion, which is superior to most naive alternatives (e.g. slicing up data). In a nutshell. MRP relies on both demographic and geographic variables in the modeling process and employs partial pooling (to the extent warranted by the data) in order to improve estimates for smaller units. There many excellent (https://en.wikipedia.org/wiki/Multilevel_regression_with_poststratification) and intuitive (https://scholar.princeton.edu/sites/default/files/jkastellec/files/mrp_primer.pdf) introductions about MRP online, so we refrain from a general introduction here. Instead, we give a brief summary of the analytical steps we took. We implemented MRP relying on the *rstanarm* R library in a fully Bayesian framework.

## Stage 1 - Estimating individual level protective behavior with multilevel models

First, we seek to build a model of individual level protective behavior. Our independent variables are the following:

- age_cat: 5year age categories from 20-24 … to 70+. (11 levels)
- female: participant sex. (2 levels)
- edu_clean: Participant's highest level of education (7 levels)
- region: Participant's place of residence (5 levels).
- date: Date of interview (~92 levels)

The R code for the full multilevel model is displayed below. Note that we add an indicator for weekends as a time smoother and that we model age conditional on gender.

Moreover, note that the type of regression is conditional on the DV. For count variables (contacts and handwashing) we run negative binomial regressions, for binary variables (avoid crowds of 10+) we run logistic regressions and for the continuous variable (level of behavior change) we run simple linear regression.

```
model <- stan_glmer(
        DV ~ 1 + weekend +
                (1 | age_cat:female) +
                (1 | edu_clean) +
                (1 | region) +
                (1 | date) +
                (1 | region:date),
        family = neg_binomial_2,
        data = data,
        seed = 3012, iter = 2000,
        control = list(adapt_delta = 0.99)
)
```

We evaluate the performance of the models with leave-one-out (LOO) cross-validation.

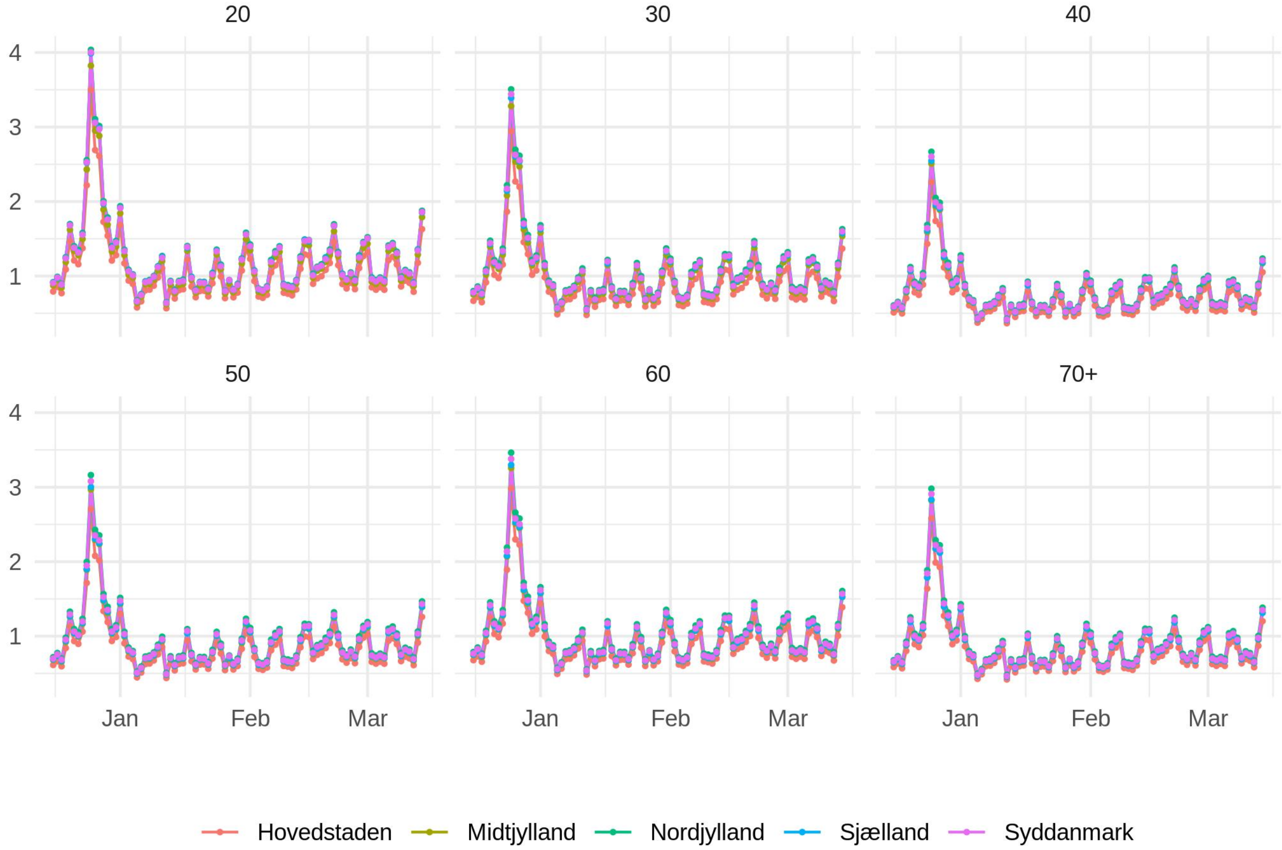
## Stage 2 - Simulation and Post-stratification

As the multilevel model informs us about the 'effects' of various demographic factors, we can next simulate the outcome for every type of person in our model. In other words, we predict the expected outcome for every possible combination of age, sex, education, region and date. $(11 * 2 * 7 * 5 = 770)$. Next, we rely on census data from the Statistics Denmark (https://www.statistikbanken.dk/statbank5a/default.asp?w=1440) to weigh the frequency of each type of respondent in each region.[2] We repeat this procedure for every day in our data. (We assume population distribution does not change between our dates.) We then aggregate these for region × 10 year age groups and estimate 50% and 90% credible intervals based on the posterior distributions.

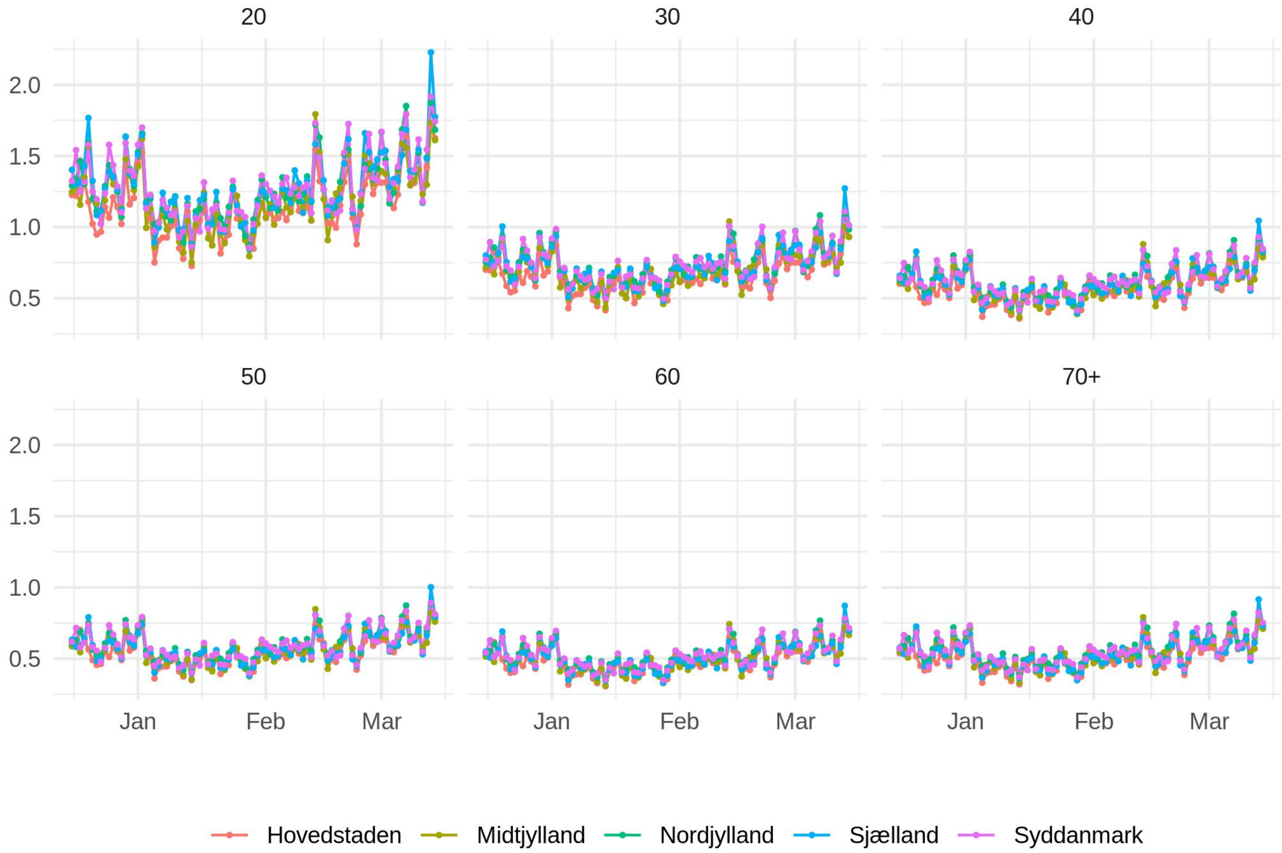
# Results

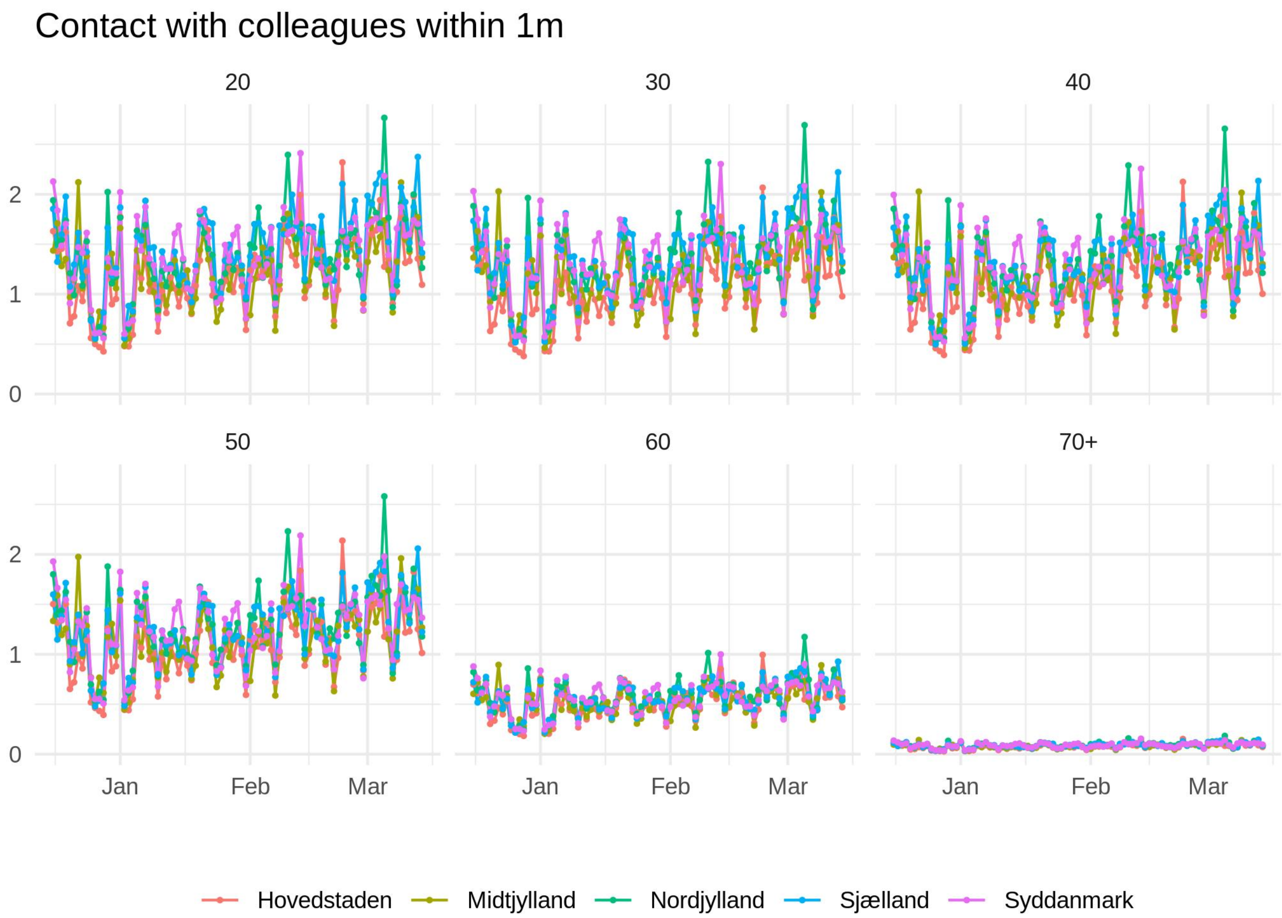
The plots below show our main results: the level of protective behavior for every age group in every region for every date in our data. We omitted credible intervals, which would be highly overlapping given the limited variation by region.

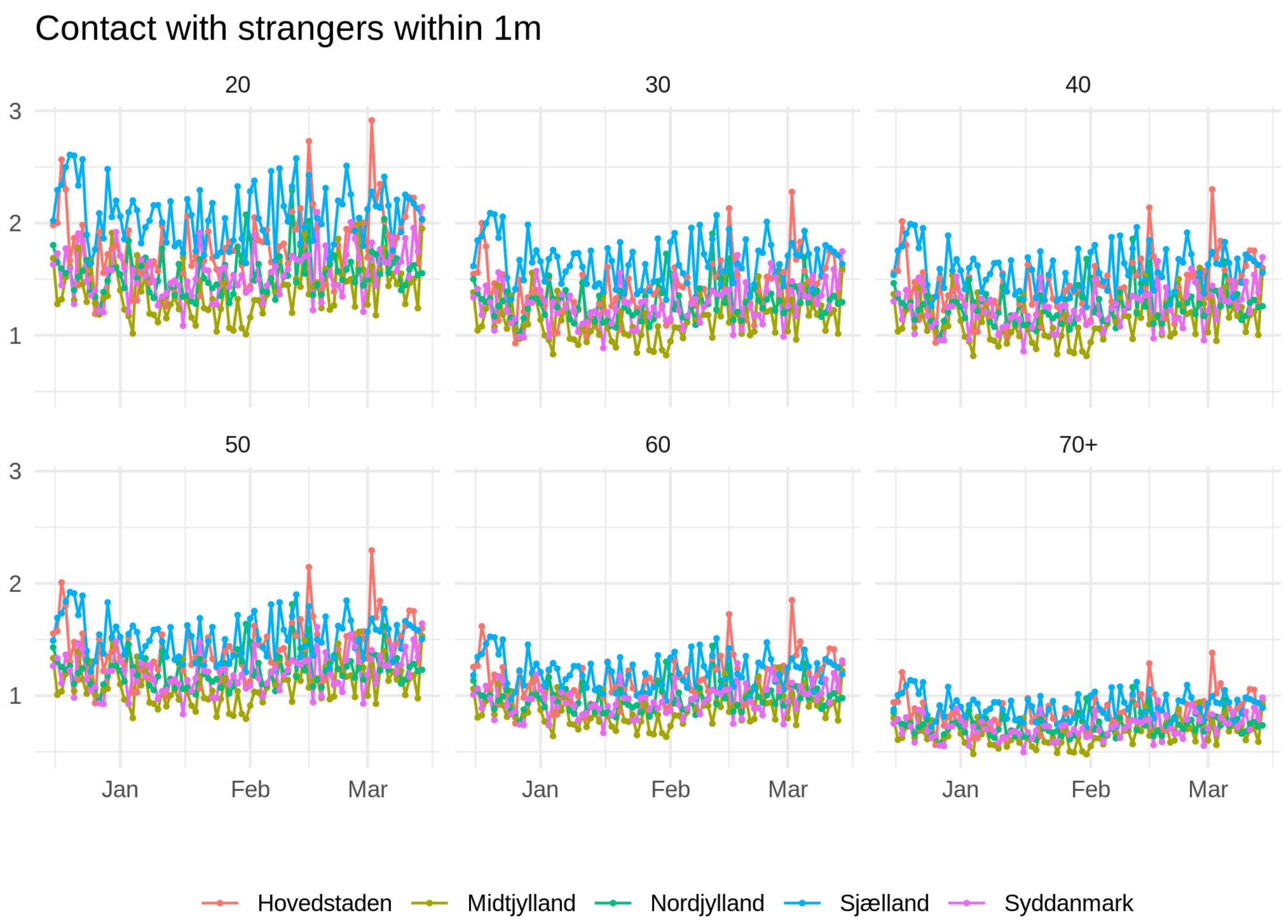
## Contact with family within 1m



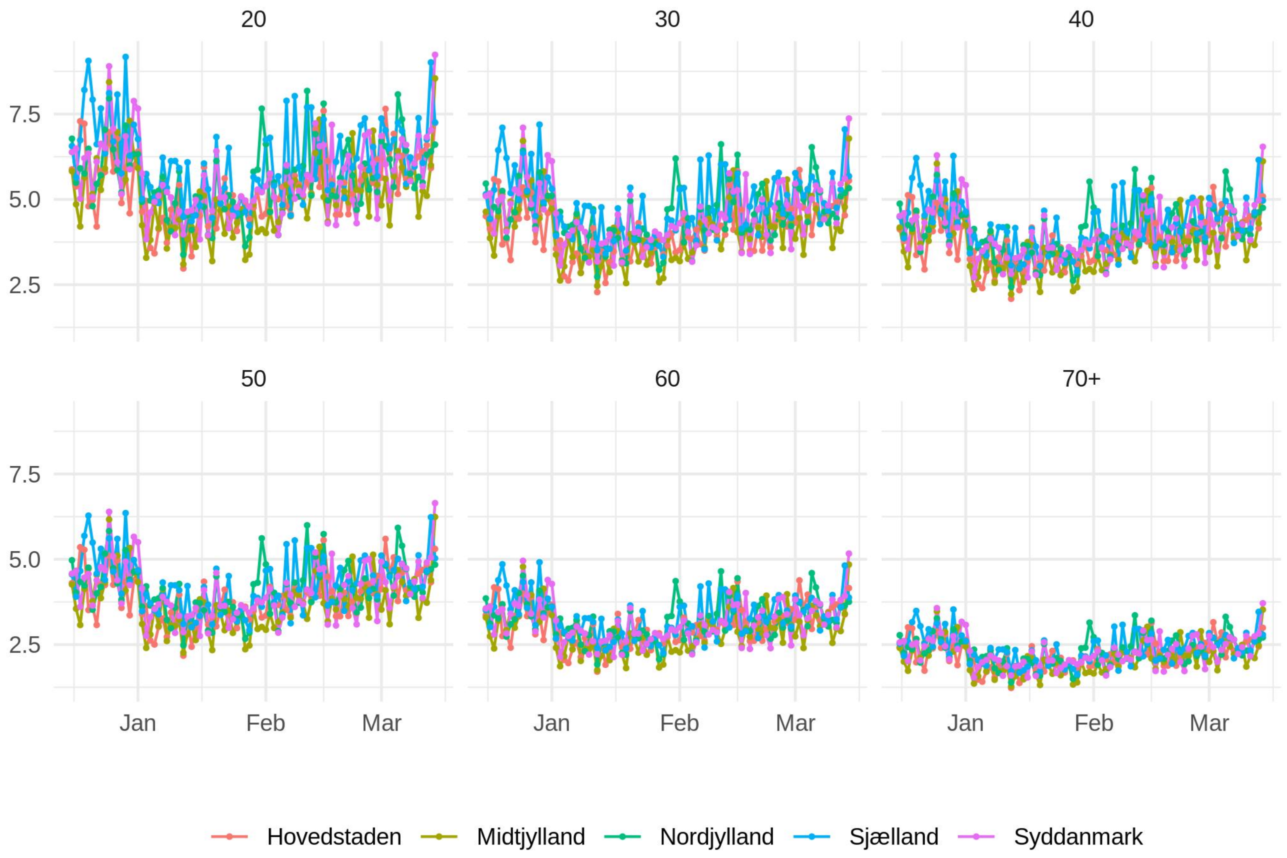

## Contact with friends within 1m



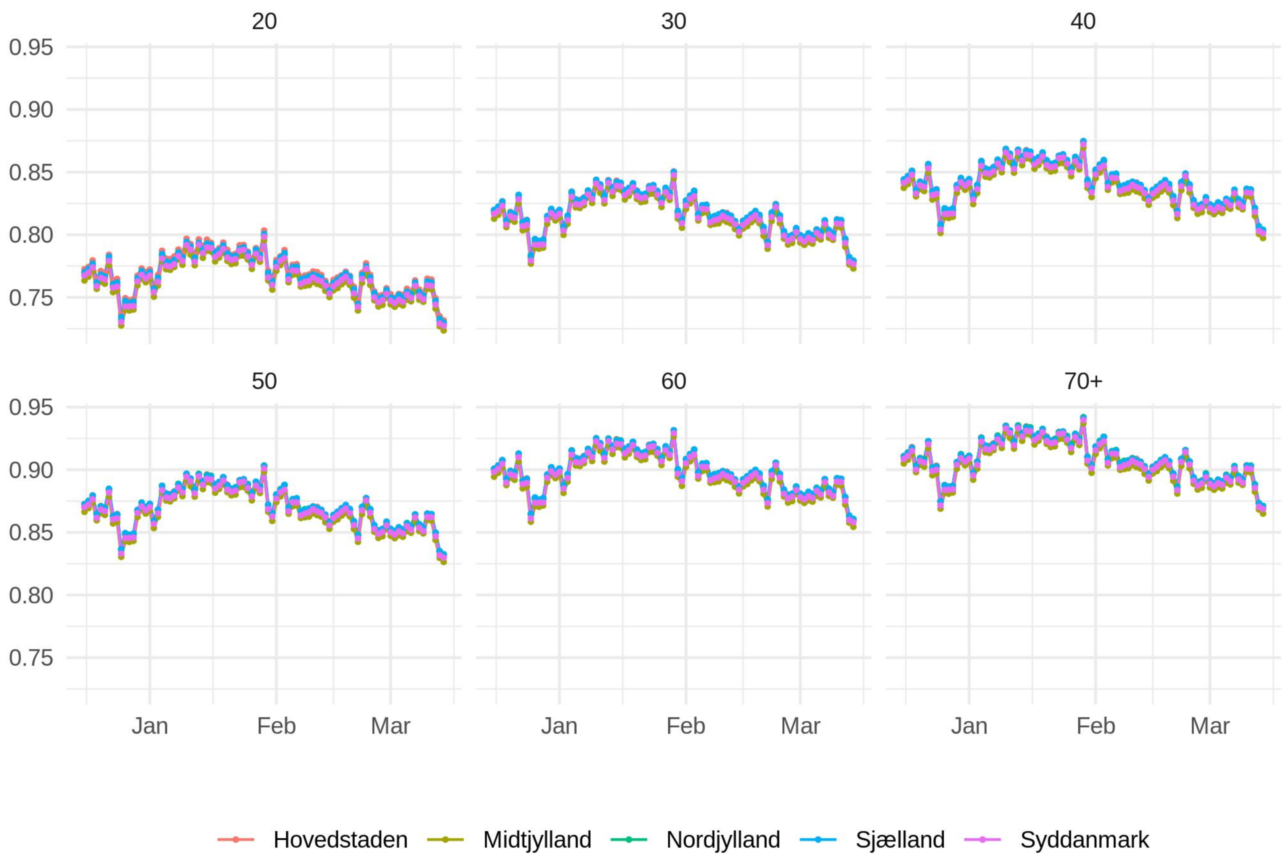

Contact with colleagues within 1m
20
30
40
50
60
70+
0
1
2
Jan
Feb
Mar
Hovedstaden
Midtjylland
Nordjylland
Sjælland
Syddanmark

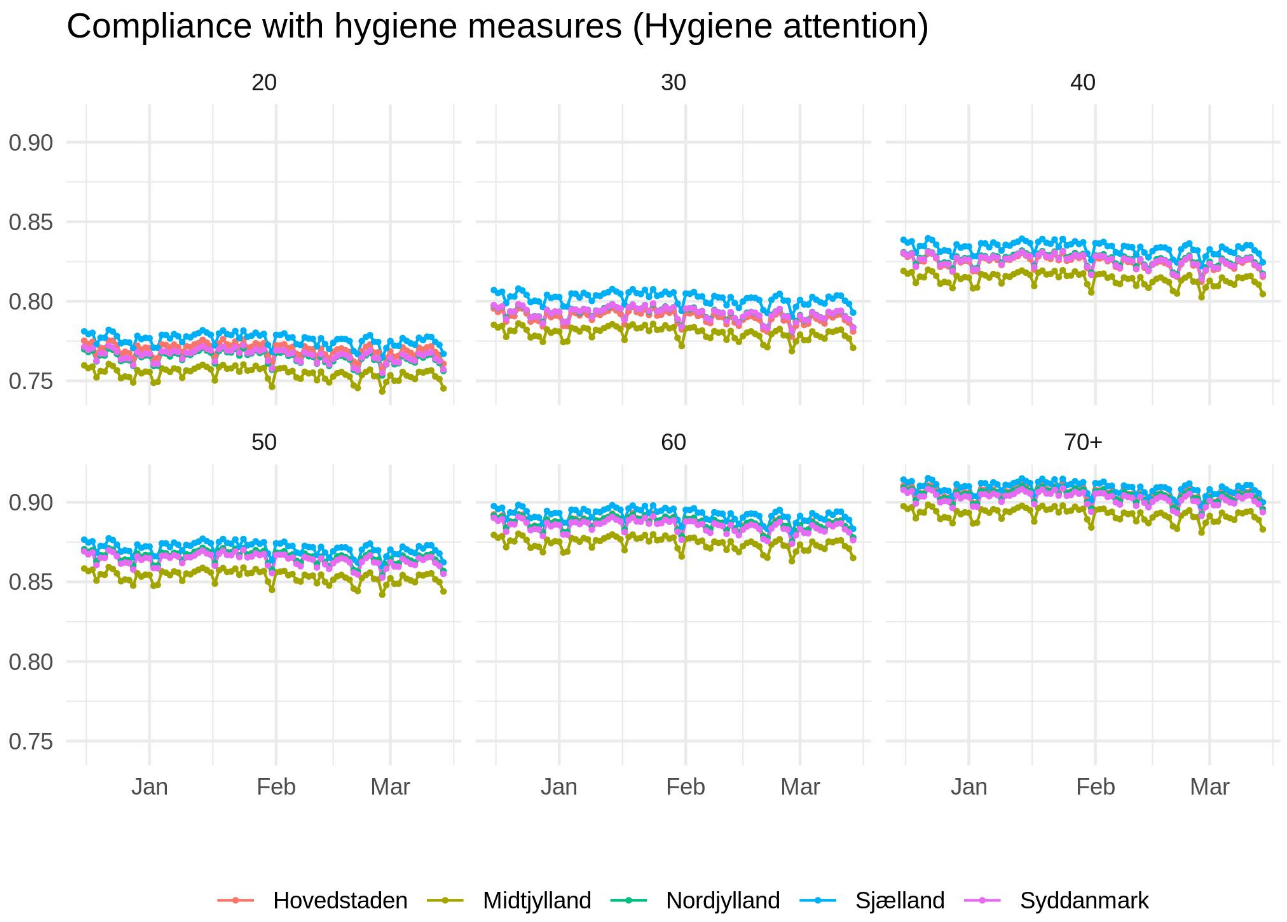

Contact with strangers within 1m
20
30
40
50
60
70+
1
2
3
Jan
Feb
Mar
Hovedstaden
Midtjylland
Nordjylland
Sjælland
Syddanmark

Sum of 1m contacts outside of household




Compliance with social distancing (Contact attention)






## Contact

If you have any comments about these models, please contact Alexander Bor (mailto:alexander.bor@ps.au.dk).

## Acknowledgements

We are grateful for help by Magnus Storm Rasmussen, Mathias Tromborg, and Gabor Simonovits.

---

1. Specifically, respondents are asked to remember “how many people they have been physically close to in the past 24 hours. Physically close is understood here as closer than 1 meters for at least 15 minutes.”↩
2. The publicly available census info is limited to 5 year age categories between 15 and 69, we purchased the Census data for the elderly (70+).↩