Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 211–226
Marseille, 20-25 June 2022

Causal Investigation of Public Opinion during the COVID-19 Pandemic via Social Media Text

Michael Jantscher, Roman Kern
Know-Center Graz
Graz University of Technology
Graz, Austria
mjantscher@know-center.at, rkern@tugraz.at

Abstract

Understanding the needs and fears of citizens, especially during a pandemic such as COVID-19, is essential for any government or legislative entity. An effective COVID-19 strategy further requires that the public understand and accept the restriction plans imposed by these entities. In this paper, we explore a causal mediation scenario in which we want to emphasize the use of NLP methods in combination with methods from economics and social sciences. Based on sentiment analysis of Tweets towards the current COVID-19 situation in the UK and Sweden, we conduct several causal inference experiments and attempt to decouple the effect of government restrictions on mobility behavior from the effect that occurs due to public perception of the COVID-19 strategy in a country. To avoid biased results we control for valid country specific epidemiological and time-varying confounders. Comprehensive experiments show that not all changes in mobility are caused by countries implemented policies but also by the support of individuals in the fight against this pandemic. We find that social media texts are an important source to capture citizens' concerns and trust in policy makers and are suitable to evaluate the success of government policies.

Keywords: Mediation Analysis, Social Media, COVID-19

1. Introduction

As the COVID-19 pandemic started to spread around the world in 2020, many countries worldwide were forced to implement stringent non-pharmaceutical interventions (NPIs) to reduce the transmission of the virus and to protect their citizens. In the absence of pharmaceutical treatments and preventative vaccines in 2020, success in containing and slowing the spread of the virus relies on a good strategic response program. The evidence shows that social distancing orders, implemented by various countries, appear to be the most effective strategy to reduce the transmission of the virus at that time (Haug et al., 2020; Badr et al., 2020). To standardize these response programs and make them comparable across countries Hale et al. (2021) provides a normalized measure called the Stringency Index. However, not only government restrictions contribute to infection reduction but also voluntary and awareness-driven behavior of individuals has an obvious impact as the research from Goolsbee and Syverson (2021; Farboodi et al. (2020) and Yang et al. (2020) clearly demonstrates.

In this paper, we contribute to an emerging field of research that emphasizes the role of social preferences and their importance for policy makers, so far mainly associated with works from the field of econometrics. Policy makers still resort to cross sectional telephone or online surveys at specific times during the pandemic. However, such surveys can get quite expensive and fail to capture public reactions towards COVID-19 containment measures, their trust in policy makers and sources of information or mental health implications across a broad population and continuously over time. Social media posts can act as a suitable proxy, offering the opportunity to study their impact on policy and decision makers in close to real-time. Previous work using social media data in the field of economics and epidemiology (Zhuravskaya et al., 2020; Al-Garadi et al., 2016; Algaba et al., 2020; Sridhar and Getoor, 2019a) opens a door to a fairly new and promising area of applications.

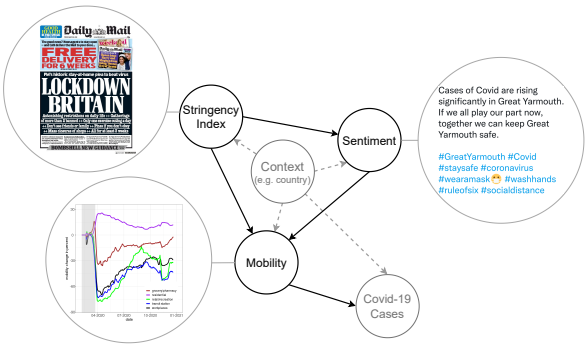


Figure 1: Causal diagram that encodes the direct relationship between government restriction (non pharmaceutical interventions, NPIs), measured via Stringency Index and mobility. In addition we observe the indirect relationship mediated by public opinion, considering social media text and its sentiment as proxy. These relations are affected by several country specific, epidemiological and temporal confounders.

Specifically, we examine the causal effects of stringent restrictions and the response and the behavior of a country's population on a daily level during the pandemic. As a base we use a geotagged Twitter dataset,

including COVID-19 related keywords and hashtags. Citizens' concerns about specific restrictions as well as their perceptions of the impact of containment measures can be estimated by analysing the sentiment of these Tweets. Public sentiment may act as a mediator between the Stringency Index of restrictions and the mobility behavior as shown in the causal graph in Figure 1. Crucially, to avoid biased results, we need to control for confounders that affect the Stringency Index as well as the mediator and the output. We hypothesize that social media texts help to estimate the success of government policies on mobility behavior in different countries. Causal mediation analysis, relying on the potential outcome framework of Imai et al. (2010), allows us to combine these data sources in multiple ways to get valid estimates of the decoupled direct and mediated effects. To test the robustness of the effect estimates, we conduct two different regression specifications and an ablation study for the mediator.

2. Related Work

Policy effectiveness In the field of epidemiology and economics, there is a substantial body of research on the effectiveness of non-pharmaceutical interventions (NPIs) on COVID-19. These public health measures are the most effective interventions when it comes to a pandemic outbreak as long as there is no effective vaccine to protect against it. (Haug et al., 2020; Badr et al., 2020; Brauner et al., 2021; Zhang et al., 2020; Baier et al., 2020; Lucchini et al., 2021) discuss and analyse the success of various implemented NPIs over different countries and phases of the pandemic. Abouk and Heydari (2021) shows that social distancing and especially stay at home orders have the strongest causal impact on the spread of the COVID-19 virus. The individual support of health measurements is usually analysed via online or telephone surveys (Sabat et al., 2020; Cowling et al., 2020). The main drawback of this approach is that it does not continuously capture the sentiment of the population. The time lag between the implementation of NPIs and the survey must also be mentioned as an additional disadvantage of this method. In other work, observational data like the daily number of Google searches for COVID-19 related terms (Alfaro et al., 2020) or social media trends (Doogan et al., 2020; Boon-Itt and Skunkan, 2020; Kruspe et al., 2020a; Porcher and Renault, 2021; Jin et al., 2021) are used to study the impact of voluntary risk behavior on mobility decisions on the country level. Although their results show statistically significant correlations, they struggle to identify the disentangled causal effects resulting from implemented NPIs or from individual risk awareness. Allcott et al. (2020) studies in his work the partisan differences in Americans' response to the COVID-19 pandemic. They show that areas with more Republicans results in less social distancing compared to areas with tend to be more Democratic.

Social media sentiment analysis Social media acts as an important source for knowledge acquisition for a large number of citizens and thus influences how they perceive and cope with the COVID-19 pandemic (Cuello-Garcia et al., 2020; Abd-Alrazaq et al., 2020). Research on sentiment of Tweets like in Abd-Alrazaq et al. (2020) and Lwin et al. (2020) show that the global sentiment concerning COVID-19 related topics are overwhelmingly negative over the first half of 2020. They also revealed that over the first five months the dominantly emotion about this pandemic is anger and fear. Abd-Alrazaq et al. (2020) uses Latent Dirichlet Allocation (LDA) to model the sentiment of semantically related topics in tweets. Topic-specific analysis could be a promising direction for future mediation analysis.

Causality in NLP Work on causality in the field of NLP can be split into two groups: (i) identify causal relationships within the text (e.g. (Zhao et al., 2017; Kyriakakis et al., 2019; Yu et al., 2019; Chen et al., 2020)), and (ii) utilizing causal methods for NLP tasks. We are concerned with the second group of works. Especially on the topic of emotion and sentiment detection, there have been recently a number of works. In Sridhar and Getoor (2019b) the authors also include the linguistic style and consider unobserved confounders, which were interpreted, e.g., as ideology. An overview of approaches on the role of confounders is given in the recent survey by Keith et al. (2020), where the authors focuses on social media. They report on a number of works, categorised by the treatment, outcome, confounder, domain, and the followed causal inference approach.

3. Background

Causal models The structural causal model, as described by Pearl and others (2000), mathematically expresses the causal mechanism of a system. In our work, we aim to identify factors that play a causal role in determining the direct effect between exposure T (a country's Stringency Index) and the output Y (mobility behavior of the citizens) as well as the mediated effect through M, which represents citizens' sentiment. Causal identification is given by controlling for the confounding vector Context as shown in Figure 1. Given this identification, causal effects can be obtained by statistical estimates. We review the estimation of (i) the total causal effect (TE), (ii) the natural direct effect (NDE) and (iii) the natural indirect effect (NIE) by introducing Pearl's do notation and the counterfactual mediation framework from Imai et al. (2010). The latter builds on the specification of two statistical models, the mediator regression model $f_M(T, C)$ and the outcome regression model $f_Y(T, M, C)$. The mediation model represents the conditional distribution of the mediator M given the treatment T and a set of confounders C. The outcome model defines the conditional distribution of the outcome Y given T, M, C.

Total effect The TE estimates the mobility behavior Y when the Stringency Index T changes from $T = 0$ to $T = 1$ while the sentiment, which represents the mediator, is allowed to follow the change in T. We can ask about the overall causal effect as follows: *"What would be the effect on mobility if we increase or decrease the Stringency Index of a country?"*

$$\begin{aligned} \text{TE} &= \mathbb{E}\left[f_Y(1, f_M(1, C), C) - f_Y(0, f_M(0, C), C)\right] \\ &= \mathbb{E}\left[Y \middle| \operatorname{do}(T = 1)\right] - \mathbb{E}\left[Y \middle| \operatorname{do}(T = 0)\right] \end{aligned} \tag{1}$$

Natural direct effect The NDE estimates the mobility behavior Y as the Stringency Index changes from $T = 0$ to $T = 1$ while setting the sentiment variable to the value which would have been obtained under the Stringency Index $T = 0$. This corresponds to the question: *"Among the actual sentiment of the population, would stronger/weaker government restrictions change mobility?"*

$$\begin{aligned} \text{NDE} &= \mathbb{E}\left[f_Y(1, f_M(0, C), C) - f_Y(0, f_M(0, C), C)\right] \\ &= \mathbb{E}\left[Y_{1,M_0} - Y_{0,M_0}\right] \end{aligned} \tag{2}$$

Natural indirect effect The NIE estimates the mobility behavior Y when the Stringency Index is held constant at $T = 0$ while the sentiment changes to the value which would have been obtained under the Stringency Index $T = 1$. In our context we can raise the question: *"How would mobility change if the public sentiment had instead been more positive or more negative while keeping everything else (i.e. the Stringency Index) the same?"*

$$\begin{aligned} \text{NIE} &= \mathbb{E}\left[f_Y(0, f_M(1, C), C) - f_Y(0, f_M(0, C), C)\right] \\ &= \mathbb{E}\left[Y_{0,M_1} - Y_{0,M_0}\right] \end{aligned} \tag{3}$$

The regression specifications for $f_Y(T, M, C)$ and $f_M(T, C)$ are stated in Section 4.5.

ADE & AME In addition, following Imai et al. (2010), the average direct effect (ADE) and the average mediated effect (AME), the main effect estimates in our analyses, can be computed based on NDE and NIE, respectively. So far we considered the case of a binary exposure. For our setting, this approach must be generalized to a continuous treatment value T which can take values in $[0, 1]$. The causal effects can be estimated for different treatment levels. To this end, a baseline treatment level $t_0 = 0$ is chosen and the causal estimates, i.e., the mediated effects defined as $\delta(t_0, t)$, are obtained for all different treatment levels t with respect to this baseline. The resulting AME is the summation of these different effects averaged over the distribution of the observed treatment levels F_T,

$$\text{AME} = \int \delta(0, t) dF_T(t). \tag{4}$$

The AME is denoted significant if the effect estimates $\delta(0, t)$ for all different treatment levels are significant within a 95% confidence interval. This approach also holds for the TE and the ADE, respectively.

4. Materials & Methods

Since the focus of this work relies on the estimation of an unbiased effect of implemented country wide policies on mobility behavior as well as the effect mediated through public perception, we need to control for potential confounders.

In this section, we describe how we standardize and aggregate NPIs, public opinion from Tweets and mobility behavior at the country level. We also introduce key confounders and discuss how they are collected and processed. Since Twitter is particularly popular in the United Kingdom (UK), we focus our empirical analysis on the UK and the countries of the UK. In addition, the analysis is also carried out with Swedish data. Sweden is a great exception for a country where less restrictive interventions are made and Tweets in English are quite common.

4.1. Treatment: Policy

Policy represents the *treatment* variable in our causal analysis model. Recent research has shown that non-pharmaceutical interventions have a major impact on mobility behaviors, leading to a reduction in COVID-19 virus transmission (Haug et al., 2020; Badr et al., 2020). This results in a direct causal relationship between interventions and mobility, as shown in Figure 1. To make policies, implemented by governments, comparable between countries, the University of Oxford developed the COVID-19 Government Response Tracker (Hale et al., 2021). In this work, they capture policies related to closure and containment, health and economic on countries' national and subnational jurisdictions. The tracker combines these indicators and provides an overall measure of the intensity of government response, called Government Response Stringency Index which takes values in $[0, 1]$. Figure 2 visualises the UK country specific Stringency Indices. The United Kingdom adopted a fairly restrictive COVID-19 policy in 2020, with two national lockdowns beginning in late March and early November.

4.2. Mediator: Public opinion on Twitter

Since we intend to analyze public opinion about the current COVID-19 situation and the impact on mobility behavior in the UK, we focus on Tweets discussing recent news about COVID-19 from January 2020 to December 2020. Public opinion acts as as *mediator* in our causal setting. We want to find out (i) whether sentiment is a suitable mediator between the Stringency Index and mobility and (ii) how robust the mediation effect is by conducting multiple causal experiments. Therefore, the sentiment mediator and the causal relationship in Figure 1 must be verified in the analysis. Abd-Alrazaq et al. (2020) and Gupta et al. (2021) collected COVID-19 related Tweet IDs and performed

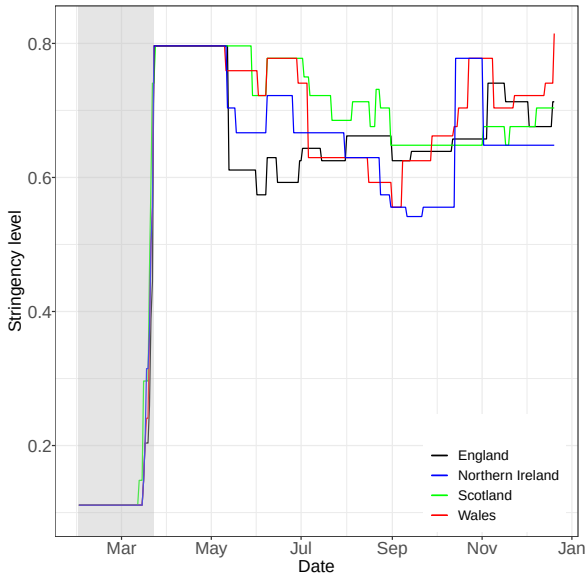


Figure 2: Daily Stringency Index per country of the UK during the year 2020. The grey shaded area visualises the time before the announcement of the first lockdown.

sentiment analysis and LDA topic modeling [1]. In addition, only Tweets from the UK and Sweden on the first two topics (Table 7 in the Appendix) have been considered when hydrating the actual Tweet text. In Table 1, we have a closer look on keywords used in these Twitter messages. In addition, keywords can be grouped into several themes within the previous mentioned topics. It can be seen that most of the posts are about statements on epidemiological measurements such as cases and deaths. In addition, Twitter users are concerned about their current work situation, the introduction of a strict lockdown, but also discuss COVID-19 testing and vaccines. To classify the opinion of Tweets, we use sentiment analysis from Gupta et al. (2021). Table 2 lists some examples of these analysed Tweets.
It can be shown that a more positive sentiment represents a more agreed opinion or acceptance of the implemented restriction orders. But also a more positive interpretation of the way the different institutions (government, NHS) deal with the current COVID-19 situation. For further causal analysis, the daily aggregated sentiment of a country has to be calculated. All necessary processing steps and visualizations can be found in the Appendix A.1.

4.3. Outcome: Mobility behavior

Mobility is defined as the *outcome* of the causal model, which is partly influenced by the policies of the country and the public's sentiment as shown in Figure 1. The Google Mobility Reports (Google Mobility, 2020) capture movement trends over different categories relative to a predefined baseline. The baseline is the median value of the same day of the week from January 3 to February 6 in 2020. The categories *workplaces*, *grocery and pharmacy*, *transit stations* and *retail and recreation* represent general mobility trends for certain activities. An increase in the categories *residential* or *parks* is indicative of decreased mobility as they suggest increased activity around the home environment. Therefore it can be understood as an indicator for the compliance of stay at home orders. In our analysis we neglect the category park because no reasonable causal direct and mediated effect is assumed. For further processing steps and visualization, see Appendix A.2

[1] https://www.openicpsr.org/openicpsr/project/120321/version/V11/view

Theme	Keywords
NPIs	lockdown, pandemic, restriction, mask, social distancing
Epidem	case, death, number, died, spread, today, week, outbreak, infection, year, time, risk, month
Medical	test, hospital, vaccine, patient, care, icu, positive, negativ, symptom, staff, disease, testing, mental health, infection
General	uk, work, nhs, nh, government, family, situation, job, school, england, information, help, crisis, support, life, children, schotland, brexit, boris johnson

Table 1: Top keywords extracted from relevant Tweets and split by theme.

4.4. Key confounding factors

As indicated in Section 4.2, Twitter users tend to tweet about epidemiological data such as cases and deaths. Because epidemiological measurements also form the basis for the decision-making process to introduce NPIs, this type of data seems to be an relevant confounder in our causal setting (Goolsbee and Syverson, 2021; Chernozhukov et al., 2021). This previous work emphasizes the causal relationship of the confounding vector, as visualized in Figure 1.

Epidemiological data We gathered reported COVID-19 cases and deaths from the COVID-19 Data Repository maintained by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) (Dong et al., 2020). We focus on running weekly cases and deaths by summing up daily measurements (from day t to $t-6$). This is done because daily measurements are strongly affected by the time of reporting and testing. Since policy makers adjust restrictions also on growth rates, we additionally added log weekly case and death growth rates as confounding factors as in the work of

Tweet	Sentiment
Happy Wednesday people. Keep smiling, keep positive, things WILL get better. #Covid_19#WearAMask #WednesdayMotivation #positivity#Lockdown2 #KeepSafe	positive
Northern England: Restrictions reimposed as COVID-19 cases surge	neutral
The NHS is a disgrace. Why on earth anybody would wish to applaud them for placing millions at risk, based on the fakery that is Covid, is beyond belief.	negative

Table 2: Examples of positive, neutral and negative Tweets

Chernozhukov et al. (2021).

$$\Delta\, CW_t^i = \log\left(\mathrm{CW}_t^i\right) - \log\left(\mathrm{CW}_{t-7}^i\right) \tag{5}$$

Were $\Delta\, CW_t^i$ denotes the log weekly case growth rate at day t for country i and CW_t^i represents the number of new confirmed cases in the past seven days for a given day t and country i. The same calculation is applied for the log weekly death growth rate.

Time varying confounders Of course, there are a lot of additional unobserved time varying confounding factors (political and socioeconomic factors, pandemic fatigue etc.) that need to be taken into account but can not be measured directly. We hypothesize that these additional, unobserved confounding factors are constant within a one week interval. Therefore, we stratify our data by this one week interval and assume a fixed effect during this period. Moreover, by introducing a weekly indicator variable, the residual autocorrelation in the variable of interest tends to be much smaller for the stratified groups than for the whole data set (Goolsbee and Syverson, 2021; Bhaskaran et al., 2013; Dundar et al., 2007).

4.5. Mediation analysis

Following previous work on policy effectiveness estimation like the ones from Lucchini et al. (2021; Haug et al. (2020; Alfaro et al. (2020) or Allcott et al. (2020), we conduct two different regression specifications (i) ordinary least squares (OLS) and (ii) mixed-effects. According to the process of mediation analysis, as introduced in Section 3, two models are defined. One represents the *mediator regression model* $f_M(T, C)$ which infers the sentiment. The other defines the *outcome regression model* $f_Y^y(T, M, C)$ which derives the change for mobility category y. The causal effects (TE, ADE, AME) are estimated independently for each of the five mobility categories. We additionally assume no time lags between treatment, mediator, output and confounders. We assume that population's response on social media is instant to the announcement of the government as well as to actual epidemiological information within one week. Behavioral adaption of individuals is also assumed to be instant within a week to COVID-19 related information and restrictions. All exogenous and endogenous variables are summarised in Table 3.

Type	Variable
Treatment	Stringency Index
Mediator	Average Sentiment
Output	Mobility trend of a certain category
Confounder	Unobserved week fixed effect
	Weekly running cases
	Weekly running deaths
	Log weekly case growth rate
	Log weekly death growth rate

Table 3: Exogenous and endogenous regression variables

Ordinary Least Square analysis We examine the effects by running a standard OLS regression model for the mediator and the output. The direct, mediated and total effects are estimated on (i) aggregated measurements from the UK and (ii) Sweden.

$$\begin{aligned} M_t &= f_M(T, C) \\ &= \alpha_0 + \alpha_1 T_t + \alpha_2 C_t + \epsilon_{t1} \\ Y_t^y &= f_Y^y(T, M, C) \\ &= \beta_0^y + \beta_1^y T_t + \beta_2^y M_t + \beta_3^y C_t + \epsilon_{t2}^y \end{aligned} \tag{6}$$

M_t represents the mediator variable and measures the output of the public's sentiment at time t. T_t indicates the stringency of the government's response. The vector C_t includes all epidemiological and time varying confounding factors mentioned in Table 3. Y_t^y represents the output variable and stands for the change in movement against a baseline for a mobility category y at time t. ϵ_{t1} and ϵ_{t2}^y are error terms associated with the mediator M_t and output Y_t^y, respectively. It is assumed that $\epsilon_{t1}, \epsilon_{t2}^y \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

Weighted mixed-effects analysis We additionally look at UK country level data to study the impact of the stringency of NPIs on sentiment and mobility. To analyse the mediation effect, a panel data mixed-effects model is used.

$$
\begin{aligned}
M_{ct} &= f_M(T, C) \\
&= \alpha_0 + \alpha_1 T_{ct} + \alpha_2 C_{ct} + \text{Country}_c + \epsilon_{ct1} \\
Y^y_{ct} &= f_Y(T, M, C) \\
&= \beta^y_0 + \beta^y_1 T_{ct} + \beta^y_2 M_{ct} + \beta^y_3 C_{ct} + \text{Country}^y_c + \beta^y_{ct}
\end{aligned} \tag{7}
$$

Here the mediator variable M_{ct} reflects the sentiment for day t in country c. The response variable Y^y_{ct} represents the change in movement in country c at day t for the mobility category y. Compared to the OLS regression approach, we additionally control for country random effects by introducing the intercept variable Country_c. More specifically, Country_c controls for unobserved time-invariant country characteristics (e.g. population density, political preferences but also Twitter and social media preferences). The confounding vector $C_{c,t}$ thus includes not only UK-wide epidemiological data but also country level measures (Yan et al., 2020). The error terms are again assumed to be $\epsilon_{c,t}, \epsilon^i_{c,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$. To account for the fact that population sizes vary greatly between countries, and thus country measures contribute differently to the overall UK statistics, each daily observation is weighted by the probability of being measured in a particular country. For example, England has a larger population than Northern Ireland and therefore observations from England are assigned greater weight.

5. Empirical Analysis

In this section, we describe our experiments. We compare the outputs in terms of the reliability and robustness of the causal effect estimates.

5.1. United Kingdom OLS analysis

For the OLS mediation approach, we examine the causal direct, indirect and total effect on aggregated UK level. This experiment utilizes the OLS regression specifications as stated in Equation 6. The analysis starts with the report of the first positive COVID-19 case in the UK. Table 4 reports the obtained results for the different mobility categories. The effect estimates depict the change of mobility in percent compared to a baseline as stated in the description of the Google Mobility Reports in Chapter 4.3. As indicated in Equation 4, the presented causal effects are the averaged estimates across the observed stringency levels. We find a strong alignment between the stringency of government restrictions and all mobility categories. As expected, more stringent NPIs lead to a decrease in mobility categories and an increase in the residential category. The same applies for all total effects. One cause of this behavior is the recommended stay at home order. When examining the mediated effect, it can be stated that for all mobility categories, except for workplaces, a reasonable effect between sentiment and mobility can be figured out.

When public opinion is more positive about the current restrictions and how the COVID-19 pandemic is handled, people tend to follow these restrictions and recommendations from the government and NHS, and therefore reduce their mobility. This is especially true for the categories *retail and recreation* and *transit stations*. As a result, people are more likely to choose to stay at home, which is reflected in the increase in mobility of the *residential* category. An exception yields for every day needs where a more positive sentiment is associated with an increase in visits to grocery stores or pharmacies.

5.2. United Kingdom mixed-effects analysis

In this experiment a multilevel mediation analysis is performed where the treatment, mediator and outcome are measured on country level. Covariates and confounders are extended with country specific epidemiological measurements. This experiment utilizes the mixed-effects regression specifications as stated in Equation 7. Table 5 presents the average causal effects for this mixed-effects examination. The absolute values of the ADE and AME for all mobility categories, with the exception of *grocery and pharmacy*, decreased slightly compared to the OLS approach. Figure 3 visualizes the effects on all mobility categories for a control level of 0 and a treatment level of 0.8.

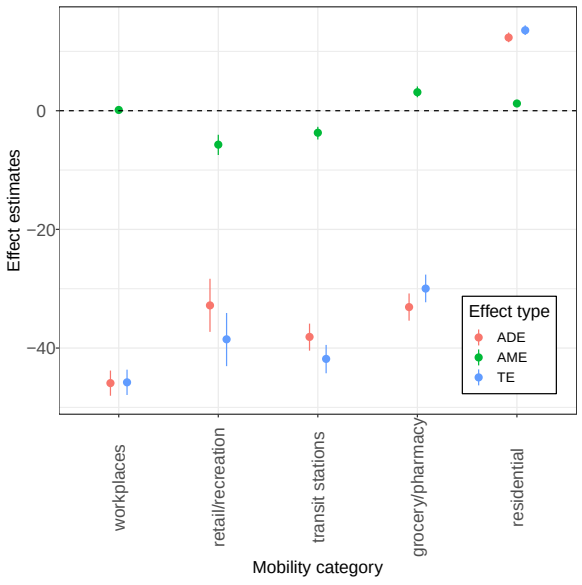


Figure 3: Causal effect estimates on all mobility categories against a treatment level (stringency index) of 0.8 in the UK. Error bars indicate the 95% confidence interval. With the exception of workplace mobility, social media sentiment explains a change in behavior.

Mediator permutation test To understand the contribution of sentiment as a mediator and to test the robustness of the analysis, an ablation study is applied. In this experiment, we randomly shuffle a predefined rate of the mediator and run the mediation analysis on this permuted dataset. It can be stated as the permutation rate increases, the mediated effect tends to go towards

	Workplaces	Retail and Recreation	Transit stations	Grocery and pharmacy	Residential
ADE	-46.0***	-47.4***	-44.3***	-24.0***	15.0***
AME	-1.1	-10.0***	-6.0***	3.2***	2.3***
TE	-47.1***	-57.4***	-50.3***	-20.8***	17.3***
Mediator regression model $f_M(T, C)$ $\bar{R}^2$	0.71	0.71	0.71	0.71	0.71
Outcome regression model $f_Y(T, M, C)$ $\bar{R}^2$	0.95	0.90	0.96	0.83	0.95

Table 4: Estimated average direct, average mediated and total effect using the OLS regression specification for different mobility categories in the United Kingdom. The effect estimates are the percent change in mobility to a predefined mobility baseline between January 3 and February 6 in 2020. ***p<0.01

	Workplaces	Retail and Recreation	Transit stations	Grocery and pharmacy	Residential
ADE	-37.0***	-26.5***	-30.7***	-26.7***	10.0***
AME	0.2	-4.7***	-3.0***	2.5***	1.0***
TE	-36.8***	-31.2***	-33.7***	-24.2***	11.0***
Mediator regression model $f_M(T, C)$ $\bar{R}^2$	0.74	0.74	0.74	0.74	0.74
Outcome regression model $f_Y(T, M, C)$ $\bar{R}^2$	0.96	0.94	0.95	0.92	0.96

Table 5: Estimated average direct, average mediated and total effect using a mixed-effects specification for different mobility categories in the United Kingdom. The effect estimates are the percent change in mobility to a predefined mobility baseline between January 3 and February 6 in 2020. ***p<0.01

zero as illustrated in Figure 6 in the Appendix. Still the effect remains statistically significant for a permutation rate lower 60% within the 95% confidence interval.

5.3. Sweden OLS analysis

We also conduct an OLS regression analysis for Sweden, where the government largely relied on voluntary risk awareness since the beginning of the pandemic. Sweden was also the only country where a lockdown was avoided. Thus the stringency of implemented NPIs is totally different compared to the UK or to other countries in Europe. The effect estimates indicate that the strategy of voluntary decision-making has a direct effect on mobility behaviour, but no significant mediated effect via the sentiment of the population. For the interested reader the results are presented in Table 8 in the Appendix. For a more detailed discussion of these results, see Section 6.

6. Discussion

Q1: Are Tweets a valid measure of public opinion about the COVID-19 pandemic? In our environment, we observe social media text as a proxy for the public opinion of the population. The experiments show that sentiment analysis and topic modeling are an irreplaceable source for the analysis of citizens' public opinion. Moreover, this valuable information can be integrated into causal analysis in epidemiology, economics or social science. The main advantage of this method over surveys is that a larger number of citizens can be analysed continuously over time. To overcome the issue of selection bias and unbalanced age distribution of Twitter users, it might be advisable to additionally include data sources that reflect the sentiment of a broader population.

Q2: What influence does the quality of sentiment have on causal analysis? From the ablation study, where the mediator is partly perturbed, we see that the mediated effect remains significant for a permutation rate lower than 60%. It can be shown that sentiment has (i) a crucial impact on the effect estimates and (ii) still provides robust estimates at this level of data aggregation, even when partially perturbed. This seems to be very important due to the complexity of sentiment analysis of social media texts (Kharde et al., 2016). Therefore, the accuracy of sentiment analysis is not as critical for social media texts, and causal analysis is still robust even when sentiment analysis is not as accurate.

Q3: Are the model assumptions valid? Although we made strong assumptions for the model specifications, the results are quite promising. The permutation test and the R^2 values, stated in Table 4 and Table 5, support the choice of the multivariate linear regression models. In addition, recent literature on COVID-19 policy estimation and verification mainly uses OLS and mixed-effects models (Haug et al., 2020; Alfaro et al., 2020; Allcott et al., 2020; Jin et al., 2021; Lucchini et al., 2021).

Q4: What influence does public opinion have on behavior in the UK compared to Sweden? In the UK, the public opinion on the COVID-19 pandemic seems to have a significant influence on the mobility behavior. The results show that as people become more comfortable with the restrictions and the situation of how the pandemic is being managed, they begin to limit their mobility for non-essential places and spend more time at home. In contrast, mobility for essential areas such as grocery stores and pharmacies increases under these conditions. Studies as the ones from (Alfaro et al., 2020), (Porcher and Renault, 2021) and (Goolsbee and Syverson, 2021) verify the evidence of this effect in the US. It can be shown that analyses at different data aggregation levels (UK wide vs UK country specific) exhibit the same statistical significance for the causal effects. Nevertheless, the trade-off between aggregation level and the availability of Twitter data at this level has to be considered.

In Sweden, most of the implemented NPIs are recommendations rather than actual restrictions, especially during the first half of 2020 (Ludvigsson, 2020). The effect estimates suggest that despite the less stringent governmental decisions, we obtain a significant mobility reduction for non-essential locations and transit stations as stated in Table 8 and visualised in Figure 8. In contrast to the UK, no significant mediated effect can be observed. It must be stated that, compared to the UK, sentiment does not fluctuate as much over time. The Stringency Index also remains constant over time, reaching a value of around 0.6 at the beginning of April. We can state that public opinion has no clearly visible influence on behavior. Swedish citizens seem to adjust their behavior regardless of their opinion on the regulations and the current COVID-19 situation. The work of Irwin (2020) and AB (2020) support our findings. They found that Swedish citizens have a high level of trust in their policy makers, health authorities, and government epidemiologists. This is especially true for the first COVID-19 wave in 2020.

Q5: What could be follow-up analysis? In a future work, the linearity assumption can be extended to a more general modelling approach to verify the estimated effects resolved in this paper. This is accompanied by gathering and examining a more heterogeneous data set of additional countries. It will also be possible to examine this mediation study from a time series perspective. In this case, it might be interesting to discover causal effects also of lagged versions of the treatment and confounders on the mobility categories.

7. Conclusions

In this paper, we explore how social media texts can be used beyond correlation analysis, and how they can be understood as indicators for causal mechanisms in our society. We propose a methodology, based on recent methods from economics and political science, that uses a sentiment analysed COVID-19 dataset in combination with heterogeneous data sources to causally investigate a real-world phenomenon. We try to decouple the effect of government-imposed NPIs from the impact resulting from public opinion on COVID-19-related issues. This requires (i) to understand and analyse the epidemiological problem setting (ii) to be familiar with the quality of NLP methods which could highly affect the reliability and credibility of the study. For the latter case, an ablation study demonstrates the robustness of the mediator in this causal setting. The empirical analysis reveals that not only the implementation of NPIs causes changes in mobility but also a positive public opinion about the implemented protection measures, leads to a decrease in mobility. Our counterfactual mediation examination exhibits promising results for the causal effect estimates and remains robust under different regression specifications for the output and mediator model, respectively. Our work clearly demonstrates that beliefs and emotions about specific topics shared on social media provide insight into peoples' behavior and can therefore serve as a valuable resource for policymakers to better understand social events such as the COVID-19 pandemic.

8. Acknowledgements

The Know-Center is funded within the Austrian COMET Program—Competence Centers for Excellent Technologies under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

9. Bibliographical References

AB, N. G. I. (2020). Corona-status 20200401. https://novus.se/coronastatus-0401/ (Accessed: 2021-01-26).

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., and Shah, Z. (2020). Top concerns of tweeters during the covid-19 pandemic: infoveillance study. *Journal of medical Internet research*, 22(4):e19016.

Abouk, R. and Heydari, B. (2021). The immediate effect of covid-19 policies on social-distancing behavior in the united states. *Public Health Reports*, 136(2):245–252.

Al-Garadi, M. A., Khan, M. S., Varathan, K. D., Mujtaba, G., and Al-Kabsi, A. M. (2016). Using online social networks to track a pandemic: A systematic review. *Journal of biomedical informatics*, 62:1–11.

Alfaro, L., Faia, E., Lamersdorf, N., and Saidi, F. (2020). Social interactions in pandemics: fear, altruism, and reciprocity. Technical report, National Bureau of Economic Research.

Algaba, A., Ardia, D., Bluteau, K., Borms, S., and Boudt, K. (2020). Econometrics meets sentiment: An overview of methodology and applications. *Journal of Economic Surveys*, 34(3):512–547.

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., and Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, 191:104254.

Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., and Gardner, L. M. (2020). Association between mobility patterns and covid-19 transmission in the usa: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(11):1247–1254.

Baier, L., Kühl, N., Schöffer, J., and Satzger, G. (2020). Utilizing concept drift for measuring the effectiveness of policy interventions: The case of the covid-19 pandemic. *arXiv preprint arXiv:2012.03728*.

Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L., and Armstrong, B. (2013). Time series regression studies in environmental epidemiology. *International journal of epidemiology*, 42(4):1187–1195.

Boon-Itt, S. and Skunkan, Y. (2020). Public perception of the covid-19 pandemic on twitter: sentiment analysis and topic modeling study. *JMIR Public Health and Surveillance*, 6(4):e21978.

Brauner, J. M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., Stephenson, A. B., Leech, G., Altman, G., Mikulik, V., et al. (2021). Inferring the effectiveness of government interventions against covid-19. *Science*, 371(6531).

Chen, X., Li, Q., and Wang, J. (2020). Conditional causal relationships between emotions and causes in texts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3111–3121.

Chernozhukov, V., Kasahara, H., and Schrimpf, P. (2021). Causal impact of masks, policies, behavior on early covid-19 pandemic in the us. *Journal of econometrics*, 220(1):23–62.

Cowling, B. J., Ali, S. T., Ng, T. W., Tsang, T. K., Li, J. C., Fong, M. W., Liao, Q., Kwan, M. Y., Lee, S. L., Chiu, S. S., et al. (2020). Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in hong kong: an observational study. *The Lancet Public Health*, 5(5):e279–e288.

Cuello-Garcia, C., Pérez-Gaxiola, G., and van Amelsvoort, L. (2020). Social media can have an impact on how we manage and investigate the covid-19 pandemic. *Journal of clinical epidemiology*, 127:198–201.

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534.

Doogan, C., Buntine, W., Linger, H., and Brunt, S. (2020). Public perceptions and attitudes toward covid-19 nonpharmaceutical interventions across six countries: A topic modeling analysis of twitter data. *Journal of medical Internet research*, 22(9):e21419.

Dundar, M., Krishnapuram, B., Bi, J., and Rao, R. B. (2007). Learning classifiers when the training data is not iid. In *IJCAI*, volume 2007, pages 756–61.

Farboodi, M., Jarosch, G., and Shimer, R. (2020). Internal and external effects of social distancing in a pandemic. Technical report, National Bureau of Economic Research.

Google Mobility. (2020). Google COVID-19 Community Mobility Report. `https://www.google.com/covid19/mobility`. (Accessed: 2021-02-01).

Goolsbee, A. and Syverson, C. (2021). Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020. *Journal of public economics*, 193:104311.

Gupta, R. K., Vishwanath, A., and Yang, Y. (2021). Global reactions to covid-19 on twitter: A labelled dataset with latent topic, sentiment and emotion attributes.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., et al. (2021). A global panel database of pandemic policies (oxford covid-19 government response tracker). *Nature Human Behaviour*, pages 1–10.

Haug, N., Geyrhofer, L., Londei, A., Dervic, E., Desvars-Larrive, A., Loreto, V., Pinior, B., Thurner, S., and Klimek, P. (2020). Ranking the effectiveness of worldwide covid-19 government interventions. *Nature human behaviour*, pages 1–10.

Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, 15(4):309.

Irwin, R. E. (2020). Misinformation and decontextualization: international media reporting on sweden and covid-19. *Globalization and health*, 16(1):1–12.

Jin, Z., Peng, Z., Vaidhya, T., Schoelkopf, B., and Mihalcea, R. (2021). Mining the cause of political decision-making from social media: A case study of covid-19 policies across the us states. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 288–301.

Keith, K., Jensen, D., and O'Connor, B. (2020). Text and causal inference: A review of using text to remove confounding from causal estimates. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5332–5344.

Kharde, V., Sonawane, P., et al. (2016). Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*.

Kruspe, A., Häberle, M., Kuhn, I., and Zhu, X. X. (2020a). Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, Online, July. Association for Computational Linguistics.

Kruspe, A., Häberle, M., Kuhn, I., and Zhu, X. X. (2020b). Cross-language sentiment analysis of european twitter messages duringthe covid-19 pandemic. *arXiv preprint arXiv:2008.12172*.

Kyriakakis, M., Androutsopoulos, I., Saudabayev, A., and i Ametllé, J. G. (2019). Transfer learning for causal sentence detection. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 292–297.

Lucchini, L., Centellegher, S., Pappalardo, L., Gallotti, R., Privitera, F., Lepri, B., and De Nadai, M. (2021). Living in a pandemic: changes in mobility routines, social activity and adherence to covid-19 protective measures. *Scientific Reports*, 11(1):1–12.

Ludvigsson, J. F. (2020). The first eight months of sweden's covid-19 strategy and the key actions and actors that were involved. *Acta Paediatrica*, 109(12):2459–2471.

Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., and Yang, Y. (2020). Global sentiments surrounding the covid-19 pandemic on twitter: analysis of twitter trends. *JMIR public health and surveillance*, 6(2):e19447.

Pearl, J. et al. (2000). Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19.

Porcher, S. and Renault, T. (2021). Social distancing beliefs and human mobility: Evidence from twitter. *PloS one*, 16(3):e0246949.

Sabat, I., Neuman-Böhme, S., Varghese, N. E., Barros, P. P., Brouwer, W., van Exel, J., Schreyögg, J., and Stargardt, T. (2020). United but divided: Policy responses and people's perceptions in the eu during the covid-19 outbreak. *Health Policy*, 124(9):909–918.

Sridhar, D. and Getoor, L. (2019a). Estimating causal effects of tone in online debates. *arXiv preprint arXiv:1906.04177*.

Sridhar, D. and Getoor, L. (2019b). Estimating Causal Effects of Tone in Online Debates. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 1872–1878. International Joint Conferences on Artificial Intelligence Organization.

Yan, Y., Malik, A. A., Bayham, J., Fenichel, E. P., Couzens, C., and Omer, S. B. (2020). Measuring voluntary and policy-induced social distancing behavior during the covid-19 pandemic. *medRxiv*.

Yang, M.-J., Looney, A., Gaulin, M., and Seegert, N. (2020). What drives the effectiveness of social distancing in combatting covid-19 across us states? Technical report, Working Paper, University of Utah.

Yu, B., Li, Y., and Wang, J. (2019). Detecting causal language use in science findings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4656–4666.

Zhang, R., Li, Y., Zhang, A. L., Wang, Y., and Molina, M. J. (2020). Identifying airborne transmission as the dominant route for the spread of covid-19. *Proceedings of the National Academy of Sciences*, 117(26):14857–14863.

Zhao, S., Wang, Q., Massung, S., Qin, B., Liu, T., Wang, B., and Zhai, C. X. (2017). Constructing and embedding abstract event causality networks from text snippets. *WSDM 2017 - Proceedings of the 10th ACM International Conference on Web Search and Data Mining*, pages 335–344.

Zhuravskaya, E., Petrova, M., and Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, 12:415–438.

A. Appendix

A.1. Twitter sentiment processing

The sentiment for day t is defined as:

$$S_t = \frac{P_t - N_t}{P_t + N_t + O_t} \tag{8}$$

P_t, N_t and O_t reflect the number of positive, negative and neutral Tweets at day t. To reduce the noise of the measurements and remove random fluctuations, a moving average filter with a window size of 7 is applied. Figure 4 shows the sentiment during the first and second COVID-19 wave in the UK. The grey shaded area represents the time before the first official lockdown was implemented. It can be shown that the sentiment strongly increased by the first announcement and implementation of more stringent NPIs. (Kruspe et al., 2020b) also confirms this increase of positive emotions in the UK at the beginning of the first lockdown. Table 6 summarises the number of Tweets per country in the UK and Sweden. It should be mentioned that the overall number of Tweets in the UK is slightly higher than the sum of the individual countries because not all Tweets could be mapped to a specific country.

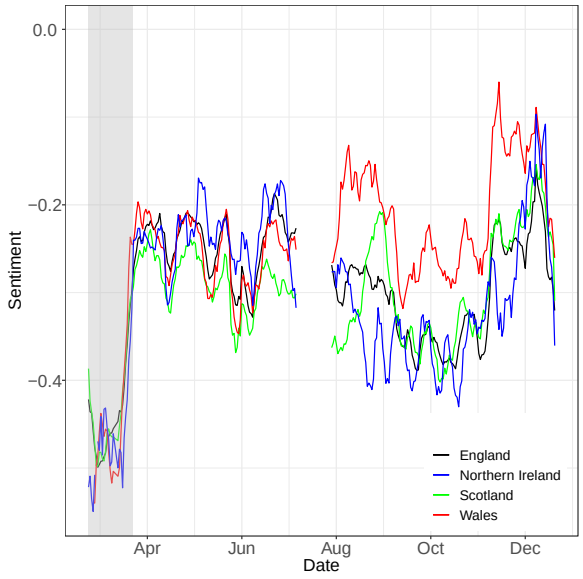


Figure 4: Average sentiment per country of the UK during the year 2020. The grey shaded area visualises the time before the announcement of the first lockdown.

Country	Number of Tweets
England	3,185,232
Scotland	434,347
Wales	140,791
Northern Ireland	88,091
United Kingdom	4,747,370
Sweden	50,643

Table 6: Number of Tweets per country in our dataset

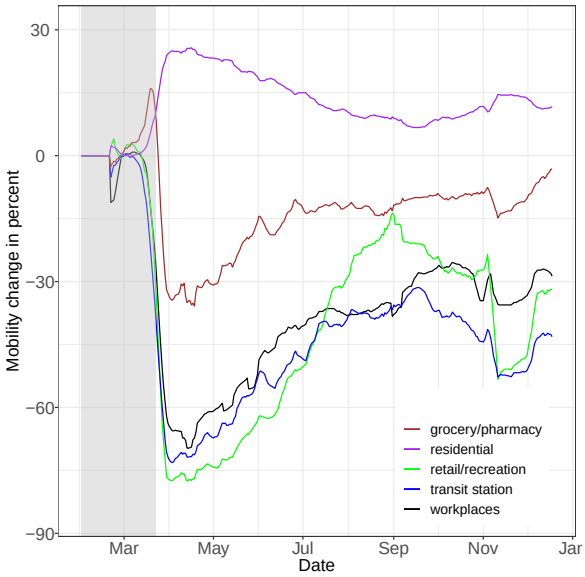


Figure 5: Evolution of the different mobility categories on UK aggregation level during the year 2020. The grey shaded area visualises the time before the announcement of the first lockdown.

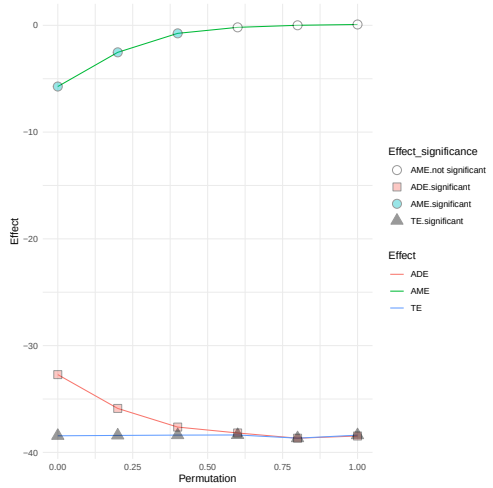
A.2. Mobility processing

The daily mobility measurements are smoothed by taking a 7 day moving average (from t to $t - 6$). In this way, we smooth idiosyncratic daily fluctuations as well as periodic fluctuations caused by days of the week.

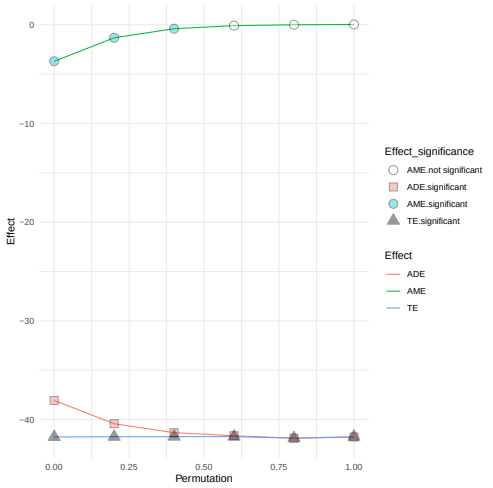
Topic	Terms
t1	people, cases, new, deaths, time, china, realdonaldtrump, lockdown, trump
t2	health, help, people, need, think, vaccine, care, fight, support
t3	pandemic, f**k, months, killed, question, wait, looks, trump, impact
t4	pay, donate, lie, focus, song, gates, page, google, caused
t5	florida, drink, named, nature, marketing, pr, ncdcgov, farmers, cr
t6	rules, bed, drtedros, speaks, privacy, parliament, physicians, strength, joke
t7	dies, pmoindia, ndtv, ai, narendramodi, mohfwindia, shoot, drharshvardhan, battle
t8	ye, ke, behaviour, brought, hidden, yup, smell, zerohedge, odds
t9	excuse, humanity, salary, wind, gtgt, rats, ice, beard, mosque
t10	internet, allah, teacher, dance, el, rona, weed, crush, fk

Table 7: Top 10 topic cluster generated from the CrystalFeel Twitter dataset (Gupta et al., 2021)

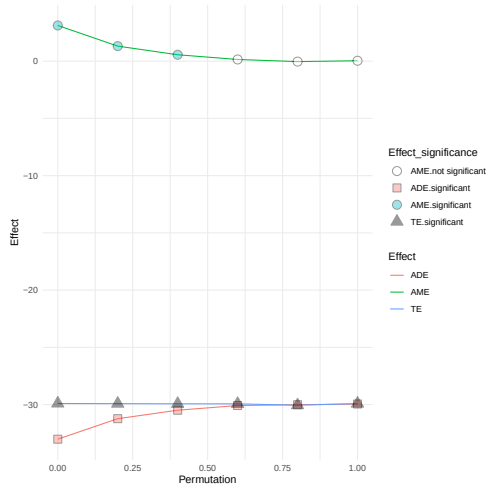
A.3. Ablation Study



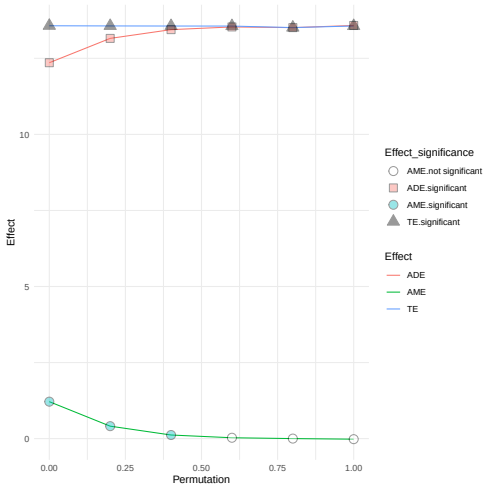
(a) Mobility category: retail and recreation



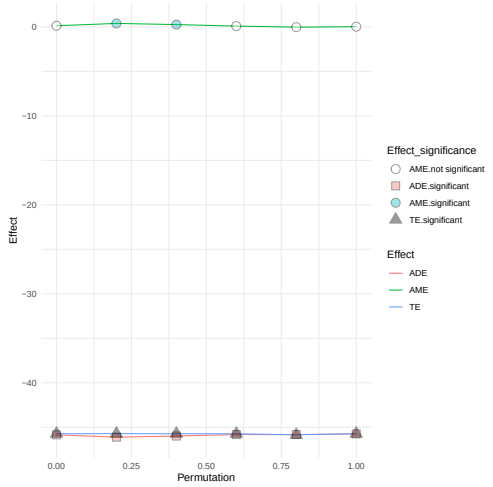
(b) Mobility category: transit stations



(c) Mobility category: grocery and pharmacy



(d) Mobility category: residential



(e) Mobility category: workplaces

Figure 6: Causal effects and their significance for all mobility categories. The x-axis represents the permutation rate for the mediator. The effects are denoted statistical significant within the 95% confidence interval.

A.4. Causal effect estimation Sweden

	Workplaces	**Retail and Recreation**	**Transit stations**	**Grocery and pharmacy**	**Residential**
ADE	22.7*	-32.0***	-27.9***	-0.2	6.4***
AME	8.3	-1.6	-4.8	-3.7	0.3
TE	31.0***	-33.6***	-32.7***	-3.9	6.7***
Mediator regression model $f_M(T, C)$ $\bar{R}^2$	0.89	0.89	0.89	0.89	0.89
Outcome regression model $f_Y(T, M, C)$ $\bar{R}^2$	0.73	0.85	0.93	0.71	0.92

Table 8: Mediated, direct and total effect of policies on the different mobility categories in Sweden during the first half of 2020. The effect estimates are the percent change in mobility to a predefined mobility baseline between January 3 and February 6 in 2020. *p<0.1; **p<0.05; ***p<0.01

A.5. Dataset UK

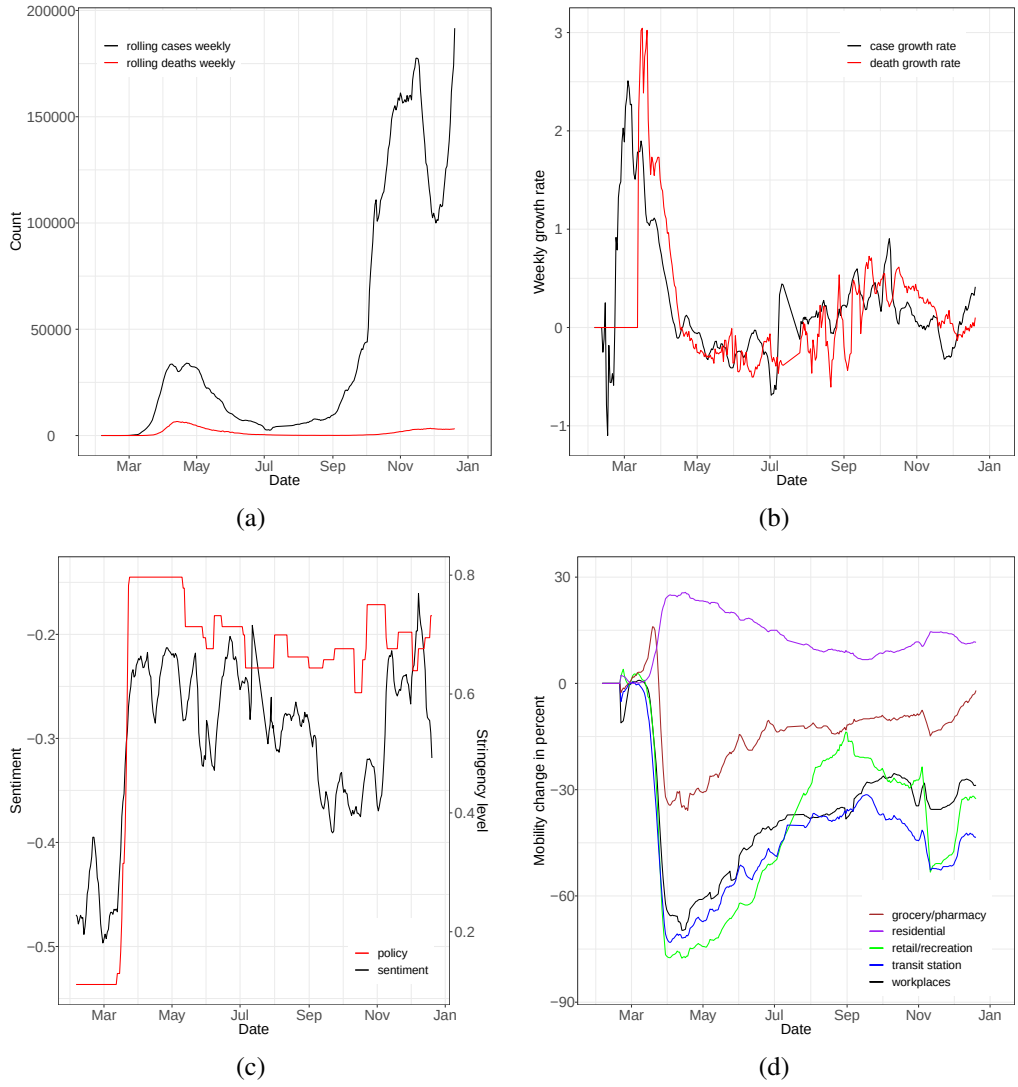


Figure 7: **Dataset UK** (a) rolling weekly reported cases and deaths (b) case and death growth rate (c) Stringency Index and sentiment (d) change in percentage for certain mobility categories to the baseline

A.6. Dataset Sweden

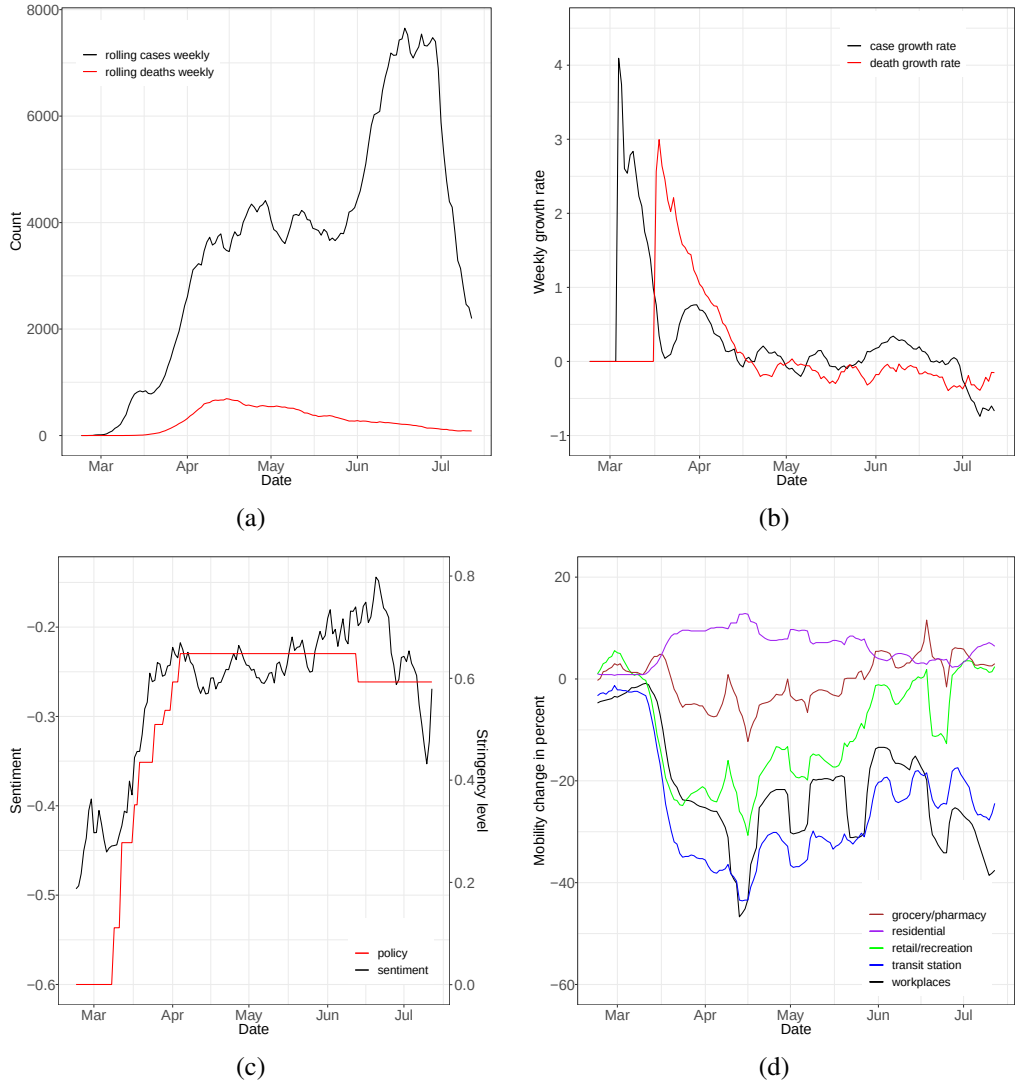


Figure 8: **Dataset Sweden** (a) rolling weekly reported cases and deaths (b) case and death growth rate (c) Stringency Index and sentiment (d) change in percentage for certain mobility categories to the baseline