Impact of misinformation on COVID-19 vaccine uptake

Team 32
Aizaan Anwar || Anne Wullenweber || Cori Campbell || Mariana Prazeres
Saran Davies || Yifan Zhang

1. Introduction

1.1 Overview

Although vaccines have been offered to entire populations in many countries by now, uptake has stagnated in some regions and within certain demographics. This stagnation is thought to be partially due to the prevalence of vaccine misinformation (Loomba et al. 2021). The phenomenon of the vaccine hesitancy (or other negative impacts) caused by misinformation is defined as ‘infodemics’ by the World Health Organisation: “*[an] overabundance of information - some accurate and some not - that occurs during an epidemic. It can lead to confusion and ultimately mistrust in governments and public health response.*” (WHO 2021).

Understanding the impact of misinformation on vaccination coverage can not only help save lives but also have a significant socio-economic impact. For example, in the absence of vaccine hesitancy, 236-305 covid-related deaths per million population can be prevented (Mesa et al. 2021). An economic analysis of 10 vaccines across 94 countries showed that $586 billion in the direct cost of illness can be saved with an investment of $34 billion. The reduction in morbidity and mortality from vaccination not only leads to long-term savings from the prevention of disease but also significant returns in investment (Rodrigues and Plotkin 2020). It is estimated that every dollar invested in vaccines over a decade can lead to a return of 16 times the original investment (Ozawa et al. 2016).

Several survey data at the individual level unambiguously conclude that misinformation has a negative impact on vaccine *intent (See Section 1.1)*. However, how vaccine *intent* translates to actual population-level vaccine *uptake* remains unknown. Therefore, the question of whether misinformation would affect the actual vaccine uptake is worth investigating. To this end, this study analyses the relationship between actual vaccine uptake and fake news occurrence in the United States based on the COVID-19 related datasets provided by Our World in Data (Mathieu et al. 2021) as well as custom datasets built by scraping data related to fake news.

The study also has important implications for future vaccine rollouts. Given recent advances in rapid vaccine development, the frequency of new vaccines entering the market may increase. Without careful consideration of how to achieve higher vaccine uptake, new means of rapid vaccine development will not be enough to combat future epidemics. Therefore, overcoming vaccine hesitancy and misinformation is of paramount importance for successful vaccination programs.

The outline of the report is as follows: the rest of the introduction reviews the studies on misinformation on vaccine intent, Section 2 discusses the data source and exploratory data analysis. Built on the datasets described in Section 2, Section 3 provides a time-series analysis on the vaccine uptake and fake news from different approaches.

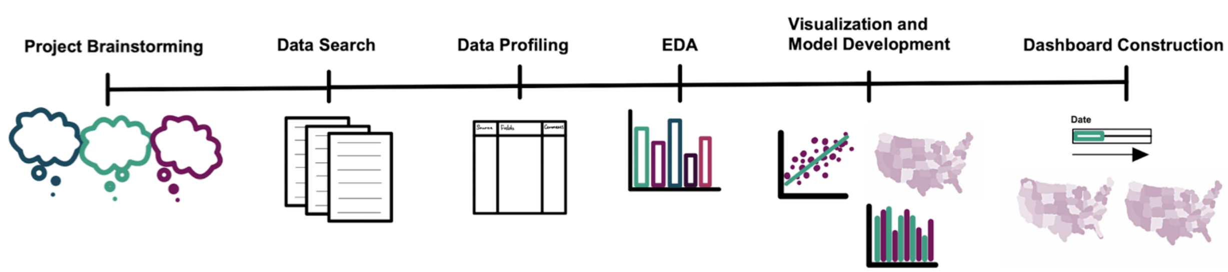
1.2 Target Audience

1. **Health authorities and policy-makers** would likely be the primary audience for this report; as vaccine hesitancy has a direct bearing on effective public health policy and government expenditure.

2. **Social media companies** should also be a target audience, as the evidence shows that their platforms are being used to spread misinformation (Wilson and Wiysonge 2020). In addition to social duty, it is also in their best interest to ensure the continued health of their target user and workforce, so as to maintain their company revenues.

1.3 Scoping process

We started by discussing shared topics of research interest and accordingly refined our question. We identified a pertinent **problem/need**: to characterise and quantify the effect of COVID-19 vaccine misinformation on vaccination uptake. From here we undertook a search for relevant datasets, identifying both summary- (country, state/province and county/municipal) and individual-level data. Datasets were then profiled including summarising structure and entry fields. Missingness and appropriate analyses were then discussed and from here an actionable analysis plan was formed. We finalised key datasets of interest and carried out further visualization and implementation of time series analysis. We also decided to build a dashboard for data presentation and user interaction. The timeline of the scoping process is below:



1.4 Misinformation on vaccine intent

Several survey data at the individual level unambiguously demonstrate that misinformation has a negative impact on vaccine *intent*.

- Based on the (Schmelz and Bowles 2021; Roozenbeek et al. 2020; Loomba et al. 2021))'s survey data in four countries (the UK, Spain, the USA, and Mexico), we found that participants' beliefs in different misinformation types are highly correlated and all these wrong beliefs have a strong negative impact on people's vaccine intent (Figure 1).
- An experimental study by (Roozenbeek et al. 2020; Schmelz and Bowles 2021; Loomba et al. 2021)) shows that individuals showed lower vaccine intent after they have been

exposed to misinformation (treatment in the study). Though the impact varies across age, income, and trust in institutions (Figure 2).

The conducted EDA on the datasets from (Roozenbeek et al. 2020; Schmelz and Bowles 2021; Loomba et al. 2021) are summarized in Appendix 4.5.

Figure 1: Participants believing in misinformation has lower vaccine intent

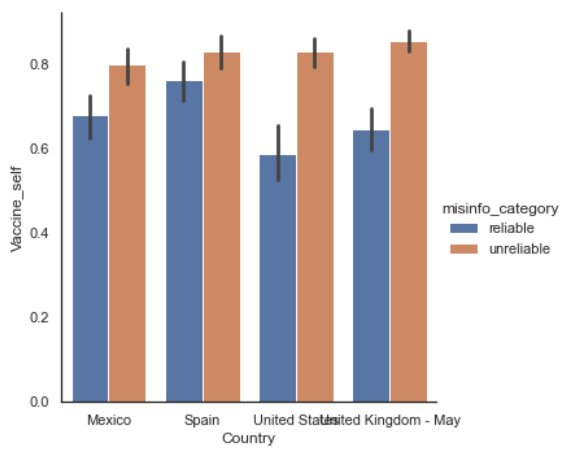
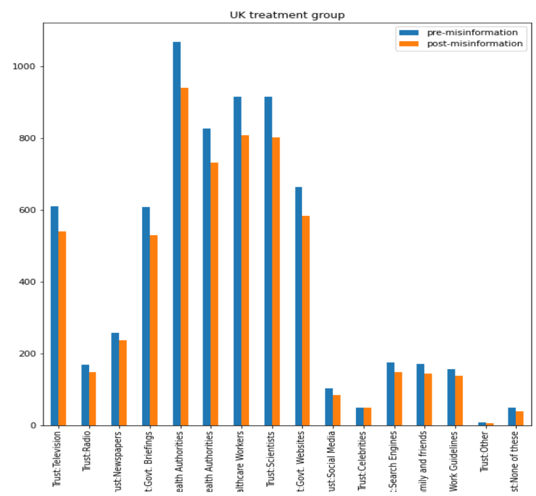


Figure 2: Vaccine intent decreased after participants were exposed to misinformation



However, despite previous survey studies/experiments concluding that misinformation has a significant negative impact on vaccine intent, **how vaccine *intent* translates to actual population-level vaccine *uptake* remains unknown.** As shown in (Roozenbeek et al. 2020; Schmelz and Bowles 2021; Loomba et al. 2021))'s, the individual effect of misinformation on vaccine intent may not translate directly to population-level vaccine uptake. **Therefore, the question of whether misinformation would affect the actual vaccine uptake is worth investigating.**

2. Data Analysis & Computation

2.1 Data

The main databases used for this project are the COVID-19 related datasets provided by Our World in Data (Mathieu et al. 2021) as well as custom datasets built by scraping data related to fake news. The reasons for focusing on those two datasets are: 1. The COVID-19 related datasets cover relevant information on the vaccine uptake (dependent variable), and COVID-19 cases, deaths, government stringency index (control variables) since the outbreak of the pandemic; 2. The fake news datasets gathered *identified falsehoods* related to the COVID vaccine (independent variable) from January 2020 to October 2021.

2.1.1 Vaccine Dataset

Data on vaccination uptake was obtained from the complete Our World in Data Covid-19 Dataset (OWIDCD) which contains metrics on vaccinations, confirmed cases, deaths, excess mortality, hospital admissions, policy responses, tests, demographics and economic profile (Mathieu et al. 2021). Each metric contains information from 38-241 countries (depending on the variable). Key variables include the number of vaccinations administered daily, daily death counts, number of daily cases, stringency index, number of diagnostic tests performed daily and gross domestic product per capita. The metrics were either fixed, updated daily or updated weekly. For this report, we chose to focus on time-series data from the United States, as this sub-dataset contained the least amount of missing information. Vaccination data from the US were available from December 2020 and updated daily.

2.1.2 Custom Fake News Dataset

Description: The custom fake news dataset is a self-built dataset that contains information about all the vaccine-related falsehoods. The dataset was constructed by scraping data from Poynter.org (“CoronaVirusFacts Alliance - Poynter” 2020). The dataset contains falsehoods that were identified by the(“CoronaVirusFacts Alliance - Poynter” 2020) in more than 70 countries from January 2020 to October 2021.

Dimension: The dataset has 2295 entries. For the U.S. subsample, there are 335 entries. One entry corresponds to one fake news, with news `title`, `rating`, `country`, and `date`. The dataset looks as follows:

	rating	title	date	location
0	Missing context	Politician Söder shown at various vaccine app...	2021/10/02	Germany
1	FALSO	Children are 50 times more likely to die from...	2021/08/11	Spain
2	FALSE	The CDC reports the deaths of two children an...	2021/08/09	Ukraine, United States
3	FALSE	COVID-19 vaccines “fight the virus wrong and ...	2021/08/08	United States
4	FALSO	Pfizer vaccines can cause baldness and sexual...	2021/08/04	Colombia

Beyond the wealth of information that this dataset provides, the true benefit of this dataset is two-fold. First, the dataset allows us to track the spread of given fake news. For example, the fake news *‘Elisa Granato, the first volunteer to try a vaccine against COVID-19, died’* was first posted on April 25th, 2020 in the United States, Nigeria, and United Kingdom, and then spread to other countries (for example Italy) in the next 1-5 days [See Section 3.1]. Second, the dataset can be aggregated across time, so that we can calculate the number of fake news across time. The time-series data will allow us to identify the impact of fake news on vaccine uptake over time.

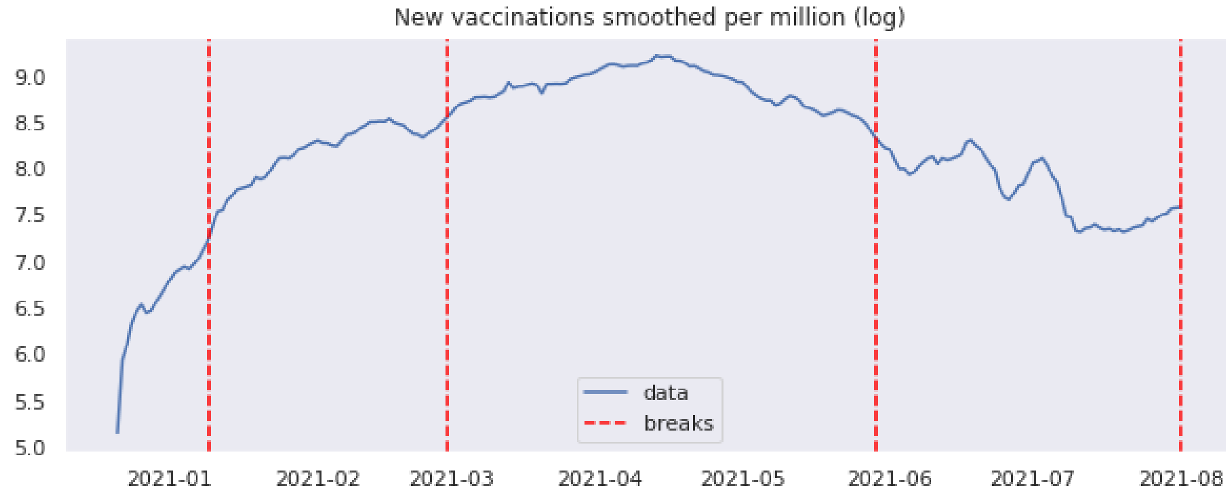
2.1.3 Data Cleaning and Sample Selection

To analyse the impact of fake news on vaccine uptake, we combined the two aforementioned datasets based on `date` and `country`. For the purpose of this study, we first removed samples when the vaccine was not available (i.e. pre-December 2020 data), and we also removed the samples when fake news information is unavailable. The selection of covariates to be used for further analysis was made based on a priori knowledge and current literature on misinformation and vaccine uptake (Roozenbeek et al. 2020; Loomba et al. 2021; Schmelz and Bowles 2021).

Furthermore, the vaccine data - like most time-series data - may be subject to structural changes (shift in levels/volatilities) in the data. As the presence of breakpoints in data may lead to errors and model instability when performing prediction, we conducted formal tests to detect breakpoints in the time series based on the dynamic programming approach. The four identified breakpoints are `['2021-01-09', '2021-02-28', '2021-05-29', '2021-10-13']`, respectively (Figure 3). Considering that before `'2021-02-28'` the daily vaccine uptake was constrained by vaccine supply rather than people’s demand, so we restricted our study to the latter two subsamples.

After doing the data cleaning, this left us with around 171 entries (dates). Despite that, to our knowledge, this dataset is the largest dataset on vaccine uptake and fake news that one can obtain. We carefully addressed the mean shift and potential small sample bias when doing modelling.

Figure 3 New vaccinations smoothed per million (log)



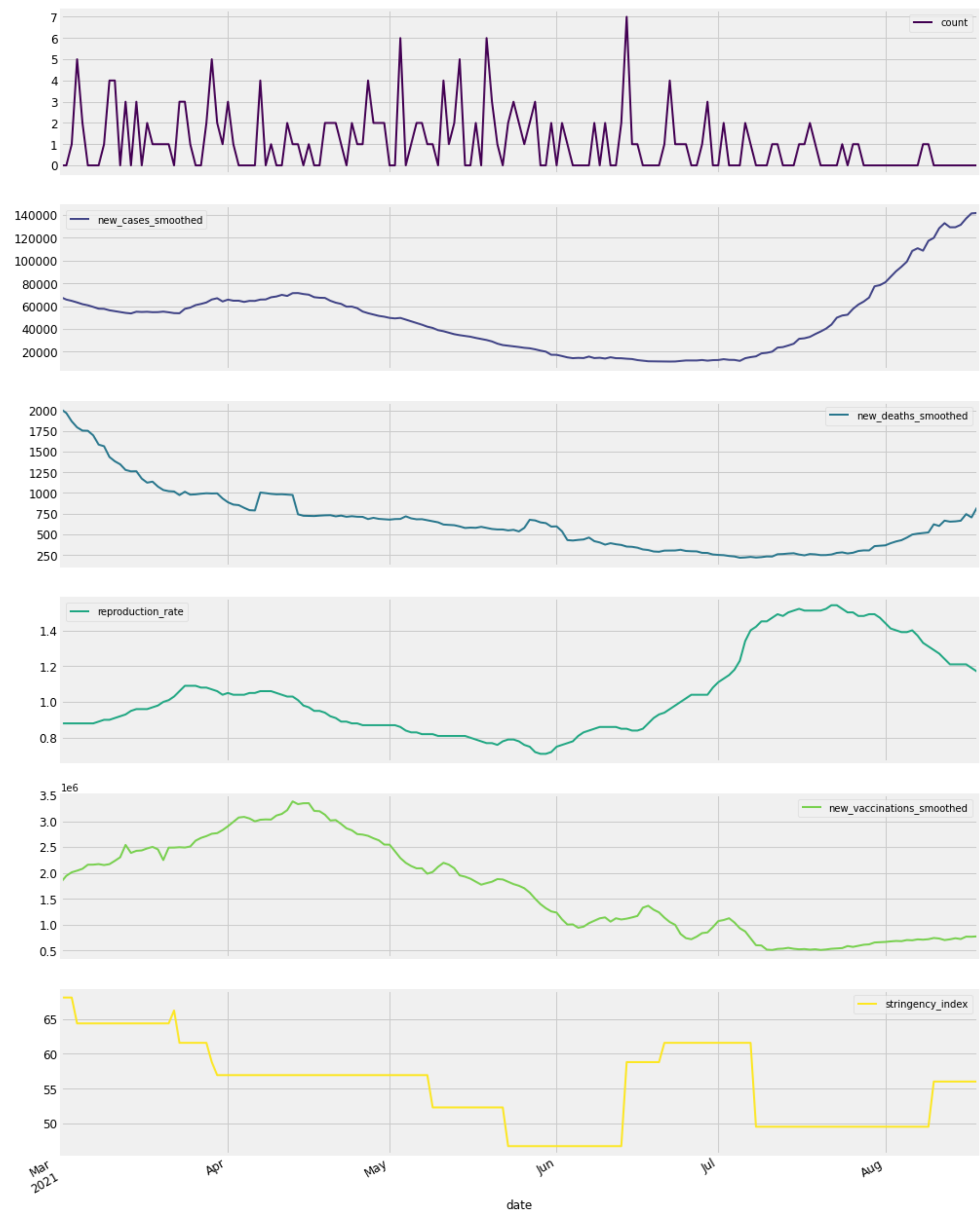
The dataset is in general non-stationary, however, all the variables should be stationary to be used for Vector Auto-Regression (VAR). Common transforms in the literature are the difference and logarithmic transform, which have been deployed in this study. After the transformation, all the variables passed the Dicky-Fuller test for stationarity.

2.2 Exploratory Data Analysis

2.2.1 Data visualisation & descriptive statistics

The plots below show the potential time-series predictors and the target variable (number of new vaccinations daily) for the interval when information on fake news was available (1st March 2021 and 18th August 2021) (see Figure 4).

Figure 4 Data visualisation



Vertical event lines were superimposed on the 'number of vaccinations' plot to represent the dates when fake news was detected to allow for better appreciation of shock events and dates of interest (Figure 5).

Figure 5 Fake news (vertical lines) and vaccine uptake (green line)

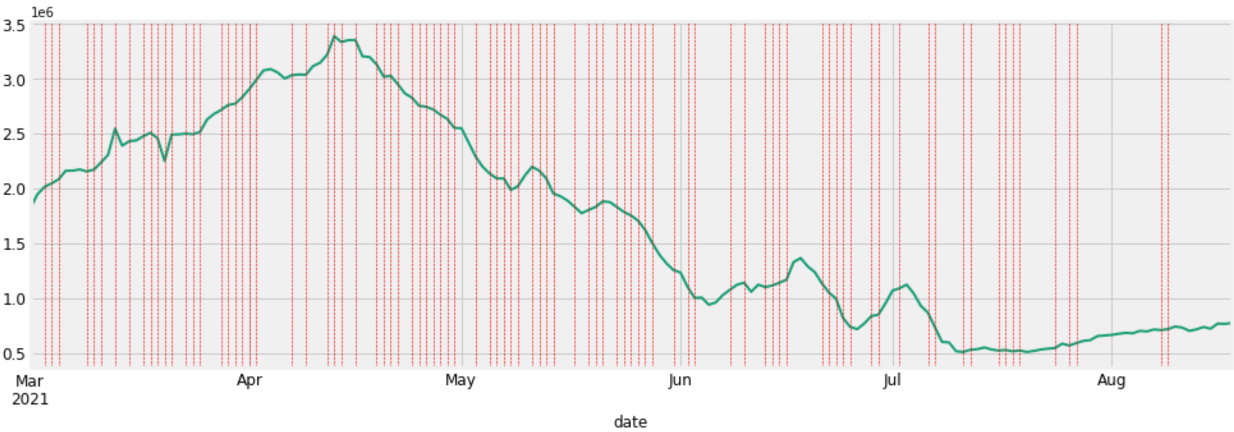


Table 1 Descriptive statistics

	Count (of fake news)	New cases smoothed	New deaths smoothed	Reproduction rate	New vaccina-tions smoothed	Stringency index
Count	171	171	171	171	171	171
Mean	1,0526	48867,52	666,71	1,0443	1681021	55,51
Std	1,3986	31218,95	389,17	0,2437	905829	5,95
Min	0	11388,57	217,86	0,71	506771	46,76
25%	0	20523,86	328,21	0,86	764495,5	49,54
50%	1	51482,29	610,71	0,97	1750524	56,94
75%	2	64700,71	806,5	1,2	2489215,5	61,57
max	7	141884	2014,29	1,54	3384387	68,06

This table summarizes the descriptive statistics of the variables of interest in the final dataset. It consists of 171 rows, each belonging to a specific day. The reported variables provide information on the number of fake news released on a specific day as well as the current number of cases, deaths and vaccinations (each smoothed by taking the 7-day average to account for outliers). The current Covid reproduction rate and the stringency index are further provided. The stringency index is a measure for the strictness of a government's current response to the pandemic - it ranges from 0 to 100 with 100 being the strictest response.

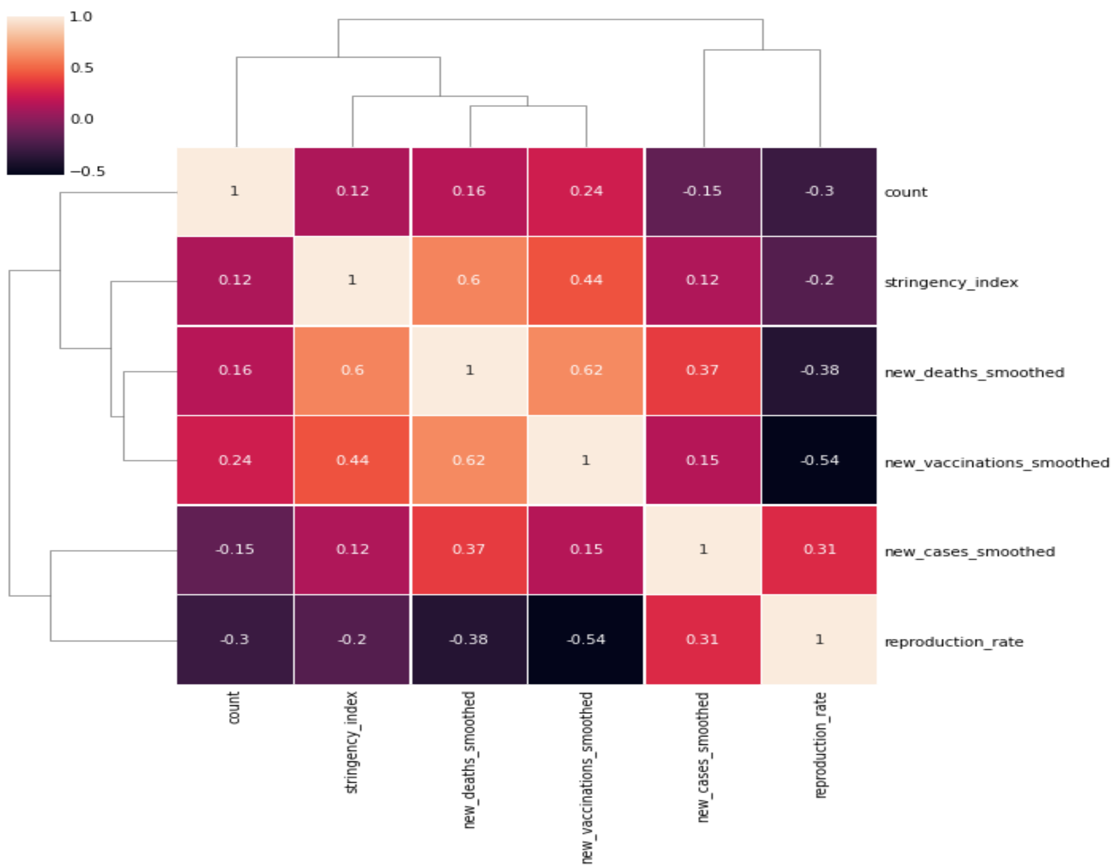
3. Statistical Analysis & Machine Learning

3.1 Feature Engineering

3.1.1 Pearson correlation test for daily variable percent change

A Pearson correlation test was done on the daily percentage change in all variables to assess the correlation of potential variables and the target variable (see Figure 6). All variables do not seem to have a strong correlation with the `count` of fake news (0.24 >r >-0.3). `Stringency_index` seems to be strongly correlated with `new_deaths_smoothed` (r = 0.6) and `new_vaccinations_smoothed` (r=0.44). `New_deaths_smoothed` is strongly correlated with `new_vaccinations_smoothed` (r=0.62). As expected, `new_cases_smoothed` is positively correlated with `reproduction_rate` (0.31) and `new_deaths_smoothed` (0.37). `Reproduction_rate` is negatively correlated with `new_vaccinations_smoothed` (r=-0.54) and `new_deaths_smoothed` (r =-0.38).

Figure 6 Pearson correlation test



3.1.2 Test for stationarity: Augmented Dickey-Fuller Test

Time-series data are often subject to non-stationarity, however, all the variables should be stationary to be used for Vector Auto-Regression (VAR). Therefore, an Augmented Dickey-Fuller (ADF) test was done on all variables to assess for stationarity (results see Table 2). All variables showed a high p-value (≥ 0.05) for the ADF test statistic, which highly suggests that all predictors are non-stationary. Common transforms in the literature are the difference and logarithmic transform, which have been deployed in this study. After the transformation, all the variables passed the Dicky-Fuller test for stationarity.

Table 2 Augmented Dickey-Fuller Test

Predictor variables	ADF test P-value
count	0.285037945
new_vaccinations_smoothed	-0.74473329
new_cases_smoothed	0.050245802
new_deaths_smoothed	0.649774608
reproduction_rate	0.424577697
stringency_index	0.134797928

3.1.2 Test for multicollinearity: Variance Inflation Factor

The variance inflation factor (VIF) was calculated to assess for multicollinearity between predictors (see Table 3 below). `Reproduction_rate` and `stringency index` were the predictors with high multicollinearity i.e. >10. `Reproduction_rate` was removed from all forecasting models as we thought this was collinear with `new_cases_smoothed`. As `stringency index` was thought to be an important predictor for policy response, this variable was retained for the first forecast model (Prophet), however, given it is quite flat for the study period, it is removed in the final Mean-adjusted Bayesian Vector Auto-Regression model.

Table 3 Variance Inflation factor

Predictor Variables	VIF
count	1.725383
new_cases_smoothed	5.793623
new_deaths_smoothed	9.584746

reproduction_rate	28.5207
stringency_index	41.97843

3.2 Multivariate Time Series Modelling and Forecasting

Various approaches were employed to model the impact of misinformation on the Covid vaccine uptake in the United States. Given that both the vaccine uptake and the amount of misinformation are time-series data, only modelling approaches suitable for time-series data were considered.

3.2.1 Facebook Prophet

Facebook Prophet is an open-source library specifically developed for time-series modelling and forecasting which is available for use with both Python and R[1]. Prophet calculates a forecast based on four main components - a function for non-periodic changes in the data g(t), a function for periodic changes in the data s(t), a function for the effect of holidays h(t) and an error term ε_t (Taylor and Letham 2018):

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

By default, Prophet performs a univariate forecast on only the time series values from previous timesteps. Additional variables can, however, be added as regressors to perform a multivariate forecast. In order to forecast with a multivariate prophet model, the values for the additional variables must be given at all time steps for which the forecast is created. The following focuses on a multivariate analysis of the vaccine uptake using Prophet, the results of a similar univariate analysis are provided in Appendix 4.1.

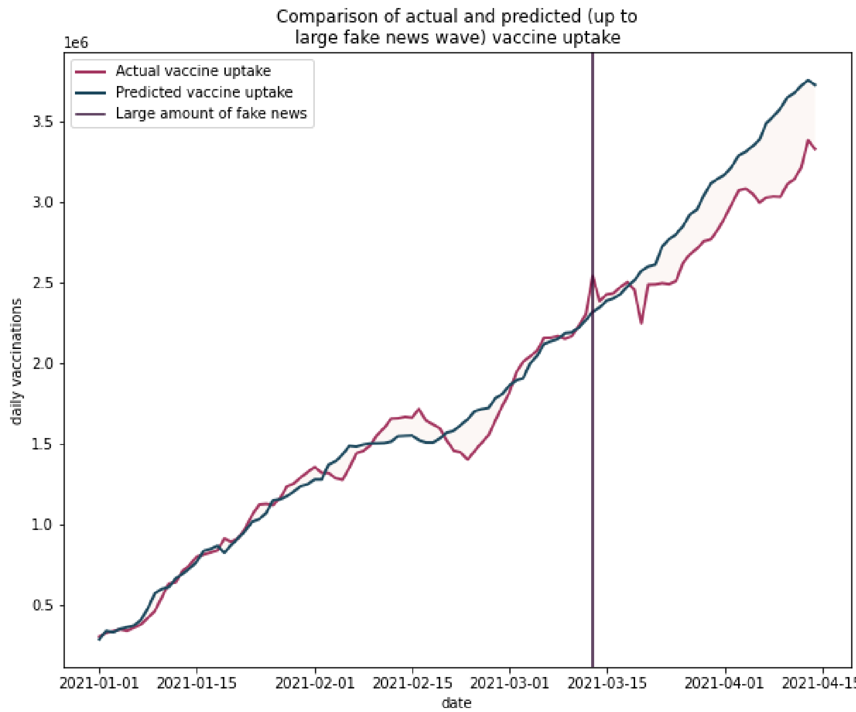
Prophet was chosen because of its capability to predict complex forecasts on non-stationary data, its wide acceptance and its incorporation of periodic trends. While weekly trends in vaccine uptake are plausible and likely, the impact of misinformation on vaccine uptake is not expected to follow a periodic pattern. As Prophet works best for data having seasonal effects and does not provide significant values of the obtained results, different models were chosen for further analysis of the impact of misinformation on vaccine uptake.

Following the EDA described in Section 2.2, three variables that seem to influence the vaccine uptake apart from misinformation and fake news, are chosen as model regressors. These are given by the daily new cases, daily new deaths and the current stringency index. The stringency index is a measure of the strictness of the current government response to the pandemic. The model was trained with data up to mid-March and a forecast of the vaccine uptake for the following 30 days was predicted. This time frame was chosen as the amount of fake news reached a maximum. As shown below, while the model predicts an (approximately) constant uptake, the actual uptake stagnates and even drops at one point shortly after an especially large

[1] https://facebook.github.io/prophet/

amount of fake news was released. This indicates a correlation between fake news and vaccine uptake and serves as motivation for further investigation of the impact of fake news on vaccine uptake.

Figure 7 Actual vaccine uptake(red) versus predicted uptake without fake news(green)



3.2.2 Vector Auto-Regression

The impact of misinformation on vaccine uptake was further modelled with Vector Auto-Regression (VAR) – a statistical model to describe the relationship of various endogenous variables over time. A VAR model is given by

$$\mathbf{Z}_t = \boldsymbol{\alpha}_1\mathbf{Z}_{t-1} + \boldsymbol{\alpha}_2\mathbf{Z}_{t-2} + \ldots + \boldsymbol{\alpha}_p\mathbf{Z}_{t-p} + u_t \qquad \text{(Biller and Nelson 2003)}$$

with $\mathbf{Z}_t$ being a vector of the values of the endogenous variables at time t and $\boldsymbol{\alpha}$ the corresponding coefficient matrices. The order of the VAR model, how many previous time steps are included in the modelling, is given by p.

To predict the impact of fake news on vaccine uptake, a multivariate VAR model was fitted to the daily vaccine uptake, the daily number of Covid cases and deaths, the daily stringency index as well as the count of the daily occurring fake news. The model was trained with data from the US between January and October 2021. As VAR models require the underlying time-series data to be stationary, the time-series data were differentiated once using pandas.diff(). The stationarity of the data was verified with the augmented Dickey-Fuller unit root test. The order of the VAR model was set to 18 since this provided a good tradeoff between the Akaike information criterion and the statistical significance of the results. The summary of the VAR model is shown below (for the sake of clarity, only coefficients with a p-value < 0.05 for the daily vaccination uptake are included in this summary).

Table 4 Vector Auto-Regression Results (selected)

```
---------------------------------------------------
  Summary of Regression Results
==================================
Model:                          VAR
Method:                         OLS
Date:                 Tue, 19, Oct, 2021
Time:                      16:48:07
--------------------------------------------------------------------
No. of Equations:      5.00000    BIC:                   55.9157
Nobs:                  285.000    HQIC:                  52.4221
Log likelihood:        -8704.04   FPE:             6.33399e+21
AIC:                   50.0845    Det(Omega_mle):  1.58476e+21

Results for equation new_vaccinations_smoothed
```

	coefficient	std. error	t-stat	prob
L1.new_vaccinations_smoothed	0.283301	0.074800	3.787	0.000
L2.count	8919.168466	4518.344130	1.974	0.048
L2.new_vaccinations_smoothed	0.155838	0.078305	1.990	0.047
L3.new_vaccinations_smoothed	0.176154	0.081230	2.169	0.030
L7.new_vaccinations_smoothed	-0.350301	0.082817	-4.230	0.000
L8.stringency_index	11959.989119	879.291056	13.602	0.000
L10.count	-21694.746871	8448.159722	-2.568	0.010
L11.count	-23210.981223	8472.456590	-2.740	0.006

L12.count	-23240.396258	8350.585580	-2.783	0.005
L13.count	-18977.054782	8098.865328	-2.343	0.019
L14.count	-15046.024365	7522.053128	-2.000	0.045
L16.count	-12361.049181	5777.733919	-2.139	0.032

The VAR model suggests that the amount of daily new vaccinations from the previous 1, 2, 3 and 7 timesteps, the stringency index from 8 time steps before as well as the count of fake news from the previous 2, 10, 11, 12, 13, 14 and 16 timesteps have a statistically significant influence on the vaccine uptake. According to the model, one new fake news article accounts for an increase of 8919 vaccinations two days later. One fake news article further accounts for a statistically significant decrease of between 12361 and 23240 vaccinations per day on days 10, 11, 12, 13, 14 and 16 after its release. The interpretation of the coefficients remains unchanged although the time-series data were differentiated once to make it stationary (Pankratz 2012). According to the modelled vaccine uptake with VAR, one fake news article accounts for a cumulative decrease of the daily new vaccinations by 105 611 over the course of 18 days.

3.2.3 Mean-adjusted Bayesian Vector Auto-Regression (BVAR)

In this section, we adopt Bayesian methods to estimate a vector autoregression (VAR) model. BVAR differs from standard VAR models in that the model parameters are treated as random variables, with prior probabilities, rather than fixed values. Moreover, as mentioned in Section 2.1.3, the vaccine time series may be subject to shifts in the mean over time. Therefore, we adapted the standard Bayesian VAR with additional lagged values of exogenous variables in its deterministic component (Villani, 2009), i.e.,

$$A(L)(y_t - Fx_t) = \varepsilon_t$$

In this representation, $A(L)$ is a lag polynomial with dimension, F is a $n * m$ matrix of coefficients with respect to the m exogenous variables. This formulation allows for the explicit inclusion of prior information about steady-state values. Incorporating it into the standard Bayesian VAR and estimates the following equation:

$$Y = XB + Z\Delta + \epsilon$$

With

$$Y = \begin{pmatrix} y_1' \\ y_2' \\ \vdots \\ y_T' \end{pmatrix}, \ X = \begin{pmatrix} y_0' & y_{-1}' & \dots & y_{1-p}' \\ y_1' & y_0' & \dots & y_{2-p}' \\ \vdots & \vdots & \ddots & \vdots \\ y_{T-1}' & y_{T-2}' & \dots & y_{T-p}' \end{pmatrix}, \ Z = \begin{pmatrix} x_1' & -x_0' & \dots & -x_{1-p}' \\ x_2' & -x_1' & \dots & -x_{2-p}' \\ \vdots & \vdots & \ddots & \vdots \\ x_T' & -x_{T-1}' & \dots & -x_{T-p}' \end{pmatrix}$$

$$B = \begin{pmatrix} A_1' \\ A_2' \\ \vdots \\ A_p' \end{pmatrix}, \ \Delta = \begin{pmatrix} F' \\ F'A_1' \\ \vdots \\ F'A_p' \end{pmatrix}, \ \mathcal{E} = \begin{pmatrix} \varepsilon_1' \\ \varepsilon_2' \\ \vdots \\ \varepsilon_T' \end{pmatrix}$$

As we were doing a Bayesian approach, we need to determine the prior distribution for the parameters.

$$\beta \sim N(\beta_0, \Omega_0)$$

$$\pi(\Sigma) \sim |\Sigma|^{-(n+1)/2}$$

$$\psi(\Sigma) \sim N(\psi_0, \Lambda_0)$$

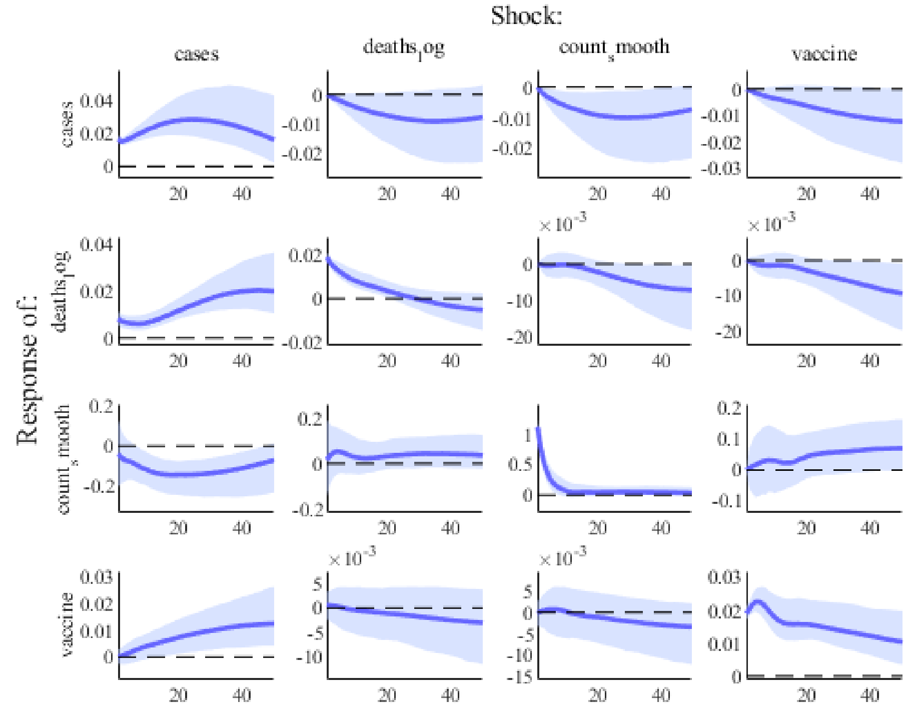
In practice, the prior parameters $\beta_0$and $\Omega_0$are set just as for the Minnesota prior. And we specify a 95% probability interval for the prior for $\psi$and to calculate the prior mean and variance retrospectively from this interval.

The advantage of this model is in terms of steady-state values - it allows the steady-state values to change over time. In this exercise, we set the prior distribution for the value of the steady states for the vaccine to follow a constant trend and vary across subsamples, as suggested by previous breakpoints estimation.

The results show that,

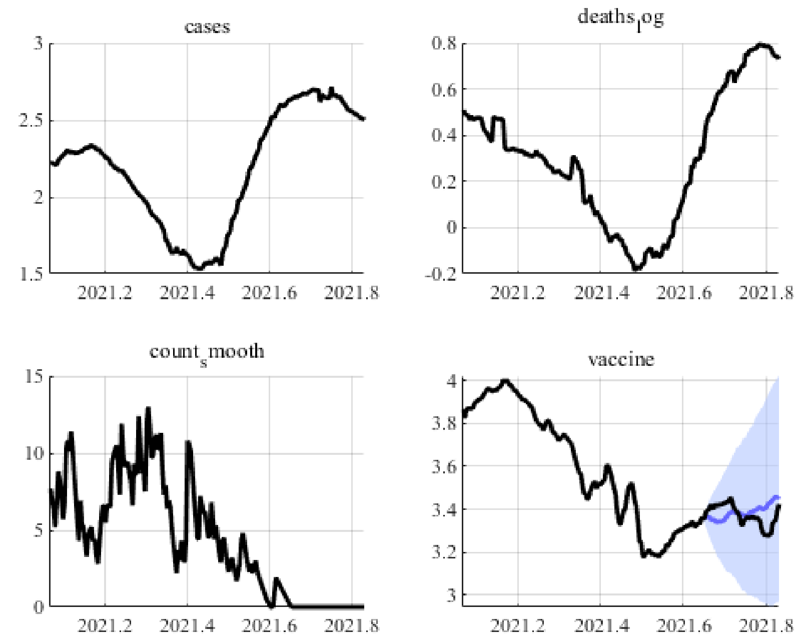
- **Following a shock in fake news, the vaccine would start to decline on day 10 and remain below the steady state for more than a month (Figure 8).** On a cumulative basis, one fake news might lead to a 12,909 decrease in vaccine uptake.
    - Caveat: the volatility around these time series are relatively high as there are some *unobservable* randomnesses around the vaccine supply and other variables.

Figure 8: Impulse Response Function (BVAR)



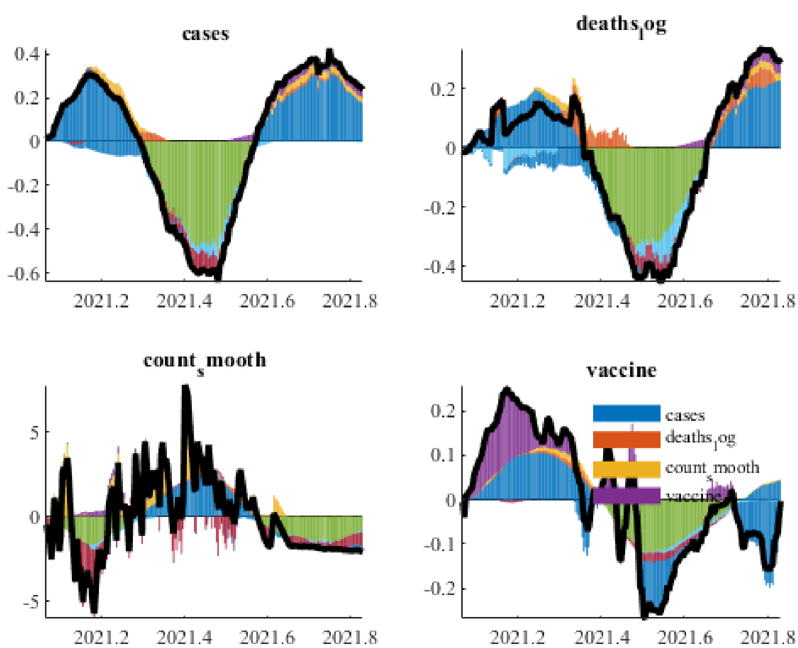
- **The model has strong predictive power.** The conditional forecast of the vaccine given the number of fake news, cases, and deaths are very close to the actual vaccine uptake in the data (Figure 9). The performance of the conditional forecast is significantly improved compared to the unconditional forecast (see Appendix).

Figure 9 Conditional Forecast of the Vaccine Uptake on cases/death and fake news



- **The negative impact of fake news is strongest in early 2021 as the vaccine started to roll out in scale.** Compared to other periods, the fake news during early 2021 causes the strongest vaccine hesitancy. This might be explained by the fact that people are more prone to fake news at the beginning of the vaccine rollout.

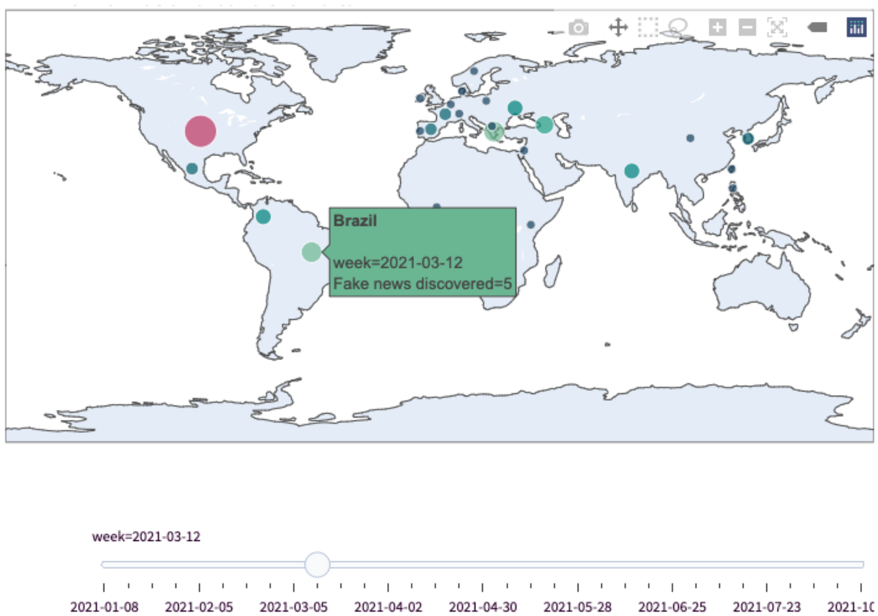
Figure 10 Historical decomposition of variance



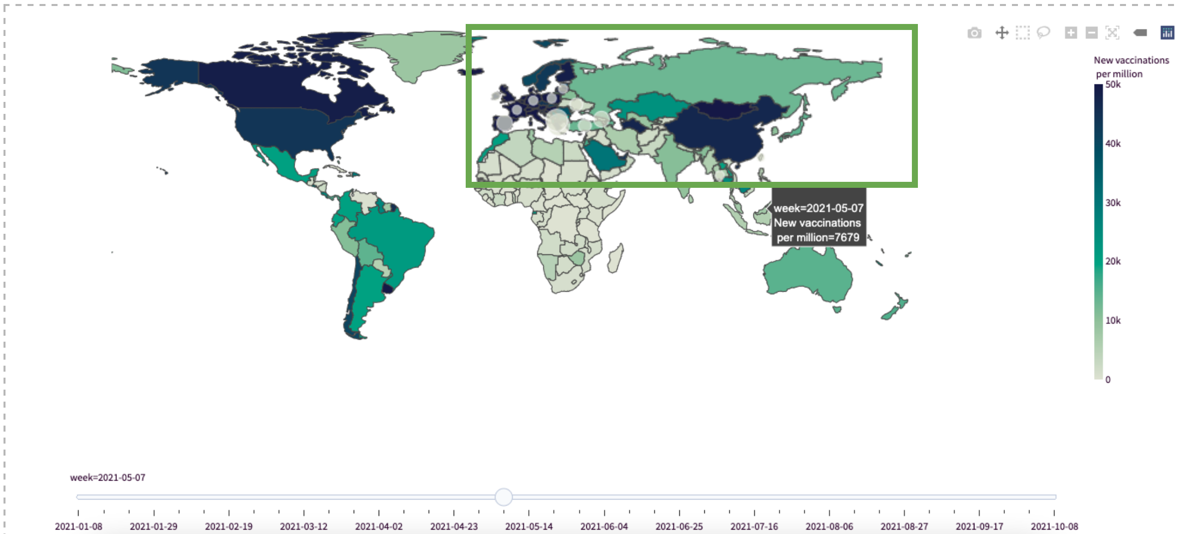
4. Dashboard

We have created a landing-page dashboard to complement our analysis. Our dashboard displays visual information about vaccine uptake and fake news over time and several countries. For details on the datasets used refer back to Section 2.1. The landing page was built to appeal to the general public and to read as an article describing our findings. We also made our supporting work - report, presentation and datafolio - easily accessible, for those readers interested in learning more.

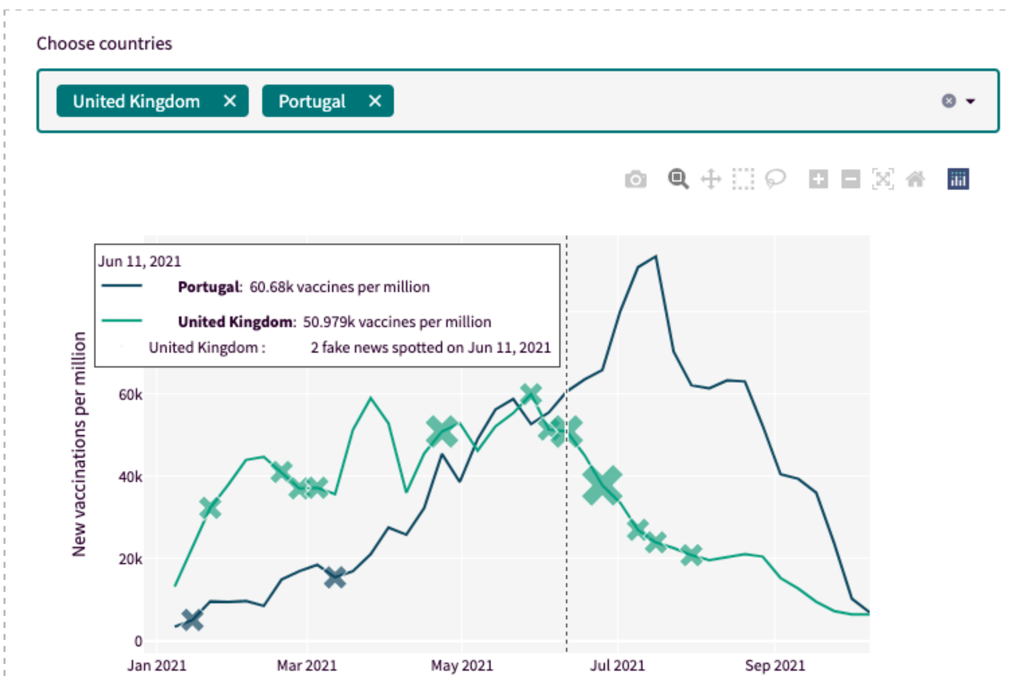
There are three interactive plots on the landing page. The first plot displays a world map where each bubble size represents the amount of fake news recorded in our dataset at the date chosen in the timeline. The exact number can be retrieved by hovering over each bubble.



The second plot is also a world map, but it is now coloured with new vaccinations per million for the date chosen. There is a supporting colour map to the right but exact values can be retrieved by hovering. There is a checkbox to overlay the previously seen bubbles for misinformation on the coloured map as in the green box.



The final interactive plot is a time series of new vaccinations where the user can select which countries to observe. Besides vaccinations, there are also x-crosses indicating misinformation events, which are detailed upon hovering.



Besides interactive plotting, we have also included a word cloud of common misinformation themes and some selected plots from the statistical analysis in 2.3.

Finally, the dashboard was implemented in Python3 using Pandas, Plotly and Streamlit and the code is publicly available at GitHub - VaXTrack.

5. Conclusions & Future Work

Based on our custom datasets and rigorous time-series analysis, we conclude that fake news can lead to vaccine hesitancy and even vaccine rejection. In particular, fake news has a severe negative impact on vaccine uptake between 10 and 16 after its release, and this impact, though dampened over time, can last roughly 30-50 days. On a cumulative basis, one fake news article could reduce new vaccinations by around 105 611 to 12,909 doses. The above results are robust to various time-series analyses.

Despite the robust results we get, our study has several limitations. First, the vaccine uptake is potentially influenced by many confounding variables (such as vaccine availability, governmental regulations and personal circumstances) which can be hard to quantify. Furthermore, classifying misinformation is a research topic on its own. Hereby, misinformation must be distinguished from disinformation. Misinformation refers to spreading factually incorrect information without taking the spreader's intent or their awareness that the information is incorrect into account. Disinformation refers to deliberately spreading misinformation knowing it is not factually correct (Guess and Lyons 2020). Our conducted study relied only on fake news instead of on misinformation on social media which is challenging to obtain due to various reasons, e.g. the abundance of data. To measure the overall impact of misinformation on vaccine intent, all these factors would need to be taken into account.

We could extend this study into various directions in the future. The statistical analysis focused on vaccine uptake and misinformation in the US - a similar analysis could be conducted for other countries to measure a variation of the impact of fake news on vaccine uptake in different countries. It could further be interesting to compare the vaccine uptake and the influence of misinformation across different demographic groups. In order to quantify the impact of social media in general on vaccine uptake, the Covid vaccine uptake could be compared with the uptakes of previous vaccines developed before social media existed in its current scale. Lastly, separate analyses for different types and topics of misinformation could be carried out and compared.

References

Biller, Bahar, and Barry L. Nelson. 2003. "Modeling and Generating Multivariate Time-Series Input Processes Using a Vector Autoregressive Technique." *ACM Transactions on Modeling and Computer Simulation* 13 (3): 211–37.

"CoronaVirusFacts Alliance - Poynter." 2020. March 18, 2020. https://www.poynter.org/coronavirusfactsalliance/.

Guess, Andrew M., and Benjamin A. Lyons. 2020. "Misinformation, Disinformation, and Online Propaganda." *Social Media and Democracy: The State of the Field, Prospects for Reform*, 10–33.

Loomba, Sahil, Alexandre de Figueiredo, Simon J. Piatek, Kristen de Graaf, and Heidi J. Larson. 2021. "Measuring the Impact of COVID-19 Vaccine Misinformation on Vaccination Intent in the UK and USA." *Nature Human Behaviour* 5 (3): 337–48.

Mathieu, Edouard, Hannah Ritchie, Esteban Ortiz-Ospina, Max Roser, Joe Hasell, Cameron Appel, Charlie Giattino, and Lucas Rodés-Guirao. 2021. "A Global Database of COVID-19 Vaccinations." *Nature Human Behaviour* 5 (7): 947–53.

Mesa, Daniela Olivera, Alexandra B. Hogan, Oliver J. Watson, Giovanni D. Charles, Katharina Hauck, Azra C. Ghani, and Peter Winskill. 2021. "Report 43: Quantifying the Impact of Vaccine Hesitancy in Prolonging the Need for Non-Pharmaceutical Interventions to Control the COVID-19 Pandemic." *Imperial College COVID-19 Response Team*. https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2021-03-24-COVID19-Report-43.pdf.

Ozawa, Sachiko, Samantha Clark, Allison Portnoy, Simrun Grewal, Logan Brenzel, and Damian G. Walker. 2016. "Return On Investment From Childhood Immunization In Low- And Middle-Income Countries, 2011–20." *Health Affairs* 35 (2): 199–207.

Pankratz, Alan. 2012. *Forecasting with Dynamic Regression Models*. John Wiley & Sons.

Rodrigues, Charlene M. C., and Stanley A. Plotkin. 2020. "Impact of Vaccines; Health, Economic and Social Perspectives." *Frontiers in Microbiology* 11: 1526.

Roozenbeek, Jon, Claudia R. Schneider, Sarah Dryhurst, John Kerr, Alexandra L. J. Freeman, Gabriel Recchia, Anne Marthe van der Bles, and Sander van der Linden. 2020. "Susceptibility to Misinformation about COVID-19 around the World." *Royal Society Open Science* 7 (10): 201199.

Schmelz, Katrin, and Samuel Bowles. 2021. "Overcoming COVID-19 Vaccination Resistance When Alternative Policies Affect the Dynamics of Conformism, Social Norms, and Crowding out." *Proceedings of the National Academy of Sciences of the United States of America* 118 (25): e2104912118.

Taylor, Sean J., and Benjamin Letham. 2018. "Forecasting at Scale." *The American Statistician* 72 (1): 37–45.

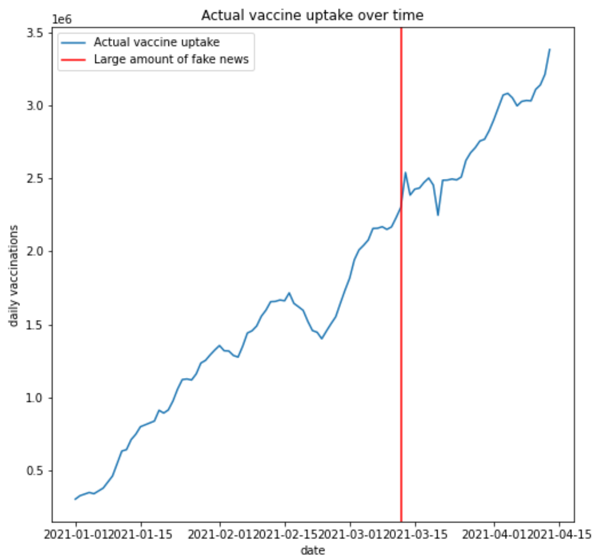
WHO. 2021. "Infodemic Management." World Health Organization Risk Communication. 2021. https://www.who.int/teams/risk-communication/infodemic-management.

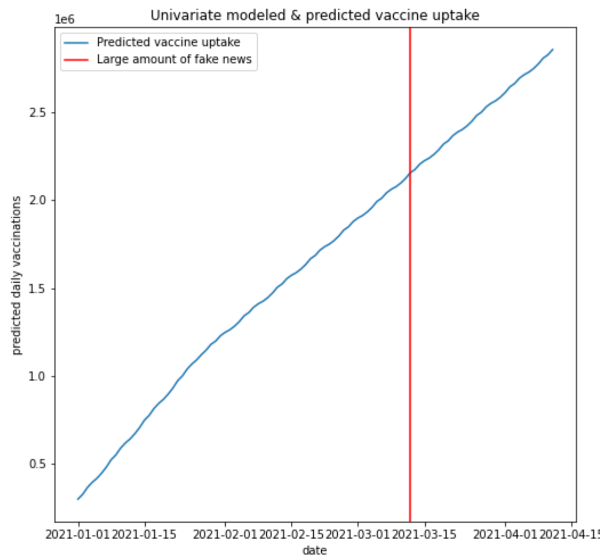
Wilson, Steven Lloyd, and Charles Wiysonge. 2020. "Social Media and Vaccine Hesitancy." *BMJ Global Health* 5 (10). https://doi.org/10.1136/bmjgh-2020-004206.

Appendix

A.1 Univariate Modeling & Forecasting with FB Prophet

A univariate analysis of the vaccine uptake over time was conducted similar to the multivariate modeling and forecast with Prophet described in Section 2.3. Similar to the multivariate model, the univariate Prophet forecast predicts a constant vaccine uptake opposed to the actual stagnating uptake at the time of the first large fake news wave on the internet. The univariate model is, however, much less capable of modeling smaller changes in the uptake than the multivariate model accounting not only for the past vaccine uptake, but also for the number of Covid cases and deaths as well as the current stringency index of the government Covid response.





A.2 Description of Datasets from Studies on Misinformation

[Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA]

orb_us.csv https://github.com/sloomba/covid19-misinfo/tree/main/dat [https://static-content.springer.com/esm/art%3A10.10]	• **Trust:** all boolean values (0 or 1), trust in different sources of information regarding Covid `'Trust:Television', 'Trust:Radio', 'Trust:Newspapers', 'Trust:White House Briefings', 'Trust:State Govt. Briefings', 'Trust:National Health Authorities', 'Trust:International Health Authorities', 'Trust:Healthcare Workers', 'Trust:Scientists', 'Trust:Govt.`	Dataset from a study in the US on whether people would get the vaccine before and after having been shown images containing misinformation on

38%2Fs41562-021-01056-1/MediaObjects/41562_2021_1056_MOESM1_ESM.pdf For some documentation	Websites', 'Trust:Social Media', 'Trust:Celebrities', 'Trust:Search Engines', 'Trust:Family and friends', 'Trust:Work Guidelines', 'Trust:Other', 'Trust:None of these', • **Reason:** all boolean values, reason not to get the vaccine if person said they were unsure about getting it or would not get it 'Reason:Unsure if safe', 'Reason:Unsure if effective', 'Reason:Not at risk', 'Reason:Wait until others', Reason:Won't be ill', 'Reason:Other effective treatments', 'Reason:Already acquired immunity', 'Reason:Approval may be rushed', 'Reason:Other', 'Reason:Do not know', • **Misinformation Images:** questions while participant sees each misinformation image (information provided for all 5 images) 'Image n:Vaccine Intent':rank ranging from -2 to 2, whether information shown in image makes participant less or more likely to get vaccine 'Image n:Agreement': rank ranging from -2 to 2, agreement with information shown in image, 'Image n:Trust': rank ranging from -2 to 2, how much do you think the information in image is trustworthy, 'Image n:Fact-check': rank ranging from -2 to 2, how likely are you to fact-check information with other sources, 'Image n:Share': rank ranging from -2 to 2, how likely to share image, 'Image n:Vaccine Intent': rank ranging from -2 to 2, influence of information provided in image on vaccine intent 2: strongly agree, 1: somewhat agree, 0: neither, -1: somewhat disagree, -2: strongly disagree • **Social media related:** 'Social media usage': rank ranging from 1 to 7, how much social media the person uses per day 1: None, 2: Less than 10 minutes per day, 3: 10–30 minutes per day, 4: 31–60 minutes per day, 5:1–2 hours per day, 6: 2–3 hours per day, 7: More than 3 hours per day, 'Seen such online content' {1, 2, 3} whether a person has seen content as in images online (yes, no, do not know), • **Vaccine Intent:** rank ranging from 1 to 4, corresponding to: [1: Yes; 2:unsure, leaning yes; 3: unsure, leaning no; 4: no], pre and post images are shown Would you have the vaccine for yourself if it became available? 'Vaccine Intent for self (Pre)', 'Vaccine Intent for self (Post)', Would you have the vaccine to protect friends, family etc.'Vaccine Intent for others (Pre)', 'Vaccine Intent for others (Post)', • **Personal Information:** 'Age':int (1: 18-24, 2: 25-34,	the vaccines. Contains information on trust in different organizations, social media usage and demographics. Dataset contains 4001 rows, 1704 of which containing at least one NaN value

	3: 35-44, 4: 45-54, 5: 55-64, 6: 65+), `'Gender'`: int (1: Male, 2: Female, 3: other), `'Education'`:int (1: no academic or professional qualifications, 2: O-level, 1 A-level or equivalents, 3: 2+ A-level or equivalents 4: undergraduate, 5: postgraduate or other professional degrees, 5: others), `'Employment'`: int (1: employed, 2: unemployed, 3: student, 4: retired, 5: other), `'Religion'`: int (1: 1: Christian, 2: Jewish, 3: Muslim, 4: Atheist, 5: Other), `'Political'`: int from 1 to 3 (1: Republican, 2: Democrat, 3: Other), `'Ethnicity'`: int from 1 to 5 (1: white, 2: hispanic, 3: black, 4: asian, 5: other), `'Income'`:int (1: under 15,000$, 2: 15,000-34,999, 3: 35,000-54,999, 4: 55,000-94,999, 5: 95,000 and over, 6: prefer not to answer), `'Treatment'`: boolean, whether misinformation or factual images were shown	
orb_uk.csv https://github.com/sloomba/covid19-misinfo/tree/main/dat https://static-content.springer.com/esm/art%3A10.1038%2Fs41562-021-01056-1/MediaObjects/41562_2021_1056_MOESM1_ESM.pdf For some documentation	**Very similar to orb_us.csv, only mentioning the differences here:** ● `'Trust:White House Briefings'`: this column only exists in the US dataset ● `'Political'`: int from 1 to 5 in the UK dataset (1: Conservative, 2: Labour, 3: Lib-Dem, 4: SNP, 5: other) ● `'Ethnicity'`: int from 1 to 4 in this dataset (1: white, 2: black, 3: asian, 4: other) ● `'Education'`:has different levels, int (1: no academic or professional qualifications/nursery or pre-school, 2: high school diploma or GED, 3: 2 year college, 4: 4 year college, 5: postgraduate or other professional degrees, 5: others) ● `"Income"`: has different levels (1: under 15,000£, 2: 15,000-24,999, 3: 25,000-34,999, 4: 35,000-54,999, 5: 55,000 and over, 6: prefer not to answer) More information on values in questionnaire:	Dataset from a study in the UK on whether people would get the vaccine before and after having been shown images containing misinformation on the vaccines. Contains information on trust in different organizations, social media usage and demographics. Dataset contains 4000 rows, 2167 of which containing at least one NaN value

[Susceptibility to misinformation about COVID-19 around the world]()

COVID-19 misinformation_final dataset.csv (https://osf.io/mf7qc/?show=revision) Downloaded: https://drive.google.com/file/d/1TEQbLkTB68u1J4RbKqq60VAIEfhAZ5pU/view?usp=sh	● `Country (string)`:name of country ● `Gender (binary)`:gender, either 'male' or 'female' ● `Age (integer)`:age at time of survey ● `Education (string)`:highest obtained degree ● `Political affiliation (string)`:'Centre right/slightly conservative', 'Middle of the road', 'Centre left/slightly liberal', 'Very left wing/liberal','Left wing/liberal', 'Right wing/conservative' ,'Very right wing/conservative' ● `Compliance (string)`:ranging from 0 to 11 ● `Trust_in_politicians_approach_effectiveness (ordinal integer)`: rank ranging from 1 to 7	Dataset contains information on susceptibility to Covid-related misinformation in studies conducted in April and May 2020. It contains data collected in Ireland, Spain, Mexico, the USA and the UK. It contains ranks for different misinformation types

aring	• `Trust_in_WHO_approach_effectiveness (ordinal integer)`: rank ranging from 1 to 7 • `Trust_in_scientists (ordinal integer)`: rank ranging from 1 to 5 • `Trust_in_journalists (ordinal integer)`: rank ranging from 1 to 5 • `Trust_in_govt (ordinal integer)`: rank ranging from 1 to 5 • `Risk perception (float)`: rank ranging from 1 to 6.17 • `Social_media_exposure(binary)`: • `WHO_media_exposure(binary)`: • `Social_media_info_trust` : rank ranging from 1 to 7 • `WHO_info_trust` : rank ranging from 1 to 7 • `Vaccine_self (binary)`: indicator of whether participant would take the vaccine themself • `misinformation_5g (integer)`: rank ranging from 1 to 7 • `misinformation_breath (integer)`: rank ranging from 1 to 7 • `misinformation_bioengineering (integer)`: rank ranging from 1 to 7 • `misinformation_hot-air (integer)`: rank ranging from 1 to 7 • `misinformation_vaccination (integer)`: rank ranging from 1 to 7 • `Misinformation (float)`: ranges 1 to 7 (7- very reliable, 1-very unreliable)	as well as ranks for trust in different organizations, gender, age, education and political affiliation and an indication on the willingness to get the vaccine. Dataset contains 5000 rows, 4473 of which containing at least one NaN value Documentation: https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.201199

Overcoming COVID-19 vaccination resistance when alternative policies affect the dynamics of conformism, social norms, and crowding out

Replication Data for: "Overcoming COVID-19 vaccination resistance when alternative policies affect the dynamics of conformism, social norms, and crowding out" Downloaded: https://drive.google.com/file/d/1Fvb9JUH2j1U3XAH-e-NPJ8YcPJqZEafd/view?usp=sharing	**Dataset was recorded in two survey waves, variables ending with 1 correspond to wave 1, variables ending with 2 to wave 2** ● `'Vacc_voluntary1'`: participant would voluntarily get vaccine, int between 0 and 4 ● `'Vacc_enforced1'`, participant would get vaccine if enforced, int between 0 and 4 ● `'Trust_gov1'`: int ranging from 1 to 7, trust in government ● `'Trust_fed1'`: int between 1 to 7, trust in federal government ● `'Trust_science1'`: int between 1 to 7, trust in science ● `'Trust_media1'`: int ranging from 1 to 7, trust in media ● `'Cov19_truth_gov1'`: int from 1 to 5, belief that government provides truthful information about covid ● `'Public_trust1'`: int ranging from 1 to 7, trust in public institutions ● `'Female1'`: int, gender ● `'TotalCov19_per100k1'`: float, current number of covid cases in area ● `'Vacc_voluntary2',`: participant would voluntarily get vaccine, int between 0 and 4 ● `'Vacc_enforced2'`: participant would get vaccine if enforced, int between 0 and 4 ● `'Trust_gov2'`, int ranging from 1 to 7, trust in government ● `'Trust_fed2'`, int between 1 to 7, trust in federal government ● `'Trust_science2'`, int between 1 to 7, trust in science ● `'Trust_media2'`,int ranging from 1 to 7, trust in media ● `'Cov19_truth_gov2'`: int from 1 to 5, belief that government provides truthful information about covid ● `'Public_trust2'`, int ranging from 1 to 7, trust in public institutions ● `'Vacc_effective2'`: belief in vaccine effectiveness, int from 1 to 4 ● `'Vacc_freedom2'`, belief that vaccine compromises individual freedom if enforced, int between 0 and 4 ● `'Altruism2'` people's willingness to help others, int from 1 to 7 ● `'Age2'` int, age ● `'FederalState_childhood2'` which federal german state the person grew up in, int from 1 to 16 ● `'East_childhood2'` bool, whether person grew up in east or west germany	Dataset on whether people would get the vaccine voluntarily or if they were forced to in Germany. Contains information on trust in media, science & politics. Dataset contains 2653 rows, 1704 of them containing at least one NaN value

	• `'FederalState_today2'`: which german federal state the person lives in, int from 1 to 16 • `'Female2'`: int, gender • `'High_education2'`: bool, whether person has higher education • `'Household_income2'`: int from 1 to 7 • `'N_household2'`: int, people in household • `'Single_household2'`: bool, whether it is a single household • `'Survey_day2'`: day of survey • `'Cov19_risk_group2'`: bool, person belongs to risk group • `TotalCov19_per100k2'`: float, current number of cases in area • `'Cov19_critical_locally2'`: whether covid status in area is critical, int from 1 to 9	

A.3 Detailed Description of Vaccine Uptake Dataset

Time-series data on vaccination uptake https://covid.ourworldindata.org/data/owid-covid-data.csv	• `'Iso_code'`: ISO 3166-1 alpha-3 – three-letter country codes • `'Continent'`: Continent of the geographical location • `'Location'`: Geographical location • `'Date'`: Date of observation • `'Total_cases'`: Total confirmed cases of COVID-19 • `'New_cases'`: New confirmed cases of COVID-19 • `'New_cases_smoothed'`: New confirmed cases of COVID-19 (7-day smoothed) • `'Total_deaths'`: Total deaths attributed to COVID-19 • `'New_deaths'`: New deaths attributed to COVID-19 • `'New_deaths_smoothed'`: New deaths attributed to COVID-19 (7-day smoothed) • `'Total_cases_per_million'`: Total confirmed cases of COVID-19 per 1,000,000 people • `'New_cases_per_million'`: New confirmed cases of COVID-19 per 1,000,000 people • `'New_cases_smoothed_per_million'`: New confirmed cases of COVID-19 (7-day smoothed) per 1,000,000 people • `'Total_deaths_per_million'`: Total deaths attributed to COVID-19 per 1,000,000 people • `'New_deaths_per_million'`: New deaths attributed to COVID-19 per 1,000,000 people	Time series dataset containing information about the vaccine uptake over the world. Further includes detailed information on current case, death and hospitalization rates and demographics information for the respective country.

	• `'New_deaths_smoothed_per_million'`: New deaths attributed to COVID-19 (7-day smoothed) per 1,000,000 people • `'Reproduction_rate'`: Real-time estimate of the effective reproduction rate (R) of COVID-19 • `'Icu_patients'`: Number of COVID-19 patients in intensive care units (ICUs) on a given day • `'Icu_patients_per_million'`: Number of COVID-19 patients in intensive care units (ICUs) on a given day per 1,000,000 people • `'Hosp_patients'`: Number of COVID-19 patients in hospital on a given day • `'Hosp_patients_per_million’`: Number of COVID-19 patients in hospital on a given day per 1,000,000 people • `'Weekly_icu_admissions'`: Number of COVID-19 patients newly admitted to intensive care units (ICUs) in a given week • `'Weekly_icu_admissions_per_million'`: Number of COVID-19 patients newly admitted to intensive care units (ICUs) in a given week per 1,000,000 people • `'Weekly_hosp_admissions'`: Number of COVID-19 patients newly admitted to hospitals in a given week • `'Weekly_hosp_admissions_per_million’`: Number of COVID-19 patients newly admitted to hospitals in a given week per 1,000,000 people • `'New_tests'`: New tests for COVID-19 (only calculated for consecutive days) • `'Total_tests'`: Total tests for COVID-19 • `'Total_tests_per_thousand'`: Total tests for COVID-19 per 1,000 people • `'New_tests_per_thousand'`: New tests for COVID-19 per 1,000 people • `'New_tests_smoothed'`: New tests for COVID-19 (7-day smoothed) • `'New_tests_smoothed_per_thousand'`: New tests for COVID-19 (7-day smoothed) per 1,000 people • `'Positive_rate'`: The share of COVID-19 tests that are positive, given as a rolling 7-day average • `'Tests_per_case'`: Tests conducted per new confirmed case of COVID-19, given as a rolling 7-day average • `'Tests_units'`: Units used by the location to report its testing data • `'Total_vaccinations'`: Total number of COVID-19 vaccination doses administered • `'People_vaccinated'`: Total number of people who received at least one vaccine dose	

	'People_fully_vaccinated': Total number of people who received all doses prescribed by the vaccination protocol 'Total_boosters': Total number of COVID-19 vaccination booster doses administered (doses administered beyond the number prescribed by the vaccination protocol) 'New_vaccinations': New COVID-19 vaccination doses administered (only calculated for consecutive days) 'New_vaccinations_smoothed': New COVID-19 vaccination doses administered (7-day smoothed) 'Total_vaccinations_per_hundred': Total number of COVID-19 vaccination doses administered per 100 people in the total population 'People_vaccinated_per_hundred': Total number of people who received at least one vaccine dose per 100 people in the total population 'People_fully_vaccinated_per_hundred': Total number of people who received all doses prescribed by the vaccination protocol per 100 people in the total population 'Total_boosters_per_hundred': Total number of COVID-19 vaccination booster doses administered per 100 people in the total population 'New_vaccinations_smoothed_per_million': New COVID-19 vaccination doses administered (7-day smoothed) per 1,000,000 people in the total population 'Stringency_index': Government Response Stringency Index: composite measure based on 9 response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest response) 'Population': Population in 2020 'Population_density': Number of people divided by land area, measured in square kilometers, most recent year available 'Median_age’: Median age of the population, UN projection for 2020 'Aged_65_older': Share of the population that is 65 years and older, most recent year available 'Aged_70_older': Share of the population that is 70 years and older in 2015 'Gdp_per_capita’: Gross domestic product at purchasing power parity (constant 2011 international dollars), most recent year available 'Extreme_poverty': Share of the population living in extreme poverty, most recent year available since 2010	

	• 'Cardiovasc_death_rate': Death rate from cardiovascular disease in 2017 (annual number of deaths per 100,000 people) • 'Diabetes_prevalence': Diabetes prevalence (% of population aged 20 to 79) in 2017 • 'Female_smokers': Share of women who smoke, most recent year available • 'Male_smokers': Share of men who smoke, most recent year available • 'Handwashing_facilities': Share of the population with basic handwashing facilities on premises, most recent year available • 'Hospital_beds_per_thousand': Hospital beds per 1,000 people, most recent year available since 2010 • 'Life_expectancy': Life expectancy at birth in 2019 • 'Human_development_index': A composite index measuring average achievement in three basic dimensions of human development—a long and healthy life, knowledge and a decent standard of living. Values for 2019 • 'Excess_mortality_cumulative_absolute': Cumulative difference between the reported number of deaths since 1 January 2020 and the projected number of deaths for the same period based on previous years • 'Excess_mortality_cumulative': Percentage difference between the cumulative number of deaths since 1 January 2020 and the cumulative projected deaths for the same period based on previous years • 'Excess_mortality': Percentage difference between the reported number of weekly or monthly deaths in 2020–2021 and the projected number of deaths for the same period based on previous years • 'Excess_mortality_cumulative_per_million': Cumulative difference between the reported number of deaths since 1 January 2020 and the projected number of deaths for the same period based on previous years, per million people	

A.4 Detailed Description of Fake News Dataset

This section describes the custom fake news dataset scrapped from Poynter.org ("CoronaVirusFacts Alliance - Poynter" 2020).

Custom fake news dataset	• `title:` the title of the news • `rating`: classification of this fake news, categorical • `Country`: country of the release of the fake news • `Date:` date of the release of the fake news	The dataset contains falsehoods that were identified by the("CoronaVirus Facts Alliance - Poynter" 2020) in more than 70 countries from January 2020 to October 2021.

A.5 EDA of Misinformation Studies

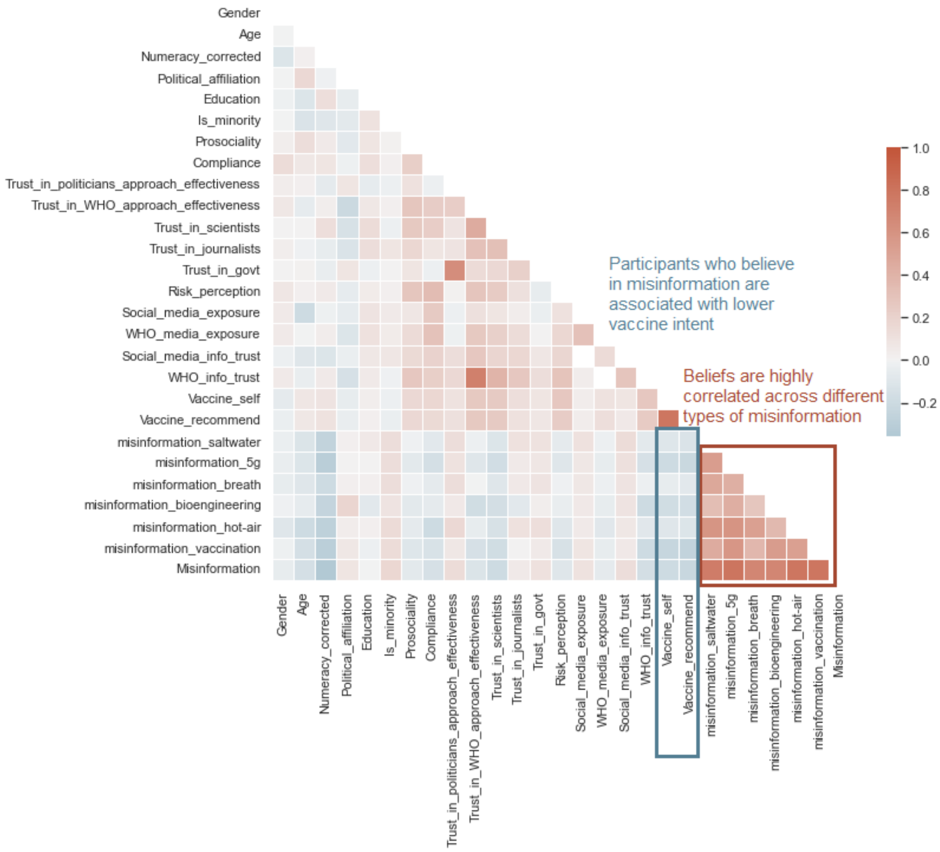
This section we explored the effects of misinformation on an individual level, by exploring the results of three studies (Schmelz and Bowles 2021; Roozenbeek et al. 2020; Loomba et al. 2021). Overall, we show that misinformation is negatively correlated with vaccine intent, while trust is highly correlated. In particular, a misinformation treatment leads to decreasing vaccine intent, no matter which institutions people trust in. Moreover, these studies were conducted before vaccines were approved thus do not indicate how vaccine intent translates into vaccine uptake.

A.5.1 Survey on misinformation and vaccine intent

Based on the dataset collected by Roozenbeek et al. 2020 which covers the COVID-19 misinformation in four countries data (the UK (May), Spain, the USA and Mexico) (Roozenbeek et al. 2020), a rough EDA concludes that:

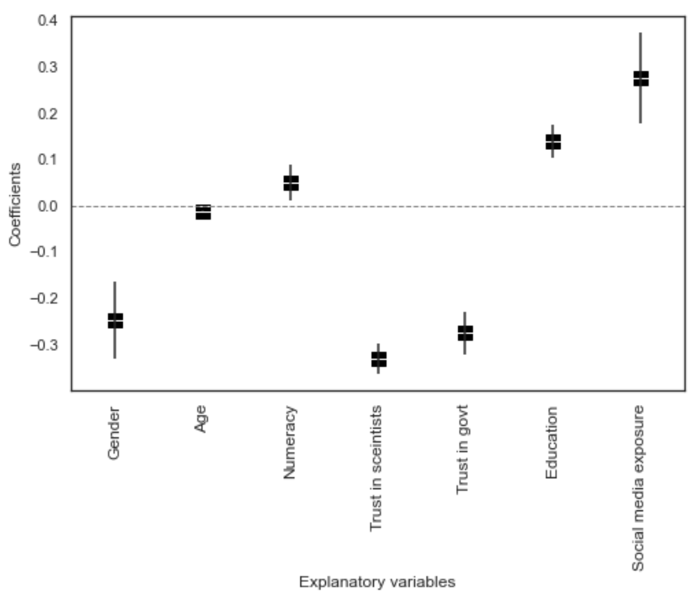
First, participants' beliefs in different misinformation types are highly correlated. To be more specific, participants who believe in '5G networks may be making people more susceptible to the coronavirus' also believe other types of misinformation about COVID-19, such as 'Gargling salt water or lemon juice reduces the risk of infection from coronavirus'.

Figure A.5.1 Beliefs in different misinformation are highly correlated



Second, a number of factors stand out as significant predictors of susceptibility to misinformation across the board. For example, education, and social media exposure are positively correlated with susceptibility to misinformation.

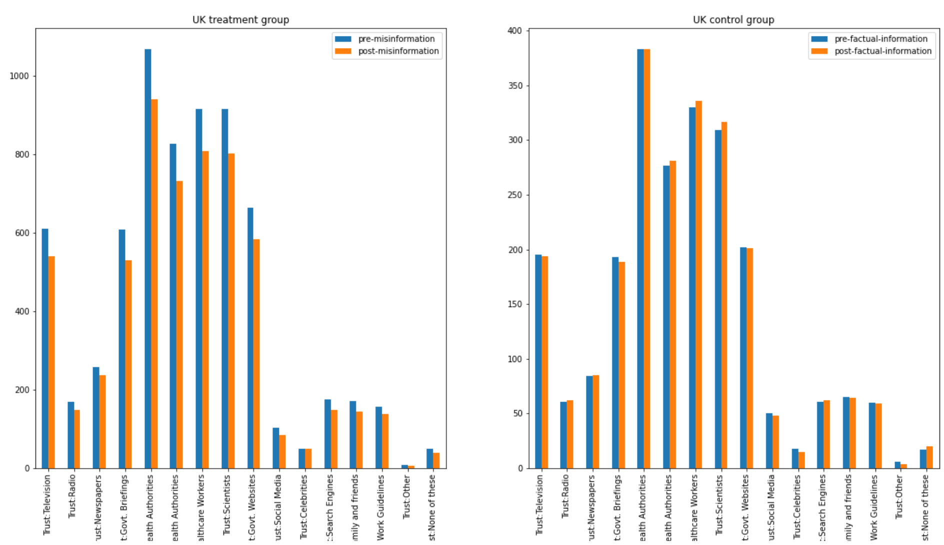
Figure A.5.2 Factors that predict susceptibility to misinformation



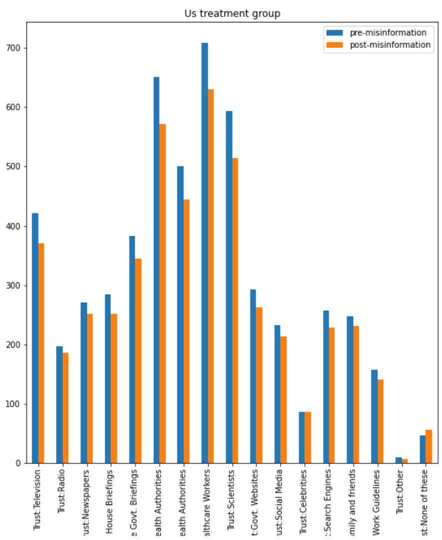
Third, misinformation beliefs have significant impacts on people's vaccine intent. On average, participants who believe in misinformation are associated with lower vaccine intent. This tends to be true across most countries. A simple logistic model confirms this argument.

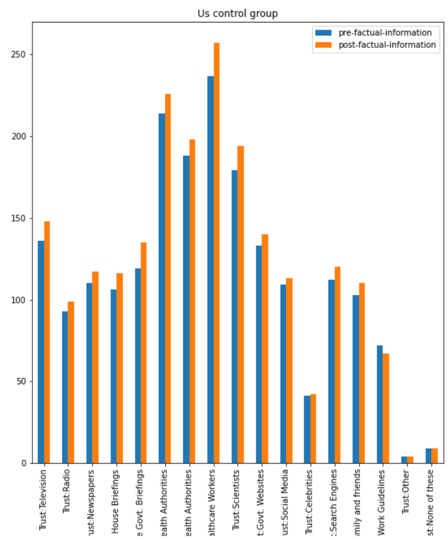
A.5.2 Experiment on misinformation and vaccine intent

(Roozenbeek et al. 2020; Schmelz and Bowles 2021; Loomba et al. 2021) measured the impact of misinformation on the individual vaccine intent in both the US and the UK by exposing study participants to five misinformation images (Treatment group) and comparing their vaccine intent before and after the exposure to the vaccine intent of participants exposed to factual information (Control group). They further record which institutions the participants trust in. The main findings of an EDA are summarized in this section.



A definite **decline in the vaccine intent in both treatment groups** is visible, except for the participants not trusting in any of the given institutions in the US. The factual information the control group participants were exposed to led to an increase in vaccine intent in the US, especially for people trusting in healthcare workers. A slight increase in vaccine willingness in the UK for the control group is also notable, however, this is less clear than in the US.





A.6 General Plots Time Series Decomposition

Figure A.6.1 Steady states for endogenous variables

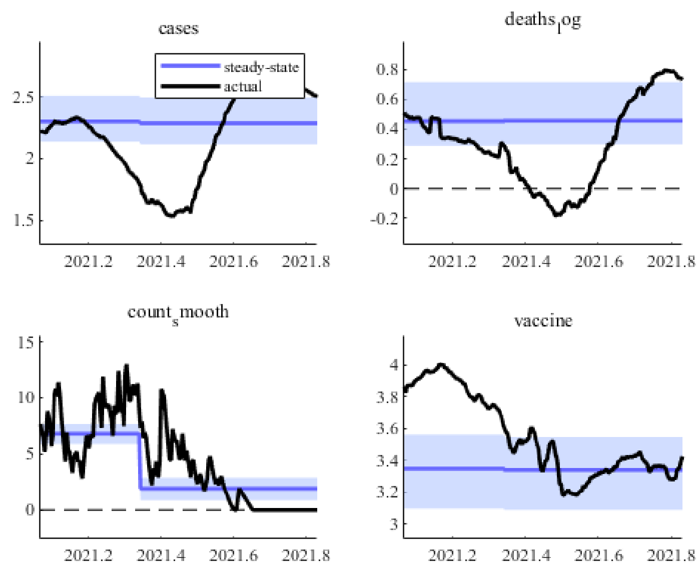


Figure A.6.2 Residuals of the endogenous variables

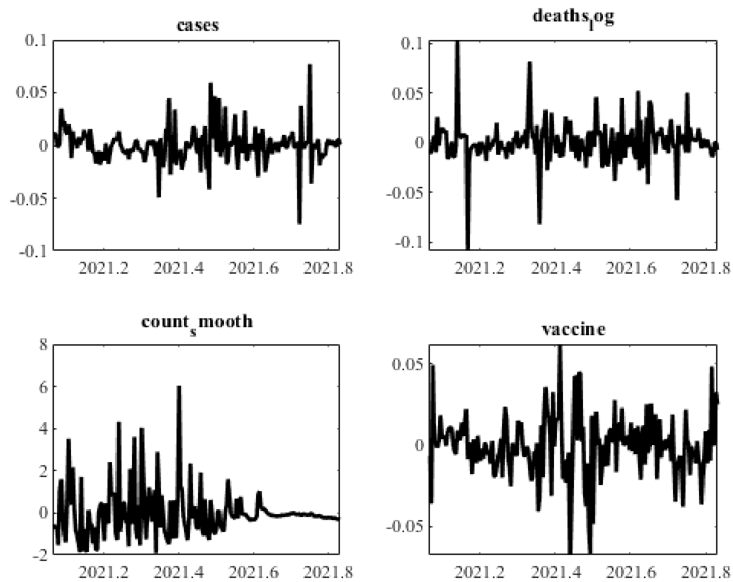


Figure A.6.3 Unconditional forecasts of endogenous variables