Utilizing Concept Drift for Measuring the Effectiveness of Policy Interventions: The Case of the COVID-19 Pandemic

Lucas Baier[1], Niklas Kühl[1,2], Jakob Schöffer[1], Gerhard Satzger[1,2]
[1]Karlsruhe Institute of Technology (KIT)
[2]IBM

Abstract: As a reaction to the high infectiousness and lethality of the COVID-19 virus, countries around the world have adopted drastic policy measures to contain the pandemic. However, it remains unclear which effect these measures, so-called non-pharmaceutical interventions (NPIs), have on the spread of the virus. In this article, we use machine learning and apply drift detection methods in a novel way to measure the effectiveness of policy interventions: We analyze the effect of NPIs on the development of daily case numbers of COVID-19 across 9 European countries and 28 US states. Our analysis shows that it takes more than two weeks on average until NPIs show a significant effect on the number of new cases. We then analyze how characteristics of each country or state, e.g., decisiveness regarding NPIs, climate or population density, influence the time lag until NPIs show their effectiveness. In our analysis, especially the timing of school closures reveals a significant effect on the development of the pandemic. This information is crucial for policy makers confronted with difficult decisions to trade off strict containment of the virus with NPI relief.

Keywords: COVID-19, pandemic, non-pharmaceutical interventions, concept drift, design science research

1 Introduction

Within just a few months in early 2020, the COVID-19 disease caused by a novel coronavirus has evolved into a global pandemic. In order to fight the spread of the pandemic, drastic policy measures with far-reaching implications for basic civil rights have been passed. Politicians have to carefully assess these so-called non-pharmaceutical interventions (NPIs) and gauge their necessity and precise timing. Typical NPIs observed across many nations include a lockdown of public life and the closure of schools. Despite consensus in the academic community on the general effectiveness of a combination of NPIs to mitigate the progression of COVID-19, their concrete timing, duration, scope, and effect have controversially been discussed in public. Therefore, fact-based evidence of NPI effectiveness is imperative for making and justifying these political decisions.

While traditional approaches from the field of epidemiology like SIR or SEIR (McCluskey, 2010) exist, these models in their original form do not include adaptive capabilities which are required for modeling changing reproduction numbers over time (Dottori & Fabricius, 2015). In this article, we introduce a novel strategy to quantify the effectiveness of interventions and illustrate it with NPI and case data from the COVID-19 pandemic: We propose to utilize the kernel theory of *concept drift detection* (Gama et al., 2014) to identify significant changes in the development of the number of infected persons. Concept drift detection is a machine learning technique usually applied to detect changes—so-called *drifts*—in the data-generating environment and trigger retraining of the corresponding model. In the field of information systems (IS), concept drift detection has been applied to, e.g., data stream analysis (Brzezinski & Stefanowski, 2017), business process mining (van Zelst et al., 2019), or innovation monitoring (Mirtalaie et al., 2017). We apply this technique in a novel way, as we do not use it to improve prediction performance on a data stream but to detect significant changes in variables of interest (e.g., COVID-19 case numbers). Specifically, these variables of interest are monitored for drifts that can be related to previous interventions, e.g., of political or medical nature. Following a Design Science Research (DSR) approach, we build an artifact applying concept drift detection for this purpose and illustrate and evaluate it for the pandemic use case.

We show that concept drift detection can be effectively applied to detect sudden changes in a target variable—in this particular case, changes of COVID-19 infection numbers related to NPIs. We base our analysis on data from 9 European countries and 28 US states obtained between 22-01-2020 and 12-05-2020, and we discover that it takes on average more than two weeks until the first NPI shows an effect. We continue to analyze differences between countries and states in light of their characteristics, including demography, medical infrastructure, GDP, climate, and initial reaction time. Our analysis reveals a significant effect of early school closures on the number of days needed to notably reduce case numbers. Furthermore, we simulate scenarios where school closures happen later—as well as the resulting effect on the development of case numbers.

These results can assist policy makers in their tradeoff between strict containments of an infectious virus and NPIs limiting the individual freedom of citizens as well as negatively affecting a country's economy. The remainder of this article is structured as follows: Section 2 introduces the underlying research design, before Section 3 covers related work and introduces the kernel theory of *concept drift*. Section 4 explains the design of our artifact which is subsequently evaluated in Section 5. With the results at hand, we counterfactually discuss the impact of different NPI enactment schedules on pandemic spread in Section 6. We summarize our findings in Section 7, which concludes this work.

2 Research design

As an overall research design, we choose Design Science Research (DSR), as it allows to consider the design-related tasks necessary when building IT artifacts (March & Smith, 1995). Moreover, it has proven to be an important and legitimate paradigm in IS research (Gregor & Hevner, 2013).

We design an artifact capable of detecting the effectiveness of interventions (e.g., policy measures) on a target variable in temporal data. Our artifact is best described as a *method*, as it consists of "actionable instructions that are conceptual" (Peffers et al., 2012, p. 401). By designing and applying our method, we aim to show its *feasibility* (Pries-Heje et al., 2008) as our key evaluation criterion. As a result, our research contributes to the two dimensions of DSR projects (Gregor & Jones, 2007), namely generalizable design knowledge ("knowledge on how to *build* the artifact") as well as a significant impact within the field of application ("knowledge resulting from the *use* of the artifact"). Regarding the former, we design an artifact utilizing concept drift algorithms to detect significant changes in a target variable like the development of case numbers during a pandemic. Therefore, we inform the design of information systems by demonstrating how the kernel theory of *concept drift* can be applied in novel ways. In terms of knowledge contribution according to Gregor & Hevner (2013), the artifact is best described as an *improvement,* as it applies a known solution (concept drift detection) to a novel problem (impact measurement of interventions).

Second, by applying the artifact to the case of NPIs for the COVID-19 pandemic, we demonstrate the effectiveness of policy interventions, providing explicit decision support for policy makers. The overview of our chosen research design as well as the integration into the existing DSR literature is depicted in Table 1.

Table 1. Overview of DSR project characteristics.

Real-world problem	Measure impact of policy interventions on COVID-19 development
Kernel theory	Concept drift detection
Artifact type (**Peffers et al. (2012)**)	Method (actionable insights)
Evaluation objective (**Pries-Heje et al. (2008)**)	Feasibility
Evaluation type (**Venable et al. (2016)**)	Technical risk & efficacy
Contribution (**Gregor & Hevner (2013)**)	Improvement; application of known solution (drift detection) to novel problem (impact measurement)

3 Related work

3.1 Concept Drift

Supervised machine learning fits a mathematical function to map input features to a corresponding, to-be-predicted, target. This function is usually learned by considering historical data as training data. The resulting model can continuously create value when deployed in information systems and delivering ongoing predictions on continuous data streams of new incoming data. However, data streams usually change over time, which also leads to changes in the underlying probability distribution (Tsymbal, 2004). This challenge of changing data over time is usually described as *concept drift* in computer science (Widmer & Kubat, 1996). A concept $p(X,y)$ is defined as the joint probability distribution of a set of features X and the corresponding label y (Gama et al., 2014). The change of a concept over time can be expressed in a mathematical definition as follows:

$$\exists X: p_{t_0}(X, y) \neq p_{t_1}(X, y)$$

Therefore, concept drift is defined as the change in the joint probability distribution between two time points t_0 and t_1. This change may entail that the machine learning model built on previous data in t_0 is no longer suitable for making predictions on new incoming data in t_1. Thus, the occurrence of concept drift requires the application of countermeasures, e.g., the frequent retraining of the underlying machine learning model.

Changes in the incoming data stream can be triggered by a multitude of different internal or external effects. In general, it is intractable to measure all possible confounding factors—which is why these factors cannot be integrated directly into the machine learning model. Those factors are often considered as hidden context of a predictive model (Widmer & Kubat, 1996). The phenomenon of concept drift is usually classified into the following categories (Žliobaitė, 2010): *Abrupt or sudden* concept drift where data structures change very quickly (e.g., a sudden drop in airline traffic during COVID-19), *gradual and incremental* concept drift (e.g., change in customers' buying preferences), or *seasonal and reoccurring* drifts (e.g., ice cream sales in summer). The more fine-grained taxonomy of Webb et al. (2016) also considers other factors, e.g., the magnitude of the drift.

Different approaches for the handling of concept drifts have been proposed: In general, the adaptation strategies for the machine learning model can be split into *blind* and *informed methods* (Gama et al., 2014). Blind adaptation strategies adapt or retrain the prediction model without any explicit drift detection strategy, usually in fixed time intervals (e.g., every month). In contrast, informed methods rely on explicit concept drift detection algorithms which are able to detect concept drifts and trigger a corresponding warning. These drift detection algorithms can further be classified into three categories (Lu et al., 2018): The first category, *error rate-based drift detection*, tracks changes by analyzing the error rate of the prediction model. If the error rate changes significantly over time, a drift alarm is triggered. The second category, *data distribution-based drift detection*, measures the dissimilarity between the distributions of historical data and more recent data. The third category, *multiple hypothesis test drift detection*, combines several techniques of the previous two categories. In general, most approaches belong to the first category (Lu et al., 2018) with Page-Hinkley-Test (Page, 1954) and ADWIN (Bifet & Gavalda, 2007) being two of the most popular algorithms. The Page-Hinkley-Test (PHT) works by continuously monitoring an input variable (e.g., the input data or the prediction accuracy). As soon as the variable differs significantly from its historical average, a change is flagged. ADWIN, in contrast, is a change detector which relies on two detection windows. As soon as the means of those two windows are distinct enough, a change alert is triggered, and the older window is dropped.

In light of the contribution of this article, it is worth noting that all aforementioned methods are usually applied to detect drift in the joint probability distribution of a set of features and a *target variable*. This information is then applied to adapt a supervised machine learning model. In this work, we apply drift detection to identify changes in a *target variable* only and subsequently link those changes back to prior interventions.

3.2 Measuring the spread of pandemics

Modeling and predicting the development of infectious diseases can be performed with different tools, such as compartmental models, agent-based models, or time series and machine learning models (Nsoesie et al., 2014). Compartmental models, such as SIR (Schoenbaum, 1924) or one of its variations, work by dividing the population in compartments based on the disease state (such as susceptible, infected, and immune/dead) and computing rates at which individuals switch between compartments (McCluskey, 2010), typically using Markov chains. Agent-based approaches model the behavior of individuals and their interactions and thereby allow for analyzing the overall transmissions. In contrast, time series and machine learning models rely on past case data and predict future values on that basis.

Regarding the spread of COVID-19, variations of the SIR model have been applied to model and predict the transmission in Hubei and other regions of China by integrating population migration data (Yang et al., 2020). Similar work has been done for India (Pandey et al., 2020). Other approaches use dynamic SIR models to account for changing reproduction numbers following NPIs for different countries (Fanelli & Piazza, 2020). Time series and machine learning approaches are also widely applied, e.g., for predicting the Italian case numbers with exponential curves (Remuzzi & Remuzzi, 2020), or by applying exponential smoothing models (Petropoulos & Makridakis, 2020). The impact of case importation from different areas on transmission rates can be investigated with generalized linear models (Kraemer et al., 2020). Furthermore, neural network-based methods such as LSTMs have been trained on earlier outbreaks of SARS and have been applied to predict the spread of COVID-19 (Yang et al., 2020). An agent-based model originally developed for flu prediction has been used for the analysis of the spread in Singapore (Koo et al., 2020).

3.3 Measuring the effectiveness of NPIs

Many approaches to measuring effects of certain actions are rooted in the field of causal inference (Pearl, 2009). In particular, Structural Causal Models (SCM) have been widely applied by social scientists (Morgan & Winship, 2014), statisticians (Cox & Wermuth, 2004), and epidemiologists (Giles et al., 2011; Robins et al., 2000; Rothman & Greenland, 2005). A Markov chain Monte Carlo approach to investigate the effectiveness of NPIs on the spread of the 2014 Ebola epidemic is proposed by Merler et al. (2015). Another attempt to determining effectiveness of public health measures in mitigating the spread of influenza is used by Lee et al. (2010), following a cohort study approach.

Regarding the effects of different NPIs on COVID-19, Ferguson et al. (2020) apply Bayesian hierarchical models to investigate the impact on the reproduction number. Another approach using a Bayesian hierarchical model is proposed by Flaxman et al. (2020), where the impact of NPIs on death numbers is examined. Other existing work adapts the SIR model to include specific terms accounting for NPIs and their effects. Those models have been used to estimate effects in China (Wang et al., 2020) and other countries (Chen & Qiu, 2020). However, it remains difficult to isolate and compute the specific effect of a particular NPI on the development of the disease as there might be many overlapping effects (Kraemer et al., 2020). Yet another important avenue of existing work is concerned with causal models and counterfactual argumentation, such as Friston et al. (2020) or Pei et al. (2020). In the latter, the authors argue that—under strong assumptions—if the US had started enforcing social distancing measures a week earlier, about 36,000 people less would have died from COVID-19. A database with the latest publications on COVID-19 is provided and constantly updated by the World Health Organization (WHO)*.

4 Artifact design

The design of our artifact is informed by knowledge from the field of concept drift detection. We argue that a change detected by a drift detection algorithm corresponds to the moment in time when an intervention shows its effectiveness in the data, thereby leading to a change in the generation process of the data, i.e., the point in time where a machine learning model would typically be retrained. As stated in Section 3.1, we use drift detection on a *target variable* to identify a substantial change caused by prior interventions.

We select the Page-Hinkley-Test (PHT) as a representative for our kernel theory of concept drift detection, as it is one of the most widely used concept drift detectors (Mitrovic et al., 2018). Based on the PHT, we build an artifact which takes as input a time series data stream as well as corresponding predictions and calculates drifts in the data. The overview of our artifact's workings is depicted in Figure 1.

The artifact calculates the time lag between NPIs and the detected drift for one country or region (output). As input, our artifact requires time series data of infected cases from the respective region as well as a prediction model trained on previous data instances. We fit an exponential smoothing model (Holt, 2004) with a seasonal component and a multiplicative trend. This allows for better modeling the spread of the disease in many different countries/states since it enables us to incorporate *exponential* rather than solely linear trends. Optimal values for the smoothing parameters are determined through grid search.

* https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/

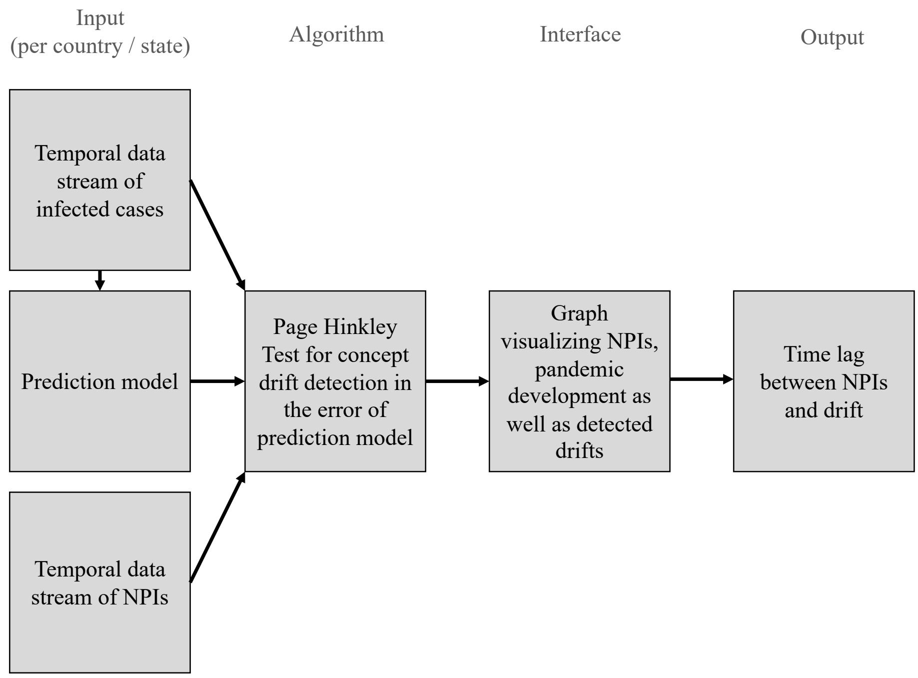

Figure 1. Artifact overview.

As an example, Figure 2 shows the daily case numbers (in grey) as well the predictions (in blue) of the exponential smoothing model for Spain. The figure confirms that the prediction model is well able to accurately represent the exponential development of the case numbers during the unrestricted spread of COVID-19.

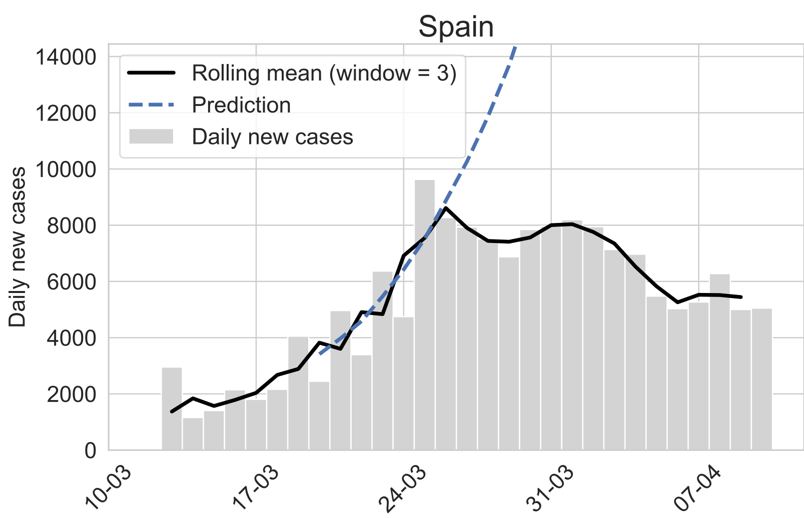

Figure 2. Daily new case numbers as well as corresponding predictions for Spain.

Subsequently, we compute the predictions for future case numbers and compare those predictions with the actual reported case numbers. To that end, we use the symmetric mean absolute percentage error (SMAPE) (Tofallis, 2015). If the error is small, the true and predicted case numbers are similar, which suggests that the pandemic evolves as expected based on the prediction model trained on historical case numbers. A large error, however, indicates that the pandemic is not evolving as predicted. We identify those dates with a significant discrepancy between true and predicted numbers—a concept drift—by applying the PHT on the SMAPE metric. We assume that in case we have identified a drift, it is associated with a change in the general trend of the case numbers. Furthermore, we assume that this change is completely associated with a previously introduced NPI. By relating the drift to the previous NPI's date of adoption, we can determine the number of days until this NPI shows an effect on the spread of the disease. The detected drifts, the development of the case numbers as well as the adoption of NPIs are then visualized within a holistic graph (Figure 3). More details on the precise implementation can be found in the appendix in Table 4.

5 Artifact evaluation

We instantiate our artifact for evaluation purposes with data from the COVID-19 pandemic. We use the daily new infections with COVID-19 between 22-01-2020 and 12-05-2020 as data input. The data is provided by the Center for Systems Science and Engineering at Johns Hopkins University (JHU CSSE, 2020), which also forms the basis for the well-known COVID-19 dashboard (Dong et al., 2020). The data set does not only contain the worldwide case numbers at the country level but also provides information at the county and state level for certain nations such as the US.

We include 9 European countries in our analysis: Austria, Belgium, Germany, Italy, Norway, Spain, Sweden, Switzerland, and UK. Regarding the US, we refrain from a country-wide analysis as the occurrence of infections differs widely across the country. Instead, we perform a more detailed analysis at the state level. However, meaningful case predictions for states with few reported COVID-19 cases are difficult to compute. Therefore, we restrict our analysis to US states with more than 10,000 cumulated cases as of 13-05-2020. This leaves 28 states to be included, with New York, New Jersey, Illinois, Massachusetts, and California being the US states with most COVID-19 cases and Mississippi the least affected one. In total, we consider 37 countries or states in our analysis. Europe NPI data are sourced from Flaxman et al. (2020), US NPI data from Keystone Strategy (2020). We gathered data on mask wearing enactment dates for each country/state individually from national news outlets.

For this analysis, we consider five types of NPIs: *gathering restrictions, social distancing measures, closure of schools, lockdowns* (closure of non-essential services), and *mask wearing*. In general, the discussion about the effectiveness of different NPIs is associated with a large amount of uncertainty (Kraemer et al., 2020). This analysis gets further complicated by the fact that NPIs, even though described by the same name, may differ significantly between countries/states. In "lockdowns", e.g., citizens in Italy were not allowed to leave their apartment for outdoor physical activities whereas in Germany they could still go for a walk with one person from a different household (Deutsche Welle, 2020). This type of lockdown is yet completely different from the lockdown in Wuhan, where public transport was shut down and inhabitants were only allowed outside for grocery shopping a few days a week (Graham-Harrison & Kuo, 2020). Additionally, the considered US states have not issued lockdown orders to date, which is why we use the closure date of non-essential services instead. In the following, we perform two analyses: First, the instantiation of our artifact with an evaluation of the time lag between NPIs and detected drift, as an indicator for the effectiveness of the NPIs. Second, we analyze meta data of each country/state and identify significant features which lead to an earlier or delayed detection of a drift, i.e., characteristics which benefit or harm the spread of the virus.

5.1 Analysis of time lag between NPIs and drift

Figure 3 illustrates the development of daily cases in both Italy and New York since the end of February 2020. Furthermore, we show the relative spread of the pandemic per country/state by computing the number of deaths in relation to its population size. We consider the number of deaths as a reference point since this number is assumed to be a more reliable indicator for the state of the pandemic than the number of infected persons (Flaxman et al., 2020). Therefore, we identify the date with one COVID-19 death per one million inhabitants (red vertical line). For instance, this date is 03-03-2020 for Italy (60 million inhabitants, 60 cumulated deaths are counted on 03-03-2020). The different NPIs are indicated by green vertical lines, e.g., in Italy school closures on 05-03-2020, gathering restrictions and social distancing measures on 09-03-2020, and the following lockdown on 11-03-2020. In blue, the prediction model is shown, which is fitted with the data before 12-03-2020. The drift as indicated by the PHT takes place on 22-03-2020. This allows to compute the difference between the drift date and respective NPIs, e.g., the difference between school closure and drift is 17 days.

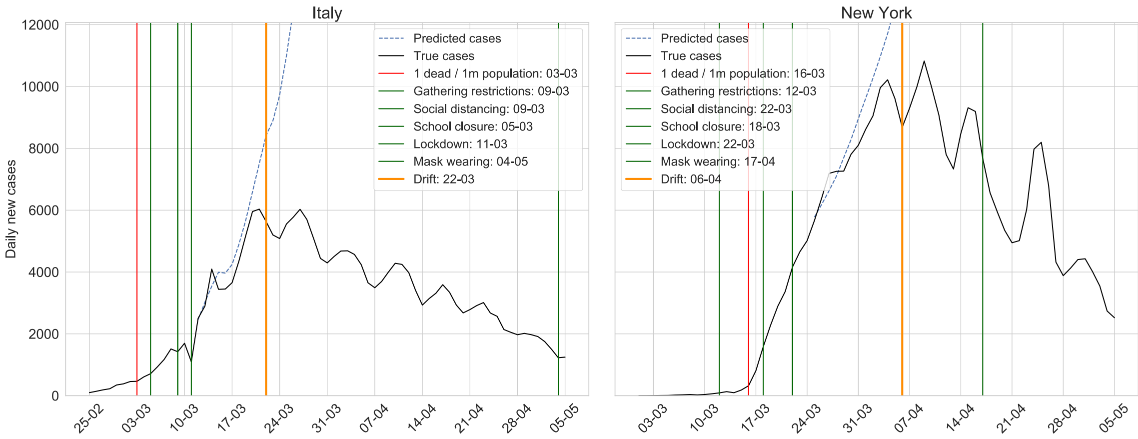

Figure 3. Development of case numbers, prediction models, NPIs, and different drifts in both Italy (left) and New York (right).

For every country/state included in our data set, we perform the same computation as depicted in Figure 3. As a result, we obtain the time lags between the NPI implementations and the time these NPIs show an effect. The mean and standard deviation for these time lags (in days) across all regions are depicted in Table 2. Note that we make the assumption that NPIs do have an effect on the spread of the pandemic—as the public reacts accordingly and adheres to them. That being said, we do not have explicit data on the *degree of adherence* to NPIs, which is why we cannot include this information as a feature in our model to measure its direct effect on the evolution of the pandemic. However, the effect of NPI adherence is implicitly embedded in the data and, therefore, captured by our approach.

Table 2. Overview of time delay between NPIs and drift.

NPI	Time between NPI and detected drift [days]	Standard deviation [days]
Gathering restriction	16.47	5.57
School closures	16.08	3.05
Social distancing	13.42	6.62
Lockdown	8.94	6.05
Mask wearing	-44.48*	29.61

On average, the NPI gathering restriction is the first NPI taken during the course of this pandemic. The mean time between this NPI and the detected drift in the data is 16.47 days, indicating how long it takes for the NPI to affect case numbers of the pandemic. This time span does not only cover the incubation time denoting the time elapsed between exposure to the virus and development of first symptoms—for COVID-19 the median incubation time is estimated to be 5.1 days (Lauer et al., 2020). It also comprises the time between occurrence of symptoms and testing (including decision time, test scheduling, and execution), the time needed to perform the analysis, and the time to report back the results. Often, official reporting of new infections is further delayed over weekends or public holidays when authorities work with less staff and on average report fewer case numbers compared to normal weekdays. Thus, it is plausible that the observed mean time between NPI and drift in the data exceeds the median incubation time by 11 days.

In addition, these results also prove to be in line with related work from experts in the field: Vogel (2020, p. 1) argues that it needs "[…] a minimum of about two weeks to see effects from mitigation measures on new cases." The leading German virologist Christian Dorsten, who has played a major role in Germany's NPI discussions, said that the effect of NPIs will take "at least ten days, rather two weeks" until changes are observable in the data (Hennig & Drosten, 2020). These estimations also lead us to the conclusion that the time lags between NPI adoption and drift detection of school closures (16.08 days) and social distancing (13.42 days) are reasonable.

While the first four NPIs were all established early in the pandemic and actually precede the detected drift (resulting in positive mean times in Table 2), the NPI of mask wearing was on average introduced 44 days *after* the drift triggered by the other four NPIs. This is indicated by a negative time difference between NPI and drift date in Table 2. Specifically, even for *each* individual country/state, the mask wearing NPI was introduced *after* the detected drift—with a time lag ranging from 9 (New Jersey) to 101 days (Switzerland). For a detailed overview, see Table 6 in the appendix*. Thus, mask wearing did not have an effect on the observed NPI-based drift identified in this study.

* Note the negative time difference between the introduction of the mask wearing NPI and the detected drift. As the NPI was introduced after the detected drift, it could not have caused and/or influenced the drift.

* Even the latest drift across all countries/states in our study (UK and New York on 2020-04-06) occurs before any of the countries/states have required mask wearing (starting with New Jersey on 2020-04-10), as Table 6 in the appendix reveals.

However, this does *not* mean that mask wearing is generally ineffective. In fact, several studies have shown that mask wearing—once adopted—is very effective in containing the pandemic (Greenhalgh et al., 2020; Howard et al., 2020).

In general, it is important to keep this time lag between NPIs and the effect on the case statistics in mind. This effect is especially important for decision makers that are currently discussing the relaxation of various NPIs to reduce societal and economic impacts. According to our approach, it takes a significant amount of time until interventions show a measurable influence on case numbers. This argument should also hold true for the opposite direction: Rising case numbers due to lifting of NPIs are only visible after significant time lags.

5.2 Impact of NPIs and country/state characteristics on drift time lag

As indicated by the standard deviations in Table 2, there are differences in the effectiveness of NPIs across various countries and states. Therefore, we perform an additional regression analysis to better understand the underlying reasons for those differences across all countries/states. As dependent variable, we choose the time lag between the date of one death per one million inhabitants (see above) and the drift in case numbers. We need this relative measure as the timing of the adopted NPIs as well as the status of the pandemic varies per country/state, and a purely date-based time series analysis would not account for this. A short time lag indicates that a country/state has reached a drift point in COVID-19 cases early, resulting in a less severe evolution of the pandemic. As for each country/state we obtain data points for NPIs and its case number development, this approach allows us to perform an overarching analysis revealing generalizable insights regarding the effectiveness of different NPIs.

We collect a set of features from different categories which we (a) hypothesize to have an influence on the spread of the pandemic and (b) are publicly available. The decisiveness of a country/state is represented by the *reaction time*, which we define as a feature which measures how early a country/state reacted with their NPIs relative to the specific development of the pandemic. Note that we do not include mask wearing as NPI in this regression analysis: In every country/state, the mask wearing NPI was introduced *after* the detected drift (negative time difference between NPI and drift in Table 2). Therefore, this NPI cannot possibly explain the different time lags across countries/states. Accordingly, we remove this NPI feature from the regression analysis, as it would add unnecessary noise.

We explain the computation of the reaction time in the following: For instance, Italy has introduced school closures (05-03-2020) two days *after* the relative death threshold of one death per one million inhabitants (03-03-2020), resulting in a reaction time of 2 days. Another example is Austria, which has reached one death per one million inhabitants on 21-03-2020. Gathering restrictions in Austria were already introduced on 10-03-2020, which means that this action was taken 11 days *before* the relative death threshold, resulting in a reaction time of -11 days. We compute the reaction time for all countries/states as well as for the NPIs gathering restrictions, social distancing measures, closure of schools, and lockdowns. Since not all countries/states have imposed all four NPIs (e.g., Sweden did not introduce a lockdown), we remove those instances for the following regression analysis, leaving us with 28 countries or states*. Since we assume a

* Austria, Belgium, Germany, Italy, Norway, Spain, Switzerland, United Kingdom, New York, New Jersey, Illinois, Massachusetts, California, Michigan, Texas, Florida, Georgia, Connecticut, Louisiana, Virginia, Ohio, Indiana, Colorado, North Carolina, Wisconsin, Alabama, Missouri, Mississippi

relationship between infections and population density, we collect the *population density per km²* as well as the *share of urban population*. General economic metrics of interest are represented by *GDP per capita in $* and the *Gini coefficient of income distribution*. Furthermore, we gather *healthcare expenditure per capita in $* and the *number of hospital beds per 100,000 inhabitants* to approximate the quality of the health care system. Climate effects are considered by including the *average temperature in March 2020*.

Before fitting a regression model to this data, we analyze the features for multicollinearity. To that end, we compute the Variance Inflation Factor (VIF). This factor measures the variance of a feature's coefficient when the full model is fitted, divided by the variance of that same coefficient if it is the only predictor. We remove all features with a VIF higher than 5, which is a commonly applied threshold (James et al., 2013). The remaining features used for the regression are shown in Table 3. We scale the data to have zero mean and a variance of one before fitting a regression model. The resulting regression coefficients and their standard errors are depicted in Table 3 (R^2 = 0.723 and adj. R^2 = 0.606).

Table 3. Regression coefficients and their standard errors.

Feature	**Coefficient**	**Std err**
Population density [inhabitants per km²]	-0.3234	0.866
Hospital beds [beds per 100,000 inhabitants]	-0.9130	0.817
Urban population [%]	-0.1745	0.988
Avg. temp. 03/2020 [°C]	-0.6218	0.921
Reaction time *gathering restrictions* [days]	1.8841	1.100
Reaction time *school closure* [days]	3.3033**	1.340
Reaction time *social distancing* [days]	0.3335	0.957
Reaction time *lockdown* [days]	-0.6470	0.890
Intercept	12.2857***	0.868

Significance level: *0.1 **0.05 ***0.01

Apart from the intercept, the only significant predictor is the reaction time regarding school closure. The corresponding coefficient is positive, which means that a higher reaction time leads to a higher time lag, as expected. In other words, an early and decisive reaction will likely result in a less severe evolution of the pandemic. It is especially interesting that in our study only school closures have a significant impact on this time lag. This might be an important finding for the current discussions on when and how to reopen schools.

Furthermore, we also test for different interaction terms within the data. For that, we derive interaction terms of size two for all NPIs considered in our study, e.g., the combined effect of school closures and gathering restrictions. This allows us to examine whether it is a combination of NPIs rather than a single one that has an effect on the shift of the drift date. This analysis, however, confirms our previous finding that school closures are the only significant effect included in the data set (R^2 = 0.722, adj. R^2 = 0.559). A best subset selection to identify the best predicting features with the highest adjusted R^2 as selection criterion leads to the same result.

6 Discussion

In this section, we analyze scenarios of different NPI responses to the pandemic in order to quantify the impacts of timely interventions. In addition, we position our results in the context of the current (controversial) discussion around the responses to the pandemic. We discuss limitations due to variations in NPI implementation and compliance and finally interpret the results for health policy makers.

6.1 Scenario analysis

We conduct a scenario analysis to investigate the effect on the pandemic development in case the NPI school closure (the only significant predictor in Table 3) had been introduced at a later time. To measure the effect of later school closures, we perform an additional regression analysis with school closures as the only feature included. Since we do not need to compare the influence of several features, we refrain from standardizing the feature which allows us to directly interpret the influence of shifting school closures on the occurrence of a drift in the case numbers. This leads to a regression coefficient of 0.9951 for school closure reaction time (R^2 = 0.664, adj. R^2 = 0.651, p-value = 0.000). Consequently, shifting school closures by one day will also influence the drift in case numbers to occur approximately one day later. Analogously, a shift of seven days will also lead to a shift of seven days in the corresponding drift.

To analyze the effect of later NPI enactment on the spread of the pandemic, we rely on the machine learning model that has been fitted on the previous case data to estimate the development of case numbers without intervention. Furthermore, we assume that as soon as the shifted drift has occurred, the pandemic follows the same pattern as the real case numbers after the observed drift.

This approach is depicted in Figure 4 by considering the situation in Italy. Due to a shift of school closures by seven days, the drift in case numbers is also shifted by seven days (from 22-03-20 indicated by a black vertical line to 29-03-20 indicated by a grey vertical line). In this time frame, the development of the case numbers is approximated by the previously trained prediction model. The drastic growth in the early stage of the pandemic is obvious. After seven days, the drift occurs, which means that the shifted school closures start to show their effectiveness. From this day on, we assume that the case numbers evolve according to the growth rates that are observed after the drift in real case numbers (depicted by the orange dotted line). Note that the projection (orange dotted line) behaves at the same relative ratio as the true cases (black line) with a shift of seven days. This approach allows us to estimate the number of additional COVID-19 cases if the NPI had been enacted with a delay by considering the difference between the projection and the true case numbers. The corresponding area is marked in light blue.

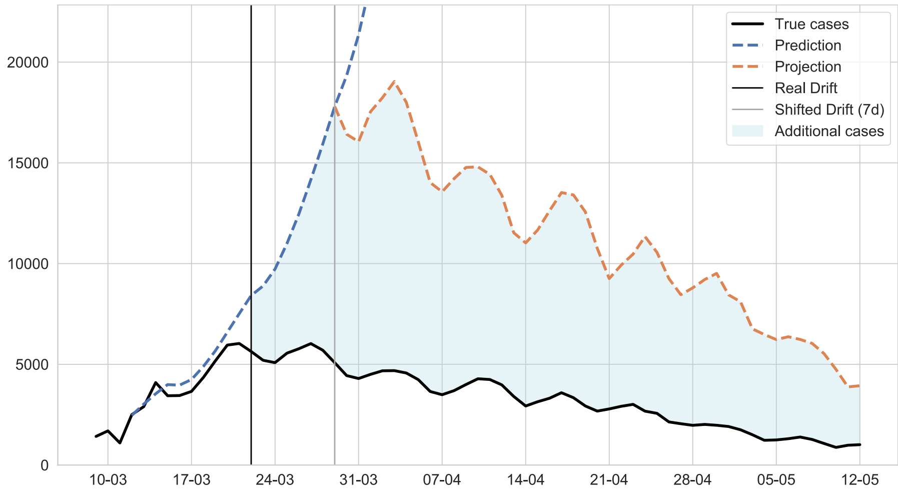

Figure 4. Estimation of additional case numbers in Italy with a seven-day shift of the NPI school closures.

We perform this analysis for each country/state included in our data set and compute the corresponding additional COVID-19 case numbers. Aggregated over all countries and states, a one-day shift for the drift date (caused by closing the schools one day later) would have resulted in an estimated 2.94 million COVID-19 cases over the period considered in this analysis (until 12-05-2020). In comparison, 1.54 million real cases were reported in this time frame according to official numbers. In summary, a one-day shift for the drift date would have led to a total of 1.40 million additional cases, or in other terms a relative increase of 92% in case numbers.

The analysis yields even more drastic results when assuming a seven-day shift of school closures. According to our prediction models, such a scenario would have resulted in a total of 7.13 million COVID-19 cases. This means an additional 5.59 million people would have been infected with COVID-19 or, equivalently, a relative increase of 462% compared to the reported case numbers. The detailed numbers for each country/state included in the analysis can be found in the appendix in Table 5.

This analysis drastically shows how important a timely reaction to the spread of the pandemic is in order to keep the active cases at a manageable level. Even the delaying of NPIs for one additional day would have had drastic consequences on the spread of the virus. Delaying school closures for seven days would have resulted in millions of additional COVID-19 cases and thereby also would have led to tens of thousands additional fatalities. Note that this number is still a conservative estimate as we do not take into account the overstraining of the health care system and corresponding excess mortality.

6.2 Relation to other studies

Many existing studies already aim to investigate the effectiveness of various NPIs on the spread of COVID-19. However, only few of them have been officially accepted in peer-reviewed journals. Many authors have published their preliminary results as preprints, presumably going through review processes. Our work differs from this existing work in two dimensions: methodologically as well as with respect to the results.

Methodologically, first, we use drift detection methods to identify significant drifts or turning points in the temporal development of case numbers—without explicitly considering any NPIs in this first modeling step. Only in a second step do we link those detected drifts back to previously introduced NPIs. This stands in contrast to various studies (Flaxman et al., 2020; Hsiang et al., 2020) that try to model the direct effect of NPIs on the development of COVID-19 case numbers or growth rates. Our approach allows us to measure the time lag between the introduction of an NPI and its effect on the case numbers, i.e., to give an estimation of how long it takes until an NPI reveals its effect in the reported case numbers. This gives stakeholders, e.g., politicians or virologists, the information to decide how long they have to wait until they can judge whether certain measures are effective. Second, we extract the impact of various NPIs via a regression analysis across multiple countries and states while accounting for additional national/state-specific characteristics.

Regarding the results, our research in particular suggests the effectiveness of early school closures as a means to reduce the spread of COVID-19. However, this does not imply that other NPIs—or factors that we did not include in our model—cannot also have a substantial effect on reducing the spread of the pandemic. A New York Times article from August 2020 (Kershner & Belluck, 2020) provides additional evidence that schools play a crucial role in containing the pandemic: It explains how re-opening schools in Israel resulted in "the largest outbreak [of COVID-19] in a single school in Israel, possibly the world […], ultimately infecting hundreds of students, teachers and relatives." While our findings are in line with additional work, e.g., Haug et al. (2020), they are different from, e.g., Banholzer et al. (2020), Flaxman et al. (2020), or Hsiang et al. (2020), who attest a somewhat limited effectiveness of school closures on the spread of the pandemic, compared to other measures such as travel bans, transit suspension, or a national lockdown. However, note that even the findings of Banholzer et al. (2020) compared to Flaxman et al. (2020) are highly inconsistent, e.g., with respect to their suggested effectiveness of gathering restrictions. This emphasizes how challenging it is to assess NPI effectiveness at the present stage of the pandemic.

6.3 NPI implementation and compliance

Throughout this discussion, it is important to keep in mind that all research related to measuring the effectiveness of different NPIs is accompanied by a high degree of uncertainty. This is in part due to different *implementations* of the various NPIs as well as to citizens' *compliance* behavior.

While we analyze impacts of five NPIs, we need to acknowledge that these are not fully comparable across countries/states as we can see, e.g., for gathering restrictions or mask wearing: We modeled gathering restrictions as a binary variable indicating whether people are allowed to gather or not. However, different adaptations with regard to gathering restrictions have been implemented—ranging from cancellation of cultural and sports events to a strict ban on any meetings with more than five people involved, even for religious gatherings (Doogan et al., 2020). Similarly, the implementation "degree" of mask wearing varies significantly across countries/states: For instance, Austria limited the mask obligation to public transport and to buildings related to the medical domain (Beer, 2020). In contrast, other countries/states rigidly require mask wearing even in outside public spaces or for children during school visits (Leffler et al., 2020). The change of testing policies over time may additionally dilute the analysis (Flaxman et al., 2020).

But even with identical implementations, NPI compliance by the public varies greatly across countries/states: Adherence is observed to be high in Asian countries and significantly lower in some European or North American states (Leffler et al., 2020). Reasons may be manifold: Stringent enforcement, higher cultural proneness to obedience, or even familiarity with the NPI (like mask wearing in Asia) may enhance compliance.

Due to those behavioral differences, we assume that the effect of NPIs with heterogeneous implementation and/or compliance would vary significantly across countries/states. To some extent we observe this in Table 2: A well-defined and enforceable school closure NPI shows a notably lower standard deviation of the time to impact than the other NPIs. We may speculate that school closures are implemented more consistently, and their compliance is more easily enforced—resulting in a more homogenous time to impact across countries/states, as opposed to “fuzzier” NPIs like gathering restrictions (showing higher standard deviations in Table 2). In fact, the variance in implementation and compliance would add noise to our variables in the analysis. We assume the same would hold true for the mask wearing NPI if it had been implemented before the detected drift: Mask wearing regulations significantly differ across countries, and monitoring and policing the wearing of masks to ensure adherence is difficult—at least without widespread camera surveillance (Kühl et al., 2020). Therefore, the time between the enactment of a mask wearing NPI and a drift will be very heterogenous across countries (resembling a variable with random distribution).

6.4 Advice to policy makers

As of November 2020, many countries are still confronted with a second wave of COVID-19 infections. Our analysis and results may inform health policy makers on a national as well as a supranational level.

First, on a national (country/state) level, policy makers may use the introduced concept drift detection methodology to identify significant turning points in the development of the pandemic in their specific territory. Such an analysis requires a reference point from which the model is trained as well as available data on the cases within this country as a prerequisite. It is then possible to quantify the time difference between an introduced NPIs and the effect, i.e., the concept drift that is detected in the data. This effectiveness measure can be used both for *individual NPIs* or a *set of NPIs*. However, at this level, our method does not allow to *isolate* the effectiveness of each NPI *within a set of NPIs*—it can only provide time spans.

Second, on a supranational level (e.g., from a WHO perspective), it is possible to perform an analysis across individual countries or states to gain insights on the individual effectiveness of NPIs within comprehensive NPI ensembles. Our regression analysis drawing on multiple country/state data demonstrates this: Within the set of applied NPIs, school closures are identified as statistically significant effective measures to contain the pandemic—an interesting result given frequent exclusion of school closures in later 2020 lockdown phases (Eddy, 2020).

For this analysis, however, policy makers need to be aware that insights on the supranational level are based on the assumption that NPIs are implemented and adhered to in a similar fashion across countries/states. As this may not be the case for some of the “fuzzier” NPIs like gathering restrictions or mask wearing (as discussed in Section 6.3), we may miss out to detect other NPIs as effective ones. Possible solutions could draw on larger international databases trying to work with clusters of countries/states that are more comparable in implementation and adherence.

7 Conclusion

The COVID-19 pandemic poses many challenges to politics and society. Especially the adopted policy measures to control the spread have been subject to heated debates. In our article, we have motivated and introduced a novel approach towards analyzing the effectiveness of different interventions in the fight of pandemics. Specifically, we propose to utilize the kernel theory of *concept drift detection* (Gama et al., 2014) to measure the time difference between the introduction of certain non-pharmaceutical interventions (NPIs) and a significant change (so-called *drift*) in the number of infected cases. To evaluate our approach, we instantiate the proposed artifact based on actual data from the COVID-19 pandemic with the goal to evaluate the general feasibility (Pries-Heje et al., 2008).

Our analysis shows that the detected amount of time between the first adopted NPI and drift amounts to 16 days on average across 37 countries/states—which is in line with experts' opinions and indicates that our approach is valid (Hennig & Drosten, 2020; Vogel, 2020). Furthermore, we analyze characteristics—including demography, medical infrastructure, GDP, and climate—of each country/state in relationship to the detected NPI effectiveness. This analysis shows that the NPI of school closure has a significant impact on the development of the COVID-19 pandemic. Specifically, the longer decision makers wait to impose school closures, the longer it takes until our method observes a significant impact on the spread of the virus. This finding should be of special interest for policy makers, as the discussion on the effectiveness of NPIs has been very controversial—our results suggest the effectiveness of timely school closures. According to our results, just a one-day shift of the drift date (caused by later school closures) would have resulted in 1.40 million additional COVID-19 infections and a seven-day shift even in a total of 5.59 million additional cases in the regarded countries/states.

The work at hand contributes to the body of knowledge in two meaningful ways: First, we generate generalizable design knowledge in a DSR project. Our research demonstrates the successful application of concept drift detection as a design principle and illustrates how it can be applied in novel ways, thus informing the design of innovative information systems: We demonstrate how concept drift can produce novel insights on the effectiveness of interventions. Specifically, we apply drift detection to identify significant changes in a *target variable* and subsequently relate those drifts back to prior interventions. Second, by applying the artifact to the case of NPI and infection data during the COVID-19 pandemic, i.e., our area of application, we analyze the effectiveness of policy interventions, providing explicit decision support. This can and should guide policy makers as they balance their options.

Future work needs to incorporate more data from additional countries/states and from other continents (e.g., Africa). Additionally, applications to other pandemics—or completely different applications of intervention effectiveness measurement—would be interesting. Examples could range, among others, from traffic regulations (e.g., speed limits), measuring the time until significant changes in traffic data (e.g., traffic fatalities) are observable, up to environmental regulations (e.g., CO_2 restrictions) and their effectiveness to combat climate change.

For the COVID-19 crisis, however, our evidence suggests that the introduction of NPIs is effective for a successful management of the pandemic—until an effective vaccine is introduced.

Appendix

Table 4. Implementation details of artifact.

Component	Details
Prediction model	Exponential smoothing model with additive seasonal component and multiplicative trend. This allows to model exponential trends.
Seasonality	Seven-day seasonality, as many countries/states exhibit a weekly pattern in their reporting of case numbers.
Grid search	We determine the optimal values for the model smoothing parameters for level, slope, and season by performing a grid search in the range (0.1, 0.2, … 0.9) and testing those values on the three days following the training set.
Training data	We use all case numbers up to seven days after the first NPI as training data. Due to the characteristics of the pandemic, it is impossible that NPIs already show an effect during this time window.
PHT parameters	Threshold = 0.3 Minimum number of instances = 3

Table 5. Overview of countries/states with additional positive case numbers based on a shift in school closures.

Country/State	Drift date	Real case numbers*	Projection with shift of 1 day	Additional case numbers with shift of 1 day	Projection with shift of 7 day	Additional case numbers with shift of 7 days
Austria	28-03-2020	8,334	22,744	14,410	65,855	57,521
Germany	30-03-2020	111,056	236,870	125,813	509,708	398,651
Italy	22-03-2020	167,799	271,507	103,708	585,554	417,755
Spain	29-03-2020	135,509	296,173	160,664	700,653	565,144
United Kingdom	06-04-2020	179,303	292,761	113,457	496,932	317,629
Belgium	31-03-2020	41,899	72,732	30,832	149,436	107,537
Switzerland	27-03-2020	18,500	27,401	8,900	42,669	24,168
New York	06-04-2020	215,797	370,874	155,076	607,383	391,585
New Jersey	01-04-2020	121,698	266,128	144,430	828,857	707,158
Illinois	01-04-2020	76,230	137,208	60,978	329,269	253,038
Massachusetts	31-03-2020	73,654	219,684	146,029	1,021,075	947,420
California	05-04-2020	57,743	102,584	44,840	190,866	133,123
Michigan	31-03-2020	41,454	79,471	38,017	274,194	232,739
Texas	05-04-2020	35,165	49,217	14,052	87,621	52,455
Florida	05-04-2020	30,383	68,848	38,464	157,068	126,684
Connecticut	31-03-2020	31,753	77,699	45,946	338,063	306,309
Louisiana	29-03-2020	28,976	74,641	45,665	233,579	204,603
Virginia	04-04-2020	23,830	44,763	20,933	119,239	95,409
Ohio	31-03-2020	23,321	39,494	16,173	91,645	68,324
Indiana	09-04-2020	19,141	32,949	13,808	47,456	28,315
Colorado	08-04-2020	14,748	47,250	32,502	54,743	39,995
North Carolina	30-03-2020	14,574	28,092	13,517	62,734	48,159
Wisconsin	12-04-2020	7,436	11,067	3,631	11,120	3,684
Alabama	01-04-2020	9,460	18,887	9,426	40,487	31,026
Missouri	13-04-2020	5,979	11,616	5,637	14,960	8,981
Mississippi	05-04-2020	8,407	13,611	5,204	21,300	12,893
Sum	-	1,538,932	2,941,220	1,402,288	7,128,020	5,589,088

*** Case numbers between drift date and 12-05-2020**

Table 6. Overview of the mask wearing NPI analysis.

Country/State	Detected drift	Introduction of mask NPI	Difference in days between drift and introduction of mask NPI
Austria	2020-03-28	2020-04-14	17 days *after* detected drift
Belgium	2020-03-31	2020-05-04	34 days *after* detected drift
Switzerland	2020-03-27	2020-07-06	101 days *after* detected drift
Germany	2020-03-30	2020-04-27	28 days *after* detected drift
Spain	2020-03-29	2020-05-21	53 days *after* detected drift
United Kingdom	2020-04-06	2020-06-15	70 days *after* detected drift
Italy	2020-03-22	2020-05-04	43 days *after* detected drift
California	2020-04-05	2020-06-18	74 days *after* detected drift
Connecticut	2020-03-31	2020-04-20	20 days *after* detected drift
Illinois	2020-04-01	2020-05-01	30 days *after* detected drift
Maryland	2020-03-30	2020-04-18	19 days *after* detected drift
Massachusetts	2020-03-31	2020-05-06	36 days *after* detected drift
Michigan	2020-03-31	2020-04-26	26 days *after* detected drift
New Jersey	2020-04-01	2020-04-10	9 days *after* detected drift
New York	2020-04-06	2020-04-17	11 days *after* detected drift
North Carolina	2020-03-30	2020-06-26	88 days *after* detected drift
Pennsylvania	2020-03-29	2020-04-19	21 days *after* detected drift
Rhode Island	2020-03-30	2020-04-18	19 days *after* detected drift
Texas	2020-04-05	2020-07-03	89 days *after* detected drift
Virginia	2020-04-04	2020-05-29	55 days *after* detected drift
Washington	2020-03-27	2020-06-26	91 days *after* detected drift

References

Banholzer, N., van Weenen, E., Kratzwald, B., Seeliger, A., Tschernutter, D., Bottrighi, P., Cenedese, A., Salles, J. P., Feuerriegel, S., & Vach, W. (2020). Estimating the impact of non-pharmaceutical interventions on documented infections with COVID-19: A cross-country analysis. *MedRxiv*.

Beer, A. (2020). *Österreich lässt Mitte Juni die Maske fallen*. https://www.tagesschau.de/ausland/oesterreich-coronavirus-lockerungen-101.html, accessed at: 21.09.2020

Bifet, A., & Gavalda, R. (2007). Learning from time-changing data with adaptive windowing. *Proceedings of the 2007 SIAM International Conference on Data Mining*, 443–448.

Brzezinski, D., & Stefanowski, J. (2017). Prequential AUC: Properties of the area under the ROC curve for data streams with concept drift. *Knowledge and Information Systems*, *52*(2), 531–562.

Chen, X., & Qiu, Z. (2020). Scenario analysis of non-pharmaceutical interventions on global COVID-19 transmissions. *ArXiv Preprint ArXiv:2004.04529*.

Cox, D. R., & Wermuth, N. (2004). Causality: A statistical view. In *International Statistical Review* (Vol. 72, Issue 3, pp. 285–305). https://doi.org/10.1111/j.1751-5823.2004.tb00237.x

Deutsche Welle. (2020). *Coronavirus: What are the lockdown measures across Europe?* https://www.dw.com/en/coronavirus-what-are-the-lockdown-measures-across-europe/a-52905137, date accessed: 2020-04-28

Dong, E., Du, H., & 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., & 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.

Dottori, M., & Fabricius, G. (2015). SIR model on a dynamical network and the endemic state of an infectious disease. *Physica A: Statistical Mechanics and Its Applications*, *434*, 25–35.

Eddy, M. (2020). *Why Is Europe Keeping Its Schools Open, Despite New Lockdowns?* https://www.nytimes.com/2020/10/29/world/europe/schools-coronavirus-europe-lockdowns.html, date accessed: 2020-11-23

Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals*, *134*, 109761.

Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunuba Perez, Z., Cuomo-Dannenburg, G., & others. (2020). *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand*.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., & others. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*, 257–361.

Friston, K. J., Parr, T., Zeidman, P., Razi, A., Flandin, G., Daunizeau, J., Hulme, O. J., Billig, A. J., Litvak, V., Moran, R. J., Price, C. J., & Lambert, C. (2020). Dynamic causal modelling of COVID-19. *Wellcome Open Research*, *5*, 89. https://doi.org/10.12688/wellcomeopenres.15881.2

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, *46*(4), 1–37.

Giles, L. V., Brauer, M., Barn, P., Künzli, N., Romieu, I., Mittleman, M. A., van Eeden, S., Allen, R., Carlsten, C., Stieb, D., Noonan, C., Smargiassi, A., Kaufman, J. D., Hajat, S., Kosatsky, T., & Brauer, M. (2011). From good intentions to proven interventions: Effectiveness of actions to reduce the health impacts of air pollution. *Environmental Health Perspectives*, *119*(1), 29–36.

Graham-Harrison, E., & Kuo, L. (2020). *China's coronavirus lockdown strategy: Brutal but effective*. https://www.theguardian.com/world/2020/mar/19/chinas-coronavirus-lockdown-strategy-brutal-but-effective, date accessed: 2020-04-28

Greenhalgh, T., Schmid, M. B., Czypionka, T., Bassler, D., & Gruer, L. (2020). Face masks for the public during the COVID-19 crisis. *The BMJ*.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximim impact. *MIS Quarterly*, *37*(2), 337–355.

Gregor, S., & Jones, D. (2007). The anatomy of a design theory. *Association for Information Systems*, *8*(5).

Haug, N., Geyrhofer, L., Londei, A., Dervic, E., Desvars-Larrive, A., Loreto, V., Pinior, B., Thurner, S., & Klimek, P. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. *MedRxiv*, 2020.07.06.20147199. https://doi.org/10.1101/2020.07.06.20147199

Hennig, K., & Drosten, C. (2020). *There is glory in prevention*. NDR Info Corona Update, Episode 48. https://www.ndr.de/nachrichten/info/coronaskript210.pdf, date accessed: 2020-06-26

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10.

Howard, J., Huang, A., Li, Z., Tufekci, Z., Zdimal, V., van der Westhuizen, H.-M., von Delft, A., Price, A., Fridman, L., & Tang, L.-H. (2020). Face masks against COVID-19: An evidence review. *Preprints*. https://www.preprints.org/manuscript/202004.0203/v1?fbclid=IwAR0h7PBSAB6ZEcr-DzBaTTIEV9kjvJiRZA7Eassb-rs75raKtOKIVKWcsFk

Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., Druckenmiller, H., Huang, L. Y., Hultgren, A., Krasovich, E., Lau, P., Lee, J., Rolf, E., Tseng, J., & Wu, T. (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, *584*, 262–267. https://doi.org/10.1038/s41586-020-2404-8

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

JHU CSSE. (2020). *Novel coronavirus (COVID-19) cases*. https://github.com/CSSEGISandData/COVID-19/, date accessed: 2020-05-15

Kershner, I., & Belluck, P. (2020). *When Covid subsided, Israel reopened its schools. It didn't go well.* https://www-nytimes-com.cdn.ampproject.org/c/s/www.nytimes.com/2020/08/04/world/middleeast/coronavirus-israel-schools-reopen.amp.html, date accessed: 2020-08-26

Keystone Strategy. (2020). *Coronavirus city and county non-pharmaceutical intervention rollout date dataset*. https://www.keystonestrategy.com/coronavirus-covid19-intervention-dataset-model/, date accessed: 2020-04-26

Koo, J. R., Cook, A. R., Park, M., Sun, Y., Sun, H., Lim, J. T., Tam, C., & Dickens, B. L. (2020). Interventions to mitigate early spread of SARS-CoV-2 in Singapore: A modelling study. *The Lancet Infectious Diseases*, *20*(6), 678–688.

Kraemer, M. U. G., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., du Plessis, L., Faria, N. R., Li, R., Hanage, W. P., & others. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*(6490), 493–497.

Kühl, N., Martin, D., Wolff, C., & Volkamer, M. (2020). "Healthy surveillance": Designing a concept for privacy-preserving mask recognition AI in the age of pandemics. *ArXiv Preprint ArXiv:2010.12026*.

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, *172*(9), 577–582.

Lee, V. J., Yap, J., Cook, A. R., Chen, M. I., Tay, J. K., Barr, I., Kelso, A., Tan, B. H., Loh, J. P., Lin, R., Cui, L., Kelly, P. M., Leo, Y. S., Chia, K. S., Kang, W. L., Tambyah, P. A., & Seet, B. (2010). Effectiveness of public health measures in mitigating pandemic influenza spread: A prospective sero-epidemiological cohort study. *The Journal of Infectious Diseases*, *202*(9), 1319–1326.

Leffler, C. T., Ing, E. B., Lykins, J. D., Hogan, M. C., McKeown, C. A., & Grzybowski, A. (2020). Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks. *MedRxiv*.

Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2018). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, *31*(12), 2346–2363.

March, S. T., & Smith, G. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266.

McCluskey, C. C. (2010). Complete global stability for an SIR epidemic model with delay—distributed or discrete. *Nonlinear Analysis: Real World Applications*, *11*(1), 55–59.

Merler, S., Ajelli, M., Fumanelli, L., Gomes, M. F. C., Piontti, A. P. y., Rossi, L., Chao, D. L., Longini, I. M., Halloran, M. E., & Vespignani, A. (2015). Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: A computational modelling analysis. *The Lancet Infectious Diseases*, *15*(2), 204–211. https://doi.org/10.1016/S1473-3099(14)71074-6

Mirtalaie, M. A., Hussain, O. K., Chang, E., & Hussain, F. K. (2017). A decision support framework for identifying novel ideas in new product development from cross-domain analysis. *Information Systems*, *69*, 59–80.

Mitrovic, T., Xue, B., & Li, X. (2018). AI 2018: Advances in artificial intelligence. *31st Australasian Joint Conference, Wellington, New Zealand, December 11-14, 2018*.

Morgan, S. L., & Winship, C. (2014). Counterfactuals and causal inference. In *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press. https://doi.org/10.1017/CBO9781107587991

Nsoesie, E. O., Brownstein, J. S., Ramakrishnan, N., & Marathe, M. V. (2014). A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and Other Respiratory Viruses*, *8*(3), 309–316.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*(1/2), 100–115.

Pandey, G., Chaudhary, P., Gupta, R., & Pal, S. (2020). SEIR and regression model based COVID-19 outbreak predictions in India. *ArXiv Preprint ArXiv:2004.00958*.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, *3*, 96–146. https://doi.org/10.1214/09-SS057

Peffers, K., Rothenberger, M., Tuunanen, T., & Vaezi, R. (2012). Design science research evaluation. *Design Science Research in Information Systems. Advances in Theory and Practice*, 398–410. https://doi.org/10.1007/978-3-642-29863-9_29

Pei, S., Kandula, S., & Shaman, J. (2020). Differential effects of intervention timing on COVID-19 spread in the United States. *MedRxiv*. https://doi.org/10.1101/2020.05.15.20103655

Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PlOS One*, *15*(3).

Pries-Heje, J., Baskerville, R., & Venable, J. R. (2008). Strategies for design science research evaluation. *ECIS*, 255–266.

Remuzzi, A., & Remuzzi, G. (2020). COVID-19 and Italy: What next? *The Lancet*, *395*(10231), 1225–1228.

Robins, J. M., Hernán, M. Á., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550–560. https://doi.org/10.1097/00001648-200009000-00011

Rothman, K. J., & Greenland, S. (2005). Causation and causal inference in epidemiology. In *American Journal of Public Health* (Vol. 95, Issue SUPPL. 1). https://doi.org/10.2105/AJPH.2004.059204

Schoenbaum, E. (1924). A contribution to the mathematical theory of pension insurance. *Časopis pro Pěstování Matematiky a Fysiky*, *053*(1), 163–173. https://doi.org/10.21136/cpmf.1924.109357

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, *66*(8), 1352–1362.

Tsymbal, A. (2004). The problem of concept drift: Definitions and related work. In *Computer Science Department, Trinity College Dublin* (Vol. 106, Issue 2).

van Zelst, S. J., Sani, M. F., Ostovar, A., Conforti, R., & La Rosa, M. (2019). Detection and removal of infrequent behavior from event streams of business processes. *Information Systems*, *90*.

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems*, *25*(1), 77–89.

Vogel, L. (2020). How long will social distancing take to work? Experts weigh in on Canada's COVID-19 response. *CMAJ*, *192*(14), E382--E383. https://doi.org/10.1503/cmaj.1095857

Wang, C., Liu, L., Hao, X., Guo, H., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., Pan, A., & others. (2020). Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in Wuhan, China. *MedRxiv*.

Webb, G. I., Hyde, R., Cao, H., Nguyen, H. L., & Petitjean, F. (2016). Characterizing concept drift. *Data Mining and Knowledge Discovery*, *30*(4), 964–994. https://doi.org/10.1007/s10618-015-0448-4

Widmer, G., & Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine Learning*, *23*(1), 69–101. https://doi.org/10.1007/BF00116900

Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., & others. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, *12*(3), 165–174.

Žliobaitė, I. (2010). Learning under concept drift: An overview. *ArXiv Preprint ArXiv:1010.4784*. http://arxiv.org/abs/1010.4784