Stockholm Institute of Transition Economics

WORKING PAPER

April 2021

No. 55

Assessing Alternative Indicators for Covid-19 Policy Evaluation, with a Counterfactual for Sweden

Chiara Latour, Franco Peracchi and Giancarlo Spagnolo

Assessing Alternative Indicators for Covid-19 Policy Evaluation, with a Counterfactual for Sweden*

Chiara Latour† Franco Peracchi‡ Giancarlo Spagnolo§

April 3, 2021

Abstract

We compare different indicators of the spread and consequences of the COVID-19 pandemic, developing a novel method to adjust daily COVID-19 deaths to match weekly excess mortality. Focusing on Sweden, the only country that has good data and did not impose a lockdown, we construct counterfactuals for what would have happened if it had imposed a lockdown, using a synthetic control method. Correcting for data problems and optimizing the synthetic control for each indicator considered, we find stronger effects than previously estimated. Most importantly, studying the ratio of positives to the number of tests, we find that a lockdown would have had sizable effects already after a week. The 3–4 weeks delay highlighted in previous studies appears mainly driven by the large changes in testing frequency that occurred in Sweden during the period considered.

Keywords: COVID-19 indicators; excess deaths; COVID-19 deaths; containment policies; lockdown; synthetic control method; Sweden.

JEL Codes: C4, F6, I18.

*We thank Stanley Cho for sharing his data and code, and Svante Strömberg for excellent research assistance.

† University of Stockholm.

‡ University of Rome Tor Vergata and EIEF.

§ SITE-Stockholm School of Economics, University of Rome Tor Vergata, EIEF, and CEPR.

1 Introduction

What indicators should be used to monitor the COVID-19 epidemic and the effects of policy interventions, such as lockdowns or other nonpharmaceutical interventions? In this paper, we argue that a single indicator is likely to miss key aspects of the epidemic. Hence, concentrating attention on a single "best" indicator (e.g., the cumulated number of COVID-19 infections or deaths), as often done in the literature and in policymaking, may be misleading. However, if several alternative indicators all point in the same direction and produce roughly similar results, this provides a much stronger basis for assessing the effects of a policy.

We illustrate our argument by focusing on the effects of a lockdown using Sweden as the reference country. There are two reasons why considering this country is interesting. First, Sweden plays an important role in the international debate because it adopted a mitigation strategy with policies much less strict than the majority of comparable countries. For this reason, it is also a unique benchmark against which to study the likely effects of the widely implemented lockdowns. Second, Sweden was the subject of several recent studies, focusing mainly on the number of recorded infections, that we can replicate with alternative indicators to identify the possible biases linked to the former. In particular, we revisit the recent studies of Born et al. (2020) and Cho (2020), who employ the synthetic control method introduced by Abadie and Gardeazabal (2003) to create a counterfactual for Sweden and thereby estimate what would have happened to the number of infections if Sweden had introduced a lockdown.

We focus on the first wave of the epidemic (February–June 2020). One reason is data availability. Another reason is the small variation of the starting date of the lockdown in the countries that adopted this policy in Spring 2020, which lends credibility to the "ceteris paribus" assumption needed by the methodology. A third reason is that the analysis of the second and third waves (Fall 2020–Spring 2021) is complicated by the spread of mutated and likely more infective variants of the virus.

Although some aspects of our approach are similar to Born et al. (2020) and Cho (2020), we depart in several important respects. Most importantly, in addition to COVID-19 infections and deaths, we consider two novel outcomes: i) adjusted COVID-19 deaths – a measure we construct to reconcile the series of daily COVID-19 deaths with the series of weekly *excess* deaths, that many view as a more reliable indicator because of the wide cross-country differences in both the intensity of testing and the recording of COVID-19 deaths; and ii) the ratio of the number of recorded infection to the number of tests, or positive rate, which is important because testing policy changed dramatically in Sweden during our sample period, possibly biasing indicators based on the absolute number of recorded infections. Further, for all outcomes, we consider both daily and cumulative values, resulting in a total of eight indicators. We also address several problems in the data that had not been previously identified.

Our results are qualitatively consistent with previous work in suggesting that the introduction of a lockdown would have greatly reduced the impact of the first wave on COVID-19 infections in Sweden. However, from the quantitative point of view, our results differ substantially, confirming that enlarging the set of indicators considered is important to improve our understanding of containment policies. After addressing data problems and re-optimizing the synthetic control method for each indicator, we find considerably stronger effects than in previous studies: we estimate that a lockdown would have reduced cumulative COVID-19 deaths by 40–47% and cumulative adjusted COVID-19 deaths by 41–43%, in contrast to 34% and 25% estimated in previous studies. Moreover, for the positive ratio we find that a lockdown would have displayed its effects almost immediately, 3–5 days after its introduction. The delay by several weeks in the effects of a lockdown estimated in previous studies focusing on the cumulative number of infections was mainly due to the large changes in the intensity of testing that took place in Sweden during the period considered (which are filtered out by the ratio of positives to tests).

The rest of the paper is organized as follows. Section 2 provides examples of the wide cross-country heterogeneity in recording COVID-19 deaths and testing intensity; and how the latter changed over time, in particular in Sweden. Section 3 describes our methodology. Section 4 presents our data and the new death measure introduced in the present study. Section 5 presents our results. Section 6 provides robustness checks. Section 7 concludes.

2 Heterogeneity in testing policies and death recording

The reason why we believe it is crucial to consider additional outcomes is the presence of large differences across countries in the way COVID-19 deaths were recorded and in the intensity of testing, as well as within-country changes in testing intensity during the relevant period, most crucially in Sweden. This section provides examples and simple quantifications of these forms of spatial and temporal heterogeneity.

2.1 Heterogeneity across countries

There has been a large debate about how different countries have been using different rules for ascribing deaths to COVID-19, suggesting a more extensive use of excess deaths as an indicator less subject to these differences. The policy snapshot "How reliable is COVID-19 mortality across countries" (Karanikolos and McKee, 2020) highlights several dimensions along which the recording of deaths as due to COVID-19 differed across countries, both before and after April 16th, when the WHO issued guidelines on certification and coding of COVID-19 as a cause of death.[1]

Quoting from their report: *"There are two main ways in which COVID-19 deaths are defined. The first, based on the WHO definition (see below), uses clinically confirmed or probable COVID-19 case (e.g. Belgium, Canada, France, Germany) and is not dependant on the availability of a laboratory test. The second, on the other hand, is reliant primarily on a positive laboratory test (e.g. Austria, Italy, the Netherlands, Spain, the United Kingdom). At the same time, there is still important variation within these two groups, as there are*

[1] The detailed comparative study by West et al. (2020) also finds considerable differences between Austria, France, Germany, Italy, Portugal, and the UK.

countries that include probable COVID-19 deaths in the definition, but still in practice require a laboratory confirmation (e.g. Cyprus, Greece, Romania, Serbia), while there are also countries that primarily use clinical diagnosis, but also include any death among positive cases (e.g. Canada)." Countries using the WHO definition are more likely to capture a greater share of COVID-19-associated deaths. Still, the same report warns us that the recording of the cause of death on the death certificates can also vary because of a different practical implementation of WHO guidelines, and different local death certification and coding practices. Quoting again from their report: *"For example, some countries using the WHO definition still require a positive test result (e.g. Greece), while others (e.g. Canada) include any death in a person with COVID-19, even if it was not triggered by the virus (e.g. trauma). There may also be changes in guidelines over time, which is particularly relevant during this pandemic, as it involves the emergence of a novel cause of death."* From what we understand, Sweden has followed the WHO definition and has been rather generous in ascribing deaths to COVID-19.[2]

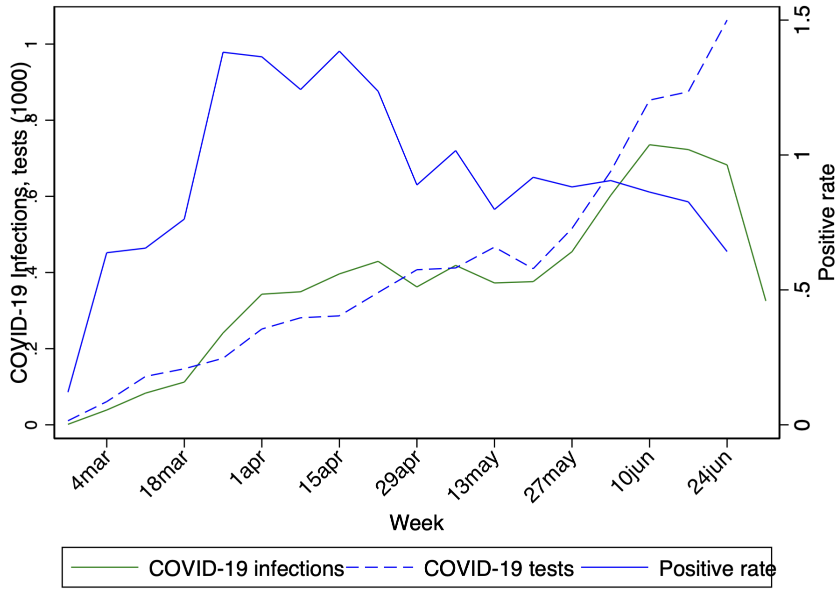
Regarding cross-country heterogeneity in testing intensity, this has been varying wildly across European countries in the period we are studying, with Sweden being one of the countries that tested less, although it has been trying to increase its testing capacity, as shown in the next subsection. For the period March 15–May 30, 2020, the average daily number of tests per 100.000 inhabitants ranged from 33 in Sweden to 55 in Norway, 88 in Ireland, and 130 in Denmark.

2.2 Heterogeneity over time

While remaining a laggard in terms of the number of performed tests for the large part of our sample period, Sweden tried to step up its testing capacity over time. This resulted in a large variation of testing intensity over time. Prior to March 12, 2020, the Swedish strategy was to test all people who had been in areas considered at risk, like China and Italy, but due to shortages in testing equipment, this strategy was rapidly changed, and testing was only targeted towards people with heavy symptoms and in need of hospitalization and medical care staff. Sweden then reversed this policy announcing on March 31, 2020, a plan to expand testing capacity to all critical services: the aim was to carry out 100,000 tests per week.

Although the Swedish government did not manage to rapidly step up its testing capacity, this policy change resulted in a constant, prolonged increase in the number of tests performed each week. As shown in Figure 1, the expansion of testing was accompanied by a parallel expansion in the number of recorded infections that had little to do with changes in the COVID-19 infection in the country, better captured by the solid blue line representing the positive rate. This parallel increase became stronger after mid-May when the positive rate stabilized and started to decline. Finally, on June 4, 2020, the Swedish government managed to implement a large expansion of its testing capacity and offered free testing to all citizens. The resulting rapid

[2] See the fact sheet "Description of data sources for deaths due to COVID-19", published by Socialstyrelsen (the National Board of Health and Welfare) on June 3, 2020 (https://www.socialstyrelsen.se/globalassets/1-globalt/covid-19-statistik/engelska-sidan/faktablad-datakallor-eng.pdf).



Notes: The figure shows the weekly number of infections, the weekly number of tests, and the positive rate in Sweden up to June 22.

Figure 1: Weekly number of infections and weekly number of tests in Sweden.

increase in the number of recorded infections is all but surprising given the steep increase in the number of citizens tested and the fact that the first people who asked to be tested when more testing became available were likely concerned with health status but unable to be tested before, hence with a higher probability to have contracted the virus.

As an example of how misleading it can be to rely upon a single indicator (in this case, the number of positive cases), a few days after Sweden finally managed to step us significantly its testing capacity while the positive rate was starting to fall, the WHO included Sweden in the set of 11 European countries with "accelerated transmission that if left unchecked will push health systems to the brink once again."[3] The Public Health Agency of Sweden publicly rejected the statement arguing that the Swedish testing policy changed dramatically in June and that the data on the number of infections were misinterpreted by the WHO.

3 Methodology

We employ the same methodology as Born et al. (2020) and Cho (2020), namely the synthetic control method introduced by Abadie and Gardeazabal (2003). This is done both for comparability with these studies and because of the simple and intuitive nature of this increasingly popular method. After briefly describing the

[3] See the statement to the press by Dr Hans Henri P. Kluge, WHO Regional Director for Europe on June 25, 2020 (https://www.euro.who.int/en/media-centre/sections/statements/2020/statement-digital-health-is-about-empowering-people).

method, we summarize its key elements. We refer to Abadie (2020) for a thorough review.

3.1 The synthetic control method

This method estimates the time-varying effect of a "treatment" (an intervention or policy) on some outcome of interest for a specific "treated unit" (an administrative district, geographical region, or country) by taking the difference in the time path of the outcome between the treated unit after the treatment and an artificial or "synthetic" unit constructed by taking a weighted average of a suitably selected set of untreated units (the "donor pool"). The weights given to the units in the donor pool are nonnegative, sum to one, and are chosen to minimize the distance between the treated and the synthetic unit in a given space of unit-specific indicators that may include pre-treatment values of the outcome of interest. In practice, these weights are usually "sparse", that is, only a few units receive positive weights. When only one donor unit receives a positive weight, the method reduces to the simple difference between two units.

As argued by Abadie (2020), "the synthetic control method is based on the idea that, when the units of observation are a small number of aggregate entities, a combination of unaffected units often provides a more appropriate comparison than any single unaffected unit alone." The method generalizes comparative case studies by formalizing the choice of the comparison units and the criteria for the comparison.

Notice that, unlike the vast literature on treatment effects, the synthetic control method estimates a time-varying individual treatment effect, not the mean or a quantile of the distribution of individual treatment effects.

3.2 Key elements of the method

The key elements of the synthetic control method are: (i) the choice of treatment (in our case, the decision to not impose a nationwide lockdown), (ii) the choice of the treated unit (in our case, Sweden), (iii) the choice of the outcome of interest (in our case, any of the indicators discussed in Sections 4.1–4.3), (iv) the length T_0 of the pre-treatment period (discussed in Section 5.1), (v) the choice of the "donor pool" (in our case, the set of countries to which Sweden is compared, also discussed in Section 5.1), (vi) the choice of unit-specific characteristics (discussed in Section 4.4), and (vii) the choice of metric to measure distance in the space of unit-specific characteristics (in our case, the same as Born et al. 2020 and Cho 2020).

Abadie (2020) argues that "the ability of a synthetic control to reproduce the trajectory of the outcome variable for the treated unit over an extended period of time [. . .] provides an indication of low bias", that "the risk of overfitting may also increase with the size of the donor pool, especially when T_0 is small", and that "each of the units in the donor pool have to be chosen judiciously to provide a reasonable control for the treated unit. Including in the donor pool units that are regarded by the analyst to be unsuitable controls [. . .] is a recipe for bias". Further, "the credibility of a synthetic control estimator depends on its ability to track the trajectory of the outcome variable for the treated unit for an extended pre-intervention period."

In practice, results from the synthetic control method tend to be quite sensitive to the choices made regarding these elements. Abadie (2020) recommends choosing a donor pool that is not too large, with units that are not too different in terms of both observable and unobservable characteristics. He also recommends choosing a pre-treatment period that is not too short. Since choices regarding these elements remain largely "ad hoc", we rely on a number of robustness checks presented in Section 6.

4 Data

This section presents our data and the new measure of COVID-19 deaths introduced in the present study.

4.1 COVID-19 infections and deaths

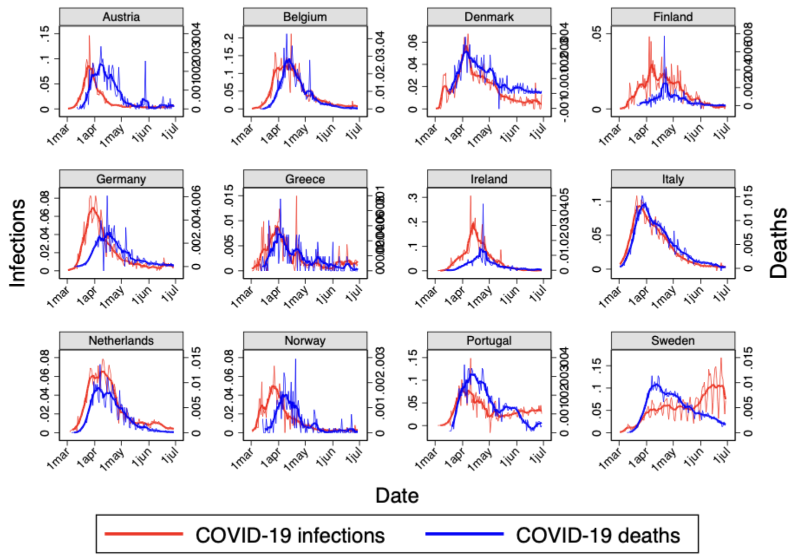
The daily and cumulative series of reported COVID-19 infections and deaths are taken from the Coronavirus Pandemic section of *Our World in Data* (Roser et al., 2020), which collects data on confirmed cases and confirmed deaths originally published by the European CDC.[4] These data are available for all countries considered. We normalize all values dividing by the estimated population size of a country at the beginning of the year 2020 and convert to cases per 1 thousand inhabitants.

Daily series are subject to strong day-of-the-week effects. They also display some small negative values for a number of countries and a few very large positive or negative values for two countries, France and Spain. The presence of implausibly large positive values and inadmissible negative values reflects periodic adjustments by the agencies issuing the data, whose nature, magnitude, and frequency vary from country to country and over time. These problems appear not to have been identified in previous studies based on the cumulative version of these data. In particular, negative values in the daily number of infections and deaths imply declines in the value of cumulative infections and deaths which may significantly affect the results of the synthetic control method.

To reduce the impact of these anomalies, we smooth the original series by taking 7-day moving averages and then divide by 7. Applying a moving average does a good job in reducing the noise in the data for all countries considered with the exception of France and Spain, for which the outliers are just too large. Because of this, we think the best course of action is to drop the two countries from the donor pool, though we add them again in one of the robustness analyses in Section 6.

Figure 2 compares the original and the smoothed daily series (respectively the thinner and the thicker lines) of COVID-19 infections and deaths in Sweden and 11 other European countries, namely those considered by Born et al. (2020) except France and Spain. While the profile of daily COVID-19 infections is quite different for Sweden, due to the very large changes in its testing policy, the profile of daily COVID-19 deaths is qualitatively similar in all countries considered, except for the much higher force of mortality in Belgium, Italy, the Netherlands, and Sweden.

[4] Downloaded from https://ourworldindata.org/coronavirus-data.



Notes: The figure shows the number of daily COVID-19 infections and deaths per 1,000 inhabitants. The thinner profiles are the original daily series, while the thicker profiles are the smoothed series obtained by taking 7-day moving averages.

Figure 2: Reported daily COVID-19 infections and deaths.

In addition to the number of COVID-19 infections per inhabitant, we also consider the positive rate, namely the ratio between the number of COVID-19 infections and the number of tests. Data on COVID-19 tests are available for all countries considered except the Czech Republic and Spain. When available, daily data on the number of new COVID-19 tests have been downloaded from the website of *Our World in Data*.[5] When daily data are not available (this is the case of Croatia, Germany, Greece, Netherlands, Poland, and Sweden), we use weekly data downloaded from the website of the European Centre for Disease Prevention and Control[6] and then construct a daily series by linear interpolation.

4.2 Mortality and excess mortality

Data on mortality from all causes are taken from the website of the *Financial Times*.[7] These data are only available at the weekly level, and are unavailable for Ireland and Romania.

Excess mortality is defined as the number of deaths recorded in a given period on top and beyond what we would have expected given mortality in the recent past. Operationally, it is computed as the difference between mortality in 2020 and average mortality in 2015–2019. Although excess mortality is only available

[5] https://ourworldindata.org/coronavirus-data.

[6] https://www.ecdc.europa.eu/en/publications-data/covid-19-testing.

[7] https://github.com/Financial-Times/coronavirus-excess-mortality-data.

on a weekly basis, we use this information to construct a simple correction of the daily series of COVID-19 death to match the weekly excess deaths. We describe this correction in the next section.

Aron and Muellbauer (2020) and Krelle et al. (2020) among others argue that excess mortality is a better measure for policy analysis because it avoids miscounting from under-reporting of COVID-19 related deaths or other health conditions left untreated. During the pandemic, we might have an increase in the number of deaths from other unrelated causes because the hospitals are overwhelmed and they work at full capacity, leading to many conditions left untreated or many people not seeking treatment. On the other hand, there might be fewer deaths from other causes such as road accidents given the mobility restrictions.

Excess mortality is also a more comparable measure because it is robust to structural differences across countries, such as the efficiency of the health system, or demographic characteristics, such as the distribution of the population by age.

4.3 Using excess mortality to estimate total COVID-19 deaths

The number of reported COVID-19 deaths is likely to represent a downward biased estimate of total COVID-19 deaths. The bias is likely to vary across countries and over time because of differences in both the testing policies and the criteria for identifying COVID-19 as the cause of death. In this section we propose a way of correcting for unreported COVID-19-related deaths, that is, deaths not attributed to COVID-19, by making use of the available weekly data on excess mortality.

Let T_{dj} denote the observable number of total deaths (i.e., deaths from all causes) on day $d = 1, 2, \ldots, 7$ of week j of 2020, and let T^0_{dj} denote the average number of total deaths on day d of week j during the baseline period 2016–2019. Excess mortality in week j of 2020 is measured by the difference $\overline{T}_j - \overline{T}^0_j$, where $\overline{T}_j$ and $\overline{T}^0_j$ are weekly averages of T_{dj} and T^0_{dj} respectively. We define excess deaths in week j of 2020 as the difference $\overline{T}_j - \overline{T}^0_j$, which can be negative whenever $\overline{T}_j < \overline{T}^0_j$, as for most countries at the beginning of 2020 and again in the Summer of 2020.

Under the assumption that COVID-19 is the only important cause of higher mortality in 2020 relative to the baseline, the positive part of excess deaths, namely $E_j = \max\{0, \overline{T}_j - \overline{T}^0_j\}$, is a measure of total (reported and unreported) daily COVID-19-related deaths in week j of 2020. If Y_{dj} denotes the smoothed number of reported COVID-19 deaths on day d of week j of 2020, obtained by taking a 7-day moving average of reported daily COVID-19 deaths, the average daily number of unreported COVID-19 deaths in week j of 2020 is measured by

$$Z_j = E_j - \overline{Y}_j,$$

where $\overline{Y}_j = \sum_{d=1}^{7} Y_{dj}/7$. We can then estimate the smoothed number of unreported COVID-19 deaths on day d of week j of 2020 by linear interpolation,

$$\widehat{Z}_{dj} = \frac{d}{7} Z_j + \left(1 - \frac{d}{7}\right) Z_{j-1}.$$

Adding the result to the smoothed daily number of reported COVID-19 deaths gives the following estimate of the daily number of total COVID-related deaths

$$\widehat{X}_{dj} = Y_{dj} + \widehat{Z}_{dj}.$$

We shall refer to $\widehat{X}_{dj}$ as adjusted COVID-19 deaths. The adjustment is sizable in countries, such as the Netherlands and Sweden, where excess mortality in the Spring of 2020 was positive and large.

4.4 Country characteristics

In constructing the synthetic control for Sweden, we initially consider the same set of country characteristics employed by Born et al. (2020), namely population size and the share of urban population. In one of the robustness analyses in Section 6, we expand this set by adding other characteristics, namely household size (also considered by Cho, 2020), GDP per capita, median population age, the fraction of people aged 70+, the number of hospital beds per inhabitants, and life expectancy. All characteristics are measured as of the latest available year.[8]

5 Results

This section presents the results from our baseline case.[9] Results from a number of robustness checks are briefly discussed in Section 6.

5.1 Implementation details

We follow Born et al. (2020) for the choice of the donor pool and the set of country indicators considered, but we exclude France and Spain for the reasons discussed in Section 4. Thus, our donor pool consists of 11 countries: 10 Western European Union (EU) countries with more than 1 million inhabitants (Austria, Belgium, Denmark, Finland, Germany, Greece, Ireland, Italy, Netherlands, and Portugal) plus Norway. Compared to Cho (2020), this gives a smaller but more homogeneous donor pool. In one of the robustness analyses in Section 6, we examine the effect of expanding the donor pool by including most of the European countries considered by Cho (2020).[10]

Unlike Born et al. (2020), and more in line with Cho (2020), we extend the length of the post-lockdown period till the end of June to fully allow for the sharp increase in testing rates that occurred in Sweden after

[8] Urban population data are from the World Bank (https://data.worldbank.org/indicator/SP.URB.TOTL?end=2019&start=2019, data on all other characteristics are from *Our World in Data* (https://ourworldindata.org/coronavirus-data.

[9] Calculations were carried out in Stata 16 using the package `synth` provided by Jens Hainmuller, Alberto Abadie, and Alexis Diamond.

[10] We exclude non-European countries and European countries with less than 1 million inhabitants.

an initial period of very low testing rates (see section 2) to fully display its effects.[11] Most importantly, as already mentioned, we increase the number of outcomes considered to include - in addition to the number of reported COVID-19 infections and deaths - also the number of adjusted COVID-19 deaths (constructed as described in Section 4.3) and the positive rate, computed as the ratio between the number of infections and the number of tests performed.[12]

We focus on cumulative infections and deaths both for consistency/comparability with previous work, and because cumulative variables tend to be less noisy than daily or weakly ones because errors of different sign tend to compensate each other across time. We add the positive rate because, as discussed in Section 2.2, the number of performed tests changed dramatically in Sweden during our sample period, and the ratio filters out this source of bias. As shown in Figure 1, changes in testing policy directly and strongly affect the absolute number of infections but not their ratio to tests performed, making the latter a more informative outcome in our context. In addition to cumulative outcomes, we also consider smoothed daily outcomes because we are interested in how fast the effect of the lockdown would have kicked in, and cumulative outcomes naturally "hide" the effects of the policy for some time.

For the countries in the donor pool, we take the pre-lockdown period to consist of the 2 weeks before the start date of the lockdown (13 days in the case of the positive rate). To improve comparability across countries, we transform time in deviations from the "treatment date", so day 0 is when the lockdown was introduced. For Sweden, that never adopted a lockdown, day 0 is set at March 17, the mean start date of the lockdown in the donor pool. As with the other papers cited, we ignore cross-country differences in the characteristics and intensity of the lockdown.

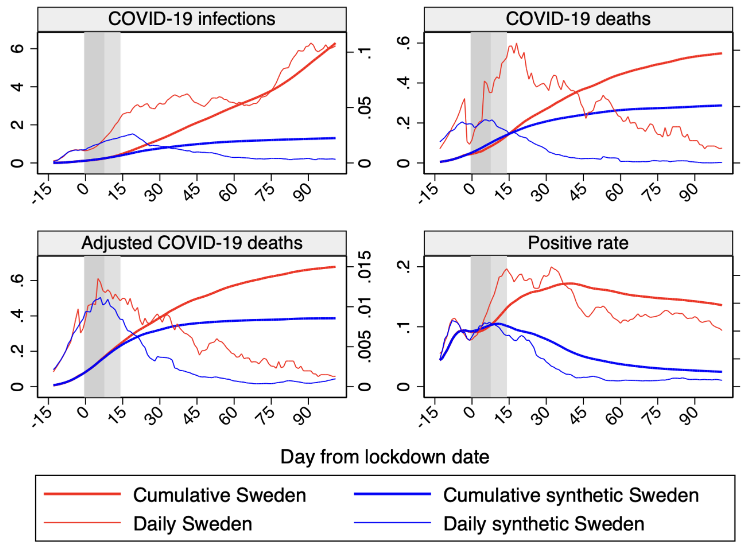
The frequency distribution of the incubation period for COVID-19 – which is the time between exposure to the virus and symptom onset – has a median of 7 days (Qin et al., 2020), while the median time from symptom onset to death ranges between 17 and 19 days (Zou et al., 2020; Wang et al., 2020). Thus, we take 7 days as the average length of the incubation period and 18 days as the average length of time from symptom onset to death. When using COVID-19 deaths and adjusted COVID-19 deaths, we shift the treatment date by 18 days from the lockdown date to take into account the expected delay with which the lockdown produces its effects on the number of deaths.

5.2 Sweden vs. synthetic Sweden

Figure 3 presents the profile of Sweden versus synthetic Sweden for all outcomes considered over our sample period, which extends until 105 days after the lockdown date (corresponding to about June 30). For all outcomes, we consider both daily and cumulative values. Notice that, unlike Cho (2020), the weights

[11] Results are sensitive to the choice of the post-lockdown period, in particular for the number of infections that are strongly influenced by the large changes in testing intensity in Sweden during our sample period. This may explain some of differences in the results between the early and the final versions of Born et al. (2020) and Cho (2020).

[12] Cho (2020) considers weekly excess deaths, but his synthetic Sweden is not optimized for this indicator and is based on the set of weights obtained for cumulative COVID-19 infections.



Notes: The profiles of daily and cumulative outcomes for Sweden and synthetic Sweden are shown in the figure. Horizontal axis measures days since the lockdown start that is normalized at day 0. The red line shows the profile for Sweden and the blue line shows the profile for synthetic Sweden. The vertical bands indicate the first 14 days after the lockdown start, with the lightest color as the first 7 and the darker as additional 7 days.

Figure 3: Profiles of daily and cumulative outcomes for Sweden and synthetic Sweden.

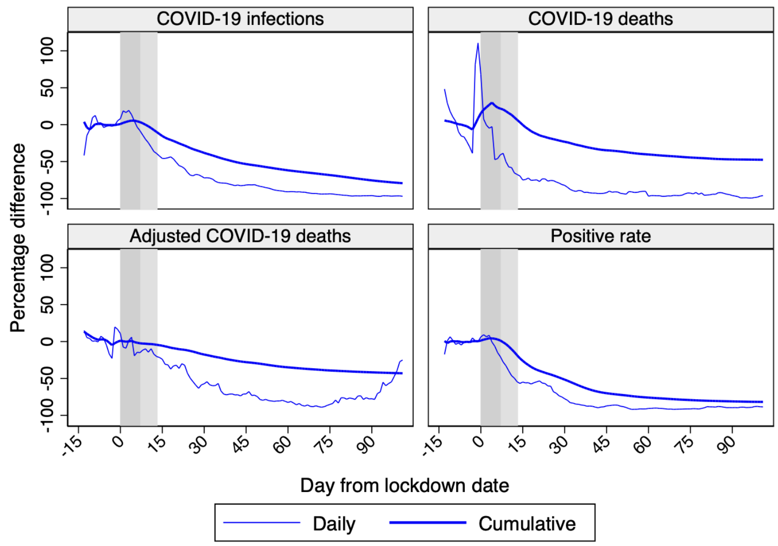
assigned to the countries in the donor pool are not constant but are re-optimized for each indicator. Although not clearly visible in the figure,[13] the fit over the pre-lockdown period is always very good.

To facilitate the assessment of the potential effects of a lockdown in Sweden, Figure 4 shows the percentage differences between synthetic Sweden and actual Sweden.

Visual inspection shows that each of our indicators consistently suggests that a lockdown would have had a strong effect in reducing the impact of COVID-19 in Sweden, confirming the sign of earlier estimates by Born et al. (2020) and Cho (2020) based on cumulative infections and deaths. Not surprising, the effects on the daily indicators are much faster.

Quantitatively, our estimates of the effects of a lockdown differ from previous work. Starting with cumulative COVID-19 infections – not our preferred outcome in the light of the evidence in Section 2.2 – we estimate a 61% reduction by May 17 (when Born et al. (2020) estimate a reduction of 48%), and a 71% reduction by June 7 (when Cho (2020) estimates a reduction of 75%). We then estimate a 40% reduction in cumulative COVID-19 deaths by May 17 (when Born et al. 2020 estimate a 34% reduction) and a 41% reduction in cumulative adjusted COVID-19 deaths by June 13 (when Cho 2020 finds a 25% reduction in excess deaths). On June 30, the end of our sample period, the reduction in cumulative adjusted COVID-19 deaths is 4 percentage points lower than for cumulative COVID-19 deaths (-43% vs -47%), which

[13] Details for the pre-treatment period are available upon request.

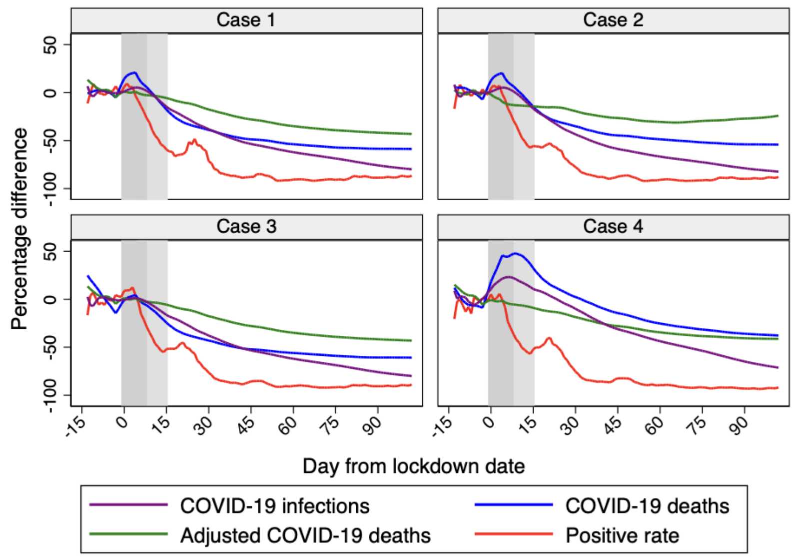


Notes: The percentage differences between synthetic Sweden and Sweden in terms of COVID-19 infections, deaths, adjusted deaths, and the positive rate are shown in the four panels. Horizontal axis measures days since the lockdown start that is normalized at day 0. The vertical bands indicate the first 14 days after the lockdown start, with the lightest color as the first 7 and the darker as additional 7 days.

Figure 4: Percentage differences between synthetic Sweden and Sweden.

is consistent with our conjecture that Sweden had a rather encompassing approach when assigning deaths to COVID-19.

Regarding the time a lockdown would have taken to display its effects, all daily (noncumulative) outcomes show that it would have sizable effects almost immediately. The curves for Sweden and synthetic Sweden start to diverge from each other less than a week after the lockdown. As we mentioned, the longer delay observed in our cumulative outcomes and in previous studies may be a natural effect of the inertia intrinsic to stock rather than flow measures. However, in the case of Sweden, the delay observed for cumulative infections could also be linked to the extremely low rate of testing that Sweden maintained in the first part of our sample period, followed by a strong increase in the last part. The inclusion of the positive rate among our outcomes allows us to shed light on the relative importance of these two possible explanations. Figure 3 shows a rather small additional delay in the observed effect of the lockdown on the cumulative positive rate relative to the daily indicator: Sweden and synthetic Sweden start to diverge less than a week after the daily indicator. This suggests that the delay in the effect of the lockdown on cumulative infections highlighted in previous studies is mainly driven by the changes in Swedish testing policy during our period, which are filtered out when using the positive rate.



Notes: The percentage differences between synthetic Sweden and Sweden in terms of COVID-19 infections, deaths, adjusted deaths, and the positive rate are shown in the four panels. The top left panel are the results for case 1 (expanded donor pool with France and Spain). The top right panel are the results for case 2 (expanded donor pool with other European countries). The bottom left panel are the results for case 3 (shifting the treatment date). The bottom right panel are the results for case 4 (adding extra control variables). Horizontal axis measures days since the lockdown start that is normalized at day 0. The vertical bands indicate the first 14 days after the lockdown start, with the lightest color as the first 7 and the darker as additional 7 days.

Figure 5: Percentage differences in cumulative outcomes between synthetic Sweden and Sweden.

6 Robustness checks

Since choices regarding key elements of the synthetic control method are somewhat "ad hoc", in this section we briefly present the results of a number of robustness checks.

We consider four cases and compare them with the baseline case in Section 5. The four cases considered are obtained by varying, one at the time, the set of countries in the donor pool (Cases 1 and 2), the treatment date (Case 3), and the predictors (Case 4). Detailed tabulations for each of the four cases are available upon request, while the percentage differences between synthetic Sweden and Sweden for each case are shown in Figure 5 with reference to the cumulative outcomes.

In Case 1 we include France and Spain ignoring the presence of negative values of daily infections and daily deaths for these two countries. This makes the results for this case more comparable with those in Born et al. (2020). The differences with respect to Section 5 are only minor.

In Case 2 we expand the donor pool to include most of the countries considered by Cho (2020). This makes the results for this case more comparable with his results. Expanding the donor pool in this way does not affect the results for COVID-19 deaths and the positive rate. When looking at cumulative COVID-19 infections and cumulative adjusted COVID-19 deaths, the curves for synthetic Sweden are higher than in the baseline case. When looking at the percentage difference, we see that the percentage difference for the case of cumulative COVID-19 deaths is higher than for COVID-19 adjusted deaths, the opposite than our baseline case.

In Case 3 we consider the effects of taking into account the average length of the incubation period of 7 days, which shifts the treatment date 7 days further. Again, the results hardly change.

Finally, in Case 4, we consider the effect of adding to population size and share of urban population other economic and socio-demographic indicators (average household size, median age, share of people aged 70+, life expectancy, GDP per capita, and hospital beds per thousands). Adding all these controls hardly changes the results.

7 Conclusions

In this paper, we compare a number of outcomes and indicators of the spread and consequences of the COVID-19 pandemic that are often used in isolation for both cross-country comparisons and policy evaluation. We focus on the highly debated case of Sweden – to our knowledge the only country with good data that did not impose a lockdown during the first wave of the pandemic. We construct counterfactuals for what would have happened if Sweden had imposed a lockdown during the Spring using a synthetic control methodology, specifically optimized for each of the indicators considered.

We address several problems in the data that had not been previously identified, and we propose a novel methodology that uses weekly data on excess mortality to correct the daily series of total COVID-19 deaths for under-reporting and cross-country heterogeneity in the definition and measurement of deaths.

We confirm the conclusions of previous work in terms of sign of the effect that a lockdown would have produced in Sweden in Spring, but after correcting for data problems and re-optimizing the synthetic control for each different indicator, we find a considerably stronger effect than previously estimated.

We also find that the use of COVID-19 deaths leads to overestimating the effect of the lockdown compared to our adjusted measure that matches excess deaths, confirming the perception that Sweden has been quite generous when counting COVID-19 deaths.

Studying the ratio of positives cases to the number of tests, we find that a lockdown would have had a sizable effect already a few days after its introduction. The long delay observed in previous studies focusing on the number of infections appears linked to the large changes in the frequency of testing that occurred in Sweden during the period considered.

Our study highlights the importance of looking at multiple indicators when evaluating policies or comparing countries, and of improving the quality of available data. The best way to produce comparable indicators for policy evaluation would of course be to have more homogeneous statistics across time and countries, possibly at a finer geographical level within each country.

Our results on Sweden do not imply that a lockdown would have been optimal or efficient, as the very high costs of a lockdown should be taken into account. Future work should address these important, complementary aspects necessary for a proper cost-benefit analysis.

References

Abadie, A. (2020). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*. Forthcoming.

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review 93*, 113–132.

Aron, J. and J. Muellbauer (2020). A pandemic primer on excess mortality statistics and their comparability across countries. *Our World in Data*. https://ourworldindata.org/excess-mortality-covid.

Born, B., A. Dietrich, and G. Müller (2020). The lockdown effect: A counterfactual for Sweden. *Centre for Economic Policy Research* (14744 (v. 2)). https://cepr.org/active/publications/discussion_papers/dp.php?dpno=14744.

Cho, S.-W. (2020). Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden. *Econometrics Journal 23*, 323–344. https://academic.oup.com/ectj/article/23/3/323/5899049.

Karanikolos, M. and M. McKee (2020). How comparable is covid-19 mortality across countries? *COVID-19 - Health System Response Monitor*. https://analysis.covid19healthsystem.org/index.php/2020/06/04/how-comparable-is-covid-19-mortality-across-countries/.

Krelle, H., C. Barclay, and C. Tallack (2020). Understanding excess mortality: what is the fairest way to compare covid-19 deaths internationally. *The Health Foundation*. https://www.health.org.uk/news-and-comment/charts-and-infographics/understanding-excess-mortality-the-fairest-way-to-make-international-comparisons.

Qin, J., C. You, Q. Lin, T. Hu, S. Yu, and X.-H. Zhou (2020). Estimation of incubation period distribution of covid-19 using disease onset forward time: A novel cross-sectional and forward follow-up study. *Science Advances 6*(33). https://advances.sciencemag.org/content/6/33/eabc1202.

Roser, M., H. Ritchie, E. Ortiz-Ospina, and J. Hasell (2020). Coronavirus pandemic (COVID-19). *Our World in Data*. https://ourworldindata.org/coronavirus.

Wang, K., Z. Qiu, J. Liu, T. Fan, C. Liu, P. Tian, Y. Wang, Z. Ni, S. Zhang, J. Luo, D. Liu, and W. Li (2020). Analysis of the clinical characteristics of 77 COVID-19 deaths. *Nature* (16384). https://https://www.nature.com/articles/s41598-020-73136-7.

West, A., T. Czypionka, M. Steffen, S. Ettelt, S. Ghislandi, and C. Mateus (2020). Reporting of COVID-19 deaths in Austria, France, Germany, Italy, Portugal and the UK. *Social Policy Working Paper*. https://www.lse.ac.uk/social-policy/Assets/Documents/PDF/working-paper-series/10-20-Anne-West.pdf.

Zou, F., T. Yu, R. Du, G. Fan, Y. Liu, and Z. Liu (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* (10229), 1054–1062. https://doi.org/10.1016/S0140-6736(20)30566-3.