COVID-19 counterfactual evidence.
Estimating the effects of school closures

Marco Giuliani

Department of Social and Political Sciences, Università degli studi di Milano

marco.giuliani@unimi.it

Full article: https://www.tandfonline.com/doi/full/10.1080/01442872.2022.2103527

Abstract

Scholars have started to estimate the effects of non-pharmaceutical interventions to reduce the health impact of COVID-19. However, the empirical evidence is highly contested, and since it is not known exactly what would have happened without those measures, political élites are left free to give credit to the voices that they prefer the most. We argue that any sensible assessment of the effectiveness of anti-COVID policies requires methodological reflection on what is actually comparable, and how to approximate the ideal "method of difference" theorized by John Stuart Mill. By evaluating the effectiveness of school closures as an anti-COVID policy, we provide two examples in which appropriate counterfactuals are inductively discovered rather than selected *a priori*. In the first one, we use Coarsened Exact Matching (CEM) in a cross-country setting, while in the second one, we implement the Synthetic Control Method in a within-country analysis. The article highlights the methodological advantages of including these techniques in the toolbox of policy scholars, while both examples confirm the effectiveness of school closures.

Keywords:

Evaluation; Comparative analysis; Counterfactual; Covid-19; School closures

Funding:

This work was supported by the Università degli studi di Milano [PSR2020]

1 Introduction

One of the main divides in the fight against the spread of COVID-19, especially in the first phases of the pandemic, has been the divide between countries that chose a more voluntary approach, based on recommendations, and countries that opted for stricter regulations, including school closures, travel bans and lockdowns (Capano et al., 2020; Ceron et al., 2021). In Europe, the benchmark for the first type of approach has been the Swedish model, which explicitly avoided imposing severe restrictions, especially regarding children's school attendance, mask wearing, and lockdowns (Andersson and Aylott, 2020; Pierre, 2020; Wenander, 2021).

The debate on the preferability of this liberal approach has assumed multiple nuances, and it is also for this reason that the model has attracted so much political interest. Assessment of its effectiveness in containing infections and deaths has obviously been at the centre of the arguments (Brusselaers et al., 2022), producing a polarized public discussion. For many, the most natural comparisons for what the media have dubbed the "Swedish experiment" are the other Nordic countries, which have adopted a more restrictive precautionary approach (Gordon et al., 2021). However, others, including the Swedish state epidemiologist Anders Tegnell, have suggested that Finland and Norway are outliers and, for a series of social and demographic reasons, Sweden should be rather compared to some West or Central European countries (Milne, 2021; Pickett, 2021; Schultz, 2021).

Clearly, "numbers are not facts [and] don't speak by themselves" (Stone, 2020), and, to attribute a causal effect to any policy, researchers need more than some arbitrary paired comparisons. The reason derives from the famous second canon of John Stuart Mill, which governs the so-called “method of difference”:

> *If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or a necessary part of the cause, of the phenomenon.* (Mill, 1843: 455)

Unfortunately, in real social life, such perfect twins rarely exist. Nonetheless, they must at least be presumed, because the same notion of causation relies on the possibility of comparing an observable with a potential outcome, what happened with what would have happened if a certain condition did not occur, or in case of a different treatment (Collins et al., 2004). This counterfactual notion of causality goes back to the works of the philosopher David Lewis (1973a, 1973b), and it has been recognized as providing relevant insights for any causal test in the field of social and political sciences (Brady, 2008).

In the absence of a perfect counterfactual, and in the impossibility of generating its surrogate as in controlled social experiments, scientists still try to approximate that ideal comparison also in observational studies. However, once again, the question is as follows: which countries are better suited to acting as counterfactuals for assessing the Swedish COVID policy? Are they Nordic countries, with similar geography and socio-economic characteristics, or other West or South-European countries, with similar population sizes? In fact, a major problem in assessing the impact of any policy, and tracing the causal relationship between a set of government decisions and their outcomes, is that of identifying appropriate counterfactuals and avoiding the risks of selection bias. Selection biases are common to both qualitative and quantitative research designs, because they simply reflect the potentially relevant differences between the compared units receiving and not receiving the treatment, such as implementing or not implementing a certain policy (Angrist and Pischke, 2009; King et al., 1994).

Furthermore, the decision of whether to introduce certain policy solutions also depends on the extent and intensity of the problem itself. For example, employing restrictive policies to limit the diffusion of COVID-19 is not independent from the spread of the virus itself – that is, from the outcome that one may want to explain. To use the experimental jargon, the treatment is not randomly assigned, but it is endogenous to the causation process. This greatly complicates the search for appropriate observational counterfactuals, which may sometimes not even exist.

Using the case of anti-COVID policies, especially the contested decision to close schools, this paper explores two different econometric options to overcome this limit, identifying or constructing appropriate counterfactuals, and eventually testing causal hypotheses. The paper is organized as follows. The next section introduces the different arguments in favour of and against closing schools to restrict the diffusion of COVID-19. In section 3, a matching strategy helps test the effectiveness of this policy by comparing 31 European countries during the first six months of 2021. Section 4 implements the synthetic research method in a within-country comparison of the effects of school openings in Italy, one of the countries that has recorded the highest coronavirus infection tolls. The final section discusses the implications of these results for the community of policy scholars and concludes. An online appendix complements the text with descriptive statistics and robustness tests.

2 The problem: to distance or not to distance (teaching)?

When, COVID-19 arrived in Europe at the beginning of 2020, one of the first policy decisions made was to shut down schools and universities. No country, not even Sweden, kept all schools entirely open during the first months of the pandemic (Pashakhanlou, 2021). Using the first component of the stringency index developed by Hale et al. (2021), specifically dedicated to the closing of schools and universities, it is possible to trace the adoption of this measure throughout the first year and a half of the pandemic. This component is an ordinal scale in which the value of zero is given to countries without restrictions; one is attributed if closures are simply recommended or if the usual organization of teaching necessitates major alterations; two when closures are decided only for certain levels or types of schools; and three is given in the case of generalized closures. Moreover, in the case of restrictions applied only to particular parts of the country, the value on the scale is reduced by half a point. Figure 1 reports the evolution of this scale, translated into a 0-100 index, for 31 European countries during the period from January 2020 to June 2021.[1]

The first mover was Italy, in which some regional governments chose to close schools at every level on 24 February 2020. The decision was shortly followed by other regions until eventually, on March 4, a national decree suspended all in-person teaching. Italy acted first because it was the first European country to be heavily hit by the virus, but the others soon made similar decisions. Twenty-six countries out of thirty-one enacted a general closure before the end of March 2020, although some kept distance teaching and learning until the end of the

[1] The sample includes all EU member states, plus the United Kingdom, Switzerland, Norway and Iceland.

school and academic year, while others gradually relaxed the restrictions, and one country, Switzerland, entirely cancelled them before June 2020.

Figure 1 The 0-100 index of school closing in 31 European countries in the period between January 2020 and June 2021

In autumn 2020, with the opening of the new school and academic year, many countries again had to resort to closures in concomitance with the upswing in infections of the second wave of the pandemic, but this time the picture was much more differentiated and nuanced, as highlighted by the large cross-country and longitudinal variation of the index in Figure 1. Governments were once again faced with the difficult choice between delaying the return to normality for millions of children and adolescents, and risking a further upsurge in the number of new infections. Some countries, such as Latvia and Finland, chose a rather stable policy halfway between full closure and full liberalization. Others, such as Portugal and Greece, constantly adapted their restrictions in response to the signals coming from the continuous monitoring of the health situation.

Political, administrative and institutional factors account for the differentiated strategies (Capano et al., 2020; McConnell and Stark, 2021; Toshkov et al., 2021), but their public justification has been facilitated by uncertain epidemiological evidence. Indeed, the first studies partially disagreed on the effectiveness of these measures, to say nothing of the feared psychological and educational consequences of protracted school closures, as well as the wider vulnerability concerns complicating the policy-makers' decisions (Flor et al., 2022).

Particularly influential, at least in supporting the strategy to keep schools open for adolescents, was the report published in July 2020 by the Swedish public health agency, which concluded that "closing of schools had no measurable effect on the number of cases of COVID-19 among children" (Folkhälsomyndigheten, 2020: 14). The analysis was based on a comparison of confirmed cases in the 1- to 19-year-old segment of the population in Finland and Sweden, with the former country implementing a generalized school closure between March and May 2020, while the latter only recommended distance-learning for higher education institutions and upper secondary schools. The lack of any effect of school openings/closings was attributed to the fact that children had comparatively lower viral loads, which reduced their risks and their transmission capacity (Ludvigsson, 2020). While the report was criticized due to the limited testing implemented at that time by Sweden, similar results were found by research based on contact tracing amongst Norwegian primary school pupils (Brandal et al., 2021).

Several other studies conducted to evaluate the effectiveness of school closures using a range of diverse methodologies have confirmed the scepticism about this kind of restrictive policy, not least because children at home may interact with each other in any case (Park et al., 2020). Simulated policy combinations on the demand for hospital services have been tested in the UK, with the conclusion that each intervention on its own is insufficient to control the circulation of the virus (Davies et al., 2020). The non-systematic contribution of school openings to the second wave of the pandemic after the summer break has been confirmed for Italy and Germany using large official datasets (Gandini et al., 2021; Isphording et al., 2020). Time-series Bayesian models have shown that school closures in Japan did not have any mitigating effect on the transmission of the infection (Iwata et al., 2020). Finally, a meta-analysis performed by Busa et al. (2021) summarizes the differences between the present COVID-19 pandemic and the better-known transmission of influenza, raising doubts as to the effectiveness of the traditional school closures usually applied to combat the latter while underlining the resulting severe psychological costs for children and adolescents.

On the other hand, studies focused on different parts of the world and adopting a broader perspective, less centred on children and teachers, have reported more promising effects of closures. Profiting from the diverse timing of school closures in spring 2020 in different US states, Auger et al. (2020) were able to attribute the variation in COVID-19 incidence and deaths to that policy. Rauscher and Burns (2021) ran a more fine-grained analysis in which they matched similar and neighbouring counties that differed mostly for the period in which school closures were implemented, and they were also able to highlight some further conditional factors for the effectiveness of the restrictive measure. A similar research design was adopted to analyse the epidemiological dynamics in Italy in autumn 2020, this time exploiting the diverse school openings in three Italian regions after the vacation period and confirming their role in spreading the virus (Tosi and Campi, 2021).

A potentially confounding element for the causal attribution of the outcomes to that specific policy is the presence of other measures or recommendations. To assess the effect of school policies separately from the impact of other measures, Stage et al. (2021) comparatively analysed school closures and openings in Norway, Sweden, Denmark and Germany by "generating a counterfactual projection of daily cases or hospital admissions, which accounts as much as possible for events prior to, but excluding, school closure"(3). Even when controlling for other measures, the number of new cases typically started to decrease one week after the enforcement of school restrictions, confirming the centrality of this type of policy in tackling the spread of the virus.

The studies reported above are just a sample of the abundant scholarly literature on the issue. With few exceptions, there is no specific attention to case selection or to the meaningfulness of the implicit or explicit comparisons because of the rush to provide an empirical foundation for evidence-based policy decisions. This lack of consideration may bias the conclusions of these quantitative analyses in the same way as the use of inappropriate reference countries for the evaluation of the Swedish anti-COVID policies discussed in the introduction.

In what follows, I introduce and exemplify two different techniques – matching and synthetic controls – that put the appropriate identification strategy at the centre of their methodological attention. I expect that closing schools – once a plausible set of covariates, including other restraining policies, has been controlled for – has a negative effect on the diffusion of the virus. The research focuses on the second and third waves of the virus in Europe to avoid any confounding factor due to the early surprise and absence of any previous experience in coping with the emergency. The hypothesis is first assessed in a cross-country comparison and then from a within-country perspective in Italy, one of the countries most affected by the virus. The various techniques used to identify the appropriate counterfactuals are introduced together with the specific research design used in each empirical test.

3 Matching cases in a cross-country comparison

As highlighted in Figure 1, some governments chose tougher restrictions on school attendance to counter the spread of the virus, while others did not. Were these measures effective? To answer this basic question, I retrieved weekly data on the incidence of new certified COVID cases during the first six months of 2021, when the different European governments already had the opportunity to define and test their strategies after the new wave of autumn 2020.

The treatment variable is the previous enforcement of highly restrictive regulations on in-person teaching, defined as more than 75% of the scale of the school closing index (Hale et al., 2021).[2] The right-hand side of the equations includes the lagged dependent variable, that is, the incidence of new cases in the previous week, to control for temporal autocorrelation and reflect the path-dependent diffusion of the virus (Beck and Katz, 1996). Models further contain the

[2] This specific threshold corresponds to the mandatory closing of schools at all levels, only potentially moderated by its geographical scope or by the weighted weekly average. More details are provided in the supplementary material.

following controls: a) the previous presence of other rigid containment policies, computed on the remaining seven subcomponents of the stringency index; b) the lagged percentage of the fully vaccinated population, which is expected to reduce the spread of the virus; c) the cumulative incidence of COVID deaths, which may impact the fears and attitudes of citizens, and on compliance with government regulations, inducing more prudent behaviours (Goolsbee and Syverson, 2021); and d) the number of COVID tests, which should be directly associated with the discovery of new cases. In Table 1, the first regression implements panel corrected standard errors (PCSE), model 2 opts for a panel regression with country fixed effects (FE) as a safeguard against omitted variables, while model 3 keeps those fixed effects together with the more conservative panel corrected standard errors.

Table 1. Estimating the effect of closing schools on the incidence of new cases

	(1) PCSE	(2) FE	(3) FE-PCSE
L. New cases (1000000)	0.88***	0.82***	0.82***
	(0.03)	(0.02)	(0.04)
L. School closing dummy	-69.10	-80.49	-80.49
	(60.57)	(67.07)	(75.34)
L. Other policies	-97.39	-168.08*	-168.08*
	(86.00)	(85.78)	(101.19)
L. Fully vaccinated (100)	-9.70**	-6.40**	-6.40
	(3.81)	(3.03)	(4.13)
Total deaths (1000000)	-0.03	-0.52***	-0.52***
	(0.03)	(0.09)	(0.16)
Tests (1000)	0.15	1.68***	1.68**
	(0.36)	(0.52)	(0.72)
Constant	241.05***	479.36**	479.36*
	(64.78)	(192.71)	(272.16)
Country fixed effects		✔	✔
N	775	775	775
Countries	31	31	31
R-squared	0.85	0.86	0.86

Note: Standard errors in parentheses *p < 0.10 **p < 0.05 ***p < 0.01

The three models return a substantially similar picture, which thanks to the inclusion of the highly significant lagged dependent variable is further supported by a large explained variance. In model 1, the signs of all the coefficients reflect the original expectations, but only the share of the fully vaccinated population is systematically associated with a lower number of infections. The incidence of COVID deaths and the number of tests are not related to the outcome, failing to reach standard levels of statistical significance. However, what is more important for the analysis is that the school closure policy does not seem to significantly reduce the diffusion of the virus, nor do the other containment policies.

The null finding regarding school closures is confirmed by the second and third models, both including country fixed effects. However, their inclusion improves the behaviour of some control variables, which, in addition to confirming the expected direction of the association, become statistically significant at least to some degree. These empirical results would represent a powerful argument for those opposing the closure of schools and partially support those resisting the imposition of restrictions *tout court*.

However, all these models potentially suffer from a relevant shortcoming, because the decision to close schools was not randomly assigned. Simply regressing the available observations assumes that they are similar in all relevant respects except for the treatment variable, whereas they are obviously not. The inclusion of control variables is a partial solution, but a more direct search for counterfactuals is needed to avoid biased interpretations of causal effects. In fact, "even when all the confounding covariates are measured […] it can be difficult to properly control for them if the distributions of the predictors are not similar across groups, [that is, if there is a] lack of *balance*" between those observations that received the treatment (the policy) and those that did not (Gelman and Hill, 2007: 200).

Let us consider the example of the impact of the incidence of new COVID cases in a certain week on that same incidence the week after. That quantity is taken into account by policy-makers to decide for or against closures: the greater the incidence of new cases, the more likely governments are to decide to close schools to reduce the further spread of the virus. This is clearly a situation of "lack of balance" among groups on some relevant factor affecting both the treatment and the outcome. If the lagged incidence of new COVID cases wholly explained the introduction of restriction policies, there would only exist either countries with many infections and consequent school closures, or countries with few contagions and open schools. In other words, there would not exist any counterfactuals, having two perfectly separated groups. In reality, other factors contribute to the adoption/rejection of such policies, so that a certain degree of overlap is likely to exist.

The $\mathcal{L}_1$ statistic introduced by Iacus et al. (2011) provides a useful measure of the level of imbalance between the treated and untreated groups. It ranges from 0, in the case of perfect balance, to 1 in the case of perfect separation, and it can be intuitively conceived as the complement to 1 of the degree of overlap between the two distributions on a single variable or on multiple ones. The top part of Table 2 provides measures of this overall measure of imbalance, together with some further details on the distributional imbalances of the original sample consisting of 177 country-week observations with school closures and another 598 without such a policy.[3]

[3] Following the good practices in these circumstances, I have included in the procedure variables that are supposed to "affect both the treatment assignment and, controlling for it, the dependent variable" (Negri, 2022), i.e., all the variables with the exclusion of testing, which is associated with the outcome but not with the policy.

Table 2. Imbalances in the raw and in the matched sample

Raw sample	$\mathcal{L}_1$	Avg	Min	Med	Max
L. New cases (1000000)	0.28	1030.80	84.98	794.7	788.16
L. Other policies	0.34	0.34	0.00	0.00	0.00
L. Fully vaccinated (100)	0.27	-5.50	0.00	-3.20	-39.35
Total deaths (1000000)	0.30	-59.90	0.00	-254.05	-347.30
Multivariate	0.81				
Balanced sample	$\mathcal{L}_1$	**Avg**	**Min**	**Med**	**Max**
L. New cases (1000000)	0.09	3.10	84.98	42.11	-119.18
L. Other policies	0.00	0.00	0.00	0.00	0.00
L. Fully vaccinated (100)	0.23	-0.57	0.00	-0.82	2.04
Total deaths (1000000)	0.01	-36.77	0.00	-66.03	-111.82
Multivariate	0.59				

The overall imbalance of the pre-treatment covariates is rather high, with $\mathcal{L}_1$ equal to 0.81. All factors present some degree of imbalance, which manifests itself in the rather large difference in average between the two parallel distributions, but also in the median values and sometimes even in the minimum and maximum ones. For example, the average incidence of new COVID cases in the instances that introduced school closures was almost double – 1031 more cases per million inhabitants – than in the observations that chose not to apply restrictions. Moreover, at that time, countries that adopted that type of restrictive policy were lagging behind in terms of vaccinations, having an average 5.5% fewer fully vaccinated inhabitants than did those that opted for a more liberal approach.

Matching – and more specifically Coarsened Exact Matching (CEM) (Blackwell et al., 2009; Iacus et al., 2012) – is a procedure with which to reduce the level of imbalance in a sample to allow a more robust test of the causal effect of the treatment, that is, the school closure policy.[4] By identifying sufficiently similar observations, and pruning those without appropriate counterfactuals, CEM "pre-process(es) raw data so that the treatment group becomes as similar as possible to the control group on a set of covariates chosen by the researcher" (Negri, 2022).

I applied CEM to the original sample using the covariates listed above, obtaining a more balanced sample consisting of 360 country-week cases, 115 experiencing the school closure policy and the other 245 not. As detailed in the lower part of Table 2, the multivariate $\mathcal{L}_1$ index in the balanced sample is now equal to 0.59, 27% smaller than before.

Including more covariates in the process to obtain a better match is not necessarily recommended. First, that strategy would have produced a much larger reduction in the number of observations, with the risk of entirely losing some of the countries covered by the analysis, whereas I managed to keep all the original 31 European nations.[5] Second, and most importantly, the univariate values reported in Table 2 show that the matching procedure has already produced a substantial reduction in imbalances on all covariates, not only in their averages but also in the whole distributions of the data, making the treated and non-treated groups much more similar in many respects.[6] For example, the average gap in the incidence of new cases has diminished from 1031 to just 3, and all the other average distances have

[4] There is a range of matching techniques aimed at solving the problem, and the choice also depends on the characteristics of the data. For a reader interested in the statistical properties of CEM compared to other matching strategies, see Iacus et al. (2011, 2019) and (Nielsen, 2020).

[5] Sample size and balance are in fact the two criteria with which the soundness of matching should be evaluated (King et al., 2017; Nielsen, 2020). A smaller sample would have probably also required a different model, like a simple OLS, while keeping the initial panel structure allows a better comparison between the original and the matched regression results.

[6] As stated by Blackwell et al. (2009: 531), "the absolute values [of $\mathcal{L}_1$] mean less than comparisons between matching solutions".

disappeared or have been greatly reduced. The same applies to the median and maximum discrepancies. Regardless, the supplementary material contains some robustness analyses that, at the cost of reducing the number of countries included in the matched sample, further increase the matching but still confirm the same results.

The last stage of the matching process is to estimate again the effect of the policy only on the matched cases and with the appropriate weights defined by CEM. Tackling the remaining imperfect balance of the sample by further including the original covariates as control variables, Table 3 reports the results of this final step of the estimation strategy.

Table 3. Estimating the effect of closing schools on the incidence of new cases using the balanced sample

	(1) PCSE	(2) FE	(3) FE-PCSE
L. New cases (1000000)	0.93***	0.98***	0.98***
	(0.03)	(0.03)	(0.05)
L. School closing dummy	-188.11***	-238.59***	-238.59***
	(52.19)	(52.45)	(49.23)
L. Other policies	-7.79	-165.01**	-165.01***
	(50.27)	(74.22)	(57.07)
L. Fully vaccinated (100)	-18.51***	-14.52**	-14.52**
	(6.83)	(6.75)	(7.46)
Total deaths (1000000)	-0.06	-0.57***	-0.57***
	(0.05)	(0.14)	(0.20)
Tests (1000)	0.13	0.80	0.80*
	(0.18)	(0.71)	(0.46)
Constant	294.00***	765.46***	765.46***
	(108.76)	(208.51)	(255.00)
Country fixed effects		✔	✔
N	360	360	360
Countries	31	31	31
R-squared	0.88	0.92	0.92

Note: Standard errors in parentheses *$p < 0.10$ **$p < 0.05$ ***$p < 0.01$

The new results show some similarities but also some important differences compared to the previous regressions. The coefficients of all variables again have the expected sign. The lagged dependent variable and the diffusion of vaccines are systematically associated with the outcome in all three models, while the number of tests turns out to be significant only in the last one. The two regressions that include fixed effects also present systematic relationships for the index summarizing several other containment policies, as well as for the cumulative incidence of COVID deaths.

However, what is most interesting is that the reductive effect of closing schools and universities on the incidence of new COVID cases now stands out as highly statistically significant and large in magnitude under all the different model specifications. Using the coefficient of the last regression, and all other things being equal, introducing the closure policy produces a decrease in the number of new weekly cases equal to 0.024% of the population, which is roughly equivalent to 4100 fewer weekly infections for a country of average size such as the Netherlands. Given the mean fatality ratio in the dataset, it would also represent approximately 80 fewer weekly deaths from COVID in that same benchmark country.

Considering that reduced contagion also implies a decreasing probability of others being infected week after week, it is easy to understand why this policy is so central to the debate on so-called “nonpharmaceutical anti-COVID interventions”. The empirical evidence provided by a matched sample, approximating the requirements of John Stuart Mill’s method of difference, reverses the initial findings and supports the choices made by the most prudent policy-makers.

4 Building counterfactuals in a within-country comparison

Coarsened exact matching is but one specific method, particularly suited to continuous variables, within the family of matching techniques (Iacus et al., 2011; Nielsen, 2020). Matching helps identify the appropriate counterfactuals on which to run comparisons, but it is certainly not the only approach useful for improving the capacity to infer the causal effects of a policy. Sometimes, counterfactuals are not discovered or identified. Instead, they are (in a sense) built by the technique itself, as in the “synthetic control method” (Abadie et al., 2015). In this approach, the counterfactual against which to compare the trajectory of the case of interest is assembled by an appropriate weighted linear combination of the other cases. It is synthetically manufactured to maximize its resemblance to the object of the analysis before the treatment (Abadie et al., 2010; Abadie and Gardeazabal, 2003).

The synthetic control method is ideal for analysing situations in which a certain policy has been formulated and implemented in one specific unit, and longitudinal data before and after that intervention are available for that treated unit and multiple others without that intervention. As such, and in more or less sophisticated versions, it has been used to investigate the effects of mandatory face masks in Germany (Mitze et al., 2020), and of lockdown measures in the United States (Friedson et al., 2021), Chile (Herrera and Godoy-Faúndez, 2021), Wuhan (Yang, 2021), Wenzhou and Shanghai in China (Tian et al., 2021). The method can also be applied to multiple treated units, comparing the average trajectory followed by those that have adopted a certain policy or experienced certain events compared to their average synthetic counterfactual (Cavallo et al., 2013).

This latter possibility is used in the within-country comparison of Italian provinces below. In Italy, legislative and administrative powers on several issues related to health and education are decentralized to the regional level, although some (e.g., the definition of the school calendar) may be further delegated to an even more local level. The pandemic has fostered a sudden, and often disordered, recentralization of many decisions. For instance, as stated in section 2, the first school closures were decided in February 2020 by some regional administrations, but then a national decree extended that decision to the entire country. The openings of schools for the new 2020-21 year were again decided locally, although mixed teaching methods were defined centrally for the universities and, after a few weeks, a new national decree again ordered distance teaching for all high schools and restricted the options for primary and lower-secondary schools according to the severity of the local epidemiological situation and the so-called "colour-coding system" (Giuliani, 2022). Thereafter, the colour differentiation was extended to upper secondary schools, fixing different percentages of class filling according to the severity of the health conditions, and eventually all schools and universities were ordered to close in the so-called "red regions".

The overall process was often confused, if not contradictory, and offered the chance for opportunistic politicized behaviour in a centre-periphery game that certainly did not help combat the pandemic (Capano, 2020). Paradoxically, if fine-tuning the strictness of the regulations according to the epidemiological conditions seems to be a sensible decision, it was a uniform national decision that somehow delayed the arrival of the second wave in Italy (Coppola and Ryan, 2020; Manica et al., 2021).

Whatever the best overall strategy should have been, I profit from the different timings of some decisions regarding the opening of schools after the summer break to again test the contested effects of school attendance on the spread of the virus. Different Italian regions decided to start the new school year on different dates. Students living in the autonomous province of Bolzano were those who returned to school earliest, on September 7; several regions opted to open the schools on September 14; and another group started almost two weeks later.[7] If the return to school had an effect on the spread of the virus, there should be an earlier increasing trend of COVID infections in the provinces that started earlier compared to those that started later.

This approach is similar to the one followed by Isphording et al. (2020), who exploited the large differences in the timing of the return to school among German *Länder* – from the beginning of August to the middle of September. They compared county cases two weeks before and three after the end of the vacation period, using as counterfactuals for the *Länder* whose students were going back in class, those who were still on vacation together with those who had re-opened at an earlier date. Isphording and colleagues counterintuitively found that "the end of summer breaks is associated with a distinct decrease [and not increase] in the number of SARS-CoV-2 cases" (14). The new organization of the school environment is prudently cited among the reasons explaining their unexpected findings, although they also admit that the results are mostly driven by "states with early summer breaks". This means that the negative effect of the return to school was mostly determined by a trend experienced in August, well before the onset of the second wave in the country.

---

[7] Having weekly data, I approximate the exact dates to the closest Monday to reflect the actual presence of students in class. In some regions, the original dates were postponed by the governors a few weeks before the start of classes, whilst in Sicily the regional government first allowed the deferral from 14 to 24 September only for schools used as polling stations for a national referendum, and then extended the possibility to all kinds of schools, producing, *de facto*, a general rescheduling.

The different timings of school openings have also been used in the Italian setting by Gandini et al. (2021), who employed the effective reproduction number R_t at the regional level as the dependent variable. Amongst other analyses, the authors provide a series of longitudinal graphs paralleling pairs of supposedly comparable cases – Trento and Bolzano, Abruzzo and Marche, Veneto and Apulia, Calabria and Sicily – whose main difference was thought to be exactly the diverse return to school by students. After a mostly visual inspection of seven-day moving averages, Gandini et al. (2021: 7) conclude that they "did not find an unequivocally constant delay between school opening and R_t rise".

Their identification strategy is questionable: Sicily *de facto* opened its schools on the same day as Calabria; Bolzano and Trento had large pre-treatment differences in their trajectories, so that they are not an ideal comparison; Veneto and Apulia are an odd pair of regions to compare, while Marche and Abruzzo, which actually show similar trends towards the end of September, have large confidence intervals before that period that make it difficult to evaluate the soundness of the comparison. I propose to improve Gandini et al.'s identification strategy by extending it to all Italian provinces and with a more systematic definition of the counterfactuals.

The outcome of interest is again the incidence of new weekly cases, considering as treatment the early opening of schools. A series of COVID-related variables is used to construct an appropriate synthetic counterfactual through a weighted linear combination of the provinces that opened schools later, located in the regions of Abruzzo, Apulia, Basilicata, Calabria, Campania, Sicily and Sardinia. Especially considering the clear North–South divide in the location of the provinces belonging to the two groups, the identification of a synthetic counterfactual as a linear combination of non-treated cases is defensible only if it can demonstrate an ability to sufficiently well approximate the treated cases in the pre-treatment period, i.e., before the opening of the schools.

After a series of multivariate panel tests, available in the online appendix, I decided to use two groups of variables as predictors associated with the outcome for the identification strategy. The first includes a set of epidemiological indices: the incidence of overall and active cases, the number of COVID tests performed per thousand persons, the positivity rate, and the reproduction number R_t. The second group involves demographic quantities: the population of the province, its density per square kilometre, and the share of students' population. Using these predictive variables, and merging with the appropriate weights the provinces in which children returned to school later, the synthetic control method is able to simulate the average trend of infections in a counterfactual province as similar as possible to those that actually opened schools earlier, except for the fact that it did not. Figure 2 compares the actual average trajectory with that of this synthetic counterfactual.

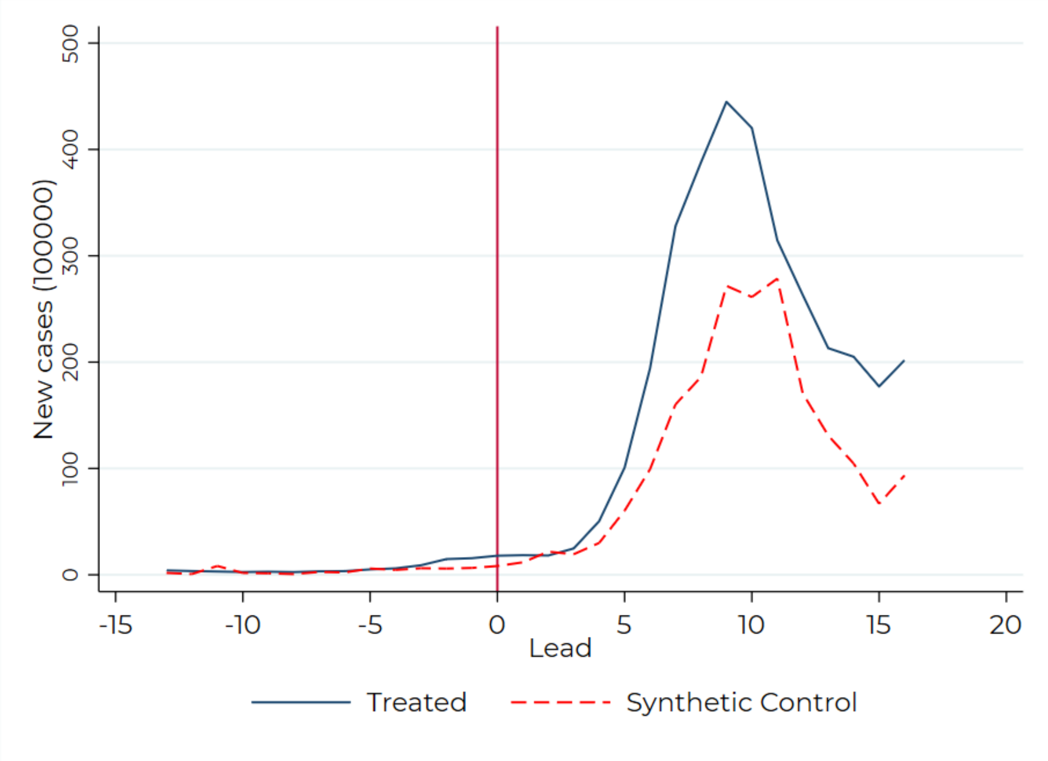

Figure 2 Incidence of new cases for school early opening provinces, and their counterfactual

The comparison starts in early June, and in the plot it is possible to see how close the two trends are until the week before the school openings, marked by the solid vertical line. The model has in fact a low pre-treatment root mean squared prediction error (RMSPE=4.77), the statistic that evaluates the soundness of the counterfactual before the policy or the event. The trajectory of the provinces with schools that opened earlier (solid blue line) started to diverge from the synthetic counterfactual (dashed red line) a couple of weeks after the early openings, anticipating the surge in new cases that involved all provinces soon after. Thereafter, the gap between the two lines increases so much that the post-treatment RMSPE is more than twenty times larger than before, which is usually considered a useful indicator of the non-random character of the deviation (Abadie et al., 2010).

Using a sample of 1000000 placebo tests, it is possible to assess the statistical significance of the gap between the trajectory of the provinces that opened schools in early September and that of their synthetic counterfactual: the difference between the two trends became systematic three weeks after the first openings. The divergence remained statistically significant until the eleventh week, i.e., fourteen days after the implementation of the new government decree of 3 November 2020, which closed all high schools in the country and contributed to the sharp common decrease in contagions clearly seen in the graph.

To complement this analysis, the same approach is replicated by selecting the "treated" unit as the sole province of Bolzano, which was the only one that opened its schools on September 7, and including in the donor pool only those provinces that opened their schools more than two weeks later.[8] Running the synthetic control method, the left panel of Figure 3 plots the actual and counterfactual trajectory of the infections for Bolzano. The pre-opening prediction error is still very low (RMSPE=4.58), while the post-treatment RMSPE is more than 60 times larger, confirming the visual departure of the actual Bolzano from its counterfactual.

[8] I lately found a work by Alfano et al. (2020), who similarly applied the synthetic research method on the early opening of schools in Bolzano. However, they defined the donor pool differently, focused on a shorter period, used daily instead of weekly data, and opted for a different set of predictive covariates and a different outcome – log total cases since August 1. Interestingly, the results are similar.

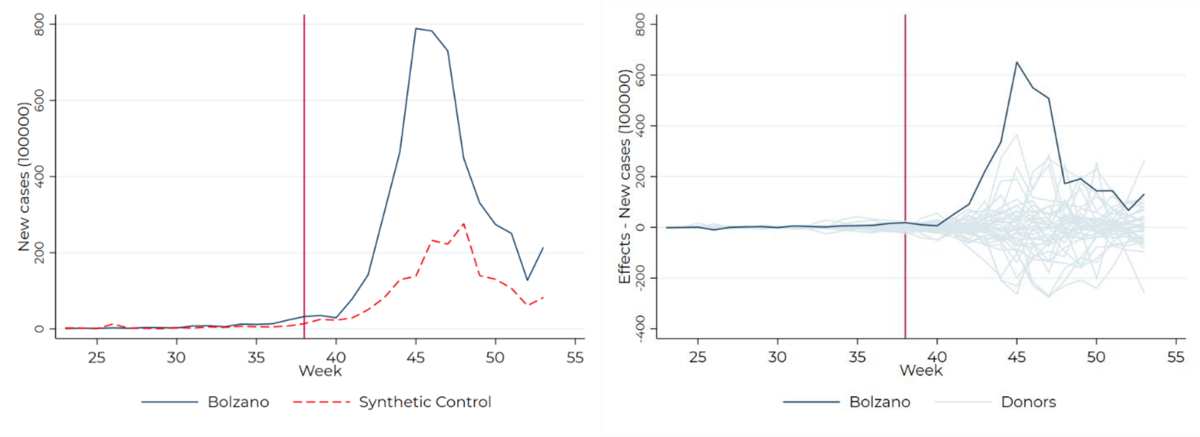

Figure 3 The incidence of new COVID cases in Bolzano and its synthetic counterfactual (left panel), and the effect of the early opening, together with the placebo effects (right panel)

The two trajectories start to diverge consistently from the 40th week, that is, three weeks after the opening of schools in Bolzano Province. By running a complete set of placebo tests, whose effects are plotted in the right panel of Figure 3, it is possible to test the statistical significance of that separation, which becomes systematic only after the fourth week following the end of the summer break. Due to the rapidly deteriorating situation, the national government later imposed a common national framework with a long list of measures to curb the spread of the virus, including curfews, school and business closures, mobility restrictions, etc. The implementation of those measures contributed to the sharp decline in new COVID cases after the 45th week, and favoured the convergence of the incidence of infections between Bolzano and the provinces that opened their schools much later.

5 Conclusion: what can we learn?

This article is a methodological exercise with some substantive results. While summarizing the latter, I want to reflect on the advantages of the former and on its far-reaching significance.

The research first implemented a matching strategy to test the effect of school closures in a comparative setting. It showed how closure produced the expected decrease in infections even after controlling for a set of other covariates, including different containment policies. The magnitude of the effect is sizeable, implying a reduction of approximately 240 weekly registered cases per million people. In our sample, and using the respective average fatality rate in that period, the value corresponds to a number of fewer COVID deaths per week ranging from less than 1 for Iceland, to 500 for Germany.

The synthetic research method was then applied to Italian sub-national data on the mirror case of early school openings. The results confirm the plausible expectation that school attendance, with all its indirect spillover effects, is a systematic channel for the spread of the virus. Provinces that opened school earlier experienced a surplus of infections that peaked after a few weeks, producing an excess of almost 2000 registered cases per million people compared to those in which students went back to school later. The gap was reabsorbed only when the Italian government approved new homogeneous restrictive rules at the beginning of the second wave of the pandemic.

Many studies point to school opening/closing as one of those non-pharmaceutical interventions with the largest direct and indirect effects (Li et al., 2021; Liu et al., 2021). There has been much debate on this issue worldwide, with systematic reviews and meta-analyses trying to summarize the findings of multiple empirical studies, and making explicit the economic, psychological and educational downside of this kind of policy (Busa et al., 2021; Krishnaratne et al., 2020). Several scholars in this research field explicitly design, look for, model, match or reconstruct some counterfactual against which to compare the cases of interest (e. g. Auger et al., 2020; Cunha et al., 2021; Gordon et al., 2021; Rauscher and Burns, 2021).

Notably, with a few exceptions, these comparisons are not performed by political scientists and, apart from some economists, not even by the more comprehensive category of social scientists.

It is probably true that there is already a large and motivating agenda for policy scholars interested in the governance of the pandemic (Dunlop et al., 2020; Weible et al., 2020), but the methodological point goes beyond COVID policies. By approximating the canons defined by Johns Stuart Mill almost two centuries ago, the tools exemplified in the article support the quest for causal relationships and evidence-based arguments, going beyond the simple identification of systematic associations (Stuart, 2010).

This is relevant for policy practitioners and academic scholars as well. They are both concerned with checking whether a problem has been solved, moderated, unaffected, or even exacerbated by the government's intervention. The first are interested in that assessment for prescriptive reasons, while the latter because evaluation remains one of the traditional phases of the policy cycle, and its insights feed back on any analyses of the preceding phases. Furthermore, that assessment is often connected to ethical and normative issues (Silverman et al., 2020), making that final stage of the cycle intrinsically political (Bovens et al., 2006).

Getting the comparisons right is a fundamental passage for those focusing on outcomes, as well as for those concerned with the politics of policy-making. Political (and policy) science was born comparative, and the issue of what is appropriate to compare has always been one of its methodological concerns (Sartori, 1991). The alternative is just what he called "parochialism" and "ad hoccism", that is, idiosyncratic explanations that cannot support scientific claims, let alone causal attribution. Going back to the topic that motivated the methodological exercise of this research, no one doubts that the effectiveness of a country's anti-COVID strategy cannot be assessed by comparing a nation in the middle of Europe to some isolated island at its antipodes – say New Zealand and Japan. However, selecting comparable cases and occurrences within a supposedly homogeneous continent is already a more complex issue, and simply running regression models on any available data is not the best strategy to avoid biased results.

Matching is a way to identify bottom-up sensible comparisons, especially when the researcher is confronted with new and complex issues that need to be approached but that are outside his or her comfort zone (and there is no doubt that COVID policies are one of those cases). It could also be a way of selecting the most similar cases for a successive qualitative analysis (Nielsen, 2014). The synthetic control method further helps produce comparable units whenever the available observations share too little and cannot even approximate the requirements of a most similar research design. The method gives its best results when only one unit is treated, a circumstance in which many other techniques fail or cannot be applied, but it can be implemented also in other settings (Abadie, 2021).

Both techniques can be conceived as some sort of "augmented comparison", helping the researcher to satisfy as close as possible those that are long known prerequisites for comparatively testing any hypothesis. Their application is not even an issue of the Qual/Quant divide (Plumper et al., 2019), since "what makes a statistical treatment theoretically significant has nothing to do with statistics" (Sartori, 1970: 1037). Counterfactuals are the essence of causal attribution (Brady, 2008; Paul, 2009), and the search for the most similar world should matter for any kind of comparison.

6 References

Abadie, A. (2021) 'Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects', *Journal of Economic Literature* 59(2): 391-425.

Abadie, A., Diamond, A. and Hainmueller, J. (2010) 'Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program', *Journal of the American Statistical Association* 105(490): 493-505.

Abadie, A., Diamond, A. and Hainmueller, J. (2015) 'Comparative Politics and the Synthetic Control Method', *American Journal of Political Science* 59(2): 495-510.

Abadie, A. and Gardeazabal, J. (2003) 'The Economic Costs of Conflict: A Case Study of the Basque Country', *The American Economic Review* 93(1): 113-132.

Alfano, V., Ercolano, S. and Cicatiello, L. (2020) 'A Synthetic Control Method Analysis of Schools Opening and Covid-19 Outbreak in Italy', *CESifo Working Papers*. Munich: Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute.

Andersson, S. and Aylott, N. (2020) 'Sweden and Coronavirus: Unexceptional Exceptionalism', *Social Sciences* 9(12).

Angrist, J.D. and Pischke, J.r.-S. (2009) *Mostly harmless econometrics : an empiricist's companion.* Princeton: Princeton University Press.

Auger, K.A., Shah, S.S., Richardson, T., et al. (2020) 'Association Between Statewide School Closure and COVID-19 Incidence and Mortality in the US', *Journal of the American Medical Association* 324(9): 859-870.

Beck, N. and Katz, J.N. (1996) 'Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models', *Political Analysis* 6: 1-36.

Blackwell, M., Iacus, S., King, G., et al. (2009) 'cem: Coarsened exact matching in Stata', *The Stata Journal* 9(4): 524-546.

Bovens, M., Hart, P.t. and Kuipers, S. (2006) The politics of policy evaluation. In: Moran M, Rein M and Goodin RE (eds) *The Oxford handbook of public policy*. Oxford: Oxford University Press, pp.319-335.

Brady, H.E. (2008) Causation and Explanation in Social Science. In: Box-Steffensmeier JMB, Henry E. and Collier D (eds) *Oxford Handbook of Political Methodology*. Oxford: Oxford University Press.

Brandal, L.T., Ofitserova, T.S., Meijerink, H., et al. (2021) 'Minimal transmission of SARS-CoV-2 from paediatric COVID-19 cases in primary schools, Norway, August to November 2020', *Euro Surveill* 26(1).

Brusselaers, N., Steadson, D., Bjorklund, K., et al. (2022) 'Evaluation of science advice during the COVID-19 pandemic in Sweden', *Humanities and Social Sciences Communications* 9(1).

Busa, F., Bardanzellu, F., Pintus, M.C., et al. (2021) 'COVID-19 and School: To Open or Not to Open, That Is the Question. The First Review on Current Knowledge', *Pediatric Reports* 13(2): 257-278.

Capano, G. (2020) 'Policy design and state capacity in the COVID-19 emergency in Italy: if you are not prepared for the (un)expected, you can be only what you already are', *Policy and Society* 39(3): 326-344.
Capano, G., Howlett, M., Jarvis, D.S.L., et al. (2020) 'Mobilizing Policy (In)Capacity to Fight COVID-19: Understanding Variations in State Responses', *Policy and Society* 39(3): 285-308.
Cavallo, E., Galiani, S., Noy, I., et al. (2013) 'Catastrophic Natural Disasters and Economic Growth', *Review of Economics and Statistics* 95(5): 1549-1561.
Ceron, M., Palermo, C.M. and Grechi, D. (2021) 'Covid-19 Response Models and Divergences Within the EU: A Health Dis-Union', *Statistics, Politics and Policy*. DOI: 10.1515/spp-2021-0003.
Collins, J., Hall, N. and Paul, L.A. (2004) Causation and Counterfactuals. Boston: The MIT Press.
Coppola, D.P. and Ryan, B.J. (2020) 'Color-coding for COVID-19 Management: An Idea Whose Time has Come', *Journal of Homeland Security and Emergency Management* 17(2).
Cunha, M., Domingos, A., Rocha, V., et al. (2021) 'How many could have been saved? Effects of social distancing on COVID-19', *Revista De Administracao Publica* 55(1): 12-26.
Davies, N.G., Kucharski, A.J., Eggo, R.M., et al. (2020) 'Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study', *The Lancet Public Health* 5(7): e375-e385.
Dunlop, C.A., Ongaro, E. and Baker, K. (2020) 'Researching COVID-19: A research agenda for public policy and administration scholars', *Public Policy and Administration* 35(4): 365-383.
Flor, L.S., Friedman, J., Spencer, C.N., et al. (2022) 'Quantifying the effects of the COVID-19 pandemic on gender equality on health, social, and economic indicators: a comprehensive review of data from March, 2020, to September, 2021', *The Lancet*. DOI: 10.1016/s0140-6736(22)00008-3.
Folkhälsomyndigheten (2020) Covid-19 in schoolchildren. A comparison between Finland and Sweden.
Friedson, A.I., McNichols, D., Sabia, J.J., et al. (2021) 'Public Policy Implications of the COVID-19 Pandemic', *Journal of Policy Analysis and Management* 40(1): 258-283.
Gandini, S., Rainisio, M., Iannuzzo, M.L., et al. (2021) 'A cross-sectional and prospective cohort study of the role of schools in the SARS-CoV-2 second wave in Italy', *The Lancet Regional Health - Europe* 5.
Gelman, A. and Hill, J. (2007) *Data Analysis Using Regression and Multilevel Hierarchical Models* Cambridge: Cambridge University Press.
Giuliani, M. (2022) 'Complying with Anti-COVID Policies. Subnational Variations and their Correlates', *Rivista Italiana di Politiche Pubbliche* (2): forthcoming.
Goolsbee, A. and Syverson, C. (2021) 'Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020', *Journal of Public Economics* 193: 104311.

Gordon, D.V., Grafton, R.Q. and Steinshamn, S.I. (2021) 'Cross-country effects and policy responses to COVID-19 in 2020: The Nordic countries', *Economic Analysis and Policy* 71: 198-210.
Hale, T., Angrist, N., Goldszmidt, R., et al. (2021) 'A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)', *Nature Human Behaviour* 5(4): 529-538.
Herrera, M. and Godoy-Faúndez, A. (2021) 'Exploring the Roles of Local Mobility Patterns, Socioeconomic Conditions, and Lockdown Policies in Shaping the Patterns of COVID-19 Spread', *Future Internet* 13(5).
Iacus, S.M., King, G. and Porro, G. (2011) 'Multivariate Matching Methods That Are Monotonic Imbalance Bounding', *Journal of the American Statistical Association* 106(493): 345-361.
Iacus, S.M., King, G. and Porro, G. (2012) 'Causal Inference without Balance Checking: Coarsened Exact Matching', *Political Analysis* 20(1): 1-24.
Iacus, S.M., King, G. and Porro, G. (2019) 'A Theory of Statistical Inference for Matching Methods in Causal Research', *Political Analysis* 27(1): 46-68.
Isphording, I.E., Lipfert, M. and Pestel, N. (2020) 'School Re-Openings after Summer Breaks in Germany Did Not Increase SARS-CoV-2 Cases', *Discussion paper.* Bonn: IZA Institute for the Study of Labor.
Iwata, K., Doi, A. and Miyakoshi, C. (2020) 'Was school closure effective in mitigating coronavirus disease 2019 (COVID-19)? Time series analysis using Bayesian inference', *International Journal of Infectious Diseases* 99: 57-61.
King, G., Keohane, R.O. and Verba, S. (1994) *Designing Social Inquiry. Scientific Inference in Qualitative Research.* Princeton: Princeton University Press.
King, G., Lucas, C. and Nielsen, R.A. (2017) 'The Balance-Sample Size Frontier in Matching Methods for Causal Inference', *American Journal of Political Science* 61(2): 473-489.
Krishnaratne, S., Pfadenhauer, L.M., Coenen, M., et al. (2020) 'Measures implemented in the school setting to contain the COVID-19 pandemic: a scoping review', *Cochrane Database Syst Rev* 12: CD013812.
Lewis, D. (1973a) 'Causation', *The Journal of Philosophy* 70(17): 556-567.
Lewis, D. (1973b) *Counterfactuals.* Oxford: Blackwell.
Li, Y., Campbell, H., Kulkarni, D., et al. (2021) 'The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries', *The Lancet Infectious Diseases* 21(2): 193-202.
Liu, Y., Morgenstern, C., Kelly, J., et al. (2021) 'The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories', *BMC Medicine* 19(1).
Ludvigsson, J.F. (2020) 'Children are unlikely to be the main drivers of the COVID-19 pandemic - A systematic review', *Acta Paediatrica* 109(8): 1525-1530.

Manica, M., Guzzetta, G., Riccardo, F., et al. (2021) 'Impact of tiered restrictions on human activities and the epidemiology of the second wave of COVID-19 in Italy', *Nature Communication* 12(1): 4570.

McConnell, A. and Stark, A. (2021) 'Understanding policy responses to COVID-19: the stars haven't fallen from the sky for scholars of public policy', *Journal of European Public Policy* 28(8): 1115-1130.

Mill, J.S. (1843) *A System of Logic, Ratiocinative and Inductive, being a Connected View of the Princip0les of Evidence, and the Methods of Scientific Investigation.* London: John W. Parker.

Milne, R. (2021) Sweden no longer stands out in pandemic, says architect of 'no lockdown' policy. *Financial Times,* 12 November.

Mitze, T., Kosfeld, R., Rode, J., et al. (2020) 'Face Masks Considerably Reduce Covid-19 Cases in Germany. A Synthetic Control Method Approach', *CESifo Working Papers.* Munich: Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute.

Negri, F. (2022) Correlation is not causation, yet… Matching and weighting for better counterfactuals. In: Damonte A and Negri F (eds) *Causality in Policy Studies. A Pluralist Toolbox*. Cham: Springer Nature.

Nielsen, R.A. (2014) 'Case Selection via Matching', *Sociological Methods & Research* 45(3): 569-597.

Nielsen, R.A. (2020) Statistical Matching with Time-Series Cross-Sectional Data: Magic, Malfeasance, or Something in between? . In: Curini L and Franzese RJ (eds) *The SAGE Handbook of Research Methods in Political Science and International Relations*. London: Sage, pp.805-821.

Park, Y.J., Choe, Y.J., Park, O., et al. (2020) 'Contact Tracing during Coronavirus Disease Outbreak, South Korea, 2020', *Emerging Infectious Diseases* 26(10): 2465-2468.

Pashakhanlou, A.H. (2021) 'Sweden's coronavirus strategy: The Public Health Agency and the sites of controversy', *World Medical Health Policy*. Epub ahead of print 2021/07/07. DOI: 10.1002/wmh3.449.

Paul, L.A. (2009) Counterfactual Theories. In: Beebee H, Hitchcock C and Menzies P (eds) *The Oxford Handbook of Causation*. Oxford: Oxford University Press.

Pickett, M. (2021) Sweden's Pandemic Experiment. *The New Yorker,* 6 April.

Pierre, J. (2020) 'Nudges against pandemics: Sweden's COVID-19 containment strategy in perspective', *Policy and Society* 39(3): 478-493.

Plumper, T., Troeger, V.E. and Neumayer, E. (2019) 'Case selection and causal inferences in qualitative comparative research', *PLoS One* 14(7): e0219727.

Rauscher, E. and Burns, A. (2021) 'Unequal Opportunity Spreaders: Higher COVID-19 Deaths with Later School Closure in the United States', *Sociological Perspectives*. DOI: 10.1177/07311214211005486.

Sartori, G. (1970) 'Concept Misformation in Comparative Politics', *The American Political Science Review* 64(4): 1033-1053.

Sartori, G. (1991) 'Comparing and Miscomparing', *Journal of Theoretical Politics* 3(3): 243-257.

Schultz, T. (2021) Do Sweden's COVID liberties cost lives? *Deutsche Welle*, 21 April.

Silverman, M., Sibbald, R. and Stranges, S. (2020) 'Ethics of COVID-19-related school closures', *Canadian Journal of Public Health* 111(4): 462-465.

Stage, H.B., Shingleton, J., Ghosh, S., et al. (2021) 'Shut and re-open: the role of schools in the spread of COVID-19 in Europe', *Philosophical Transactions of the Royal Society B-Biological Sciences* 376.

Stone, D. (2020) *Counting. How We Use Numbers to Decide What Matters.* New York: Liveright.

Stuart, E.A. (2010) 'Matching methods for causal inference: A review and a look forward', *Stat Sci* 25(1): 1-21.

Tian, T., Tan, J., Luo, W., et al. (2021) 'The Effects of Stringent and Mild Interventions for Coronavirus Pandemic', *Journal of the American Statistical Association* 116(534): 481-491.

Toshkov, D., Carroll, B. and Yesilkagit, K. (2021) 'Government capacity, societal trust or party preferences: what accounts for the variety of national policy responses to the COVID-19 pandemic in Europe?', *Journal of European Public Policy*. DOI: 10.1080/13501763.2021.1928270.

Tosi, D. and Campi, A.S. (2021) 'How Schools Affected the COVID-19 Pandemic in Italy: Data Analysis for Lombardy Region, Campania Region, and Emilia Region', *Future Internet* 13(109): 1-12.

Weible, C.M., Nohrstedt, D., Cairney, P., et al. (2020) 'COVID-19 and the policy sciences: initial reactions and perspectives', *Policy Sciences* 53(2): 225-241.

Wenander, H. (2021) 'Sweden: Non-binding Rules against the Pandemic - Formalism, Pragmatism and Some Legal Realism', *European Journal of Risk Regulation* 12(1): 127-142.

Yang, X. (2021) 'Does city lockdown prevent the spread of COVID-19? New evidence from the synthetic control method', *Global Health Research and Policy* 6(1).