Discuss this article at JournalTalk:
https://journaltalk.net/articles/6137

ECON JOURNAL WATCH 23(1)
March 2026: 69–87

A Critique of Synthetic Control Method Studies on Covid-19 Policy—Evidence from Sweden

Jonas Herby[1]

LINK TO ABSTRACT

Sweden's low excess mortality and the SCM context

In the spring of 2020, Sweden, led by state epidemiologist Anders Tegnell,[2] opted against a mandatory lockdown, a decision that drew significant international attention and academic scrutiny. During the pandemic, Sweden became a focal point in the public debate over lockdowns, with both proponents and opponents of mandatory restrictions invoking the Swedish case to support their positions. The five synthetic control method (SCM) studies examined here can be seen as attempts to formalize one side of that debate with econometric methods. Although Sweden experienced substantial COVID-19-related mortality during the initial phase of the pandemic, particularly in early 2020, its cumulative excess mortality over 2020–2021 appears low by European standards and broadly comparable to the range implied for the Nordic countries across most estimation approaches. Comparative analyses show that Sweden's excess mortality estimates are sensitive to modelling choices and baseline assumptions, but that—outside of clear outlier models—Sweden's

[1] Economist (MSc.), Copenhagen. jherby@gmail.com.

[2] Sweden's constitution protected its Public Health Agency from government overruling (Jonung and Hanke 2020). This reflects a broader pattern where constitutional incentives—specifically executive power gains rather than health data—drove emergency declarations globally (Bjørnskov and Voigt 2022).

excess mortality does not stand out as high relative to its neighbours (Kepp et al. 2022; Herby 2025b, 7.1).[3]

Nevertheless, the Swedish case has spurred a substantial literature attempting to quantify the effect of a counterfactual lockdown in the short term. Five studies use the Synthetic Control Method (SCM). They typically focus on early pandemic outcomes and treat Sweden as a natural experiment, yet they frequently abstract from longer-run mortality patterns and the considerable methodological uncertainty surrounding excess-mortality estimation:

- Martin J. Conyon and Steen Thomsen (2021) conclude—based on a 24-day pre-intervention window—that "stricter lock-down policies are associated with fewer COVID-19 infections and deaths."
- Sang-Wook Cho (2020) uses a 25-day pre-intervention period to find that "lockdown measures would have been associated with a lower excess mortality rate in Sweden by 25 percentage points".
- Benjamin Born, Alexander M. Dietrich, and Gernot J. Müller (2021) employ a 13-day pre-intervention period and find that "a 9-week lockdown in the first half of 2020 would have reduced [...] deaths by about 38%".
- Chiara Latour, Franco Peracchi, and Giancarlo Spagnolo (2022), based on a 15-day pre-intervention window, conclude that "a lockdown would have had sizable effects within one week".
- Florian Ege, Giovanni Mellace, and Seetha Menon (2023), using excess mortalities over a 20-week pre-intervention window, found that "not imposing a mandatory lockdown resulted in [...] a substantial increase in mortality".

Table 1 summarizes the key features of these five studies.

I present an empirical argument that challenges the pro-lockdown narrative: the large reported effects of lockdowns in SCM studies are artifacts of unbalanced time-varying confounding and early-shock omitted-variable bias, mechanically absorbed into model fittings. Their findings are artifacts of short pre-lockdown windows, which are epidemiologically uninformative. My critique helps us understand the conditions under which SCM is a useful method to derive causal estimates of rare or unique events and policies.

I concur that Sweden was an extraordinary case well worth studying. Also, SCM-based designs may present an improvement over dire (and misleading) early

[3] See for example: link.

forecasts, though as the analysis below suggests, the method's reliability in pandemic settings is far from assured. The early forecasts embedded strong behavioral null assumptions, including predictions from Imperial College London's Report 9 estimating that over two million could die in the United States absent lockdown (Ferguson et al. 2020). And SCM studies may advance the debate beyond the influential estimations in which identifying effects requires the strong assumption that voluntary behavioral responses are negligible (Flaxman et al. 2020; The Royal Society 2023); these influential studies attributing large effects to lockdowns are distinguished by lacking a control group altogether.[4]

I contend that, with the short pre-intervention window, the SCM mistakes noise in the data for large treatment effects. Constrained to a few weeks of pre-lockdown data, the SCM fails to account for the fact that the spread of the virus—and thus the timing of the outbreak—varied significantly across the donor countries. Without sufficient data to correct for these unobserved variables, the SCM cannot reliably identify the true impact of the policy. The claimed effects are likely driven by this lack of data and the specific timing of the outbreak in the donor pool, rather than by the lockdown itself. The design is unable to distinguish the average treatment effect on the treated (ATT) from confounded early pandemic shock composition.

The Synthetic Control Method (SCM)

The SCM is a specialized variant of difference-in-difference analysis which serves as an econometric tool for policy evaluation. Noted for its interpretability and transparency, the SCM is particularly valuable in constructing counterfactual scenarios to assess the impact of interventions without randomized control trials, provided that a sufficiently long pre-intervention period is available to establish a credible synthetic control (Abadie 2021).

In the case of a counterfactual lockdown in Sweden, the SCM is employed to construct a 'synthetic control' (often termed a 'doppelgänger') that hypothetically adheres to lockdown measures. The doppelgänger enables a comparison between the actual pandemic trajectories observed in Sweden and those that SCM estimates would have occurred under lockdown conditions. It is worth noting that this SCM design involves a conceptual inversion: the donor countries are those that implemented the intervention (lockdown), while Sweden—nominally the "treated"

[4] Despite asserting that non-pharmaceutical interventions (NPIs) "can provide powerful, effective and prolonged reductions in viral transmission," the Royal Society did not assess the causal impact of compulsory COVID-19 mandates in isolation. Instead, its analysis considered the overall societal response to the pandemic. As a consequence, the reported estimates conflate the effects of legally mandated restrictions with contemporaneous voluntary behavioral adjustments (Hanke et al. 2023).

unit—is the jurisdiction that maintained the status quo.

The construction of the synthetic control involves selecting a donor pool of potential control units (countries) and then determining a set of non-negative weights that minimize the difference between the treated unit and the synthetic control in terms of pre-intervention characteristics and outcomes. This weighted combination forms a synthetic control that serves as a benchmark for comparison during the post-intervention period. The causal effect of the intervention is estimated by comparing the post-intervention outcomes of the treated unit with those of the synthetic control.

SCM has been widely applied due to its ability to handle complex longitudinal data and provide suggestive results. Alberto Abadie (2021) underscores the significance of a comprehensive pre-intervention period: "it is of crucial importance to collect information on the affected unit and the donor pool for a large pre-intervention window" (Abadie 2021). The term "large pre-intervention window" is inherently ambiguous; however, treating a pre-intervention window of a few weeks as "large" warrants critical examination.

Beyond the application-specific concerns, recent methodological research has identified several structural vulnerabilities of the SCM itself, including susceptibility to p-hacking via donor pool selection (Francis 2025a), an index number problem (Francis 2025b), specification searching through pre-treatment indicator selection (Ferman et al. 2020), and the risk of spurious regression (Masini and Medeiros 2022). The criticisms do not necessarily apply to every SCM exercise, but they suggest that the SCM is structurally vulnerable to producing misleading results in settings with short pre-intervention windows and highly volatile outcomes.

The inadequacy of short pre-intervention windows

Studies assessing the impact of a counterfactual lockdown in Sweden using COVID-19 data, such as positive cases or deaths, are constrained by the reality of the pandemic to rely on very short pre-intervention windows. This limitation arises because countries in the donor pool typically implemented lockdowns within weeks after their first national COVID-19 case/death was reported. The short pre-intervention periods severely restrict the effectiveness of the SCM.

I have identified five studies applying the SCM to assess the consequences of Sweden's choices during the COVID-19 pandemic. The studies differ considerably in their methodological choices, including the definition of the treatment date (e.g. average lockdown date (Born et al. 2021) and peak stringency date (Cho 2020)), treatment of dates (some studies use calendar dates for the intervention

while others (Born et al. 2021; Cho 2020) align pandemic curves to epidemiological milestones such as reaching a threshold of cases per capita), the outcome variable (confirmed cases, COVID-19 deaths, or excess mortality), and the length of the pre- and post-intervention windows (See Table 1 for an overview). These choices can materially affect the SCM estimates.

TABLE 1: Overview of SCM-based estimates of lockdown effects on COVID-19 outcomes in Sweden

Study	Publication	Outcome variable(s)	Donor pool	Predictor variables	Period covered	Key findings & inference
Born, Dietrich & Müller (2021)	*PLoS ONE*	(A) Daily COVID-19 infections (log); (B) Daily COVID-19 deaths (log)	13 countries: 12 Western EU (pop. >1 million) + Norway. Key donors (Spec. A): Denmark (30.0%), Finland (25.3%), Netherlands (25.8%), Norway (15.0%), Spain (3.9%)	Log infections/deaths for 13 pre-lockdown days, population size, share of population >65 years, share of urban population	Pre-intervention: 13 days. Post-intervention: to 1 Sep 2020 (first wave). Counterfactual lockdown: 9 weeks (15 Mar–17 May 2020). Stringency = 68	Spec. A (infections): ~75% reduction. Spec. B (deaths): ~38% reduction. Range across specifications: –27% to –77% (infections), –26% to –82% (deaths). Effect delay: 3–4 weeks. RMSPE values reported in S4 Table (S1 File), but permutation p-value never computed. No proper significance test.
Cho (2020)	*Econometrics Journal*	Cumulative confirmed cases per million; excess mortality (weekly, by age group)	30 European countries (EU members excl. Malta, plus Iceland, Israel, Norway, Switzerland, UK). Aligned to day when cases/million exceed 1. Key donors: Finland (0.49), Greece (0.24), Norway (0.22), Denmark (0.03), Estonia (0.02)	Lagged infection cases (Days 7, 14, 23), average deaths, population density, urban population fraction, average household size	Pre-intervention: 25 days. Post-intervention: ~75 days (Feb–Jun 2020, first wave). Treatment date: Day 25 (peak stringency index)	~75% reduction in infections at 100 days. 25 pp. reduction in excess mortality (total pop.), up to 29 pp. for 85+ cohort. p = 0.097 (benchmark). Effect delay: ~5 weeks. Statistical significance on infections emerges after day 100.
Conyon & Thomsen (2021)	Working paper, CBS / Bentley University	Total cases per million; total deaths per million	15 European countries: Austria, Belgium, Czechia, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, UK. Key donors: Finland (0.475), Denmark (0.219), Greece (0.203), Norway (0.103)	Total deaths/million (days 1–20), lagged total cases/million (Days 7, 14, 21), population density, median age, diabetes prevalence, stringency index	Pre-intervention: 24 days (wave 1), 25 days (wave 2). Post-intervention: 180 days (wave 1, Mar–Sep 2020). Also models second wave (Nov 2020). Covers both waves of 2020	Wave 1: Sweden ~160–200% higher mortality than synthetic control at 100–180 days. Wave 2 (Nov 2020) also shows divergence. RMSPE = 2.81 (pre-treatment fit, infections). No proper significance test.

Ege, Mellace & Menon (2023)	*Scientific Reports* (Nature)	Cumulative weekly excess mortality (5-year avg. death counts) and excess death rate (Lee-Carter model)	29 countries from HMD STMF data (New Zealand, South Korea, Israel, Canada and European countries). Key donors: Norway (29.6%), Denmark (19.6%), New Zealand (17.0%), Lithuania (17.5%), Belgium (12.3%), Finland (3.8%)	Pre-treatment cumulative excess deaths (average and weekly values), GDP per capita, cardiovascular death rate, life expectancy, population density, urban population	Pre-intervention: 20 weeks (Nov 2019–Mar 2020). Post-intervention: to end Sep 2020 (first wave). Slow response: treatment at Week 22 (late Mar); rapid response: Week 20	~3,420 avoidable deaths (avg. effect ≈ 30 deaths/100,000). Peak at ~4,411 deaths (Week 35). Effect driven by 65+ age group. Results confirmed by augmented SCM. Results insignificant (placebo test: p ≈ 0.17).
Latour, Perac-chi & Spag-nolo (2022)	*PLoS ONE*	8 indicators: cumulative and daily COVID-19 infections, COVID-19 deaths, adjusted COVID-19 deaths (corrected by excess mortality), and positive rate	Baseline: 11 European countries (Denmark, Finland, Germany, Greece, Ireland, Netherlands, Norway, Portugal, Austria, Belgium, Italy—France and Spain excluded due to data problems). Weights re-optimized per indicator	Pre-treatment outcome values (15 days; 13 for positive rate), population size, share of urban population. Weights re-optimized separately for each of the 8 outcome indicators	Pre-intervention: 15 days (13 days for positive rate). Treatment date: 17 Mar 2020 (mean lockdown start). Post-intervention: 105 days to 30 Jun 2020 (first wave). 18-day lag applied for death outcomes	Sizable effects within ~1 week (not 3–5 weeks). Longer delay in prior studies driven by Sweden's low testing frequency. 61% reduction in infections by 17 May; 40% reduction in COVID-19 deaths; 41% reduction in adjusted COVID-19 deaths. Results robust across different donor pools, treatment dates, and predictor sets (4 robustness cases). No formal placebo test or p-values reported.

Notes: SCM = Synthetic Control Method. DD = Difference-in-Differences. RMSPE = Root Mean Square Prediction Error. pp. = percentage points. All studies focus on Sweden's decision not to impose a mandatory lockdown during the first wave of COVID-19 (Spring 2020). Donor pool weights reported are for the baseline/benchmark specification of each study. The table only includes SCM, but some studies also report DD estimates. Only Cho and Ege compute permutation p-values from placebo tests; Born reports RMSPE values without computing p-values; Conyon & Thomsen and Latour report no formal significance test.

Some studies seek to extend the pre-intervention window by using excess mortality, which can be estimated decades prior to COVID-19 (Ege et al. 2023). However, this approach raises several issues. First, there is little reason to believe that the initial spread of COVID-19 during the early days of the pandemic correlates with cumulative excess mortality. The COVID-19 pandemic in Europe began in March, which is not comparable to earlier seasons, where deaths from influenza or other respiratory viruses in a typical winter season usually peak months before COVID-19 deaths peaked.

Attempts to mitigate the short pre-intervention window by aligning pandemic curves based on specific milestones—such as the date when infections exceed 1 per 1 million inhabitants (Born et al. 2021; Cho 2020)—introduce a critical flaw: while they align[5] the epidemiological data, they fundamentally misalign

[5] It is important to note that the number of confirmed positive tests was heavily contingent on each country's specific testing strategy and capacity at the time. Consequently, aligning curves based on reported

the information available to decision-makers. At the onset of the pandemic, governments possessed very little actionable information about local spread due to severely limited testing capacities; they could not react based on internal infection metrics. Instead, several studies demonstrate that government policies were strongly driven by the policies initiated in neighboring countries rather than by the severity of the pandemic in their own nations (Sebhatu et al. 2020; Engler et al. 2021; Mistur et al. 2023). This finding indicates that political responses were largely a product of imitating neighbors and reacting to shared international interpretation. In the same vein, private responses—voluntary behavioral changes in which individuals take action to protect themselves regardless of mandates—were likely driven by this same diffusion of interpretation. By temporally shifting pandemic curves to match epidemiological starting points, researchers inadvertently treat countries as isolated experiments, ignoring that a country acting later in March 2020 did so with significantly more knowledge of the unfolding crisis—triggering earlier voluntary behavioral changes—than a country acting weeks earlier.

The challenge of a very short pre-intervention window also brings to the forefront a seemingly simple yet complex question: When did the lockdown in donor countries begin? Taking Denmark as an example, the timeline of lockdown measures is multifaceted. Governmental indoor cultural institutions, libraries, and leisure facilities were closed by Friday, March 13. Schools followed by Monday, March 16, and bars and restaurants were closed and public gatherings limited to 10 people by Wednesday 18, at 10:00 AM. Consequently, pinpointing the exact start date of the Danish lockdown—whether it is Friday the 13, Monday the 16, or Wednesday the 18—is non-trivial. Given the very short pre-intervention window, the choice of lockdown start date can significantly influence the SCM estimates.

There is no viable workaround for this limitation of the Synthetic Control Method within the context of the early pandemic. For COVID-19 outcomes, we simply do not possess anything resembling a large pre-intervention window—a fundamental requirement that the SCM was never designed to function without. In a COVID-19 setting, the method carries an inherent and high risk of failing to balance unobserved, time-varying confounders, such as the variable pre-lockdown viral spread (Abadie 2021). The likely result is estimates that reflect model-fit artifacts driven by short-term noise rather than true, policy-induced mortality effects.

The proof of inadequacy

A simple way to illustrate the problems with the SCM is to observe the method's estimated effect of a counterfactual lockdown in regions with the exact

cases does not necessarily reflect comparable levels of actual viral prevalence.

same no-lockdown policy but differences in viral spread before the intervention was implemented.

I use a simple instrument to divide Sweden into four hypothetical countries. The timing of the winter holiday significantly influenced the spread of the virus in Europe because ski tourists brought the virus home from the Alps. Regions with winter holidays in week 9 have been shown to experience higher viral spread because of a large unrecognized outbreak at the end of February (Arnarson 2021; Björk et al. 2021). In a recent paper, I use this result as an instrument to capture variations in pre-lockdown viral transmission across different Swedish regions (län) by dividing Sweden into four hypothetical countries—"Sweden w7," "Sweden w8," "Sweden w9," and "Sweden w10"—based on each region's winter holiday schedule (Herby 2025a).

All four hypothetical countries maintained the same COVID-19 policies, notably abstaining from lockdowns. If the decision against lockdowns were an important cause of excess mortality, we would expect the SCM-derived treatment effects (ATTs) to be of similar magnitude across all four hypothetical Swedens, even if mortality levels differ due to regional factors such as hospital capacity or demographic composition. Conversely, if lockdowns had little effect, we would not expect similar ATTs across regions. Large dispersion in ATT estimates across the four hypothetical Swedens would indicate that the SCM estimator is sensitive to variation in pre-lockdown viral seeding and other time-varying unobservables, rather than capturing a genuine policy effect.

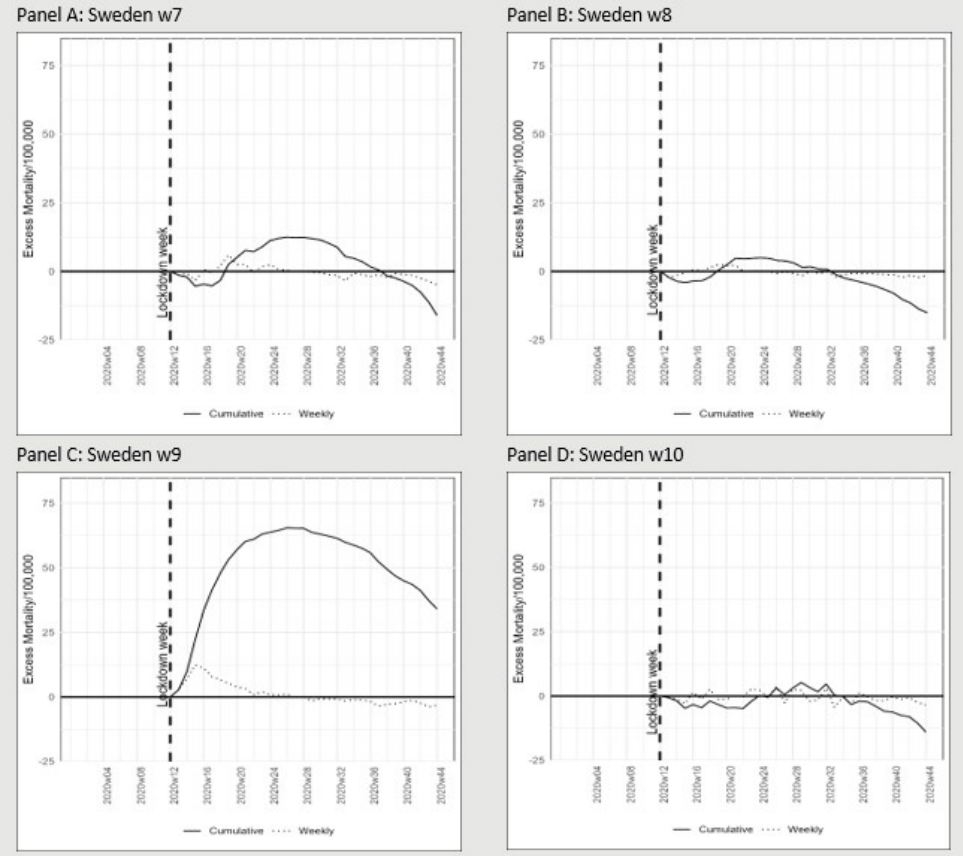
Figure 1 presents the outcomes when the SCM is run for each of the four hypothetical Swedens. My donor pool consists of 30 Northern Hemisphere countries (excluding the UK, whose early pandemic approach resembled Sweden's, and Italy, where the first regional lockdowns began in February). The pre-intervention period extends from week 27 of 2016 to week 8 of 2020, and the matching variables include mortality rates, hospital beds per 1,000 people, urban population share, GDP per capita (PPP), share of population aged 65 and above, migrant share, and seasonal mortality rates for 2016/17 through 2019/20 (see Herby (2025a) for full details). Note that the naïve specification, where I run the SCM for Sweden as a whole, finds effects similar to Ege et al. and other SCM studies (Herby 2025a).

Figure 1 shows that the estimated effect of a counterfactual lockdown is substantial in Sweden w9, but notably smaller and even negative by the end of the summer in Sweden w7, Sweden w8, and Sweden w10. If SCM were robust, the estimated effects should be similar. The fact that they diverge substantially and only show a notable (yet insignificant) effect for Sweden w9 demonstrates that the method is most likely capturing the timing of viral seeding, not the effect of lockdowns.

The findings in Figure 1 suggest that the estimates in SCM studies are likely

influenced by unobserved factors, such as the extent of COVID-19 transmission prior to lockdown. Full methodological details—including donor weights, predictor balance tables, and robustness checks—are available in Herby (2025a).

Figure 1: Post-intervention excess mortalities for hypothetical countries



Note: The vertical dashed line illustrates the timing of the lockdown in synthetic Sweden. Excess mortalities have been normalized to zero in week 12 of 2020. Y-axes have been standardized across all four panels to facilitate comparison.
Source: Figure 7 (Herby 2025a).

The seemingly large effect in Sweden w9 appears to confirm the significance of the winter holiday's timing. This result is closely linked to observed excess mortalities in the hypothetical countries, as most of the excess mortality in the spring of 2020 was related to the Stockholm area, where the winter holiday falls in week 9. The Stockholm region's early and intense viral seeding—possibly driven by ski tourism in week 9 (Arnarson 2021; Björk et al. 2021)—would produce divergent mortality trajectories regardless of lockdown policy.

Furthermore, one rationale for lockdowns was to delay deaths until vaccines

became available; the naïve specification accordingly extends the post-intervention period through week 26 of 2021 (when vaccines were widely distributed), and Panel C of Figure 3 in Herby (2025a) shows that the cumulative effect for Sweden reverses over the longer term. The small effects in Sweden w7, Sweden w8, and Sweden w10 are in line with results from studies establishing a control group (Herby et al. 2024).

While the regional analysis demonstrates the method's sensitivity to viral seeding, an examination of the reported effect timing reveals further inconsistencies with biological reality.

Neglecting biological reality and statistical rigor

Several of the SCM studies mentioned above exhibit clear signs of misspecification. I will here point to two obvious signs: biologically implausible effects and failing a proper test for significance.

Biologically implausible effects

It is well established that there is considerable variability in the time from infection with COVID-19 to death. Virtually no one dies within the first week of infection, but thereafter the probabilities climb quite steeply (Wood 2022). On average, it takes 26 days from infection to death, and 29% die within the third week after infection (day 15 to 21) while 26% die within the fourth week after infection (day 22 to 28). It is implausible that one would see a large effect of lockdowns on deaths within the first 14 days (Herby et al. 2024, app. supplementary file2).[6]

Nevertheless, several studies find large and immediate effects. For example, one study finds that nationwide school closures reduce COVID-19 deaths by 75% after just 10 days without even commenting on the biological reality that COVID-19 deaths cannot respond markedly to policy within this time frame (Neidhöfer and Neidhöfer 2020). Another study believes their "results are comparable to more recent studies using Covid-19 deaths as outcomes, where effects are seen within a week after the introduction of lockdown in synthetic Sweden" (Ege et al. 2023, 7). Here Ege et al. cite another study finding very early effects also using Sweden as a counterfactual. A much more obvious explanation is that when such effects appear, SCM is not estimating prevented deaths, but instead fitting to omitted variables.

[6] García-García et al. (2021) find a somewhat different distribution, where approximately 20% die within 14 days ((Wood 2022): 11%), 42% die on day 15 to 21 (29%), and 25% on day 22 to 28 (26%). However, the overall conclusion that the effect of lockdowns is not visible in mortality data during the first 14 days remains.

A plausible explanation is that the SCM, constrained by short and epidemiologically uninformative pre-intervention windows, is capturing variation in unobserved confounders—such as viral seeding intensity—rather than genuine policy effects. That would account for the appearance of "effects" before the virus could plausibly have killed anyone.

Failing a proper test for significance

In SCM policy evaluations, statistical significance is assessed through placebo tests: The SCM is re-estimated under pseudo-treatment assignment to each donor unit, generating a distribution of placebo ATTs. The treated unit's ATT is then compared against this distribution. Standard practice requires either screening or weighting placebo units by pretreatment fit quality, or using a fit-adjusted statistic, such as the post/pre RMSPE ratio, to avoid mechanically large gaps from poorly fitting placebos (Abadie et al. 2015, 2010).

Two of the cited SCM studies conduct donor-wise placebo tests by re-estimating SCM under sequential pseudo-treatment assignment to each country in the donor pool, generating a distribution of placebo SCM effect estimates. It should be noted that such placebo tests are themselves vulnerable to specification searching and p-hacking, as the researcher retains discretion over the donor pool composition, the outcome variable, and the time horizon at which significance is assessed (Francis 2025a).

Ege et al. interpret their placebo exercise as evidence that Sweden's estimated treatment effect is unlikely to arise by chance and is "clearly driven" by the intervention (Ege et al. 2023). However, their conclusion does not respect the criterion that is relevant for making an inference, namely, *that Sweden's estimated ATT must differ materially from the ATT magnitudes that SCM would have assigned to donors under identical estimation design (including identical intervention timing)*. Their results show, in fact, that the relevant criterion is not met.

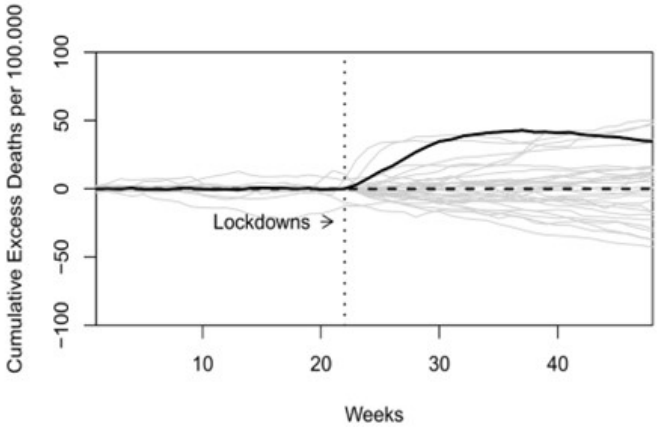
The key metric for policy evaluation is the end-of-period ATT—the cumulative impact by the time the treated and synthetic units can be compared over a meaningful horizon—not the peak transitory divergence that Ege et al. emphasize.

Examining the end-of-period treatment effect, at least four of 29 donor countries produce placebo effect estimates as large as (or larger than) Sweden's own estimate. Following Abadie (2021) and including Sweden in the denominator yields a p-value of approximately $5/30 = 0.17$—far above conventional significance thresholds. The mortality ATT reported in Ege et al. does not pass the usual significance threshold (< 5% more-extreme placebos) defined in standard statistical practice. While Ege et al. claim statistical significance for their mortality estimate, their own placebo exercise contradicts their stated conclusion.

Cho estimates a counterfactual lockdown effect on confirmed cases (Cho

2020). Cho's donor-pool placebo test shows that Sweden's estimated ATT on case trajectories is statistically indistinguishable from placebo effects for the first 90+ days, with significance emerging only after 100 days, in mid-June 2020.

Figure 2: Placebo tests on excess deaths from Ege et al.



Source: Figure 2 (Ege et al. 2023)

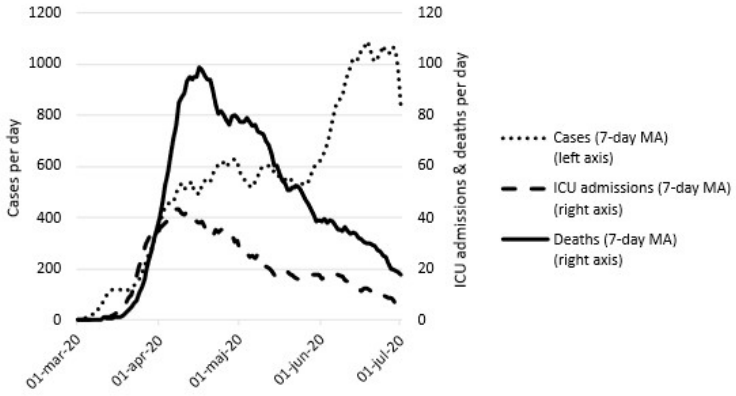
The result is internally coherent within the SCM permutation framework, but its policy relevance is debatable. As Figure 3 illustrates, cases rose sharply in June while deaths and ICU admissions were continuously falling. Is the model detecting a causal policy signal in mortality-relevant infections, or merely the mechanical consequences of a changing case-detection regime?

The remaining three SCM studies provide no donor-based placebo distributions to assess whether Sweden's estimated NPI effect could be a chance result of donor composition.

To summarize, early-wave COVID-19 mortality shocks were not unique to countries that skipped lockdowns. Belgium—which had its winter holiday in week 9, the same week as the Stockholm region—experienced among the highest first-wave mortality in Europe, not because of inadequate lockdowns, but rather because of intense pre-lockdown viral seeding. A falsification that instructs SCM to treat nationwide lockdown absence for Belgium, Italy, Spain, or the United Kingdom would, by design, yield large "death-mitigating" ATT gaps driven by pandemic shock timing and unobserved confounders.

The placebo tests do not establish statistical significance of the policy effect. They establish the opposite: SCM's dependence on narrow pre-windows and early-shock fitting in a pandemic setting leaves the design highly exposed to omitted variables. Perhaps that unfortunate dependence reflects a fundamental limitation

Figure 3: Development in COVID-19 cases, ICU admissions, and deaths in Sweden (Mar–Jun, 2020)



Source: Folkhälsomyndigheten

of the SCM, or perhaps it reflects its misapplication in pandemic settings. In any case, the "when" and "where" of mortality acceleration are generated, not by the (absence of) lockdown, but by the dependence on narrow pre-windows and early-shock fitting.

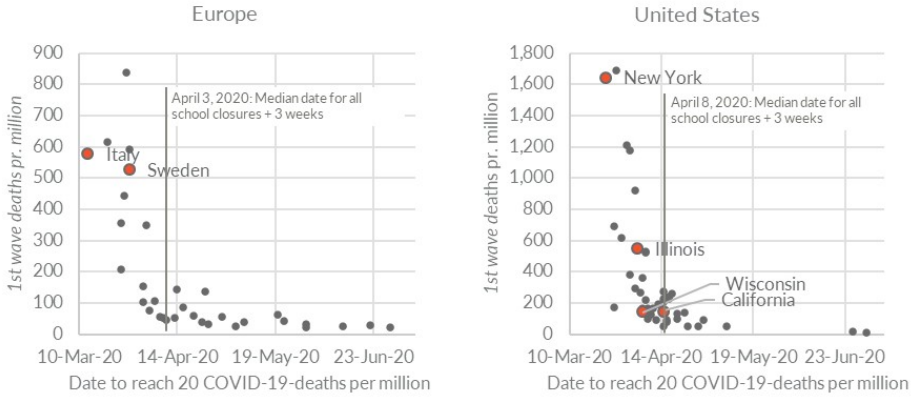
Amplifying biases

A biased estimate for a single jurisdiction would not, in principle, invalidate the broader literature if studies existed that were symmetrically biased in the opposite direction. If SCM were applied indiscriminately across many heterogeneous units, stochastic misspecifications could cross-cancel, allowing the law of large numbers to moderate model error. One might even imagine a deliberate placebo design: a synthetic "no-lockdown Belgium" to test whether SCM mechanically produces large positive policy effects when the counterfactual is known to be false.

Such a theoretical safeguard holds, however, only under one condition: geographically broad and variation-rich application. In practice, SCM pandemic studies fall far short of this ideal. A recent meta-analysis identified eleven estimates of the effect of lockdowns (Herby et al. 2024). Adding (Ege et al. 2023) to the group, nine of twelve SCM analyses examined areas with many COVID-19 deaths during the first wave such as Sweden (five studies), Italy (three studies), and New York (one study). Figure 4 illustrates that these studies tend to examine jurisdictions that were initially hit early and hard by the pandemic.

Only one of the four subnational SCM analyses that examine lockdowns in jurisdictions not characterized by early, severe pandemic shocks reports a statistically significant effect of NPIs (Neidhöfer and Neidhöfer 2020)—and that effect is biologically implausible given known infection-to-death lag structures, as discussed above. The other three analyses find no significant effect (Friedson et al. 2021; Dave et al. 2020; Reinbold 2021).

Figure 4: Selection bias toward early, high-mortality outliers in Europe and U.S.



Note: The figure illustrates the relationship between early pandemic strength and total 1st wave of COVID-19 mortality. On the X-axis is "Days to reach 20 COVID-19-deaths per million (measured from February 15, 2020)." The Y-axis shows mortality (deaths per million) by June 30, 2020. Countries and states covered by the excluded SCM studies are marked with red.
Source: Replication of Figure 4 in (Herby et al. 2024, app. supplementary file2).

Likewise, a cross-national GSCM study deliberately includes 169 countries from 1 July 2020 to 1 September 2021 to attenuate identification problems stemming from low variation in treatment (timing) and data quality. The authors conclude that they "do not find substantial and consistent mitigating effects of any NPI" on COVID-19 deaths per capita (Mader and Rüttenauer 2022).

Taken together, these findings reinforce a material falsification signal: directionally diversified SCM evidence is effectively absent, while case selection is highly concentrated on the hardest and earliest shocked mortality outliers in the North Atlantic policy literature—chiefly Italy, Sweden, and New York. This raises a clear and unresolved risk that many of the large reported SCM "lockdown effects" are not policy effects at all, but rather unbalanced time-varying confounding and early-shock omitted-variable bias, mechanically absorbed into model fittings.

Conclusion

In the introduction I described the results from five SCM studies examining the effect of a counterfactual lockdown in Sweden. The estimated effects in these SCM studies suggest that the lockdown would have averted between 70,000 and 140,000 deaths when scaled to a U.S.-sized population. That number is markedly smaller than the more than two million predicted by some epidemiological studies (Ferguson et al. 2020), or estimated by studies that do not use a control group approach and, thus, cannot distinguish between the effect of voluntary behavioral changes and lockdown measures (Flaxman et al. 2020).

That projected saving is illusory. Rather than capturing a causal policy effect, these SCM estimates mechanically absorb the stochastic variation in viral seeding and outbreak timing—heterogeneity that exists across countries regardless of their decision to lock down. I substantiate this by showing that the SCM yields highly divergent outcomes across different Swedish regions. Notably, the method finds no treatment effect in regions where viral seeding and outbreak timing aligned with the low-mortality countries in the donor pool, suggesting that the "effects" found elsewhere are artifacts of temporal mismatch rather than policy. It should be noted that this regional analysis does not rule out the possibility that a counterfactual lockdown could have reduced excess mortality in the Stockholm region (Sweden w9), where early and intense viral seeding may have made the epidemic phase more amenable to intervention. However, the fact that no comparable effect appears in the other regions suggests that the aggregate SCM estimate is driven by this particular regional pattern rather than a uniform national policy effect.

Despite the inherent unsuitability of the SCM for pandemic trajectories, it has been widely deployed to evaluate lockdowns, with a notable selection bias toward early, high-mortality outliers such as Sweden, Italy, and New York. Relying on pre-treatment windows of mere weeks, these authors extrapolated multi-week forecasts to make sweeping policy claims. Would these same researchers be willing to estimate a new business strategy's effect on revenue for the following two quarters based on 15–20 days of volatile data, particularly when the timing of the strategy is ambiguous? The question is acute given the warning signs that were ignored: many studies report "lockdown effects" that defy biological lag times, while failing to rigorously test for statistical significance against placebo distributions.

The application of SCM to the first wave of the pandemic serves as a cautionary tale. By forcing a method designed for long-term stability onto a chaotic, short-term exponential process, researchers did not uncover the causal effect of lockdowns. Instead, they inadvertently fitted their models to the noise of viral timing. In doing so, they lent unearned statistical credibility to a false narrative, reinforcing the narrative that strict mandates were the primary driver of survival.

I invite the authors of the five Swedish SCM studies summarized in Table 1

to respond to this critique, so that readers may assess competing interpretations of the evidence.

Data and Code

Data and code used in this research is available at this link

References

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

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 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. Link

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2015. Comparative Politics and the Synthetic Control Method. *American Journal of Political Science* 59(2): 495–510. Link

Arnarson, Björn Thor. 2021. How a School Holiday Led to Persistent COVID-19 Outbreaks in Europe. *Scientific Reports* 11: 24390. Link

Björk, Jonas, Kristoffer Mattisson, and Anders Ahlbom. 2021. Impact of Winter Holiday and Government Responses on Mortality in Europe during the First Wave of the COVID-19 Pandemic. *European Journal of Public Health* 31(2): 272–77. Link

Bjørnskov, Christian, and Stefan Voigt. 2022. This Time Is Different?—On the Use of Emergency Measures during the Corona Pandemic. *European Journal of Law and Economics* 54: 63–81. Link

Born, Benjamin, Alexander M. Dietrich, and Gernot J. Müller. 2021. The Lockdown Effect: A Counterfactual for Sweden. *PLoS ONE* 16(4): e0249732. Link

Cho, Sang-Wook (Stanley). 2020. Quantifying the Impact of Nonpharmaceutical Interventions during the COVID-19 Outbreak: The Case of Sweden. *The Econometrics Journal* 23(3): 323–344. Link

Conyon, Martin J., and Steen Thomsen. 2021. COVID-19 in Scandinavia. Preprint, February 16. Link

Dave, Dhaval, Andrew I. Friedson, Kyutaro Matsuzawa, Drew McNichols, and Joseph J. Sabia. 2020. Are the Effects of Adoption and Termination of Shelter-in-Place Orders Symmetric? Evidence from a Natural Experiment. *NBER Working Paper* 27322. National Bureau of Economic Research (Cambridge, MA). Link

Ege, Florian, Giovanni Mellace, and Seetha Menon. 2023. The Unseen Toll: Excess Mortality during Covid-19 Lockdowns. *Scientific Reports* 13: 18745. Link

Engler, Sarah, Palmo Brunner, Romane Loviat, et al. 2021. Democracy in Times of the Pandemic: Explaining the Variation of COVID-19 Policies across European Democracies. *West European Politics* 44(5–6): 1077–102. Link

Ferguson, Neil M., Daniel Laydon, Gemma Nedjati-Gilani, et al. 2020. Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand. *Imperial College COVID-19 Response Team Report* 9. Imperial College London (London, UK). Link

Ferman, Bruno, Cristine Pinto, and Vitor Possebom. 2020. Cherry Picking with Synthetic Controls. *Journal of Policy Analysis and Management* 39(2): 510–32. Link

Flaxman, Seth, Swapnil Mishra, Axel Gandy, et al. 2020. Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe. *Nature* 584: 257–261. Link

Francis, Joseph. 2025a. A P-Hacker's Guide to the Synthetic Control Method. Unpublished paper. Link

Francis, Joseph. 2025b. A Replication of "Comparative Politics and the Synthetic Control Method" by Abadie et al. (2015). *I4R Discussion Paper Series* No. 271. The Institute for Replication (I4R). Link

Friedson, Andrew I., Drew McNichols, Joseph J. Sabia, and Dhaval Dave. 2021. Shelter-in-Place Orders and Public Health: Evidence from California During the Covid-19 Pandemic. *Journal of Policy Analysis and Management* 40(1): 258–283. Link

García-García, David, María Isabel Vigo, Eva S. Fonfría, Zaida Herrador, et al. 2021. Retrospective Methodology to Estimate Daily Infections from Deaths (REMEDID) in COVID-19: The Spain Case Study. *Scientific Reports* 11: 11274. Link

Hanke, Steve H., Jonas Herby, and Lars Jonung. 2023. The Puzzling Royal Society Report on Covid-19. *National Review*, September 21. Link

Herby, Jonas. 2025a. Don't Jump to Faulty Conclusions: Using the Synthetic Control Method to Evaluate the Effect of a Counterfactual Lockdown in Sweden. *Nationaløkonomisk Tidsskrift* 2025:1. Link

Herby, Jonas. 2025b. *Prisen værd?: nedlukningerne ingen taler om*. 2. udgave. Kolbein Publishing House.

Herby, Jonas, Lars Jonung, and Steve H. Hanke. 2024. Were COVID-19 Lockdowns Worth It? A Meta-Analysis. *Public Choice* 203: 337–367. Link

Jonung, Lars, and Steve H. Hanke. 2020. Freedom and Sweden's Constitution. *Wall Street Journal*, May 20. Link

Kepp, Kasper Planeta, Jonas Björk, Vasilis Kontis, et al. 2022. Estimates of Excess Mortality for the Five Nordic Countries during the COVID-19 Pandemic 2020–2021. *International Journal of Epidemiology* 51(6): 1722–1732. Link

Latour, Chiara, Franco Peracchi, and Giancarlo Spagnolo. 2022. Assessing Alternative Indicators for Covid-19 Policy Evaluation, with a Counterfactual for Sweden. *PLoS ONE* 17(3): e0264769. Link

Mader, Sebastian, and Tobias Rüttenauer. 2022. The Effects of Non-Pharmaceutical Interventions on COVID-19 Mortality: A Generalized Synthetic Control Approach Across 169 Countries. *Frontiers in Public Health* 10: 820642. Link

Masini, Ricardo, and Marcelo C. Medeiros. 2022. Counterfactual Analysis and Inference With Nonstationary Data. *Journal of Business & Economic Statistics* 40(1): 227–39. Link

Mistur, Evan M., John Wagner Givens, and Daniel C. Matisoff. 2023. Contagious COVID-19 Policies: Policy Diffusion during Times of Crisis. *Review of Policy Research* 40(1): 36–62. Link

Neidhöfer, Guido, and Claudio Neidhöfer. 2020. The Effectiveness of School Closures and Other Pre-Lockdown COVID-19 Mitigation Strategies in Argentina, Italy, and South Korea. *ZEW - Centre for European Economic Research Discussion Paper* No. 20-034. ZEW - Leibniz Centre for European Economic Research (Mannheim, Germany). Link

Reinbold, Gary W. 2021. Effect of Fall 2020 K-12 Instruction Types on COVID-19 Cases, Hospital Admissions, and Deaths in Illinois Counties. *American Journal of Infection Control* 49: 1146–1151. Link

Sebhatu, Abiel, Karl Wennberg, Stefan Arora-Jonsson, and Staffan I. Lindberg. 2020. Explaining the Homogeneous Diffusion of COVID-19 Nonpharmaceutical Interventions across Heterogeneous Countries. *PNAS* 117(35): 21201–21208. Link

The Royal Society. 2023. COVID-19: Examining the Effectiveness of Non-Pharmaceutical Interventions. Report, August. Link

Wood, Simon N. 2022. Inferring UK COVID-19 Fatal Infection Trajectories from Daily Mortality Data: Were Infections Already in Decline before the UK Lockdowns? *Biometrics* 78(3): 1127–1140. Link

About the Author

Jonas Herby (MSc. in Economics) is Managing Director of IRES (Institute for Regulatory Studies), an independent Danish institute focused on regulatory economics. His meta-analysis "Were COVID-19 Lockdowns Worth It?" (with Lars Jonung and Steve H. Hanke), published in *Public Choice* (2024), generated global headlines and controversy when released as a Johns Hopkins working paper in early 2022, becoming one of the most widely discussed studies of the pandemic and fueling a fierce international debate over lockdown efficacy. He has also published "Don't Jump to Faulty Conclusions" in *Nationaløkonomisk Tidsskrift* (2025) and is the author of *Prisen værd? Nedlukningerne ingen taler om* (Kolbein, 2025), a book-length evaluation of the Danish COVID-19 lockdowns. He is based in Copenhagen. His email is jherby@gmail.com.

Go to archive of Comments section
Go to March 2026 issue