SAE./No.200/January 2022

Studies in Applied Economics

A LITERATURE REVIEW AND META-ANALYSIS OF THE EFFECTS OF LOCKDOWNS ON COVID-19 MORTALITY

Jonas Herby, Lars Jonung, and Steve H. Hanke

Johns Hopkins Institute for Applied Economics,
Global Health, and the Study of Business Enterprise

A Literature Review and Meta-Analysis of the Effects of Lockdowns on COVID-19 Mortality

By Jonas Herby, Lars Jonung, and Steve H. Hanke

About the Series

The *Studies in Applied Economics* series is under the general direction of Prof. Steve H. Hanke, Founder and Co-Director of The Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise (hanke@jhu.edu). The views expressed in each working paper are those of the authors and not necessarily those of the institutions that the authors are affiliated with.

About the Authors

Jonas Herby (herby@cepos.dk) is special advisor at Center for Political Studies in Copenhagen, Denmark. His research focuses on law and economics. He holds a master's degree in economics from University of Copenhagen.

Lars Jonung (lars.jonung@nek.lu.se) is professor emeritus in economics at Lund University, Sweden. He served as chairperson of the Swedish Fiscal Policy Council 2012-13, as research advisor at the European Commission 2000-2010, and as chief economic adviser to Prime Minister Carl Bildt in 1992-94. He holds a PhD in Economics from the University of California, Los Angeles.

Steve H. Hanke is a Professor of Applied Economics and Founder & Co-Director of The Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise. He is a Senior Fellow and Director of the Troubled Currencies Project at the Cato Institute, a contributor at National Review, a well-known currency reformer, and a currency and commodity trader. Prof. Hanke served on President Reagan's Council of Economic Advisers, has been an adviser to five foreign heads of state and five foreign cabinet ministers, and held a cabinet-level rank in both Lithuania and Montenegro. He has been awarded seven honorary doctorate degrees and is an Honorary Professor at four foreign institutions. He was President of Toronto Trust Argentina in Buenos Aires in 1995, when it was the world's best-performing mutual fund. Currently, he serves as Chairman of the Supervisory Board of Advanced Metallurgical Group N.V. in Amsterdam. In 1998, he was named one of the twenty-five most influential people in the world by World Trade Magazine. In 2020, Prof. Hanke was named a Knight of the Order of the Flag.

Abstract

This systematic review and meta-analysis are designed to determine whether there is empirical evidence to support the belief that "lockdowns" reduce COVID-19 mortality. Lockdowns are defined as the imposition of at least one compulsory, non-pharmaceutical intervention (NPI). NPIs are any government mandate that directly restrict peoples' possibilities, such as policies that limit internal movement, close schools and businesses, and ban international travel. This study employed a systematic search and screening procedure in which 18,590 studies are identified that could potentially address the belief posed. After three levels of screening, 34 studies ultimately qualified. Of those 34 eligible studies, 24 qualified for inclusion in the meta-analysis. They were separated into three groups: lockdown stringency index studies, shelter-in-place-order (SIPO) studies, and specific NPI studies. An analysis of each of these three groups support the conclusion that lockdowns have had little to no effect on COVID-19 mortality. More specifically, stringency index studies find that lockdowns in Europe and the United States only reduced COVID-19 mortality by 0.2% on average. SIPOs were also ineffective, only reducing COVID-19 mortality by 2.9% on average. Specific NPI studies also find no broad-based evidence of noticeable effects on COVID-19 mortality.

While this meta-analysis concludes that lockdowns have had little to no public health effects, they have imposed enormous economic and social costs where they have been adopted. In consequence, lockdown policies are ill-founded and should be rejected as a pandemic policy instrument.

Acknowledgements

The authors thank Line Andersen, Troels Sabroe Ebbesen, Nicholas Hanlon, and Anders Lund Mortensen for their research assistance.

The authors also with to thank Douglas Allen, Fredrik N. G. Andersson, Jonas Björk, Christian Bjørnskov, Joakim Book, Gunnar Brådvik, Kristoffer Torbjørn Bæk, Ulf Gerdtham, Daniel B. Klein, Fredrik Charpentier Ljungqvist, Christian Heebøl-Nielsen, Martin Paldam, Jonas Ranstam, Spencer Ryan, John Strezewski, Roger Svensson, Ulf Persson, Anders Waldenström, and Joakim Westerlund for their comments.

Key Words: COVID-19, lockdown, non-pharmaceutical interventions, mortality, systematic review, meta-analysis

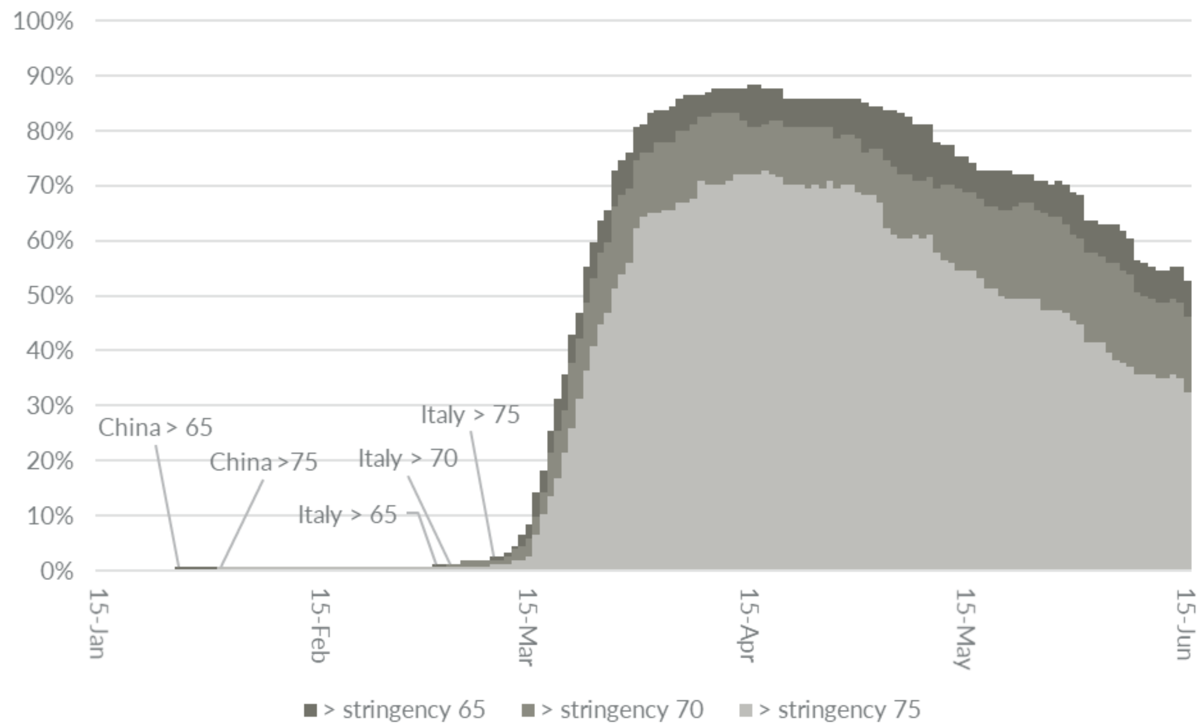
JEL Classification: I18; I38; D19

1 Introduction

The global policy reaction to the COVID-19 pandemic is evident. Compulsory non-pharmaceutical interventions (NPIs), commonly known as "lockdowns" – policies that restrict internal movement, close schools and businesses, and ban international travel – have been mandated in one form or another in almost every country.

The first NPIs were implemented in China. From there, the pandemic and NPIs spread first to Italy and later to virtually all other countries, see Figure 1. Of the 186 countries covered by the Oxford COVID-19 Government Response Tracker (OxCGRT), only Comoros, an island country in the Indian Ocean, did not impose at least one NPI before the end of March 2020.

Figure 1: Share of countries with OxCGRT stringency index above thresholds, January - June 2020

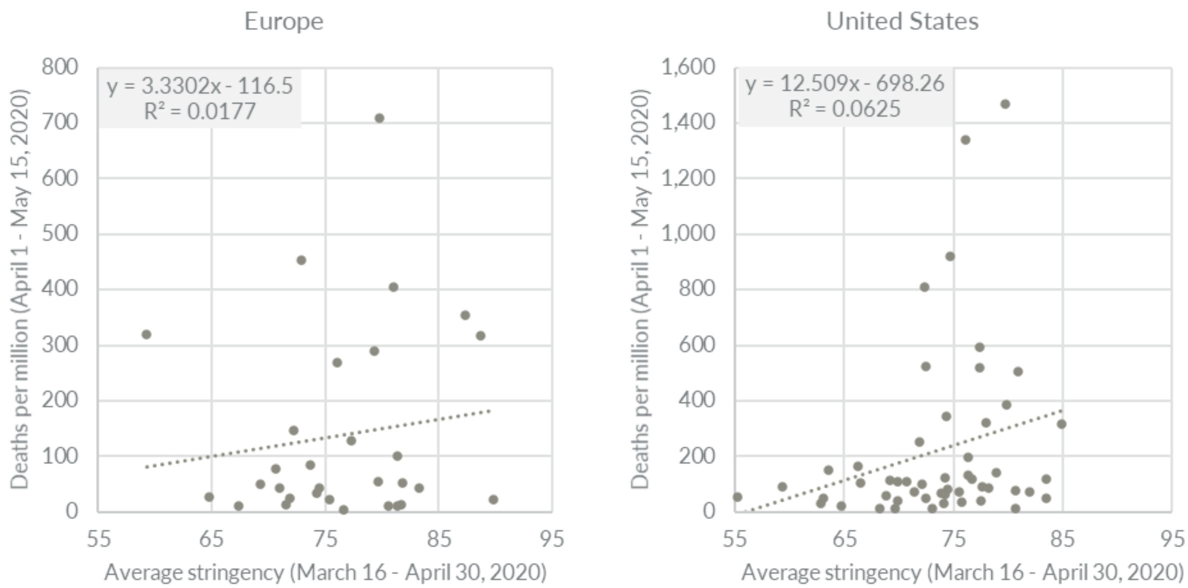


Comment: The figure shows the share of countries, where the OxCGRT stringency index on a given date surpassed index 65, 70 and 75 respectively. Only countries with more than one million citizens are included (153 countries in total). The OxCGRT stringency index records the strictness of NPI policies that restrict people's behavior. It is calculated using all ordinal containment and closure policy indicators (i.e., the degree of school and business closures, etc.), plus an indicator recording public information campaigns.
Source: Our World in Data.

Early epidemiological studies predicted large effects of NPIs. An often cited model simulation study by researchers at the Imperial College London (Ferguson et al. (2020)) predicted that a suppression strategy based on a lockdown would reduce COVID-19 mortality by up to 98%.[1] These predictions were questioned by many scholars. Our early interest in the subject was spurred by two studies. First, Atkeson et al. (2020) showed that "across all countries and U.S. states that we study, the growth rates of daily deaths from COVID-19 fell from a wide range of initially high levels to levels close to zero within 20-30 days after each region experienced 25 cumulative deaths." Second, Sebhatu et al. (2020) showed that "government policies are strongly driven by the policies initiated in other countries," and less by the specific COVID-19-situation of the country.

A third factor that motivated our research was the fact that there was no clear negative correlation between the degree of lockdown and fatalities in the spring of 2020 (see Figure 2). Given the large effects predicted by simulation studies such as Ferguson et al. (2020), we would have expected to at least observe a simple negative correlation between COVID-19 mortality and the degree to which lockdowns were imposed.[2]

Figure 2: Correlation between stringency index and COVID-19 mortality in European countries and U.S. states during the first wave in 2020



Source: Our World in Data

[1] With R0 = 2.4 and trigger on 60, the number of COVID-19-deaths in Great Britain could be reduced to 8,700 deaths from 510,000 deaths (-98%) with a policy consisting of case isolation + home quarantine + social distancing + school/university closure, cf. Table 4 in Ferguson et al. (2020). R0 (the basic reproduction rate) is the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection.

[2] In addition, the interest in this issue was sparked by the work Jonung did on the expected economic effects of the SARS pandemic in Europe in 2006 (Jonung and Röger, 2006). In this model-based study calibrated from Spanish flu data, Jonung and Röger concluded that the economic effects of a severe pandemic would be rather limited—a sharp contrast to the huge economic effects associated with lockdowns during the COVID-19 pandemic.

Today, it remains an open question as to whether lockdowns have had a large, significant effect on COVID-19 mortality. We address this question by evaluating the current academic literature on the relationship between lockdowns and COVID-19 mortality rates.[3] We use "NPI" to describe *any government mandate which directly restrict peoples' possibilities*. Our definition does *not* include governmental recommendations, governmental information campaigns, access to mass testing, voluntary social distancing, etc., but *do* include mandated interventions such as closing schools or businesses, mandated face masks etc. We define *lockdown* as any policy consisting of at least one NPI as described above.[4]

Compared to other reviews such as Herby (2021) and Allen (2021), the main difference in this meta-analysis is that we carry out a systematic and comprehensive search strategy to identify all papers potentially relevant to answer the question we pose. We identify 34 eligible empirical studies that estimate the effect of mandatory lockdowns on COVID-19 mortality using a counterfactual difference-in-difference approach. We present our results in such a way that they can be systematically assessed, replicated, and used to derive overall meta-conclusions.[5]

2 Identification process: Search strategy and eligibility criteria

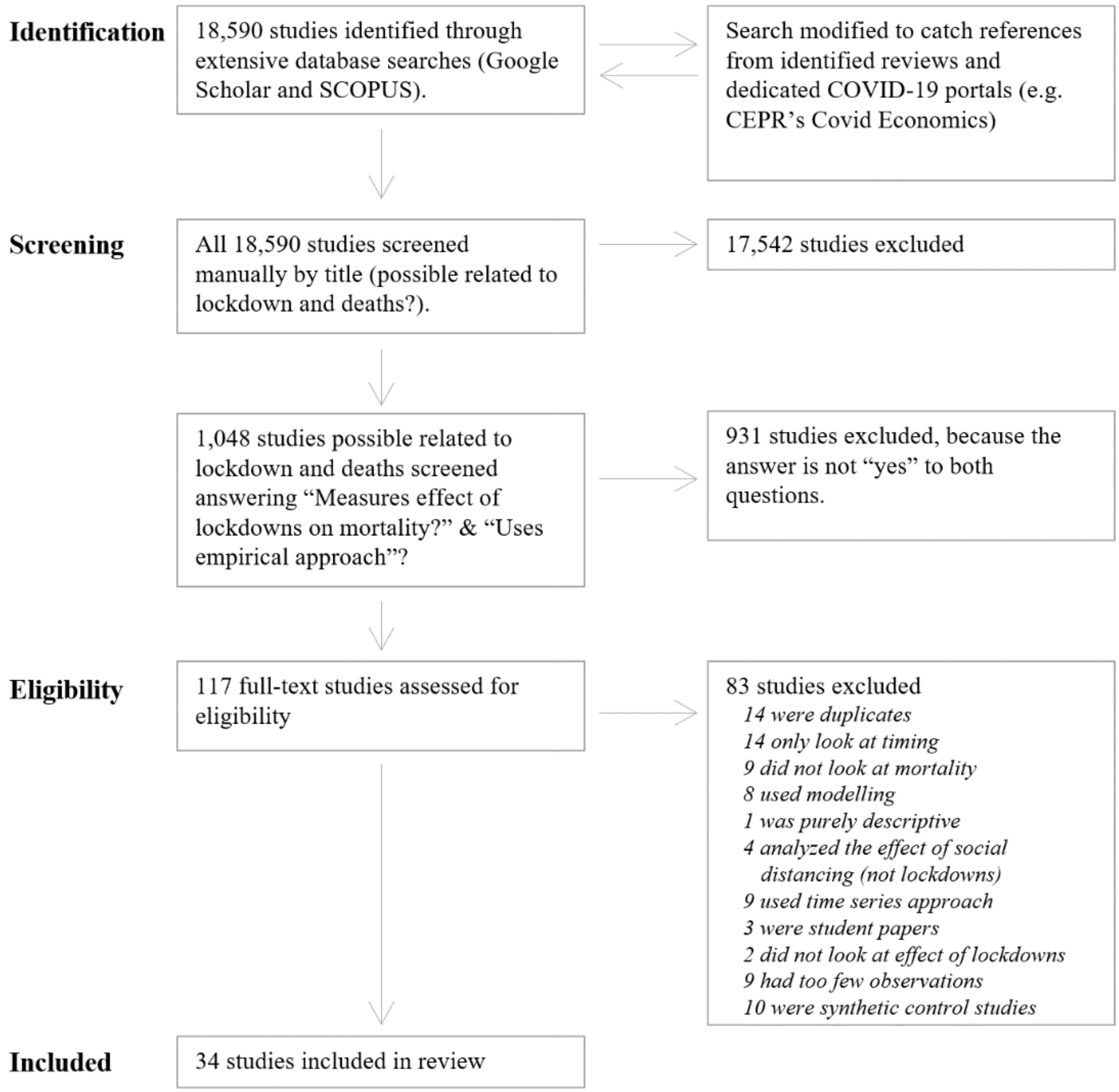
Figure 3 shows an overview of our identification process using a flow diagram designed according to PRISMA guidelines (Moher et al. (2009). Of 18,590 studies identified during our database searches, 1,048 remained after a title-based screening. Then, 931 studies were excluded, because they either did not measure the effect of lockdowns on mortality or did not use an empirical approach. This left 117 studies that were read and inspected. After a more thorough assessment, 83 of the 117 were excluded, leaving 34 studies eligible for our meta-analysis. A table with all 83 studies excluded in the final step can be found in Appendix B, Table 8.

[3] We use "mortality" and "mortality rates" interchangeably to mean COVID-19 deaths per population.

[4] For example, we will say that Country A introduced the *non-pharmaceutical interventions* school closures and shelter-in-place-orders as part of the country's *lockdown.*

[5] An interesting question is, "What damage lockdowns do to the economy, personal freedom and rights, and public health in general?" Although this question is important, it requires a full cost-benefit study, which is beyond the scope of this study.

Figure 3: PRISMA flow diagram for the selection of studies.



Below we present our search strategy and eligibility criteria, which follow the PRISMA guidelines and are specified in detail in our protocol Herby et al. (2021).

2.1 Search strategy

The studies we reviewed were identified by scanning *Google Scholar* and *SCOPUS* for English-language studies. We used a wide range of search terms which are combinations of three search strings: a disease search string ("covid," "corona," "coronavirus," "sars-cov-2"), a government response search string[6], and a methodology search string[7]. We identified papers based on 1,360 search terms. We also required mentions of "deaths," "death," and/or "mortality." The search terms were continuously updated (by adding relevant terms) to fit this criterion.[8]

We also included all papers published in *Covid Economics*. Our search was performed between July 1 and July 5, 2021 and resulted in 18,590 unique studies.[9] All studies identified using SCOPUS and Covid Economics were also found using Google Scholar. This made us comfortable that including other sources such as VOXeu and SSRN would not change the result. Indeed, many papers found using Google Scholar were from these sources.

All 18,590 studies were first screened based on the title. Studies clearly not related to our research question were deemed irrelevant.[10]

After screening based on the title, 1,048 papers remained. These papers were manually screened by answering two questions:

1. Does the study measure the effect of lockdowns on mortality?
2. Does the study use an empirical *ex post* difference-in-difference approach (see eligibility criteria below)?

Studies to which we could not answer "yes" to both questions were excluded. When in doubt, we made the assessment based on reading the full paper, and in some cases, we consulted with colleagues.[11]

After the manual screening, 117 studies were retrieved for a full, detailed review. These studies were carefully examined, and metadata and empirical results were stored in an Excel

---

[6] The government response search string used was: "non-pharmaceutical," "nonpharmaceutical," "NPI," "NPIs," "lockdown," "social distancing orders," "statewide interventions," "distancing interventions," "circuit breaker," "containment measures," "contact restrictions," "social distancing measures," "public health policies," "mobility restrictions," "covid-19 policies," "corona policies," "policy measures."

[7] The methodology search string used was: ("fixed effects," "panel data," "difference-in-difference," "diff-in-diff," "synthetic control," "counterfactual" , "counter factual," "cross country," "cross state," "cross county," "cross region," "cross regional," "cross municipality," "country level," "state level," "county level," "region level," "regional level," "municipality level," "event study."

[8] If a potentially relevant paper from one of the 13 reviews (see eligibility criteria) did not show up in our search, we added relevant words to our search strings and ran the search again. The 13 reviews were: Allen (2021); Brodeur et al. (2021); Gupta et al. (2020); Herby (2021); Johanna et al. (2020); Nussbaumer-Streit et al. (2020); Patel et al. (2020); Perra (2020); Poeschl and Larsen (2021); Pozo-Martin et al. (2020); Rezapour et al. (2021); Robinson (2021); Zhang et al. (2021).

[9] SCOPUS was continuously monitored between July 5th and publication using a search agent. Although the search agent returned several hits during this period, only one of them, An et al. (2021), was eligible according to our eligibility criteria. The study is not included in our review, but the conclusions are in line with our conclusions, as An et al. (2021) conclude that "The analysis shows that the mask mandate is consistently associated with lower infection rates in the short term, and its early adoption boosts the long-term efficacy. By contrast, the other five policy instruments— domestic lockdowns, international travel bans, mass gathering bans, and restaurant and school closures—show weaker efficacy."

[10] This included studies with titles such as "COVID-19 outbreak and air pollution in Iran: A panel VAR analysis" and "Dynamic Structural Impact of the COVID-19 Outbreak on the Stock Market and the Exchange Rate: A Cross-country Analysis Among BRICS Nations."

[11] Professor Christian Bjørnskov of University of Aarhus was particularly helpful in this process.

spreadsheet. All studies were assessed by at least two researchers. During this process, another 64 papers were excluded because they did not meet our eligibility criteria. Furthermore, nine studies with too little jurisdictional variance (< 10 observations) were excluded,[12] and 10 synthetic control studies were excluded.[13] A table with all 83 studies excluded in the final step can be found in Appendix B, Table 8. Below we explain why these studies are excluded.

2.2 Eligibility criteria

Focus on mortality and lockdowns

We only include studies that attempt to establish a relationship (or lack thereof) between lockdown policies and COVID-19 mortality or excess mortality. We exclude studies that use cases, hospitalizations, or other measures.[14]

Counterfactual difference-in-difference approach

We distinguish between two methods used to establish a relationship (or lack thereof) between mortality rates and lockdown policies. The first uses registered cross-sectional mortality data. These are *ex post* studies. The second method uses simulated data on mortality and infection rates.[15] These are *ex ante* studies.

We include all studies using a counterfactual difference-in-difference approach from the former group but disregard all *ex ante* studies, as the results from these studies are determined by model assumptions and calibrations.

Our limitation to studies using a "counterfactual difference-in-difference approach" means that we exclude all studies where the counterfactual is based on forecasting (such as a SIR-model) rather than derived from a difference-in-difference approach. This excludes studies like Duchemin et al. (2020) and Matzinger and Skinner (2020). We also exclude all studies based on interrupted time series designs that simply compare the situation before and after lockdown, as

---

[12] The excluded studies with too few observations were: Alemán et al. (2020), Berardi et al. (2020), Conyon et al. (2020a), Coccia (2021), Gordon et al. (2020), Juranek and Zoutman (2021), Kapoor and Ravi (2020), Umer and Khan (2020), and Wu and Wu (2020).

[13] The excluded synthetic control studies were: Conyon and Thomsen (2021), Dave et al. (2020), Ghosh et al. (2020), Born et al. (2021), Reinbold (2021), Cho (2020), Friedson et al. (2021), Neidhöfer and Neidhöfer (2020), Cerqueti et al. (2021), and Mader and Rüttenauer (2021).

[14] Analyses based on cases may pose major problems, as testing strategies for COVID-19 infections vary enormously across countries (and even over time within a given country). In consequence, cross-country comparisons of cases are, at best, problematic. Although these problems exist with death tolls as well, they are far more limited. Also, while cases and death tolls are correlated, there may be adverse effects of lockdowns that are not captured by the number of cases. For example, an infected person who is isolated at home with family under a SIPO may infect family members with a higher viral load causing more severe illness. So even if a SIPO reduces the number of cases, it may theoretically increase the number of COVID-19-deaths. Adverse effects like this may explain why studies like Chernozhukov et al. (2021) finds that SIPO reduces the number of cases but have no significant effect on the number of COVID-19-deaths. Finally, mortality is hierarchically the most important outcome, cf. GRADEpro (2013)

[15] These simulations are often made in variants of the SIR-model, which can simulate the progress of a pandemic in a population consisting of people in different states (Susceptible, Infectious, or Recovered) with equations describing the process between these states.

the effect of lockdowns in these studies might contain time-dependent shifts, such as seasonality. This excludes studies like Bakolis et al. (2021) and Siedner et al. (2020).

Given our criteria, we exclude the much-cited paper by Flaxman et al. (2020), which claimed that lockdowns saved three million lives in Europe. Flaxman et al. assume that the pandemic would follow an epidemiological curve unless countries locked down. However, this assumption means that the only interpretation possible for the empirical results is that lockdowns are the only thing that matters, even if other factors like season, behavior etc. caused the observed change in the reproduction rate, Rt. Flaxman et al. are aware of this and state that "our parametric form of R_t assumes that changes in Rt are an immediate response to interventions rather than gradual changes in behavior." Flaxman et al. illustrate how problematic it is to force data to fit a certain model if you want to infer the effect of lockdowns on COVID-19 mortality.[16]

The counterfactual difference-in-difference studies in this review generally exploit variation across countries, U.S. states, or other geographical jurisdictions to infer the effect of lockdowns on COVID-19 fatalities. Preferably, the effect of lockdowns should be tested using randomized control trials, natural experiments, or the like. However, there are very few studies of this type.[17]

Synthetic control studies

The synthetic control method is a statistical method used to evaluate the effect of an intervention in comparative case studies. It involves the construction of a synthetic control which functions as the counter factual and is constructed as an (optimal) weighted combination of a pool of donors. For example, Born et al. (2021) create a synthetic control for Sweden which consists of 30.0% Denmark, 25.3% Finland, 25.8% Netherlands, 15.0% Norway, and 3.9% Sweden. The effect of the intervention is derived by comparing the actual developments to those contained in the synthetic control.

We exclude synthetic control studies because of their inherent empirical problems as discussed by Bjørnskov (2021b). He finds that the synthetic control version of Sweden in Born et al. (2021) deviates substantially from "actual Sweden," when looking at the period before mid-March 2020, when Sweden decided not to lock down. Bjørnskov estimates that *actual Sweden* experienced

[16] Several scholars have criticized Flaxman et al. (2020), e.g. see Homburg and Kuhbandner (2020), Lewis (2020), and Lemoine (2020).

[17] Kepp and Bjørnskov (2021) is one such study. They use evidence from a quasi-natural experiment in the Danish region of Northern Jutland. After the discovery of mutations of Sars-CoV-2 in mink – a major Danish export – seven of the 11 municipalities of the region went into extreme lockdown in early November, while the four other municipalities retained the moderate restrictions of the remaining country. Their analysis shows that while infection levels decreased, they did so before lockdown was in effect, and infection numbers also decreased in neighbor municipalities without mandates. They conclude that efficient infection surveillance and voluntary compliance make full lockdowns unnecessary, at least in some circumstances. Kepp and Bjørnskov (2021) is not included in our review, because they focus on cases and not COVID-19 mortality. Dave et al. (2020) is another such study. They see the Wisconsin Supreme Court abolishment of Wisconsin's "Safer at Home" order (a SIPO) as a natural experiment and find that "the repeal of the state SIPO impacted social distancing, COVID-19 cases, or COVID-19-related mortality during the fortnight following enactment." Dave et al. (2020) is not included in our review, because they use a synthetic control method.

approximately 500 fewer deaths the first 11 weeks of 2020 and 4,500 fewer deaths in 2019 compared to *synthetic Sweden*.

This problem is inherent in all synthetic control studies of COVID-19, Bjørnskov argues, because the synthetic control should be fitted based on a long period of time before the intervention or the event one is studying the consequences of – i.e., the lockdown Abadie (2021). However, this is not possible for the coronavirus pandemic, as there clearly *is* no long period with coronavirus before the lockdown. Hence, the synthetic control study approach is *by design* not appropriate for studying the effect of lockdowns.

Jurisdictional variance - few observations

We exclude all interrupted time series studies which simply compare mortality rates before and after lockdowns. Simply comparing data from before and after the imposition of lockdowns could be the result of time-dependent variations, such as seasonal effects. For the same reason, we also exclude studies with little jurisdictional variance.[18] For example, we exclude Conyon et al. (2020b) who "exploit policy variation between Denmark and Norway on the one hand and Sweden on the other" and, thus, only have one jurisdictional area in the control group. Although this *is* a difference-in-difference approach, there is a non-negligible risk that differences are caused by much more than just differences in lockdowns. Another example is Wu and Wu (2020), who use all U.S. states, but pool groups of states so they end with basically three observations. None of the excluded studies cover more than 10 jurisdictional areas.[19] One study is a special case of the jurisdictional variance criteria (Auger et al. (2020). Those researchers analyze the effect of school closures in U.S. states and find that those closures reduce mortality by 35%. However, all 50 states closed schools between March 13, 2020, and March 23, 2020, which means that all difference-in-difference is based on maximum 10 days. Given the long lag between infection and death, there is a risk that Auger et al.'s approach is an interrupted time series analysis where they compare United States before and after school closures, rather than a true difference-in-difference approach. However, we choose to include this study, as it is eligible under our protocol Herby et al. (2021).

Publication status and date

We include all *ex post* studies regardless of publication status and date. That is, we cover both working papers and papers published in journals. We include the early papers because the knowledge of the COVID-19-pandemic grew rapidly in the beginning, making later papers able to stand on the shoulders of previous work. Also, in the early days of COVID-19, speed was

[18] A jurisdictional area can be countries, U.S. states, or counties. With "jurisdictional variance" we refer to variation in mandates across jurisdictional areas.

[19] All studies excluded on this criterion are listed in footnote 12.

crucial which may have affected the quality of the papers. Including them makes it possible to compare the results of early studies to studies carried out at a later stage.[20]

The role of optimal timing

We exclude papers which analyze the effect of early lockdowns in contrast to later lockdowns. There's no doubt that being prepared for a pandemic and knowing when it arrives at your doorstep is vital. However, at least two problems arise with respect to evaluating the effect of well-timed lockdowns.

First, when COVID-19 hit Europe and the United States, it was virtually impossible to determine the right timing. The World Health Organization declared the outbreak a pandemic on March 11, 2020, but at that date, Italy had already registered 13.7 COVID-19 deaths per million. On March 29, 2020, 18 days after the WHO declared the outbreak a pandemic and the earliest a lockdown response to the WHO's announcement could potentially have an effect, the mortality rate in Italy was a staggering 178 COVID-19 deaths per million with an additional 13 per million dying each day.[21]

Secondly, it is extremely difficult to differentiate between the effect of public awareness and the effect of lockdowns when looking at timing because people and politicians are likely to react to the same information. As Figure 4 illustrates, all European countries and U.S. states that were hit hard and early by COVID-19 experienced high mortality rates, whereas all countries hit relatively late experienced low mortality rates. Björk et al. (2021) illustrate the difficulties in analyzing the effect of timing. They find that a 10-stringency-points-stricter lockdown would reduce COVID-19 mortality by a total of 200 deaths per million[22] if done in week 11, 2020, but would only have approximately 1/3 of the effect if implemented one week earlier or later and no effect if implemented three weeks earlier or later. One interpretation of this result is that lockdowns do not work if people either find them unnecessary and fail to obey the mandates or if people voluntarily lock themselves down. This is the argument Allen (2021) uses for the ineffectiveness of the lockdowns he identifies. If this interpretation is true, what Björk et al. (2021) find is that information and signaling is far more important than the strictness of the lockdown. There may be other interpretations, but the point is that studies focusing on timing cannot differentiate between these interpretations. However, if lockdowns have a notable effect, we should see this effect regardless of the timing, and we should identify this effect more correctly by excluding studies that exclusively analyze timing.

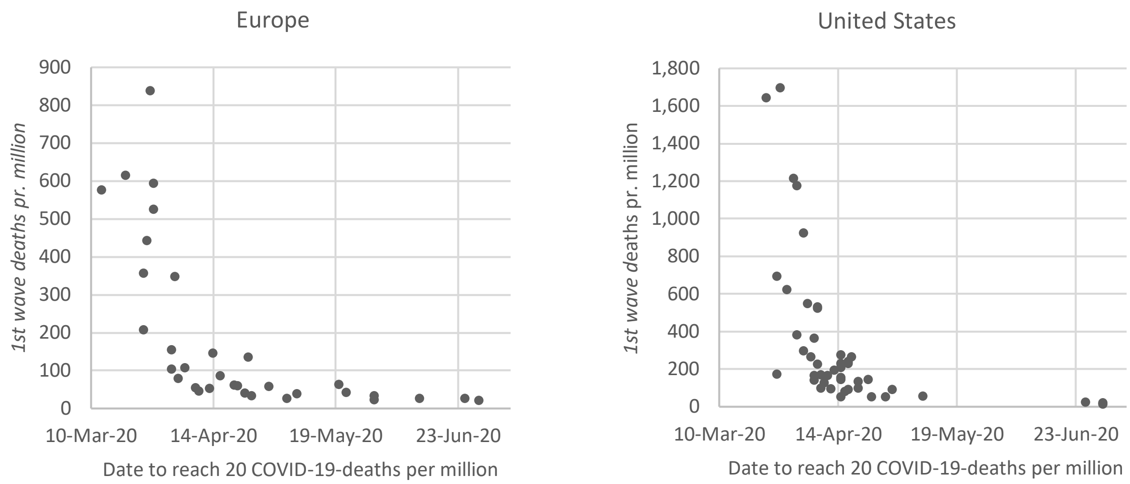
---

[20] We also intended to exclude studies which were primarily based on data from 2021 (as these studies would be heavily affected by vaccines) and studies that did not cover at least one EU-country, the United States, one U.S. U.S. state or Latin America, and where at least one country/state was not an island. However, we did not find any such studies.

[21] There's approximately a two-to-four-week gap between infection and deaths. See footnote 29.

[22] They estimate that 10-point higher stringency will reduce excess mortality by 20 "per week and million" in the 10 weeks from week 14 to week 23.

Figure 4: **Taken by surprise. The importance of having time to prepare**



Comment: The figure shows the relationship between early pandemic strength and total 1st wave of COVID-19 death toll. 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.
Source: Reported COVID-19 deaths and OxCGRT stringency for European countries and U.S. states with more than one million citizens. Data from Our World in Data.

We are aware of one meta-analysis by Stephens et al. (2020), which looks into the importance of timing. The authors find 22 studies that look at policy and timing with respect to mortality rates, however, only four were multi-country, multi-policy studies, which could possibly account for the problems described above. Stephens et al. conclude that "the timing of policy interventions across countries relative to the first Wuhan case, first national disease case, or first national death, is not found to be correlated with mortality." (See Appendix A for further discussion of the role of timing.)

3 The empirical evidence

In this section we present the empirical evidence found through our identification process. We describe the studies and their results, but also comment on the methodology and possible identification problems or biases.

3.1 Preliminary considerations

Before we turn to the eligible studies, we present some considerations that we adopted when interpreting the empirical evidence.

Empirical interpretation

While the policy conclusions contained in some studies are based on statistically significant results, many of these conclusions are ill-founded due to the tiny impact associated with said statistically significant results. For example, Ashraf (2020) states that "social distancing measures has proved effective in controlling the spread of [a] highly contagious virus." However, their estimates show that the average lockdown in Europe and the U.S only reduced COVID-19 mortality by 2.4%.[23] Another example is Chisadza et al. (2021). The authors argue that "less stringent interventions increase the number of deaths, whereas more severe responses to the pandemic can lower fatalities." Their conclusion is based on a negative estimate for the squared term of *stringency* which results in a total negative effect on mortality rates (i.e. fewer deaths) for stringency values larger than 124. However, the stringency index is limited to values between 0 and 100 by design, so the conclusion is clearly incorrect. To avoid any such biases, we base our interpretations solely on the empirical estimates and not on the authors' own interpretation of their results.

Handling multiple models, specifications, and uncertainties

Several studies adopt a number of models to understand the effect of lockdowns. For example, Bjørnskov (2021a) estimates the effect after one, two, three, and four weeks of lockdowns. For these studies, we select the longest time horizon analyzed to obtain the estimate closest to the long-term effect of lockdowns.

Several studies also use multiple specifications including and excluding potentially relevant variables. For these studies, we choose the model which the authors regard as their main specification. Finally, some studies have multiple models which the authors regard as equally important. One interesting example is Chernozhukov et al. (2021), who estimate two models with and without national case numbers as a variable. They show that including this variable in their model alters the results substantially. The explanation could be that people responded to national conditions. For these studies, we present both estimates in Table 1, but – following Doucouliagos and Paldam (2008) – we use an average of the estimates in our meta-analysis in order to not give more weight to a study with multiple models relative to studies with just one principal model.

For studies looking at different classes of countries (e.g. rich and poor), we report both estimates in Table 1 but use the estimate for rich Western countries in our meta-analysis, where we derive common estimates for Europe and the United States.

Effects are measured "relative to Sweden in the spring of 2020"

Virtually all countries in the world implemented mandated NPIs in response to the COVID-19 pandemic. Hence, most estimates are relative to "doing the least," which in many Western countries means relative to doing as Sweden has done, especially during the first wave, when Sweden, do to constitutional constraints, implemented very few restrictions compared to other western countries (Jonung and Hanke 2020). However, some studies *do* compare the effect of doing something to the effect of doing absolutely nothing (e.g. Bonardi et al. (2020)).

The consequence is that some estimates are relative to "doing the least" while others are relative to "doing nothing." This may lead to biases if "doing the least" works as a signal (or warning)

[23] We describe how we arrive at the 2.4% in Section 4.

which alters the behavior of the public. For example, Gupta et al. (2020) find a large effect of emergency declarations, which they argue "are best viewed as an information instrument that signals to the population that the public health situation is serious and they act accordingly," on social distancing but not of other policies such as SIPOs (shelter-in-place orders). Thus, if we compare a country issuing a SIPO to a country doing nothing, we may overestimate the effect of a SIPO, because it is the sum of the signal *and* the SIPO. Instead, we should compare the country issuing the SIPO to a country "doing the least" to estimate the *marginal* effect of the SIPO.

To take an example, Bonardi et al. (2020) find relatively large effects of doing *something* but no effect of doing *more*. They find no extra effect of stricter lockdowns relative to less strict lockdowns and state that "our results point to the fact that people might adjust their behaviors quite significantly as partial measures are implemented, which might be enough to stop the spread of the virus." Hence, whether the baseline is Sweden, which implemented a ban on large gatherings early in the pandemic, or the baseline is "doing nothing" can affect the magnitude of the estimated impacts. There is no obvious right way to resolve this issue, but since estimates in most studies are relative to doing less, we report results as compared to "doing less" when available. Hence, for Bonardi et al. we state that the effect of lockdowns is zero (compared to Sweden's "doing the least").

3.2 Overview of the findings of eligible studies

Table 1 covers the 34 studies eligible for our review.[24] Out of these 34 studies, 22 were peer-reviewed and 12 were working papers. The studies analyze lockdowns during the first wave. Most of the studies (29) use data collected before September 1st, 2020 and 10 use data collected before May 1st, 2020. Only one study uses data from 2021. All studies are cross-sectional, ranging across jurisdictions. Geographically, 14 studies cover countries worldwide, four cover European countries, 13 cover the United States, two cover Europe and the United States, and one covers regions in Italy. Seven studies analyze the effect of SIPOs, 10 analyze the effect of stricter lockdowns (measured by the OxCGRT stringency index), 16 studies analyze specific NIP's independently, and one study analyzes other measures (length of lockdown).

Several studies find no statistically significant effect of lockdowns on mortality. For example, this includes Bjørnskov (2021a) and Stockenhuber (2020) who find no significant effect of stricter lockdowns (higher OxCGRT stringency index), Sears et al. (2020) and Dave et al. (2021), who find no significant effect of SIPOs, and Chaudhry et al. (2020), Aparicio and Grossbard (2021) and Guo et al. (2021) who find no significant effect of any of the analyzed NIP's, including business closures, school closures and border closures.

Other studies find a significant negative relationship between lockdowns and mortality. Fowler et al. (2021 find that SIPOs reduce COVID-19 mortality by 35%, while Chernozhukov et al.

[24] The following information can be found for each study in Table 2.

(2021) find that employee mask mandates reduces mortality by 34% and closing businesses and bars reduces mortality by 29%.

Some studies find a significant positive relationship between lockdowns and mortality. This includes Chisadza et al. (2021), who find that stricter lockdowns (higher OxCGRT stringency index) increases COVID-19 mortality by 0.01 deaths/million per stringency point and Berry et al. (2021), who find that SIPOs increase COVID-19 mortality by 1% after 14 days.

Most studies use the number of official COVID-19 deaths as the dependent variable. Only one study, Bjørnskov (2021a), looks at total excess mortality which – although is not perfect – we perceive to be the best measure, as it overcomes the measurement problems related to properly reporting COVID-19 deaths.

Several studies explicitly claim that they estimate the actual causal relationship between lockdowns and COVID-19 mortality. Some studies use instrumental variables to justify the causality associated with their analysis, while others make causality probable using anecdotal evidence.[25] But, Sebhatu et al. (2020) show that government policies are strongly driven by the policies initiated in neighboring countries rather than by the severity of the pandemic in their own countries. In short, it is not the severity of the pandemic that drives the adoption of lockdowns, but rather the propensity to copy policies initiated by neighboring countries. The Sebhatu et al. conclusion throws into doubt the notion of a causal relationship between lockdowns and COVID-19 mortality.

Table 1: Summary of eligible studies

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
Alderman and Harjoto (2020); "COVID-19: U.S. shelter-in-place orders and demographic characteristics linked to cases, mortality, and recovery rates"	COVID-19 mortality	Use State-level data from the COVID-19 Tracking Project data all U.S. states, and a multivariate regression analysis to empirically investigate the impacts of the duration of shelter-in-place orders on mortality.	Find that shelter-in-place orders are - for the average duration - associated with 1% (insignificant) fewer deaths per capita.	
Aparicio and Grossbard (2021); "Are Covid Fatalities in the U.S. Higher than in the EU, and If so, Why?"	COVID-19 mortality	Their main focus is to explain the gap in COVID-19-fatalities between Europe and the United States based on COVID-deaths and other data from 85 nations/states. They include status for "social events" (ban on public gatherings, cancellation of major events and conferences), school closures, shop closures "partial lockdowns" (e.g. night curfew) and "lockdowns" (all-day curfew) 100 days after the pandemic onset in a country/state. None of these interventions have a significant effect on COVID-19 mortality. They also find no	Find no effect of "social events" (ban on public gatherings, cancellation of major events and conferences), school closures, shop closures "partial lockdowns" (e.g. night curfew) and "lockdowns" (all-day curfew) 100 days after the pandemic onset.	In the abstract the authors states that "various types of social distance measures such as school closings and lockdowns, and how soon they were implemented, help explain the U.S./EUROPE gap in cumulative deaths measured 100 days after the pandemic's onset in a state or country" although their estimates are insignificant.

[25] E.g. Dave et al. (2021) states that "estimated case reductions accelerate over time, becoming largest after 20 days following enactment of a SIPO. These findings are consistent with a causal interpretation."

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
		significant effect of early cancelling of social events, school closures, shop closures, partial lockdowns and full lockdowns.		
Ashraf (2020); "Socioeconomic conditions, government interventions and health outcomes during COVID-19"	COVID-19 mortality	Their main focus is on the effectiveness of policies targeted to diminish the effect of socioeconomic inequalities (economic support) on COVID-19-deaths. They use data from 80 countries worldwide and include the OxCGRT stringency as a control variable in their models. The paper finds a significant negative (fewer deaths) effect of stricter lockdowns. The effect of lockdowns is insignificant, when they include an interaction term between the socioeconomic conditions index and the economic support index in their model.	For each 1-unit increase in OxCGRT stringency index, the cumulative mortality changes by -0.326 deaths per million (fewer deaths). The estimate is -0.073 deaths per million but insignificant, when including an interaction term between the socioeconomic conditions index and the economic support index.	
Auger et al. (2020); "Association between statewide school closure and COVID-19 incidence and mortality in the U.S."	COVID-19 mortality	U.S. population-based observational study which uses interrupted time series analyses incorporating a lag period to allow for potential policy-associated changes to occur. To isolate the association of school closure with outcomes, state-level nonpharmaceutical interventions and attributes were included in negative binomial regression models. Models were used to derive the estimated absolute differences between schools that closed and schools that remained open. The main outcome of the study is COVID-19 daily incidence and mortality per 100000 residents.	State that they adjust for several factors (e..g percentage of state's population aged 15 years and 65 years, CDC's social vulnerability index, stay-at-home or shelter-in-place order, restaurant and bar closure, testing rate per 1000 residents etc.), but does not specify how and do not present estimates.	All 50 states closed schools between March 13, 2020, and March 23, 2020. Hence, all difference-in-difference is based on maximum 10 days, and given the long lag between infection and death, there is a risk that their approach is more an interrupted time series analysis, where they compare United States before and after school closures, rather than a true difference-in-difference approach. However, we choose to include the study in our review as it - objectively speaking - lives up to the eligibility criteria specified in our protocol.
Berry et al. (2021); "Evaluating the effects of shelter-in-place policies during the COVID-19 pandemic"	COVID-19 mortality	The authors use U.S. county data on COVID-19 deaths from Johns Hopkin and SIPO data from the University of Washington to estimate the effect of SIPO's. They find no detectable effects of SIPO on deaths. The authors stress that their findings should not be interpreted as evidence that social distancing behaviors are not effective. Many people had already changed their behaviors before the introduction of shelter-in-place orders, and shelter-in-place orders appear to have been ineffective precisely because they did not meaningfully alter social distancing behavior.	SIPO increases the number of deaths by 0,654 per million after 14 days (see Fig. 2)	The authors conclude that "We do not find detectable effects of these policies [SIPO] on disease spread or deaths." However, this statement does not correspond to their results. In figure 2 they show that the effect on deaths is significant after 14 days. Looks at the effect 14 days after SIPO's are implemented which is a short lag given that the time between infection and deaths is at least 2-3 weeks.
Bjørnskov (2021a); "Did Lockdown Work? An Economist's Cross-Country Comparison"	Excess mortality	Uses excess mortality and OxCGRT stringency from 24 European countries to estimate the effect of lockdown on the number of deaths one, two, three and four weeks later. Finds no effect (negative but insignificant) of (stricter) lockdowns. The author's specification using instrument variables yields similar results.	A stricter lockdown (OxCGRT stringency) does not have a significant effect on excess mortality.	Finds a positive (more deaths) effect after one and two weeks, which could indicate that other factors (omitted variables) affect the results.
Blanco et al. (2020); "Do Coronavirus Containment Measures Work? Worldwide Evidence"	COVID-19 mortality	Use data for deaths and NPIs from Hale et al. (2020) covering 158 countries between January and August 2020 to evaluate the effect of eight different NPIs (stay at home, bans on gatherings, bans on public	When using the naïve dummy variable approach, all parameters are statistically	Run the same model four times for each of the different NPIs (stay at home-orders, ban on meetings, ban on public events and mobility restrictions). These NPIs were often introduced almost simultaneously so there is a high risk of

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
		events, closing schools, lockdowns of workplaces, interruption of public transportation services, and international border closures. They address the possible endogeneity of the NPIs by using instrumental variables.	insignificant. On the contrary, estimates using the instrumental variable approach indicate that NPIs are effective in reducing the growth rate in the daily number of deaths 14 days later.	multicollinearity with each run capturing the same underlying effect. Indeed, the size and standard errors of the estimates are worryingly similar. Looks at the effect 14 days after NPIs are implemented which is a fairly short lag given the time between infection and deaths is 2-3 weeks, cf. e.g. Flaxman et al. (2020), which according to Bjørnskov (2020) appears to be the minimum typical time from infection to death).
Bonardi et al. (2020); "Fast and local: How did lockdown policies affect the spread and severity of the covid-19"	Growth rates	Use NPI data scraped from news headlines from LexisNexis and death data from Johns Hopkins University up to April 1st 2020 in a panel structure with 184 countries. Controls for country fixed effects, day fixed effects and within-country evolution of the disease.	Find that certain interventions (SIPO, regional lockdown and partial lockdown) work (in developed countries), but that stricter interventions (SIPO) do not have a larger effect than less strict interventions (e.g. restrictions on gatherings). Find no effect of border closures.	Find a positive (more deaths) effect on day 1 after lockdown which may indicate that their results are driven by other factors (omitted variables). We rely on their publicly available version submitted to CEPR Covid Economics, but estimates on the effect of deaths can be found in Supplementary material, which is available in an updated version hosted on the Danish Broadcasting Corporation's webpage: https://www.dr.dk/static/documents/2021/03/04/managing_pandemics_e3911c11.pdf
Bongaerts et al. (2021); "Closed for business: The mortality impact of business closures during the Covid-19 pandemic"	COVID-19 mortality	Uses variation in exposure to closed sectors (e.g. tourism) in municipalities within Italy to estimate the effect of business closures. Assuming that municipalities with different exposures to closed sectors are not inherently different, they find that municipalities with higher exposure to closed sectors experienced subsequently lower mortality rates.	Business shutdown saved 9,439 Italian lives by April 13th 2020. This corresponds to a reduction of deaths by 32%, as there were 20,465 COVID-19-deaths in Italy by mid April 2020.	They (implicitly) assume that municipalities with different exposures to closed sectors are not inherently different. This assumption could be problematic, as more touristed municipalities can be very different from e.g. more industrialized municipalities.
Chaudhry et al. (2020); "A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes"	COVID-19 mortality	Uses information on COVID-19 related national policies and health outcomes from the top 50 countries ranked by number of cases. Finds no significant effect of any NPI on the number of COVID-19-deaths.	Finds no significant effect on mortality of any of the analyzed interventions (partial border closure, complete border closure, partial lockdown (physical distancing measures only), complete lockdown (enhanced containment measures including suspension of all non-essential services), and curfews).	
Chernozhukov et al. (2021); "Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S."	Growth rates	Uses COVID-deaths from the New York Times and Johns Hopkins and data for U.S. States from Raifman et al. (2020) to estimate the effect of SIPO, closed nonessential businesses, closed K-12 schools, closed restaurants except takeout, closed movie theaters, and face mask mandates for employees in public facing businesses.	Finds that mandatory masks for employees and closing K-12 schools reduces deaths. SIPO and closing business (average of closed businesses, restaurants and movie theaters) has no statistically significant effect. The effect of school closures is highly sensitive to the	States that "our regression specification for case and death growths is explicitly guided by a SIR model although our causal approach does not hinge on the validity of a SIR model." We are uncertain if this means that data are managed to fit an SIR-model (and thus should fail our eligibility criteria).

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
			inclusion of national case and death data.	
Chisadza et al. (2021); "Government Effectiveness and the COVID-19 Pandemic"	COVID-19 mortality	Uses COVID-19-deaths and OxCGRT stringency from 144 countries to estimate the effect of lockdown on the number of COVID-19-deaths. Find a significant positive (more deaths) non-linear association between government response indices and the number of deaths.	An increase by 1 on "stringency index" increases the number of deaths by 0.0130 per million. The sign of the squared term is negative, but the combined non-linear estimate is positive (increases deaths) and larger than the linear estimate for all values of the OxCGRT stringency index.	The author states that "less stringent interventions increase the number of deaths, whereas more severe responses to the pandemic can lower fatalities." However, according to their estimates this is not correct, as the combined non-linear estimate cannot be negative for relevant values of the OxCGRT stringency index (0 to 100).
Dave et al. (2021); "When Do Shelter-in-Place Orders Fight Covid-19 Best? Policy Heterogeneity Across States and Adoption Time"	COVID-19 mortality	Uses smartphone location tracking and state data on COVID-19 deaths and SIPO data (supplemented by their own searches) collected by the New York Times to estimate the effect of SIPO's. Finds that SIPO was associated with a 9%–10% increase in the rate at which state residents remained in their homes full-time, but overall they do not find an significant effect on mortality after 20+ days (see Figure 4). Indicate that the lacking significance may be due to long term estimates being identified of a few early adopting states.	Finds no overall significant effect of SIPO on deaths but does find a negative effect (fewer deaths) in early adopting states.	Find large effects of SIPO on deaths after 6-14 days in early adopting states (see Table 8), which is before an SIPO-related effect would be seen. This could indicate that other factors rather than SIPO's drive the results.
Dergiades et al. (2020); "Effectiveness of government policies in response to the COVID-19 outbreak"	COVID-19 mortality	Uses daily deaths from the European Centre for Disease Prevention and Control and OxCGRT stringency from 32 countries worldwide (including U.S.) to estimates the effect of lockdown on the number of deaths.	Finds that the greater the strength of government interventions at an early stage, the more effective these are in slowing down or reversing the growth rate of deaths.	Focus is on the effect of early stage NPIs and thus does not absolutely live up to our eligibility criteria. However, we include the study as it differentiates between lockdown strength at an early stage.
Fakir and Bharati (2021); "Pandemic catch-22: The role of mobility restrictions and institutional inequalities in halting the spread of COVID-19"	COVID-19 mortality	Uses data from 127 countries. combining high-frequency measures of mobility data from Google's daily mobility reports, country-date-level information on the stringency of restrictions in response to the pandemic from Oxford's Coronavirus Government Response Tracker (OxCGRT), and daily data on deaths attributed to COVID-19 from Our World In Data and the Johns Hopkins University. Instrument stringency using day-to-day changes in the stringency of the restrictions in the rest of the world.	Find large causal effects of stricter restrictions on the weekly growth rate of recorded deaths attributed to COVID-19. Show that more stringent interventions help more in richer, more educated, more democratic, and less corrupt countries with older, healthier populations and more effective governments.	Finds a larger effect on deaths after 0 days than after 14 and 21 days (Table 3). This is surprising given that it takes 2-3 weeks from infection to death, and it may indicate that their results are driven by other factors.
Fowler et al. (2021); "Stay-at-home orders associate with subsequent decreases in COVID-19 cases and fatalities in the United States"	COVID-19 mortality	Uses U.S. county data on COVID-19 deaths and SIPO data collected by the New York Times to estimate the effect of SIPO's using a two-way fixed-effects difference-in-differences model. Find a large and early (after few days) effect of SIPO on COVID-19 related deaths.	Stay-at-home orders are also associated with a 59.8 percent (18.3 to 80.2) average reduction in weekly fatalities after three weeks. These results suggest that stay-at-home orders	Finds the largest effect of SIPO on deaths after 10 days (see Figure 4), before a SIPO-related effect could possibly be seen as it takes 2-3 weeks from infection to death. This could indicate that other factors drive their results.

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
			might have reduced confirmed cases by 390,000 (170,000 to 680,000) and fatalities by 41,000 (27,000 to 59,000) within the first three weeks in localities that implemented stay-at-home orders.	
Fuller et al. (2021); "Mitigation Policies and COVID-19–Associated Mortality — 37 European Countries, January 23–June 30, 2020"	COVID-19 mortality	Uses COVID-19-deaths and OxCGRT stringency in 37 European countries to estimate the effect of lockdown on the number of COVID-19-deaths. Find a significant negative (fewer deaths) effect of stricter lockdowns after mortality threshold is reached (the threshold is a daily rate of 0.02 new COVID-19 deaths per 100,000 population (based on a 7-day moving average))	For each 1-unit increase in OxCGRT stringency index, the cumulative mortality decreases by 0.55 deaths per 100,000.	
Gibson (2020); "Government mandated lockdowns do not reduce Covid-19 deaths: implications for evaluating the stringent New Zealand response"	COVID-19 mortality	Uses data for every county in the United States from March through June 1, 2020, to estimate the effect of SIPO (called "lockdown") on COVID-19 mortality. Policy data are acquired from American Red Cross reporting on emergency regulations. His control variables include county population and density, the elder share, the share in nursing homes, nine other demographic and economic characteristics and a set of regional fixed effects. Handles causality problems using instrument variables (IV).	Find no statistically significant effect of SIPO.	Gibson use the word "lockdown" as synonym for SIPO (writes "technically, government-ordered community quarantine")
Goldstein et al. (2021); "Lockdown Fatigue: The Diminishing Effects of Quarantines on the Spread of COVID-19 "	COVID-19 mortality	Uses panel data from 152 countries with data from the onset of the pandemic until December 31, 2020. Finds that lockdowns tend to reduce the number of COVID-19 related deaths, but also that this benign impact declines over time: after four months of strict lockdown, NPIs have a significantly weaker contribution in terms of their effect in reducing COVID-19 related fatalities.	Stricter lockdowns reduce deaths for the first 60 days, whereafter the cumulative effect begins to decrease. If reintroduced after 120, the effect of lockdowns is smaller in the short run, but after 90 days the effect is almost the same as during first lockdown (only app. 10% lower).	There is little documentation in the study (e.g. no tables with estimates).
Guo et al. (2021); "Mitigation Interventions in the United States: An Exploratory Investigation of Determinants and Impacts"	COVID-19 mortality	Uses policy data from 1,470 executive orders from the state–government websites for all 50 states and Washington DC and COVID-19-deaths from Johns Hopkins University in a random-effect spatial error panel model to estimate the effect of nine NPIs (SIPO, strengthened SIPO, public school closure, all school closure, large-gathering ban of more than 10 people, any gathering ban, restaurant/bar limit to dining out only, nonessential business closure, and mandatory self-quarantine of travelers) on COVID-19 deaths.	Two mitigation strategies (all school closure and mandatory self-quarantine of travelers) showed positive (more deaths) impact on COVID-19-deaths per 10,000. Six mitigation strategies (SIPO, public school closure, large gathering bans (>10), any gathering ban, restaurant/bar limit to dining out only, and nonessential business	Only conclude on NPIs which reduce mortality. However, the conclusion is based on one-tailed tests, which means that all positive estimates (more deaths) are deemed insignificant. Thus, in their mortality-specification (Table 3, Proportion of Cumulative Deaths Over the Population), the estimate of all school closures (.204) and mandatory self-quarantine of travelers (0.363) is deemed insignificant based on schools CI [.029, .379] and quarantine CI [.193, .532]. We believe, these results should be interpreted as a significant increase in mortality, and that these results should have been part of their conclusion.

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
			closure) did not show any impact (Table 3, "Proportion of Cumulative Deaths Over the Population).	
Hale et al. (2020); "Global assessment of the relationship between government response measures and COVID-19 deaths"	COVID-19 mortality	Uses the OxCGRT stringency and COVID-19-deaths from the European Centre for Disease Prevention and Control for 170 countries. Estimates both cross-sectional models in which countries are the unit of analysis, as well as longitudinal models on time-series panel data with country-day as the unit of analysis (including models that use both time and country fixed effects).	Finds that higher stringency in the past leads to a lower growth rate in the present, with each additional point of stringency corresponding to a 0.039%-point reduction in daily deaths growth rates six weeks later.	
Hunter et al. (2021); "Impact of non-pharmaceutical interventions against COVID-19 in Europe: A quasi-experimental non-equivalent group and time-series"	COVID-19 mortality	Uses death data from the European Centre for Disease Prevention and Control (ECDC) and NPI-data from the Institute of Health Metrics and Evaluation. Argues that they use a quasi-experimental approach to identify the effect of NPIs because no analyzed intervention was imposed by all European countries and interventions were put in place at different points in the development of the epidemics.	Finds that mass gathering restrictions and initial business closures (businesses such as entertainment venues, bars and restaurants) reduces the number of deaths, whereas closing educational facilities and issuing SIPO increases the number of deaths. Finds no effect of closing non-essential services and mandating/recommending masks (Table 3)	Finds an effect of closing educational facilities and non-essential services after 1-7 days before lockdown could possibly have an effect on the number of deaths. This may indicate that other factors are driving their results.
Langeland et al. (2021); "The Effect of State Level COVID-19 Stay-at-Home Orders on Death Rates"	COVID-19 mortality	Estimates the effect of state-level lockdowns on COVID-19 deaths using multiple quasi-Poisson regressions with lockdown time length as the explanatory variable. Does not specify how lockdown is defined and what their data sources are.	Finds no significant effect of SIPO on the number of deaths after 2-4, 4-6 and 6+ weeks.	They write that "6+ weeks of lockdown is the only setting where the odds of dying are statistically higher than in the no lockdown case." However, all estimates are insignificant in Table C. Looks as if lockdown duration may cause a causality problem, because politicians may be less likely to ease restrictions when there are many cases/deaths.
Leffler et al. (2020); "Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks"	COVID-19 mortality	Use COVID-19 deaths from Worldometer and info about NPIs (mask/mask recommendations, international travel restrictions and lockdowns (defined as any closure of schools or workplaces, limits on public gatherings or internal movement, or stay-at-home orders) from Hale et al. (2020) for 200 countries to estimate the effect of the duration of NPIs on the number of deaths.	Finds that masking (mask recommendations) reduces mortality. For each week that masks were recommended the increase in per-capita mortality was 8.1% (compared to 55.7% increase when masks were not recommended). Finds no significant effect of the number of weeks with internal lockdowns and international travel restrictions (Table 2).	Their "mask recommendation" category includes some countries, where masks were mandated (see Supplemental Table A1) and may (partially) capture the effect of mask mandates. Looks at duration which may cause a causality problem, because politicians may be less likely to ease restrictions when there are many cases/deaths.
Mccafferty and Ashley (2021); "Covid-19 Social Distancing Interventions by Statutory Mandate and Their Observational	Other	Use data from 27 U.S. states and 12 European countries to analyze the effect of NPIs on peak morality rate using general linear mixed effects modelling.	Finds that no mandate (school closures, prohibition on mass gatherings, business closures, stay at home	

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
Correlation to Mortality in the United States and Europe"			orders, severe travel restrictions, and closure of non-essential businesses) was effective in reducing the peak COVID-19 mortality rate.	
Pan et al. (2020); "Covid-19: Effectiveness of non-pharmaceutical interventions in the united states before phased removal of social distancing protections varies by region"	COVID-19 mortality	Uses county-level data for all U.S. states. Mortality is obtained from Johns Hopkins, while policy data are obtained from official governmental websites. Categorizes 12 policies into 4 levels of disease control; Level 1 (low) - State of Emergency; Level 2 (moderate) - school closures, restricting access (visits) to nursing homes, or closing restaurants and bars; Level 3 (high) - non-essential business closures, suspending non-violent arrests, suspending elective medical procedures, suspending evictions, or restricting mass gatherings of at least 10 people; and Level 4 (aggressive) - sheltering in place / stay-at-home, public mask requirements, or travel restrictions. Use stepped-wedge cluster randomized trial (SW-CRT) for clustering and negative binomial mixed model regression.	Concludes that only (duration of, see comment in next column) level 4 restrictions are associated with reduced risk of death, with an average 15% decline in the COVID-19 death rate per day. Implementation of level 3 and level 2 restrictions increased death rates in 6 of 6 regions, while longer duration increased death rates in 5 of 6 regions.	They focus on the negative estimate of duration of Level 4. However, their implementation estimate is large and positive, and the combined effect of implementation and duration is unclear.
Pincombe et al. (2021); "The effectiveness of national-level containment and closure policies across income levels during the COVID-19 pandemic: an analysis of 113 countries"	COVID-19 mortality	Uses daily data for 113 countries on cumulative COVID-19 death counts over 130 days between February 15, 2020, and June 23, 2020, to examine changes in mortality growth rates across the World Bank's income group classifications following shelter-in-place recommendations or orders (they use one variable covering both recommendations and orders).	Finds that shelter-in-place recommendations/orders reduces mortality growth rates in high income countries (although insignificant) but increases growth rates in countries in other income groups.	
Sears et al. (2020); "Are we #stayinghome to Flatten the Curve?"	COVID-19 mortality	Uses cellular location data from all 50 states and the District of Columbia to investigate mobility patterns during the pandemic across states and time. Adding COVID-19 death tolls and the timing of SIPO for each state they estimate the effect of stay-at-home policies on COVID-19 mortality.	Find that SIPOs lower deaths by 0.13- 0.17 per 100,000 residents, equivalent to death rates 29-35% lower than in the absence of policies. However, these estimates are insignificant at a 95% confidence interval (see Table 4). The study also finds reductions in activity levels prior to mandates. Human encounter rate fell by 63 percentage points and nonessential visits by 39 percentage points relative to pre-COVID-19 levels, prior to any state implementing a statewide mandate	In the abstract the authors state that death rates would be 42-54% lower than in the absence of policies. However, this includes averted deaths due to pre-mandate social distancing behavior (p. 6). The effect of SIPO is a reduction in deaths by 29%-35% compared to a situation without SIPO but with pre-mandate social distancing. These estimates are insignificant at a 95% confidence interval.

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
Shiva and Molana (2021); "The Luxury of Lockdown"	COVID-19 mortality	Uses COVID-19-deaths and OxCGRT stringency from 169 countries to estimate the effect of lockdown on the number of deaths 1-8 weeks later. Finds that stricter lockdowns reduce COVID-19-deaths 4 weeks later (but insignificant 8 weeks later) and have the greatest effect in high income countries. Finds no effect of workplace closures in low-income countries.	A stricter lockdown (1 stringency point) reduces deaths by 0,1% after 4 weeks. After 8 weeks the effect is insignificant.	
Spiegel and Tookes (2021); "Business restrictions and Covid-19 fatalities"	COVID-19 mortality	Use data for every county in the United States from March through December 2020 to estimate the effect of various NPIs on the COVID-19-deaths growth rate. Derives causality by 1) assuming that state regulators primarily focus on the state's most populous counties, so state regulation in smaller counties can be viewed as a quasi randomized experiment, and 2) conducting county pair analysis, where similar counties in different states (and subject to different state policies) are compared.	Finds that some interventions (e.g. mask mandates, restaurant and bar closures, gym closures, and high-risk business closures) reduces mortality growth, while other interventions (closures of low- to medium-risk businesses and personal care/spa services) did not have an effect and may even have increased the number of deaths.	In total they analyze the lockdown effect of 21 variables. 14 of 21 estimates are significant, and of these 6 are negative (reduces deaths) while 8 are positive (increases deaths). Some results are far from intuitive. E.g. mask recommendations increases deaths by 48% while mask mandates reduces deaths by 12%, and closing restaurants and bars reduces deaths by 50%, while closing bars but not restaurants only reduces deaths by 5%.
Stockenhuber (2020); "Did We Respond Quickly Enough? How Policy-Implementation Speed in Response to COVID-19 Affects the Number of Fatal Cases in Europe"	COVID-19 mortality	Uses data for the number of COVID-19 infections and deaths and policy information for 24 countries from OxCGRT to estimate the effect of stricter lockdowns on the number of deaths using principal component analysis and a generalized linear mixed model.	Finds no significant effect of stricter lockdowns on the number of fatalities (Table 4).	Groups data on lockdown strictness into four groups and lose significant information and variation.
Stokes et al. (2020); "The relative effects of non-pharmaceutical interventions on early Covid-19 mortality: natural experiment in 130 countries"	COVID-19 mortality	Uses daily Covid-19 deaths for 130 countries from the European Centre for Disease Prevention and Control (ECDC) and daily policy data from the Oxford COVID-19 Government Response Tracker (OxCGRT). Looks at all levels of restrictions for each of the nine sub-categories of the OxCGRT stringency index (school, work, events, gatherings, transport, SIPO, internal movement, travel).	Of the nine sub-categories in the OxCGRT stringency index, only travel restrictions are consistently significant (with level 2 "Quarantine arrivals from high-risk regions" having the largest effect, and the strictest level 4 "Total border closure" having the smallest effect). Restrictions on very large gatherings (>1,000) has a large significant negative (fewer deaths) effect, while the effect of stricter restrictions on gatherings are insignificant. Authors recommend that the closing of schools (level 1) has a very large (in absolute terms it's twice the effect of border quarantines) positive	Their results are counter intuitive and somewhat inconclusive. Why does limiting very large gatherings (>1,000) work, while stricter limits do not? Why do recommending school closures cause more deaths? Why is the effect of border closures before 1st death insignificant, while the effect of closing borders after 1st death is significant (and large)? And why does quarantining arrivals from high-risk regions work better than total border closures? With 23 estimated parameters in total these counter intuitive and inconclusive results could be caused by multiple test bias (we correct for this in the meta-analysis), but may also be caused by other factors such as omitted variable bias.

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
			effect (more deaths) while stricter interventions on schools have no significant effect. Required cancelling of public events also has a significant positive (more deaths) effect. We focus on their 14-38 days results, as they catch the longest time frame (their 0-24 day model returns mostly insignificant results).	
Toya and Skidmore (2020); "A Cross-Country Analysis of the Determinants of Covid-19 Fatalities"	COVID-19 mortality	Uses COVID-19-deaths and lockdown info from various sources from 159 countries in a cross-country event study. Controls for country specifics by including socio-economic, political, geographic, and policy information. Finds little evidence for the efficacy of NPIs.	Complete travel restrictions prior to April 2020 reduced deaths by -0.226 per 100.000 by April 1st 2021, while mandatory national lockdown prior to April 2020 increased deaths by 0.166 by April 1st 2021. Recommended local lockdowns reduced deaths but results are based on one observation. Partial travel restrictions, mandatory local lockdowns and recommended national lockdowns did not have a significant effect on deaths.	The study looks at the lockdown status prior to April 2020 and the effect on deaths the following year (until April 1st 2021). The authors state this is to reduce concerns about endogeneity but do not explain why the lockdowns in the spring of 2020 are a good instrument for lockdowns during later waves are.
Tsai et al. (2021); "Coronavirus Disease 2019 (COVID-19) Transmission in the United States Before Versus After Relaxation of Statewide Social Distancing Measures"	Reproduction rate, Rt	Uses data for NPIs that were implemented and/or relaxed in U.S. states between 10 March and 15 July 2020. Using segmented linear regression, they estimate the extent to which relaxation of social distancing affected epidemic control, as indicated by the time-varying, state-specific effective reproduction number (Rt). Rt is based on death tolls.	Finds that in the 8 weeks prior to relaxing NPIs, Rt was declining, while after relaxation Rt started to increase.	Their Figure 1 shows that Rt on average increases app. 10 days before relaxation, which could indicate that other factors (omitted variables) affect the results.

Note: All comments on the significance of estimates are based on a 5% significance level unless otherwise stated.

It is difficult to make a conclusion based on the overview in Table 1. Is -0.073 to -0.326 deaths/million per stringency point, as estimated by Ashraf (2020), a large or a small effect relative to. the 98% reduction in mortality predicted by the study published by the Imperial College London (Ferguson et al. (2020). This is the subject for our meta-analysis in the next section. Here, it turns out that -0.073 to -0.326 deaths/million per stringency point is a relatively modest effect and only corresponds to a 2.4% reduction in COVID-19 mortality on average in the U.S. and Europe.

4 Meta-analysis: The impact of lockdowns on COVID-19 mortality

We now turn to the meta-analysis, where we focus on the impact of lockdowns on COVID-19 mortality.

In the meta-analysis, we include 24 studies in which we can derive the relative effect of lockdowns on COVID-19 mortality, where mortality is measured as COVID-19-related deaths per million. In practice, this means that the studies we included estimate the effect of lockdowns on mortality or the effect of lockdowns on mortality growth rates, while using a counterfactual estimate.[26]

Our focus is on the effect of compulsory non-pharmaceutical interventions (NPI), policies that restrict internal movement, close schools and businesses, and ban international travel, among others. We do not look at the effect of voluntary behavioral changes (e.g. voluntary mask wearing), the effect of recommendations (e.g. recommended mask wearing), or governmental services (voluntary mass testing and public information campaigns), but only on mandated NPIs.

The studies we examine are placed in three categories. Seven studies analyze the effect of stricter lockdowns based on the OxCGRT stringency indices, 13 studies analyze the effect of SIPOs (6 studies only analyze SIPOs, while seven analyze SIPOs among other interventions), and 11 studies analyze the effect of specific NPIs independently (lockdown vs. no lockdown).[27] Each of these categories is handled so that comparable estimates can be made across categories. Below, we present the results for each category and show the overall results, as well as those based on various quality dimensions.

Quality dimensions

We include quality dimensions because there are reasons to believe that can affect a study's conclusion. Below we describe the dimensions, as well as our reasons to believe that they are necessary to fully understand the empirical evidence.

- *Peer-reviewed vs. working papers*: We distinguish between peer-reviewed studies and working papers as we consider peer-reviewed studies generally being of higher quality than working papers.[28]

- *Long vs. short time period*: We distinguish between studies based on long time periods (with data series ending *after* May 31, 2020) and short time periods (data series ending at or before May 31, 2020), because the first wave did not fully end before late June in the U.S. and Europe. Thus, studies relying on short data periods lack the last part of the first wave and may yield biased results if lockdowns only "flatten the curve" and do not prevent deaths.

[26] As a minimum requirement, one needs to know the effect on the top of the curve.
[27] The total is larger than 21 because the 11 SIPO studies include seven studies which look at multiple measures.
[28] Vetted papers from CEPR Covid Economics are considered as working papers in this regard.

- *No early effect on mortality*: On average, it takes approximately three weeks from infection to death.[29] However, several studies find effects of lockdown on mortality almost immediately. Fowler et al. (2021) find a significant effect of SIPOs on mortality after just four days and the largest effect after 10 days. An early effect may indicate that other factors (omitted variables) drive the results, and, thus, we distinguish between studies which find an effect on mortality sooner than 14 days after lockdown and those that do not.[30] Note that many studies do not look at the short term and thus fall into the latter category by default.

- *Social sciences vs. other sciences*: While it is true that epidemiologists and researchers in natural sciences should, in principle, know much more about COVID-19 and how it spreads than social scientists, social scientists are, in principle, experts in evaluating the effect of various policy interventions. Thus, we distinguish between studies published by scholars in social sciences and by scholars from other fields of research. We perceive the former as being better suited for examining the effects of lockdowns on mortality. For each study, we have registered the research field for the corresponding author's associated institute (e.g., for a scholar from "Institute of economics" research field is registered as "Economics"). Where no corresponding author was available, the first author has been used. Afterwards, all research fields have been classified as either from the "Social Science" or "Other.""[31]

We also considered including a quality dimension to distinguish between studies based on excess mortality and studies based on COVID-19 mortality, as we believe that excess mortality is potentially a better measure for two reasons. First, data on total deaths in a country is far more precise than data on COVID-19 related deaths, which may be both underreported (due to lack of tests) or overreported (because some people die *with* – but not *because of* – COVID-19). Secondly, a major purpose of lockdowns is to save lives. To the extend lockdowns shift deaths *from* COVID-19 *to* other causes (e.g. suicide), estimates based on COVID-19 mortality will overestimate the effect of lockdowns. Likewise, if lockdowns save lives in other ways (e.g. fewer traffic accidents) lockdowns' effect on mortality will be underestimated. However, as only one

[29] Leffler et al. (2020) writes, "On average, the time from infection with the coronavirus to onset of symptoms is 5.1 days, and the time from symptom onset to death is on average 17.8 days. Therefore, the time from infection to death is expected to be 23 days." Meanwhile, Stokes et al. (2020) writes that "evidence suggests a mean lag between virus transmission and symptom onset of 6 days, and a further mean lag of 18 days between onset of symptoms and death."

[30] Some of the authors are aware of this problem. E.g. Bjørnskov (2021a) writes "when the lag length extends to three or fourth weeks, that is, the length that is reasonable from the perspective of the virology of Sars-CoV-2, the estimates become very small and insignificant" and "these results confirm the overall pattern by being negative and significant when lagged one or two weeks (the period when they cannot have worked) but turning positive and insignificant when lagged four weeks."

[31] Research fields classified as social sciences were economics, public health, management, political science, government, international development, and public policy, while research fields not classified as social sciences were ophthalmology, environment, medicine, evolutionary biology and environment, human toxicology, epidemiology, and anesthesiology.

of the 34 studies (Bjørnskov (2021a)) is based on excess mortality, we are unfortunately forced to disregard this quality dimension.

Meta-data used for our quality dimensions as well as other relevant information are shown in Table 2.

Table 2: Metadata for the studies included in the meta-analysis

1. Study (Author & title)	2. Included in meta-analysis	3. Publication status	4. End of data period	5. Earliest effect	6. Field of research	7. Lockdown measure	8. Geographical coverage
Alderman and Harjoto (2020); "COVID-19: U.S. shelter-in-place orders and demographic characteristics linked to cases, mortality, and recovery rates"	Yes	Peer-review	11-Jun-20	n/a	Economics (Social science)	SIPO	United States
Aparicio and Grossbard (2021); "Are Covid Fatalities in the U.S. Higher than in the EU, and If so, Why?"	Yes	Peer-review	22-Jul-20	n/a	Economics (Social science)	Specific NPIs	Europe and United States
Ashraf (2020); "Socioeconomic conditions, government interventions and health outcomes during COVID-19"	Yes	WP	20-May-20	n/a	Economics (Social science)	Stringency	World
Auger et al. (2020); "Association between statewide school closure and COVID-19 incidence and mortality in the U.S."	Yes	Peer-review	07-May-20	>21 days	Medicine (Other)	Specific NPIs	United States
Berry et al. (2021); "Evaluating the effects of shelter-in-place policies during the COVID-19 pandemic"	Yes	Peer-review	30-May-20	8-14 days	Public policy (Social science)	SIPO	United States
Bjørnskov (2021a); "Did Lockdown Work? An Economist's Cross-Country Comparison"	Yes	Peer-review	30-Jun-20	<8 days	Economics (Social science)	Stringency	Europe
Blanco et al. (2020); "Do Coronavirus Containment Measures Work? Worldwide Evidence"	No	WP	31-Aug-20	8-14 days	Economics (Social science)	Specific NPIs	World
Bonardi et al. (2020); "Fast and local: How did lockdown policies affect the spread and severity of the covid-19"	Yes	WP	13-Apr-20	<8 days	Economics (Social science)	Specific NPIs	World
Bongaerts et al. (2021); "Closed for business: The mortality impact of business closures during the Covid-19 pandemic"	Yes	Peer-review	13-Apr-20	8-14 days	Management (Social science)	Specific NPIs	One country
Chaudhry et al. (2020); "A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes"	Yes	Peer-review	01-Apr-20	n/a	Anesthesiology (Other)	Specific NPIs	World
Chernozhukov et al. (2021); "Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S."	Yes	Peer-review	03-Aun-20	n/a	Economics (Social science)	Specific NPIs	United States
Chisadza et al. (2021); "Government Effectiveness and the COVID-19 Pandemic"	Yes	Peer-review	01-Sep-20	n/a	Economics (Social science)	Stringency	World
Dave et al. (2021); "When Do Shelter-in-Place Orders Fight Covid-19 Best? Policy Heterogeneity Across States and Adoption Time"	Yes	Peer-review	20-Apr-20	Finds no effect	Economics (Social science)	SIPO	United States
Dergiades et al. (2020); "Effectiveness of government policies in response to the COVID-19 outbreak"	No	WP	30-Apr-20	n/a	Management (Social science)	Stringency	World
Fakir and Bharati (2021); "Pandemic catch-22: The role of mobility restrictions and institutional inequalities in halting the spread of COVID-19"	No	Peer-review	30-Jul-20	<8 days	Economics (Social science)	Stringency	World

1. Study (Author & title)	2. Included in meta-analysis	3. Publication status	4. End of data period	5. Earliest effect	6. Field of research	7. Lockdown measure	8. Geographical coverage
Fowler et al. (2021); "Stay-at-home orders associate with subsequent decreases in COVID-19 cases and fatalities in the United States"	Yes	Peer-review	07-May-20	<8 days	Public Health (Social science)	SIPO	United States
Fuller et al. (2021); "Mitigation Policies and COVID-19–Associated Mortality — 37 European Countries, January 23–June 30, 2020"	Yes	WP	30-Jun-20	n/a	Epidemiology (Other)	Stringency	Europe
Gibson (2020); "Government mandated lockdowns do not reduce Covid-19 deaths: implications for evaluating the stringent New Zealand response"	Yes	Peer-review	01-Jun-20	Finds no effect	Economics (Social science)	SIPO	United States
Goldstein et al. (2021); "Lockdown Fatigue: The Diminishing Effects of Quarantines on the Spread of COVID-19 "	Yes	WP	31-Dec-20	<8 days	International Development (Social science)	Stringency	World
Guo et al. (2021); "Mitigation Interventions in the United States: An Exploratory Investigation of Determinants and Impacts"	Yes	Peer-review	07-Apr-20	n/a	Social work (Social science)	Specific NPIs	United States
Hale et al. (2020); "Global assessment of the relationship between government response measures and COVID-19 deaths"	No	WP	27-May-20	n/a	Government (Social science)	Stringency	World
Hunter et al. (2021); "Impact of non-pharmaceutical interventions against COVID-19 in Europe: A quasi-experimental non-equivalent group and time-series"	No	Peer-review	24-Apr-20	<8 days	Medicine (Other)	Specific NPIs	Europe
Langeland et al. (2021); "The Effect of State Level COVID-19 Stay-at-Home Orders on Death Rates"	No	WP	Not specified	Finds no effect	Political Science (Social science)	Other	United States
Leffler et al. (2020); "Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks"	Yes	Peer-review	09-May-20	n/a	Ophthalmology (Other)	Specific NPIs	World
Mccafferty and Ashley (2021); "Covid-19 Social Distancing Interventions by Statutory Mandate and Their Observational Correlation to Mortality in the United States and Europe"	No	Peer-review	12-Apr-20	Finds no effect	Ophthalmology (Other)	Specific NPIs	Europe and United States
Pan et al. (2020); "Covid-19: Effectiveness of non-pharmaceutical interventions in the united states before phased removal of social distancing protections varies by region"	No	WP	29-May-20	n/a	Environment (Other)	Specific NPIs	United States
Pincombe et al. (2021); "The effectiveness of national-level containment and closure policies across income levels during the COVID-19 pandemic: an analysis of 113 countries"	No	Peer-review	23-Jun-20	n/a	Health Science (Social science)	SIPO	World
Sears et al. (2020); "Are we #stayinghome to Flatten the Curve?"	Yes	WP	29-Apr-20	Finds no effect	Economics (Social science)	SIPO	United States
Shiva and Molana (2021); "The Luxury of Lockdown"	Yes	Peer-review	08-Jun-20	15-21 days	Government (Social science)	Stringency	World
Spiegel and Tookes (2021); "Business restrictions and Covid-19 fatalities"	Yes	Peer-review	31-Dec-20	<8 days	Management (Social science)	Specific NPIs	United States
Stockenhuber (2020); "Did We Respond Quickly Enough? How Policy-Implementation Speed in Response to COVID-19 Affects the Number of Fatal Cases in Europe"	Yes	Peer-review	12-Jul-20	n/a	Evolutionary Biology and Environment (Other)	Stringency	Europe
Stokes et al. (2020); "The relative effects of non-pharmaceutical interventions on early	Yes	WP	01-Jun-20	n/a	Economics (Social science)	Specific NPIs	World

1. Study (Author & title)	2. Included in meta-analysis	3. Publication status	4. End of data period	5. Earliest effect	6. Field of research	7. Lockdown measure	8. Geographical coverage
Covid-19 mortality: natural experiment in 130 countries"							
Toya and Skidmore (2020); "A Cross-Country Analysis of the Determinants of Covid-19 Fatalities"	Yes	WP	01-Apr-21	n/a	Economics (Social science)	Specific NPIs	World
Tsai et al. (2021); "Coronavirus Disease 2019 (COVID-19) Transmission in the United States Before Versus After Relaxation of Statewide Social Distancing Measures"	No	Peer-review	15-Jul-20	<8 days	Psychiatry (Social science)	Specific NPIs	United States

Note: Research fields classified as social sciences were economics, public health, health science, management, political science, government, international development, and public policy, while research fields not classified as social sciences were ophthalmology, environment, medicine, evolutionary biology and environment, human toxicology, epidemiology and anesthesiology.

Interpreting and weighting estimates

The estimates used in the meta-analysis are not always readily available in the studies shown in Table 2. In Appendix B Table 9, we describe for each paper how we interpret the estimates and how they are converted to a common estimate (the relative effect of lockdowns on COVID-19 mortality) which is comparable across all studies.

Following Paldam (2015) and Stanley and Doucouliagos (2010), we also convert standard errors[32] and use the precision of each estimate (defined as 1/SE) to calculate the precision-weighted average of all estimates and present funnel plots. The precision-weighted average is our primary indicator of the efficacy of lockdowns, but we also report arithmetic averages and medians in the meta-analysis.

In the following sections, we present the meta-analysis for each of the three groups of studies (stringency index-studies, SIPO-studies, and studies analyzing specific NPIs).

4.1 Stringency index studies

Seven eligible studies examine the link between lockdown stringency and COVID-19 mortality. The results from these studies, converted to common estimates, are presented in Table 3 below. All studies are based on the COVID-19 Government Response Tracker's (OxCGRT) stringency index of Oxford University's Blavatnik School of Government (Hale et al. (2020)).

The OxCGRT stringency index neither measures the expected effectiveness of the lockdowns nor the expected costs. Instead, it describes the stringency based on nine equally weighted parameters.[33] Many countries followed similar patterns and almost all countries closed schools,

[32] Standard errors are converted such that the t-value, calculated based on common estimates and standard errors, is unchanged. When confidence intervals are reported rather than standard errors, we calculate standard errors using t-distribution with ∞ degrees of freedom (i.e. 1.96 for 95% confidence interval).

[33] The nine parameters are "C1 School closing," "C2 Workplace closing," "C3 Cancel public events," "C4 Restrictions on gatherings," "C5 Close public transport," "C6 Stay at home requirements," "C7 Restrictions on internal movement," "C8 International travel controls" and "H1 Public information campaigns." The latter, "H1

while only a few countries issued SIPOs without closing businesses. Hence, it is reasonable to perceive the stringency index as continuous, although not necessarily linear. The index includes recommendations (e.g. "workplace closing" is 1 if the government recommends closing (or work from home), cf. Hale et al. (2021)), but the effect of including recommendations in the index is primarily to shift the index parallelly upward and should not alter the results relative to our focus on mandated NPIs. It is important to note that the index is not perfect. As pointed out by Book (2020), it is certainly possibly to identify errors and omissions in the index. However, the index is objective and unbiased and as such, useful for cross-sectional analysis with several observations, even if not suitable for comparing the overall strictness of lockdowns in two countries.

Since the studies examined use different units of estimates, we have created common estimates for Europe and United States to make them comparable. The common estimates show the effect of the average lockdown in Europe and United States (with average stringencies of 76 and 74, respectively, between March 16th and April 15th, 2020, compared to a policy based solely on recommendations (stringency 44)). For example, Ashraf (2020) estimates that the effect of stricter lockdowns is -0.073 to -0.326 deaths/million per stringency point. We use the average of these two estimates (-0.200) in the meta-analysis (see Table 9 in Appendix B for a description for all studies). The average lockdown in Europe between March 16th and April 15th, 2020, was 32 points stricter than a policy solely based on recommendations (76 vs. 44). In United States, it was 30 points. Hence, the total effect of the lockdowns compared to the recommendation policy was -6.37 deaths/million in Europe (32 x -0.200) and -5.91 deaths/million in United States. With populations of 748 million and 333 million, respectively the total effect as estimated by Ashraf (2020) is 4,766 averted COVID-19 deaths in Europe and 1,969 averted COVID-19 deaths in United States. By the end of the study period in Ashraf (2020), which is May 20, 2020, 164,600 people in Europe and 97,081 people in the United States had died of COVID-19. Hence, the 4,766 averted COVID-19 deaths in Europe and the 1,969 averted COVID-19 deaths in the United States corresponds to 2.8% and 2.0% of all COVID-19 deaths, respectively, with an arithmetic average of 2.4%. Our common estimate is thus -2.4%, cf. Table 3. So, this means that Ashraf (2020) estimates that without lockdowns, COVID-19 deaths in Europe would have been 169,366 and COVID-19 deaths in the U.S. would have been 99,050. Our approach is not unproblematic. First of all, the level of stringency varies over time for all countries. We use the stringency between March 16th and April 15th, 2020 because this period covers the main part of the first wave which most of the studies analyze. Secondly, OxCGRT has changed the index over time and a 10-point difference today may not be exactly the same as a 10-point difference when the studies were finalized. However, we believe these problems are unlikely to significantly alter our results.

Public information campaigns," is not an intervention following our definition, as it is not a mandatory requirement. However, of 97 European countries and U.S. States in the OxCGRT database, only Andorra, Belarus, Bosnia and Herzegovina, Faeroe Islands, and Moldova – less than 1.6% of the population – did not get the maximum score by March 20, 2020, so the parameter simply shifts the index parallelly upward and should not have notable impact on the analyzes.

Table 3 demonstrates that the studies find that lockdowns, on average, have reduced COVID-19 mortality rates by 0.2% (precision-weighted). The results yield a median of -2.4% and an arithmetic average of -7.3%. Only one of the seven studies, Fuller et al. (2021), finds a significant *and* (relative to the effect predicted in studies like Ferguson et al. (2020)) substantial effect of lockdowns (-35%). The other six studies find much smaller effects. Hence, based on the stringency index studies, we find little to no evidence that mandated lockdowns in Europe and the United States had a noticeable effect on COVID-19 mortality rates. And, as will be discussed in the next paragraph, the fifth column of Table 3 displays the number of quality dimensions (out of 4) met by each study.

Table 3: Overview of common estimates from studies based on stringency indexes

Effect on COVID-19 mortality	**Estimate (Estimated Averted Deaths / Total Deaths)**	**Standard error**	**Weight (1/SE)**	**Quality dimensions**
Bjørnskov (2021)	-0.3%	0.8%	119	3
Shiva and Molana (2021)	-4.1%	0.4%	248	4
Stockenhuber (2020)*	0.0%	n/a	n/a	3
Chisadza et al. (2021)	0.1%	0.0%	7,390	4
Goldstein et al. (2021)	-9.0%	3.8%	26	2
Fuller et al. (2021)	-35.3%	9.1%	11	2
Ashraf (2020)	-2.4%	0.4%	256	2
Precision-weighted average (arithmetic average / median)	**-0.2% (-7.3%/-2.4%)**			

Note: The table shows the estimates for each study converted to a common estimate, i.e. the implied effect on COVID-19 mortality in Europe and United States. A negative number corresponds to fewer deaths, so -5% means 5% lover COVID-19 mortality. For studies which report estimates in deaths per million, the common estimate is calculated as: (COVID-19 mortality with "common area's" policy) / (COVID-19 mortality with recommendation policy) -1, where (COVID-19 mortality with recommendation policy) is calculated as ((COVID-19 mortality with "common area's" policy) - Estimate x Difference in stringency x population). Stringencies in Europe and United States are equal to the average stringency from March 16th to April 15th 2020 (76 and 74 respectively) and the stringency for the policy based solely on recommendations is 44 following Hale et al. (2020). For the conversion of other studies see Table 9 in appendix B.

* *It is not possible to calculate a common estimate for Stockenhuber (2020). When calculating arithmetic average / median, the study is included as 0%, because estimates are insignificant and signs of estimates are mixed (higher strictness can cause both lower and higher COVID-19 mortality).*

We now turn to the quality dimensions. Table 4 presents the results differentiated by the four quality dimensions. Two studies, Shiva and Molana (2021) and Chisadza et al. (2021), meet all quality dimensions. The precision-weighted average for these studies is 0.0%, meaning that lockdowns had no effect on COVID-19 mortality. Two studies live up to 3 of 4 quality dimensions (Bjørnskov (2021a) and Stockenhuber (2020)). The precision-weighted average for these studies is -0.3%, meaning that lockdowns reduced COVID-19 mortality by 0.3%. Three studies lack at least two quality dimensions.[34] These studies find that lockdowns reduce COVID-19 mortality by 4.2%. To sum up, we find that the studies that meet at least 3 of 4 quality measures find that lockdowns have little to no effect on COVID-19 mortality, while studies that

[34] In fact, the working papers by P. Goldstein et al. (2021), Fuller et al. (2021) and Ashraf (2020) all lack exactly two quality parameters.

meet 2 of 4 quality measures find a small effect on COVID-19 mortality. These results are far from those estimated with the use of epidemiological models, such as the Imperial College London (Ferguson et al. (2020).

Table 4: Overview of common estimates split on quality dimensions for studies based on stringency indexes

Values show effect on COVID-19 mortality	**Precision-weighted average***	**Arithmetic average**	**Median**
Peer-reviewed vs. working papers			
Peer-reviewed [4]	0.0%	-1.1%	-0.2%
Working paper [3]	-4.2%	-15.6%	-9.0%
Long vs. short time period			
Data series ends after 31 May 2020 [6]	-0.1%	-8.1%	-0.2%
Data series ends before 31 May 2020 [1]	-2.4%	-2.4%	-9.0%
No early effect on mortality			
Does not find an effect within the first 14 days (including n/a) [5]	-0.2%	-8.3%	-2.4%
Finds effect within the first 14 days [2]	-1.9%	-4.7%	-4.7%
Social sciences vs. other sciences			
Social sciences [5]	-0.1%	-3.1%	-2.4%
Other sciences [2]	-35.3%	-17.7%	-17.7%
4 of 4 quality dimensions [2]	**0.0%**	**-2.0%**	**-2.0%**
3 of 4 quality dimensions [2]	**-0.3%**	**-0.2%**	**-0.2%**
2 of 4 quality dimensions or fewer [3]	**-4.2%**	**-15.6%**	**-9.0%**

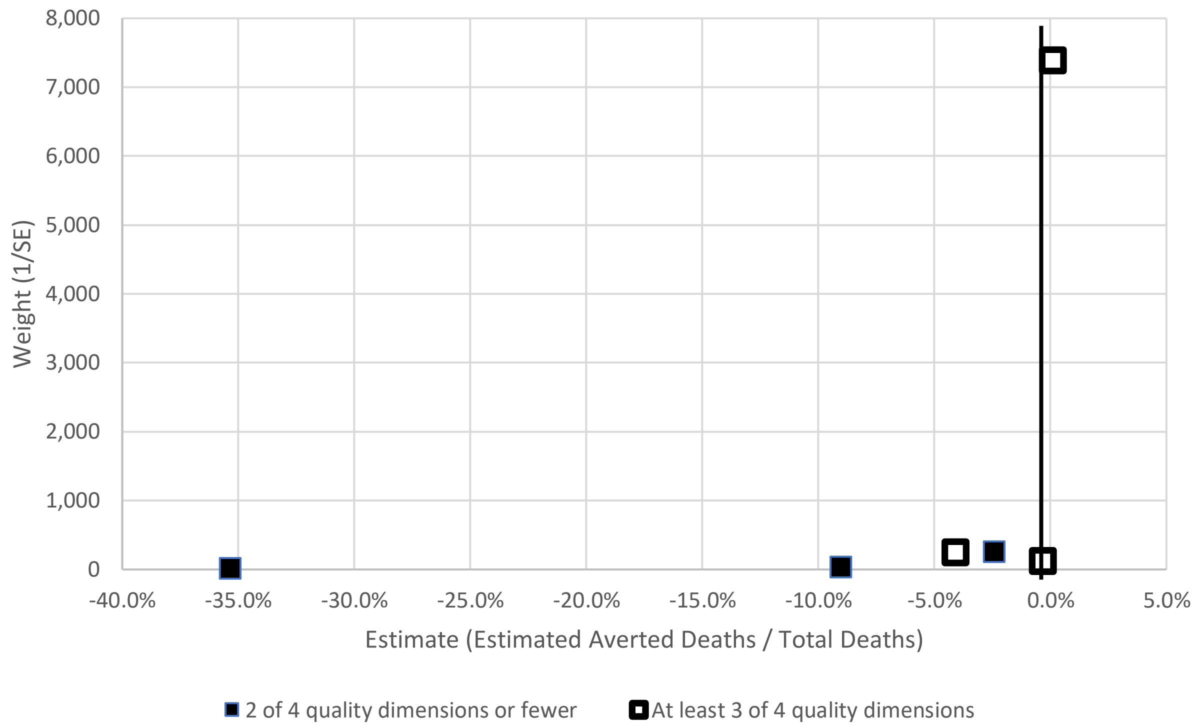
*Note: The table shows the common estimate as described in Table 3 for each quality dimension. The number of studies in each category is in square brackets. * The precision-weighted average does not include studies where no common standard error is available, cf. Table 3.*

Figure 5 shows a funnel plot for the studies in Table 3, except Stockenhuber (2020), where common estimate standard errors cannot be derived. Chisadza et al. (2021) has a far higher precision than the other studies (1/SE is 7,398 and the estimate is 0.1%)[35], and there are indications that the estimate from Fuller et al. (2021) (the bottom left) is an imprecise outlier.[36] Figure 5 The plot also shows that the studies with at least 3 of 4 quality dimensions are centered around zero and generally have higher precision than other studies.

[35] Excluding Chisadza et al. (2021) from the precision-weighted average changes the average to -3.5%.

[36] Excluding Fuller et al. (2021) from the precision-weighted average only marginally changes the average because the precision is very low.

Figure 5: Funnel plot for estimates from studies based on stringency indexes



Note: The figure displays all estimates and the precision of the estimate defined as one over the standard error. Studies where standard errors are not available are not included. Studies which live up to at least 3 of 4 quality dimensions are marked with white, while studies which lives up to 2 of 3 quality dimensions or less are marked with black. The vertical line illustrates the precision-weighted average.

Overall conclusion on stringency index studies

Compared to a policy based solely on recommendations, we find little evidence that lockdowns had a noticeable impact on COVID-19 mortality Only one study, Fuller et al. (2021), finds a substantial effect, while the rest of the studies find little to no effect. Indeed, according to stringency index studies, lockdowns in Europe and the United States reduced only COVID-19 mortality by 0.2% on average.

In the following section we will look at the effect of SIPOs. The section follows the same structure as this section.

4.2 Shelter-in-place order (SIPO) studies

We have identified 13 eligible studies which estimate the effect of Shelter-In-Place Orders (SIPOs) on COVID-19 mortality, cf. Table 5. Seven of these studies look at multiple NPIs of which a SIPO is just one, while six studies estimate the effect of a SIPO vs. no SIPO in the United States. According to the containment and closure policy indicators from OxCGRT, 41 states in the U.S. issued SIPOs in the spring of 2020. But usually, these were introduced after implementing other NPIs such as school closures or workplace closures. On average, SIPOs were issued 7½ days after *both* schools and workplaces closed, and 12 days after the first of the two closed. Only one state, Tennessee, issued a SIPO before schools and workplaces closed. The 10 states that did not issue SIPOs all closed schools. Moreover, of those 10 states, three closed some non-essential businesses, while the remaining 7 closed all non-essential businesses. Because of this, we perceive estimates for SIPOs based on U.S.-data as the marginal effect of SIPOs on top of other restrictions, although we acknowledge that the estimates may capture the effects of other NPI measures as well.

The results of eligible studies based on SIPOs are presented in Table 5. The table demonstrates that the studies generally find that SIPOs have reduced COVID-19 mortality by 2.9% (on a precision-weighted average). There is an apparent difference between studies in which a SIPO is one of multiple NPIs, and studies in which a SIPO is the only examined intervention. The former group generally finds that SIPOs *increase* COVID-19 mortality *marginally*, whereas the latter finds that SIPOs *decrease* COVID-19 mortality. As we will see below, this difference could be explained by differences in the quality dimensions, and especially the time period covered by each study.

Table 5: Overview of estimates from studies based on SIPOs

Values show effect on COVID-19 mortality	**Estimate (Estimated Averted Deaths / Total Deaths)**	**Standard error**	**Weight (1/SE)**	**Quality dimensions**
Studies where SIPO is one of several examined interventions and not (as) likely to capture the effect of other interventions				
Chernozhukov et al. (2021)	-17.7%	14.3%	7	4
Chaudhry et al. (2020) *	0.0%	n/a	n/a	2
Aparicio and Grossbard (2021)	2.6%	2.8%	35	4
Stokes et al. (2020)	0.8%	11.1%	9	3
Spiegel and Tookes (2021)	13.1%	6.6%	15	3
Bonardi et al. (2020)	0.0%	n/a	n/a	1
Guo et al. (2021)	4.6%	14.8%	4	3
Average (median) where SIPO is one of several variables	2.8% (0.5%/0.8%)			
Studies where SIPO is the only examined intervention and may capture the effect of other interventions				
Sears et al. (2020)	-32.2%	17.6%	6	2
Alderman and Harjoto (2020)	-1.0%	0.6%	169	4
Berry et al. (2020)	1.1%	n/a	n/a	2
Fowler et al. (2021)	-35.0%	7.0%	14	2
Gibson (2020)	-6.0%	24.3%	4	4
Dave et al. (2020)	-40.8%	36.1%	3	3
Average (median) where SIPO is the only variable	-5.1% (-19.0%/-19.1%)			
Precision-weighted average (arithmetic average / median) for all studies	**-2.9% (-8.5%/0.0%)**			

*Note: * Chaudhry et al. (2020) does not provide an estimate but states that SIPO is insignificant. We use 0% when calculating the arithmetic average and median. Chaudhry et al. (2020) and Berry et al. (2021) do not affect the precision-weighted average, as we do not know the standard errors.*

Table 6 presents the results differentiated by quality dimensions. Four studies (Chernozhukov et al. (2021), Aparicio and Grossbard (2021), Alderman and Harjoto (2020) and Gibson (2020)) meet all quality dimensions but find vastly different effects of SIPOs on COVID-19 mortality. The precision weighted average of the four studies is -1.0%. Four studies meet 3 of 4 quality dimensions. They overall find that SIPOs *increase* COVID-19 mortality, as the precision-weighted average is positive (3.7%). The five studies that meet 2 of 4 quality dimensions or fewer[37] find a substantial reduction in COVID-19-mortality (-34.2%). This substantial reduction seems to be driven by relatively short data series. The latest data point for the three studies which find large effects of lockdowns (Sears et al. (2020), Fowler et al. (2021), and Dave et al. (2021)) are April 29, May 7, and April 20, respectively. This may indicate that SIPOs can delay deaths but not eliminate them completely. Disregarding these studies with short data series, the precision-weighted average is -0.1%.

Table 6: Quality dimensions for studies based on SIPOs

Values show effect on COVID-19 mortality	Precision-weighted average*	Arithmetic average	Median
Peer-reviewed vs. working papers			
Peer-review [10]	-2.4%	-7.9%	-0.5%
Working paper [3]	-12.0%	-10.5%	0.0%
Long vs. short time period			
Data serie ends after 31 May 2020 [6]	-0.1%	-1.4%	-0.1%
Data serie ends before 31 May 2020 [7]	-25.9%	-14.6%	0.0%
No early effect on mortality			
Finds effect within the first 14 days [9]	-2.0%	-10.0%	-1.0%
Does not find an effect within the first 14 days (including n/a) [4]	-10.3%	-5.2%	0.0%
Social sciences vs. other sciences			
Social sciences [12]	-2.9%	-9.2%	-0.5%
Other sciences [1]	n/a	0.0%	0.0%
4 of 4 quality dimensions [4]	-1.0%	-5.5%	-3.5%
3 of 4 quality dimensions [4]	3.7%	-5.6%	2.7%
2 of 4 quality dimensions or fewer [5]	-34.2%	-13.2%	0.0%

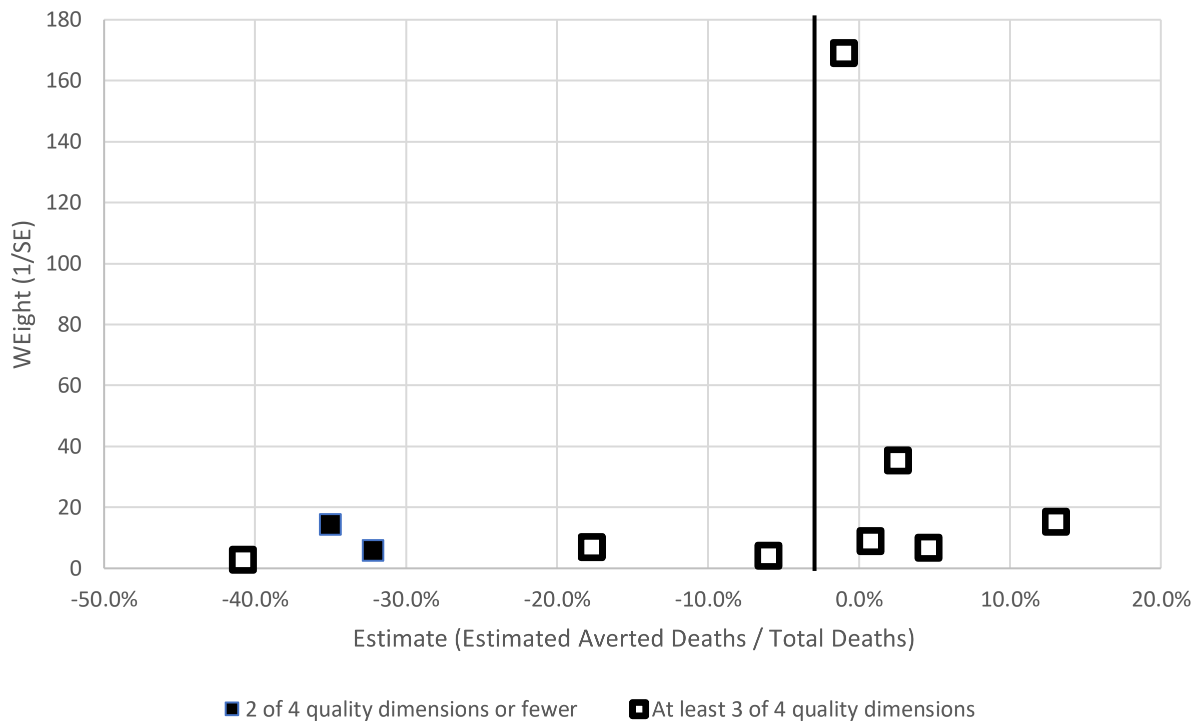
*Note: The table shows the common estimate as described in Table 5 for each quality dimension. The number of studies in each category is in square brackets. * The precision-weighted average does not include studies where no common standard error is available, cf. Table 5.*

Figure 6 shows a funnel plot for the studies in Table 5, except Chaudhry et al. (2020) and Berry et al. (2021), where common standard errors cannot be derived. Sears et al. (2020) stands out with a precision far higher than those of the other studies. But generally, the precisions of the studies are low and the estimates are placed on both sides of the zero-line with some ‘tail’ to the

[37] Bonardi et al. (2020) only meet one quality dimension (social science).

left.[38] Figure 5 also shows that four of eight studies with at least 3 of 4 quality dimensions find that SIPOs *increase* COVID-19 mortality by 0.8% to 13.1%.

Figure 6: Funnel plot for estimates from SIPO studies



Note: The figure displays all estimates and the precision of the estimate defined as one over the standard error. Studies where standard errors are not available are not included. Studies which live up to at least 3 of 4 quality dimensions are marked with white, while studies which lives up to 2 of 4 quality dimensions or less are marked with black. The vertical line illustrates the precision-weighted average.

Overall conclusion on SIPO studies

We find no clear evidence that SIPOs had a noticeable impact on COVID-19 mortality. Some studies find a large negative relationship between lockdowns and COVID-19 mortality, but this seems to be caused by short data series which does not cover a full COVID-19 'wave'. Several studies find a small positive relationship between lockdowns and COVID-19 mortality. Although this appears to be counterintuitive, it could be the result of an (asymptomatic) infected person being isolated at home under a SIPO can infect family members with a higher viral load causing more severe illness.[39] The overall effect measured by the precision-weighted average is -2.9%. The result is in line with Nuzzo et al. (2019), who state that "In the context of a high-impact

[38] This could indicate some publication bias, but the evidence is weak and with only 13 estimates, this cannot be formally tested

[39] E.g. see Guallar et al. (2020), who concludes, "Our data support that a greater viral inoculum at the time of SARS-CoV-2 exposure might determine a higher risk of severe COVID-19."

respiratory pathogen, quarantine may be the least likely NPI to be effective in controlling the spread due to high transmissibility" and World Health Organization Writing Group (2006), who conclude that "forced isolation and quarantine are ineffective and impractical."[40]

In the following section, we will look at the effect found in studies analyzing specific NPIs.

4.3 Studies of specific NPIs

A total of 11 eligible studies look at (multiple) specific NPIs independently or simply lockdown vs. no lockdown.[41] The definition of the specific NPIs varies from study to study and are somewhat difficult to compare. The variety in the definitions can be seen in the analysis of non-essential business closures and bar/restaurant closures. Chernozhukov et al. (2021) focus on a combined parameter (the average of business closure and bar/restaurant closure in each state), Aparicio and Grossbard (2021) look at business closure but not bar/restaurant closure, Spiegel and Tookes (2021) examine bar/restaurant closure but not business closure, and Guo et al. (2021) look at both business closures and bar/restaurant closures independently.

Some studies include several NPIs (e.g. Stokes et al. (2020) and Spiegel and Tookes (2021)), while others cover very few. Bongaerts et al. (2021) only study business closures, and Leffler et al. (2020) look at internal lockdown and international travel restrictions). Few NPIs in a model are potentially a problem because they can capture the effect of excluded NPIs. On the other hand, several NPIs in a model increase the risk of multiple test bias.

The differences in the choice of NPIs and in the number of NPIs make it challenging to create an overview of the results. In Table 7, we have merged the results in six overall categories but note that the estimates may not be fully comparable across studies. In particular, the lockdown-measure varies from study to study and in some cases is poorly defined by the authors. Also, there are only a few estimates within some of the categories. For instance, the estimate of the effect of facemasks is based on only two studies.

Table 7 illustrates that generally there is no evidence of a noticeable relationship between the most-used NPIs and COVID-19. Overall, lockdowns and limiting gatherings seem to increase COVID-19 mortality, although the effect is modest (0.6% and 1.6%, respectively) and border closures has little to no effect on COVID-19 mortality, with a precision-weighted average of -0.1% (removing the imprecise outlier from Guo et al. (2021) changes the precision-weighted average to -0.2%). We find a small effect of school closure (-4.4%), but this estimate is mainly driven by Auger et al. (2020), who – as noted earlier – use an "interrupted time series study"

[40] Both Nuzzo et al. (2019) and World Health Organization Writing Group (2006) focus on quarantining infected persons. However, if quarantining infected persons is not effective, it should be no surprise that quarantining uninfected persons could be ineffective too.

[41] Note that we – according to our search strategy – did not search on specific measures such as "school closures" but on words describing the overall political approach to the COVID-19 pandemic such as "non-pharmaceutical," "NPIs," "lockdown" etc.

approach and may capture other effects such as seasonal and behavioral effects. The absence of a notable effect of school closures is in line with Irfan et al. (2021), who – based on a systematic review and meta-analysis of 90 published or preprint studies of transmission in children – concluded that "risks of infection among children in educational-settings was lower than in communities. Evidence from school-based studies demonstrate it is largely safe for young children (<10 years of age) to be at schools; however, older children (between 10 and 19 years of age) might facilitate transmission." UNICEF (2021) and ECDC (2020) reach similar conclusions.[42]

Mandating facemasks – an intervention that was not widely used in the spring of 2020, and in many countries was even discouraged – seems to have a large effect (-21.2%), but this conclusion is based on only two studies.[43] Again, our categorization may play a role, as the larger mask-estimate from Chernozhukov et al. (2021) is in fact "employee facemasks," not a general mask mandate. Our findings are somewhat in contrast to the result found in a review by Liu et al. (2021), who conclude that "fourteen of sixteen identified randomized controlled trials comparing face masks to no mask controls failed to find statistically significant benefit in the intent-to-treat populations." Similarly, a pre-COVID Cochrane review concludes, "There is low certainty evidence from nine trials (3507 participants) that wearing a mask may make little or no difference to the outcome of influenza-like illness (ILI) compared to not wearing a mask (risk ratio (RR) 0.99, 95% confidence interval (CI) 0.82 to 1.18). There is moderate certainty evidence that wearing a mask probably makes little or no difference to the outcome of laboratory-confirmed influenza compared to not wearing a mask (RR 0.91, 95% CI 0.66 to 1.26; 6 trials; 3005 participants)" (Jefferson et al. (2020)).[44] However, it should be noted that even if no effect is found in controlled settings, this does not necessarily imply that mandated face masks does not reduce mortality, as other factors may play a role (e.g. wearing a mask may function as a tax on socializing if people are bothered by wearing a face masks when they are socializing).

---

[42] UNICEF (2021) concludes, "The preliminary findings thus far suggest that in-person schooling – especially when coupled with preventive and control measures – had lower secondary COVID-19 transmission rates compared to other settings and do not seem to have significantly contributed to the overall community transmission risks." Whereas, ECDC (2020) conclude, "School closures can contribute to a reduction in SARS-CoV-2 transmission, but by themselves are insufficient to prevent community transmission of COVID-19 in the absence of other nonpharmaceutical interventions (NPIs) such as restrictions on mass gathering," and states, "There is a general consensus that the decision to close schools to control the COVID-19 pandemic should be used as a last resort. The negative physical, mental health and educational impact of proactive school closures on children, as well as the economic impact on society more broadly, would likely outweigh the benefits."

[43] Note again, that we – according to our search strategy – did not search on the specific measures such as "masks," "face masks," "surgical masks" but on words describing the overall political approach to the COVID-19 pandemic such as "non-pharmaceutical," "NPIs," "lockdown" etc. Thus, we do not include most of the studies in mask reviews such as Liu et al. (2021) and Jefferson et al. (2020).

[44] Lipp and Edwards (2014) also find no evidence of an effect and – looking at disposable surgical face masks for preventing surgical wound infection in clean surgery – conclude, "Three trials were included, involving a total of 2113 participants. There was no statistically significant difference in infection rates between the masked and unmasked group in any of the trials." Meanwhile, Li et al. (2021) – based on six case-control studies – conclude, "In general, wearing a mask was associated with a significantly reduced risk of COVID-19 infection (OR = 0.38, 95% CI: 0.21-0.69, I^2 = 54.1%).

Only business closure consistently shows evidence of a negative relationship with COVID-19 mortality, but the variation in the estimated effect is large. Three studies find little to no effect, and three find large effects. Two of the larger effects are related to closing bars and restaurants. The "close business" category in Chernozhukov et al. (2021) is an average of closed businesses, restaurants, and movie theaters, while that same category is "closing restaurants and bars" in Spiegel and Tookes (2021). The last study finding a large effect is Bongaerts et al. (2021), the only eligible single-country study.[45]

As a final observation on Table 7, studies with fewer quality dimensions seem to find larger effects, but the pattern is not systematic.[46]

Table 7: Overview of estimates from studies of specific NPIs

	Lockdown (complete/ partial)	Facemasks/ Employee face masks	Business closure (/bars & restaurants)	Border closure (/quarantine)	School closures	Limiting gatherings	Quality dimensions
Chernozhukov et al. (2021)		-34.0%	-28.6%				4
Bongaerts et al. (2021)			-31.6%				2
Chaudhry et al. (2020)*	0.0%			0.0%			2
Toya & Skidmore (2021)	0.5%			-0.1%			3
Aparicio & Grossbard (2021)			-1.3%		0.5%	0.8%	4
Auger et al. (2020)					-58.0%		2
Leffler et al. (2020)	1.7%			-15.6%			2
Stokes et al. (2020)			0.3%	-24.6%	-0.1%	-6.3%	3
Spiegel & Tookes (2021)		-13.5%	-50.2%			11.8%	3
Bonardi et al. (2020)*	0.0%			0.0%			1
Guo et al. (2021)			-0.4%	36.3%	-0.2%	5.7%	3
Precision-weighted average	**0.6%**	**-21.2%**	**-10.6%**	**-0.1%**	**-4.4%**	**1.6%**	
Arithmetic average	**0.6%**	**-23.8%**	**-18.6%**	**-0.7%**	**-14.4%**	**3.0%**	
Median	**0.3%**	**-23.8%**	**-14.9%**	**0.0%**	**-0.1%**	**3.2%**	
4 of 4 quality dimensions	**n/a [0]**	**-34.0% [1]**	**-2.9% [2]**	**n/a [0]**	**0.5% [1]**	**0.8% [1]**	
3 of 4 quality dimensions	**0.5% [1]**	**-13.5% [1]**	**-21.5% [3]**	**0.0% [3]**	**-0.1% [2]**	**5.6% [3]**	
2 of 4 quality dimensions or fewer	**1.7% [2]**	**n/a [1]**	**-31.6% [2]**	**-15.6% [2]**	**-58.0% [1]**	**n/a [1]**	

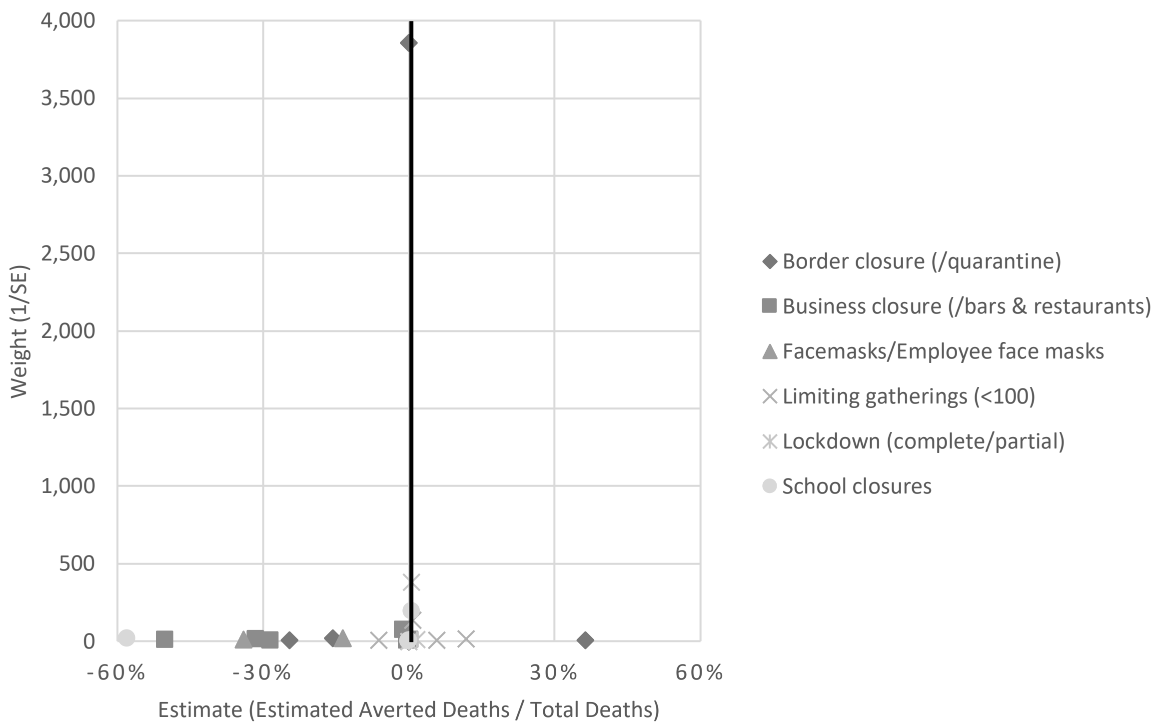
*Note: * It is not possible to derive common estimates and standard errors from Chaudhry et al. (2020) and Bonardi et al. (2020). Chaudhry et al. (2020) states that the effect of the various NPIs is insignificant without listing the estimates and standard errors. Bonardi et al. (2020) states that partial or regional lockdowns are as effective as stricter NPIs but does not provide information to calculate common estimates. Instead, we assume the estimate is 0% when calculating arithmetic average and median, while the estimates are excluded from the calculation of precision-weighted averages because there are no standard errors.*

[45] Bongaerts et al. (2021) (implicitly) assume that municipalities with different exposures to closed sectors are not inherently different, which may be a relatively strong assumption and could potentially drive their results.

[46] We saw with SIPOs that studies based on short data series tended to find larger effects than studies based on short data series. This is also somewhat true for studies examining multiple specific measures. If we focus on studies with long data series (>May 31st, 2020), the precision-weighted estimates are as follows (average for all studies in parentheses for easy comparison): Lockdown (complete/partial): 0.5% (0.6%), Facemasks/Employee face masks: -21.2% (-21.2%), Business closures (/bars & restaurants): -8.1% (-10.6%), Border closures (/quarantine): -0.1% (-0.1%), School closures: 0.5% (-4.4%), Limiting gatherings: 1.4% (1.6%).

Figure 7 shows a funnel plot for all estimates in Table 7, except Chaudhry et al. (2020) and Bonardi et al. (2020), where common standard errors cannot be derived. Two estimates from Toya and Skidmore (2020) stands out with a precision far higher than those of other studies, and estimates are placed with some 'tail' to the left, which could indicate some publication bias, i.e. reluctance to publish results that show large positive (more deaths) effects of lockdowns. The most precise estimates are gathered around 0%, while less precise studies are spread out between -58% and 36%. The precision-weighted average of all estimates across all NPIs is -0.6%.

Figure 7: Funnel plot for estimates from studies of specific NPIs



Note: The figure displays all estimates except two (se text in figure) of specific NPIs and the precision of the estimate defined as one over the standard error. Studies where standard errors are not available are not included.

Overall conclusion on specific NPIs

Because of the heterogeneity in NPIs across studies, it is difficult to draw strong conclusions based on the studies of multiple specific measures. We find no evidence that lockdowns, school closures, border closures, and limiting gatherings have had a noticeable effect on COVID-19 mortality. There is some evidence that business closures reduce COVID-19 mortality, but the variation in estimates is large and the effect seems related to closing bars. There may be an effect of mask mandates, but just two studies look at this, one of which one only looks at the effect of employee mask mandates.

5 Concluding observations

Public health experts and politicians have – based on forecasts in epidemiological studies such as that of Imperial College London (Ferguson et al. (2020) – embraced compulsory lockdowns as an effective method for arresting the pandemic. But, have these lockdown policies been effective in curbing COVID-19 mortality? This is the main question answered by our meta-analysis.

Adopting a systematic search and title-based screening, we identified 1,048 studies published by July 1st, 2020, which potentially look at the effect of lockdowns on mortality rates. To answer our question, we focused on studies that examine the actual impact of lockdowns on COVID-19 mortality rates based on registered cross-sectional mortality data and a counterfactual difference-in-difference approach. Out of the 1,048 studies, 34 met our eligibility criteria.

Conclusions

Overall, our meta-analysis fails to confirm that lockdowns have had a large, significant effect on mortality rates. Studies examining the relationship between lockdown strictness (based on the OxCGRT stringency index) find that the average lockdown in Europe and the United States only reduced COVID-19 mortality by 0.2% compared to a COVID-19 policy based solely on recommendations. Shelter-in-place orders (SIPOs) were also ineffective. They only reduced COVID-19 mortality by 2.9%.

Studies looking at specific NPIs (lockdown vs. no lockdown, facemasks, closing non-essential businesses, border closures, school closures, and limiting gatherings) also find no broad-based evidence of noticeable effects on COVID-19 mortality. However, closing non-essential businesses seems to have had some effect (reducing COVID-19 mortality by 10.6%), which is likely to be related to the closure of bars. Also, masks may reduce COVID-19 mortality, but there is only one study that examines universal mask mandates. The effect of border closures, school closures and limiting gatherings on COVID-19 mortality yields precision-weighted estimates of -0.1%, -4.4%, and 1.6%, respectively. Lockdowns (compared to no lockdowns) also do not reduce COVID-19 mortality.

Discussion

Overall, we conclude that lockdowns are not an effective way of reducing mortality rates during a pandemic, at least not during the first wave of the COVID-19 pandemic. Our results are in line with the World Health Organization Writing Group (2006), who state, "Reports from the 1918 influenza pandemic indicate that social-distancing measures did not stop or appear to dramatically reduce transmission […] In Edmonton, Canada, isolation and quarantine were instituted; public meetings were banned; schools, churches, colleges, theaters, and other public gathering places were closed; and business hours were restricted without obvious impact on the epidemic." Our findings are also in line with Allen's (2021) conclusion: "The most recent research has shown that lockdowns have had, at best, a marginal effect on the number of Covid-19 deaths." Poeschl and Larsen (2021) conclude that "interventions are generally effective in mitigating COVID-19 spread". But, 9 of the 43 (21%) results they review find "no or uncertain association" between lockdowns and the spread of COVID-19, suggesting that evidence from that own study contradicts their conclusion.

The findings contained in Johanna et al. (2020) are in contrast to our own. They conclude that "for lockdown, ten studies consistently showed that it successfully reduced the incidence, onward transmission, and mortality rate of COVID-19." The driver of the difference is three-fold. First, Johanna et al. include modelling studies (10 out of a total of 14 studies), which we have explicitly excluded. Second, they included interrupted time series studies (3 of 14 studies), which we also exclude. Third, the only study using a difference-in-difference approach (as we have done) is based on data collected before May 1st, 2020. We should mention that our results indicate that early studies find relatively larger effects compared to later studies.

Our main conclusion invites a discussion of some issues. Our review does not point out *why* lockdowns did not have the effect promised by the epidemiological models of Imperial College London (Ferguson et al. (2020). We propose four factors that might explain the difference between our conclusion and the view embraced by some epidemiologists.

First, people respond to dangers outside their door. When a pandemic rages, people believe in social distancing regardless of what the government mandates. So, we believe that Allen (2021) is right, when he concludes, "The ineffectiveness [of lockdowns] stemmed from individual changes in behavior: either non-compliance or behavior that mimicked lockdowns." In economic terms, you can say that the demand for costly disease prevention efforts like social distancing and increased focus on hygiene is high when infection rates are high. Contrary, when infection rates are low, the demand is low and it may even be morally and economically rational not to comply with mandates like SIPOs, which are difficult to enforce. Herby (2021) reviews studies which distinguish between mandatory and voluntary behavioral changes. He finds that – on average – voluntary behavioral changes are 10 times as important as mandatory behavioral changes in combating COVID-19. If people voluntarily adjust their behavior to the risk of the pandemic, closing down non-essential businesses may simply reallocate consumer visits away from "nonessential" to "essential" businesses, as shown by Goolsbee and Syverson (2021), with limited impact on the total number of contacts.[47] This may also explain why epidemiological model simulations such as Ferguson et al. (2020) – which do not model behavior endogenously – fail to forecast the effect of lockdowns.

Second, mandates only regulate a fraction of our potential contagious contacts and can hardly regulate nor enforce handwashing, coughing etiquette, distancing in supermarkets, etc. Countries like Denmark, Finland, and Norway that realized success in keeping COVID-19 mortality rates relatively low allowed people to go to work, use public transport, and meet privately at home during the first lockdown. In these countries, there were ample opportunities to legally meet with others.

[47] In economic terms, lockdowns are substitutes for – not complements to – voluntary behavioral changes.

Third, even if lockdowns are successful in initially reducing the spread of COVID-19, the behavioral response may counteract the effect completely, as people respond to the lower risk by changing behavior. As Atkeson (2021) points out, the economic intuition is straightforward. If closing bars and restaurants causes the prevalence of the disease to fall toward zero, the demand for costly disease prevention efforts like social distancing and increased focus on hygiene also falls towards zero, and the disease will return.[48]

Fourth, unintended consequences may play a larger role than recognized. We already pointed to the possible unintended consequence of SIPOs, which may isolate an infected person at home with his/her family where he/she risks infecting family members with a higher viral load, causing more severe illness. But often, lockdowns have limited peoples' access to safe (outdoor) places such as beaches, parks, and zoos, or included outdoor mask mandates or strict outdoor gathering restrictions, pushing people to meet at less safe (indoor) places. Indeed, we do find some evidence that limiting gatherings was counterproductive and increased COVID-19 mortality.

One objection to our conclusions may be that we do not look at the role of timing. If timing is very important, differences in timing may empirically overrule any differences in lockdowns. We note that this objection is not necessarily in contrast to our results. If timing is very important relative to strictness, this suggests that well-timed, but very mild, lockdowns should work as well as, or better than, less well-timed but strict lockdowns. This is not in contrast to our conclusion, as the studies we reviewed analyze the effect of lockdowns compared but to doing very little (see Section 3.1 for further discussion). However, there is little solid evidence supporting the timing thesis, because it is inherently difficult to analyze (see Section 2.2 for further discussion). Also, even if it can be empirically stated that a well-timed lockdown is effective in combating a pandemic, it is doubtful that this information will ever be useful from a policy perspective.

But, what explains the differences between countries, if not differences in lockdown policies? Differences in population age and health, quality of the health sector, and the like are obvious factors. But several studies point at less obvious factors, such as culture, communication, and coincidences. For example, Frey et al. (2020) show that for the same policy stringency, countries with more obedient and collectivist cultural traits experienced larger declines in geographic mobility relative to their more individualistic counterpart. Data from Germany Laliotis and Minos (2020) shows that the spread of COVID-19 and the resulting deaths in predominantly Catholic regions with stronger social and family ties were much higher compared to non-Catholic ones at the local NUTS 3 level.[49]

Government communication may also have played a large role. Compared to its Scandinavian neighbors, the communication from Swedish health authorities was far more subdued and embraced the idea of public health vs. economic trade-offs. This may explain why Helsingen et

[48] This kind of behavior response may also explain why Subramanian and Kumar (2021) find that increases in COVID-19 cases are unrelated to levels of vaccination across 68 countries and 2947 counties in the United States. When people are vaccinated and protected against severe disease, they have less reason to be careful.

[49] The NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing up the economic territory of the EU and the UK. There are 1215 regions at the NUTS 3-level.

al. (2020), found, based on questionnaire data collected from mid-March to mid-April, 2020, that even though the daily COVID-19 mortality rate was more than four times higher in Sweden than in Norway, Swedes were less likely than Norwegians to not meet with friends (55% vs. 87%), avoid public transportation (72% vs. 82%), and stay home during spare time (71% vs. 87%). That is, despite a more severe pandemic, Swedes were less affected in their daily activities (legal in both countries) than Norwegians.

Many other factors may be relevant, and we should not underestimate the importance of coincidences. An interesting example illustrating this point is found in Arnarson (2021) and Björk et al. (2021), who show that areas where the winter holiday was relatively late (in week 9 or 10 rather than week 6, 7 or 8) were hit especially hard by COVID-19 during the first wave because the virus outbreak in the Alps could spread to those areas with ski tourists. Arnarson (2021) shows that the effect persists in later waves. Had the winter holiday in Sweden been in week 7 or week 8 as in Denmark, the Swedish COVID-19 situation could have turned out very differently.[50]

Policy implications

In the early stages of a pandemic, before the arrival of vaccines and new treatments, a society can respond in two ways: mandated behavioral changes or voluntary behavioral changes. Our study fails to demonstrate significant positive effects of mandated behavioral changes (lockdowns). This should draw our focus to the role of voluntary behavioral changes. Here, more research is needed to determine how voluntary behavioral changes can be supported. But it should be clear that one important role for government authorities is to provide information so that citizens can voluntarily respond to the pandemic in a way that mitigates their exposure.

Finally, allow us to broaden our perspective after presenting our meta-analysis that focuses on the following question: "What does the evidence tell us about the effects of lockdowns on mortality?" We provide a firm answer to this question: The evidence fails to confirm that lockdowns have a significant effect in reducing COVID-19 mortality. The effect is little to none.

The use of lockdowns is a unique feature of the COVID-19 pandemic. Lockdowns have not been used to such a large extent during any of the pandemics of the past century. However, lockdowns during the initial phase of the COVID-19 pandemic have had devastating effects. They have contributed to reducing economic activity, raising unemployment, reducing schooling, causing political unrest, contributing to domestic violence, and undermining liberal democracy. These costs to society must be compared to the benefits of lockdowns, which our meta-analysis has shown are marginal at best. Such a standard benefit-cost calculation leads to a strong conclusion: lockdowns should be rejected out of hand as a pandemic policy instrument.

[50] Another case of coincidence is illustrated by Shenoy et al. (2022), who find that areas that experienced rainfall early in the pandemic realized fewer deaths because the rainfall induced social distancing.

6 Appendix A. The role of timing

Some of the included papers study the importance of the timing of lockdowns, while several other papers only looking at timing of (but not on the inherent effect of) lockdowns have been excluded from the literature list in this review. There's no doubt that being prepared for a pandemic and knowing when it arrives at your doorstep is vital. However, two problems arise with respect to imposing early lockdowns.

First of all, it was virtually impossible to determine the right timing when COVID-19 hit Europe and the United States. The World Health Organization declared the outbreak of a pandemic on 11 March 2020, but at that date Italy had already registered 13.7 COVID-19-deaths per million (all infected before approximately 22 February, because of the roughly 18 day gap between infection and death, c.f. e.g.. Bjørnskov (2021a)). On 29 March 2020, 18 days after WHO declared the outbreak a pandemic and the earliest a lockdown response to WHO's announcement could have an effect, the death toll in Italy was a staggering 178 COVID-19-deaths per million with an additionally 13 per million dying each day.

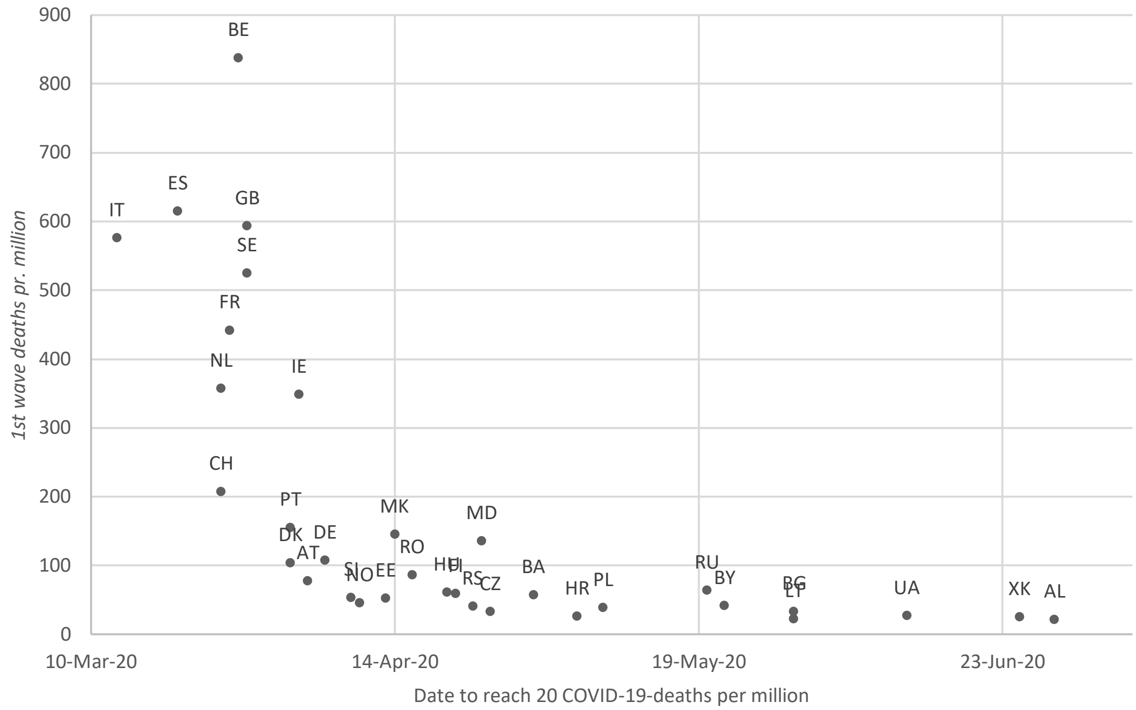
There are reasons to believe that many countries and regions were hit particularly hard during the first wave of COVID, because they had no clue about how bad it really was. This point is illustrated in Figure 8 (and Figure 9), which show that countries (and states), which were hit hard and early, experienced large death tolls compared to countries where the pandemic had a slower start. Björk et al. (2021) and Arnarson (2021) show that areas with a winter holiday in week 10 and – especially – week 9 were hit hard, because they imported cases from the Alps *before* they knew the pandemic was wide spread at the ski resorts. Hence, while acting early by warning citizens and closing business may be an effective strategy; this was not a feasible strategy for most countries in the spring of 2020.

The second problem is that it is extremely difficult to differentiate between the effect of public awareness and the effect of lockdowns. If people and politicians react to the same information, for example deaths in geographical neighboring countries (many EU-countries reacted to deaths in Italy) or in another part of the same country, the effect of lockdowns cannot easily be separated from the effect of voluntary social distancing or, use of hand sanitizers. Hence, we find it problematic to use national lockdowns and differences in the progress of the pandemic in different regions to say anything about the effect of early lockdowns on the pandemic, as the estimated effect might just as well come from voluntary behavior changes, when people in Southern Italy react to the situation in Northern Italy.

We have seen no studies which we believe credibly separate the effect of early lockdown from the effect of early voluntary behavior changes. Instead, the estimates in these studies capture the effects of lockdowns *and* voluntary behavior changes. As Herby (2021) illustrates, voluntary behavior changes are essential to a society's response to an pandemic and can account for up to 90% of societies' total response to the pandemic.

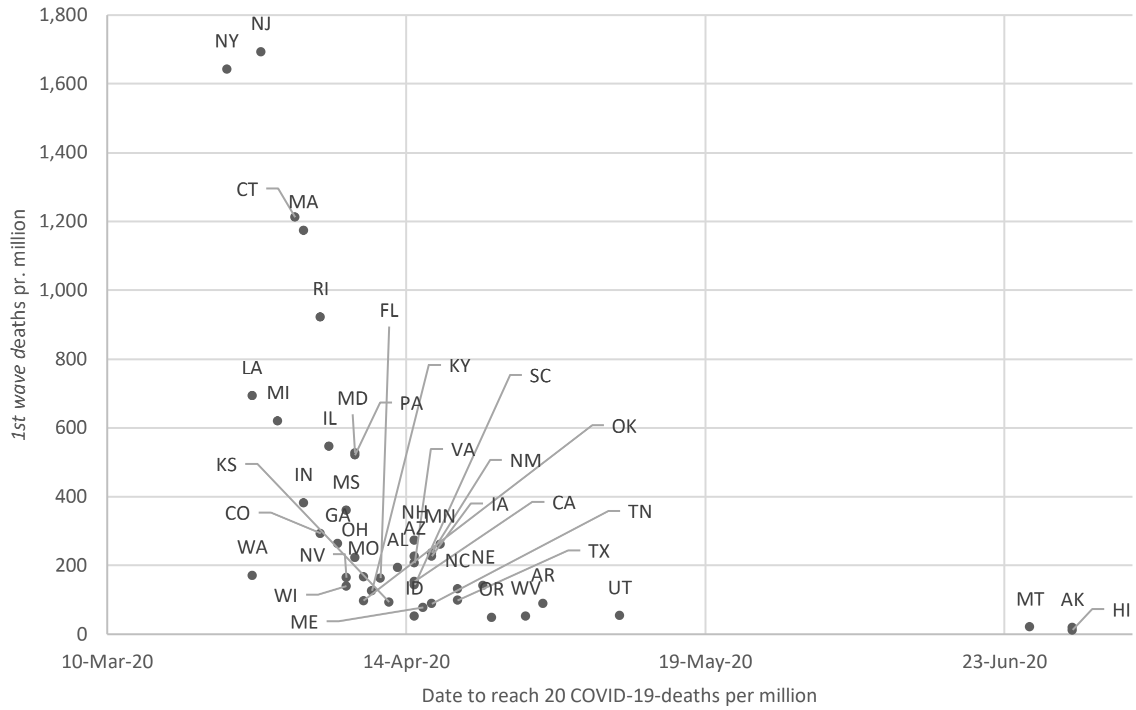
Including these studies will greatly overestimate the effect of lockdowns, and, hence, we chose not to include studies focusing on timing of lockdowns in our review.

Figure 8: Taken by surprise. The importance of having time to prepare in Europe



Description: European countries with more than one million citizens.
Source: Our World in Data

Figure 9: Taken by surprise. The importance of having time to prepare in U.S. states



Description: U.S. states with more than one million citizens.
Source: Our World in Data

7 Appendix B. Supplementary information

7.1 Excluded studies

Below is a list will the studies excluded during the eligibility phase of our identification process and a short description of our basis for excluding the study.

Table 8: Studies excluded during the eligibility phase of our identification process

1. Study (Author & title)	2. Reason for exclusion
Alemán et al. (2020); "Evaluating the effectiveness of policies against a pandemic"	Too few observations
Alshammari et al. (2021); "Are countries' precautionary actions against COVID-19 effective? An assessment study of 175 countries worldwide"	Is purely descriptive
Amuedo-Dorantes et al. (2020); "Timing is Everything when Fighting a Pandemic: COVID-19 Mortality in Spain"	Duplicate
Amuedo-Dorantes et al. (2021); "Early adoption of non-pharmaceutical interventions and COVID-19 mortality"	Only looks at timing
Amuedo-Dorantes, Kaushal and Muchow (2020); "Is the Cure Worse than the Disease? County-Level Evidence from the COVID-19 Pandemic in the United States"	Duplicate
Amuedo-Dorantes, Kaushal and Muchow (2021); "Timing of social distancing policies and COVID-19 mortality: county-level evidence from the U.S."	Only looks at timing
Arruda et al. (2021); "ASSESSING THE IMPACT OF SOCIAL DISTANCING ON COVID-19 CASES AND DEATHS IN BRAZIL: AN INSTRUMENTED DIFFERENCE-IN-	Social distancing (not
Bakolis et al. (2021); "Changes in daily mental health service use and mortality at the commencement and lifting of COVID-19 'lockdown' policy in 10 UK sites: a regression	Uses a time series approach
Bardey, Fernández and Gravel (2021); "Coronavirus and social distancing: do non-pharmaceutical-interventions work (at least) in the short run?"	Only looks at timing
Berardi et. Al. (2020); "The COVID-19 pandemic in Italy: policy and technology impact on health and non-health outcomes"	Too few observations
Bhalla (2020); "Lockdowns and Closures vs COVID–19: COVID Wins"	Uses modelling
Björk et al. (2021); "Impact of winter holiday and government responses on mortality in Europe during the first wave of the COVID-19 pandemic"	Only looks at timing
Bongaerts, Mazzola and Wagner (2020); "Closed for business"	Duplicate
Born, Dietrich and Müller (2021); "The lockdown effect: A counterfactual for Sweden"	Synthetic control study
Born, Dietrich and Müller (2021); "The lockdown effect: A counterfactual for Sweden"	Duplicate
Bushman et al. (2020); "Effectiveness and compliance to social distancing during COVID-19"	Social distancing (not
Castaneda and Saygili (2020); "The effect of shelter-in-place orders on social distancing and the spread of the COVID-19 pandemic: a study of Texas"	Uses a time series approach
Cerqueti et al. (2021); "The sooner the better: lives saved by the lockdown during the COVID-19 outbreak. The case of Italy"	Synthetic control study
Chernozhukov, Kasahara and Schrimpf (2021); "Mask mandates and other lockdown policies reduced the spread of COVID-19 in the U.S."	Duplicate
Chin et al. (2020); "Effects of non-pharmaceutical interventions on COVID-19: A Tale of Three Models"	Uses modelling
Cho (2020); "Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden"	Synthetic control study
Coccia (2020); "The effect of lockdown on public health and economic system: findings from first wave of the COVID-19 pandemic for designing effective strategies to cope	Only looks at timing
Coccia (2021); "Different effects of lockdown on public health and economy of countries: Results from first wave of the COVID-19 pandemic"	Too few observations
Conyon and Thomsen (2021); "COVID-19 in Scandinavia"	Synthetic control study
Conyon et al. (2020); "Lockdowns and COVID-19 deaths in Scandinavia"	Too few observations
Dave et al. (2020); "Did the Wisconsin Supreme Court restart a COVID-19 epidemic? Evidence from a natural experiment"	Synthetic control study
Delis, Iosifidi and Tasiou (2021); "Efficiency of government policy during the COVID-19 pandemic"	Do not look at mortality
Dreher et al. (2021); "Policy interventions, social distancing, and SARS-CoV-2 transmission in the United States: a retrospective state-level analysis"	Do not look at mortality
Duchemin, Veber and Boussau (2020); "Bayesian investigation of SARS-CoV-2-related mortality in France"	Uses modelling
Fair et. Al. (2021); "Estimating COVID-19 cases and deaths prevented by non-pharmaceutical interventions in 2020-2021, and the impact of individual actions: a retrospective	Uses modelling
Filias (2020); "The impact of government policies effectiveness on the officially reported deaths attributed to covid-19."	Student paper
Fowler et al. (2021); "Stay-at-home orders associate with subsequent decreases in COVID-19 cases and fatalities in the United States"	Duplicate
Friedson et al. (2020); "Did California's shelter-in-place order work? Early coronavirus-related public health effects"	Duplicate
Friedson et al. (2020); "Shelter-in-place orders and public health: evidence from California during the COVID-19 pandemic"	Synthetic control study
Fuss, Weizman and Tan (2020); "COVID19 pandemic: how effective are interventive control measures and is a complete lockdown justified? A comparison of countries and	Do not look at mortality
Ghosh, Ghosh and Narymanchi (2020); "A Study on The Effectiveness of Lock-down Measures to Control The Spread of COVID-19"	Synthetic control study
Glogowsky et al. (2021); "How Effective Are Social Distancing Policies? Evidence on the Fight Against COVID-19"	Only looks at timing
Glogowsky, Hansen and Schächtele (2020); "How effective are social distancing policies? Evidence on the fight against COVID-19 from Germany"	Duplicate
Glogowsky, Hansen and Schächtele (2020); "How Effective Are Social Distancing Policies? Evidence on the Fight Against COVID-19 from Germany"	Duplicate
Gordon, Grafton and Steinshamn (2021); "Cross-country effects and policy responses to COVID-19 in 2020: The Nordic countries"	Do not look at mortality
Gordon, Grafton and Steinshamn (2021); "Statistical Analyses of the Public Health and Economic Performance of Nordic Countries in Response to the COVID-19 Pandemic"	Too few observations
Guo et al. (2020); "Social distancing interventions in the United States: An exploratory investigation of determinants and impacts"	Duplicate
Huber and Langen (2020); "The impact of response measures on COVID-19-related hospitalization and death rates in Germany and Switzerland"	Duplicate
Huber and Langen (2020); "Timing matters: the impact of response measures on COVID-19-related hospitalization and death rates in Germany and Switzerland"	Only looks at timing
Jain et al. (2020); "A comparative analysis of COVID-19 mortality rate across the globe: An extensive analysis of the associated factors"	Do not look at mortality
Juranek and Zoutman (2021); "The effect of non-pharmaceutical interventions on the demand for health care and mortality: evidence on COVID-19 in Scandinavia"	Too few observations
Kakpo and Nuhu (2020); "Effects of Social Distancing on COVID-19 Infections and Mortality in the U.S."	Social distancing (not
Kapoor and Ravi (2020); "Impact of national lockdown on COVID-19 deaths in select European countries and the U.S. using a Changes-in-Changes model"	Too few observations
Khatiwada and Chalise (2020); "Evaluating the efficiency of the Swedish government policies to control the spread of Covid-19."	Student paper
Korevaar et al. (2020); "Quantifying the impact of U.S. state non-pharmaceutical interventions on COVID-19 transmission"	Do not look at mortality
Kumar et. Al. (2020); "Prevention-Versus Promotion-Focus Regulatory Efforts on the Disease Incidence and Mortality of COVID-19: A Multinational Diffusion Study Using	Do not look at mortality
Le et al. (2020); "Impact of government-imposed social distancing measures on COVID-19 morbidity and mortality around the world"	Uses a time series approach
Liang et al. (2020); "Covid-19 mortality is negatively associated with test number and government effectiveness"	Not effect of lockdowns
Mader and Rütternauer (2021); "The effects of non-pharmaceutical interventions on COVID-19-related mortality: A generalized synthetic control approach across 169 countries"	Synthetic control study
Matzinger and Skinner (2020); "Strong impact of closing schools, closing bars and wearing masks during the Covid-19 pandemic: results from a simple and revealing analysis"	Uses modelling
Mccafferty and Ashley (2020); "Covid-19 Social Distancing Interventions by State Mandate and their Correlation to Mortality in the United States"	Duplicate
Medline et al. (2020); "Evaluating the impact of stay-at-home orders on the time to reach the peak burden of Covid-19 cases and deaths: does timing matter?"	Only looks at timing

1. Study (Author & title)	2. Reason for exclusion
Mu et al. (2020); "Effect of social distancing interventions on the spread of COVID-19 in the state of Vermont"	Uses modelling
Nakamura (2020); "The Impact of Rapid State Policy Response on Cumulative Deaths Caused by COVID-19"	Student paper
Neidhöfer and Neidhöfer (2020); "The effectiveness of school closures and other pre-lockdown COVID-19 mitigation strategies in Argentina, Italy, and South Korea"	Synthetic control study
Oliveira (2020); "Does' Staying at Home'Save Lives? An Estimation of the Impacts of Social Isolation in the Registered Cases and Deaths by COVID-19 in Brazil"	Social distancing (not
Palladina et al. (2020); "Effect of Implementation of the Lockdown on the Number of COVID-19 Deaths in Four European Countries"	Uses a time series approach
Palladina et al. (2020); "Effect of timing of implementation of the lockdown on the number of deaths for COVID-19 in four European countries"	Duplicate
Palladino et al. (2020); "Excess deaths and hospital admissions for COVID-19 due to a late implementation of the lockdown in Italy"	Uses a time series approach
Peixoto et al. (2020); "Rapid assessment of the impact of lockdown on the COVID-19 epidemic in Portugal"	Uses modelling
Piovani et. Al. (2021); "Effect of early application of social distancing interventions on COVID-19 mortality over the first pandemic wave: An analysis of longitudinal data from 37	Only looks at timing
Reinbold (2021); "Effect of fall 2020 K-12 instruction types on CoViD-19 cases, hospital admissions, and deaths in Illinois counties"	Synthetic control study
Renne, Roussellet and Schwenkler (2020); "Preventing COVID-19 Fatalities: State versus Federal Policies"	Uses modelling
Siedner et al. (2020); "Social distancing to slow the U.S. COVID-19 epidemic: Longitudinal pretest–posttest comparison group study"	Duplicate
Siedner et al. (2020); "Social distancing to slow the U.S. COVID-19 epidemic: Longitudinal pretest–posttest comparison group study"	Uses a time series approach
Silva, Filho and Fernandes (2020); "The effect of lockdown on the COVID-19 epidemic in Brazil: evidence from an interrupted time series design"	Uses a time series approach
Stamam et al. (2020); "IMPACT OF LOCKDOWN MEASURE ON COVID-19 INCIDENCE AND MORTALITY IN THE TOP 31 COUNTRIES OF THE WORLD."	Uses a time series approach
Steinegger et al. (2021); "Retrospective study of the first wave of COVID-19 in Spain: analysis of counterfactual scenarios"	Only looks at timing
Stephens et al. (2020); "Does the timing of government COVID-19 policy interventions matter? Policy analysis of an original database."	Only looks at timing
Supino et al. (2020); "The effects of containment measures in the Italian outbreak of COVID-19"	Uses a time series approach
Timelli and Girardi (2021); "Effect of timing of implementation of containment measures on Covid-19 epidemic. The case of the first wave in Italy"	Only looks at timing
Trivedi and Das (2020); "Effect of the timing of stay-at-home orders on COVID-19 infections in the United States of America"	Only looks at timing
Umer and Khan (2020); "Evaluating the Effectiveness of Regional Lockdown Policies in the Containment of Covid-19: Evidence from Pakistan"	Too few observations
VoPham et al. (2020); "Effect of social distancing on COVID-19 incidence and mortality in the U.S."	Do not look at mortality
Wu and Wu (2020); "Stay-at-home and face mask policies intentions inconsistent with incidence and fatality during U.S. COVID-19 pandemic"	Too few observations
Xu et al. (2020); "Associations of Stay-at-Home Order and Face-Masking Recommendation with Trends in Daily New Cases and Deaths of Laboratory-Confirmed COVID-19 in	Do not look at mortality
Yehya, Venkataramani and Harhay (2020); "Statewide Interventions and Coronavirus Disease 2019 Mortality in the United States: An Observational Study"	Only looks at timing
Ylli et al. (2020); "The lower COVID-19 related mortality and incidence rates in Eastern European countries are associated with delayed start of community circulation Alban	Not effect of lockdowns

7.2 Interpretation of estimates and conversion to common estimates

In Table 9, we describe for each study used in the meta-analysis how we interpret their results and convert the estimates to our common estimate. Standard errors are converted such that the t-value, calculated based on common estimates and standard errors, is unchanged. When confidence intervals are reported rather than standard errors, we calculate standard errors using t-distribution with ∞ degrees of freedom (i.e. 1.96 for 95% confidence interval).

Table 9: Notes on studies included in the meta-analysis

1. Study (Author & title)	2. Date Published	3. Journal	4. Comments regarding meta-analysis
Alderman and Harjoto (2020); "COVID-19: U.S. shelter-in-place orders and demographic characteristics linked to cases, mortality, and recovery rates"	26-Nov-20	Transforming Government: People, Process and Policy	We use the 1% effect noted by the authors in "We find that the natural log of the duration (in days) that the state instituted shelter-in-place reduces percentages of mortality by 0.0001%, or approximately 1% of the means of percentages of deaths per capita in our sample. The standard error is calculated on basis of the t-value in Table 3.
Aparicio and Grossbard (2021); "Are Covid Fatalities in the U.S. Higher than in the EU, and If so, Why?"	16-Jan-21	Review of Economics of the Household	We use estimates from Table 3, model 5. For each estimate the common estimate is calculated as (difference in COVID-19 mortality with NPI)/(difference in COVID-19 mortality without NPI)-1, where (difference in COVID-19 mortality with NPI) is 237.89 (Table 2 states that deaths per million is 406.99 in U.S. and 169.10 in Europe) and (difference in COVID-19 mortality without NPI) is estimated as exp(ln(difference in COVID-19 mortality with NPI)-estimate).
Ashraf (2020); "Socioeconomic conditions, government interventions and health outcomes during COVID-19"	1-Jul-20	ResearchGate	It is unclear whether they prefer the model with or without the interaction term. In the meta-analysis, we use an average of -0.326 (Table 3, without) and -0.073 (Table 6, with) deaths per million per stringency point (i.e. -0.200). The common estimate is the average effect in Europe and United States respectively calculated as (Actual COVID-19 mortality) / (COVID-19 mortality with recommendation policy) -1, where (COVID-19 mortality with recommendation policy) is calculated as ((Actual COVID-19 mortality) - Estimate x Difference in stringency x population). Stringencies in Europe and United States are equal to the average stringency from March 16th to April 15th 2020 (76 and 74 respectively) and the stringency for the policy based solely on recommendations is 44 following Hale et al. (2020).

1. Study (Author & title)	2. Date Published	3. Journal	4. Comments regarding meta-analysis
Auger et al. (2020); "Association between statewide school closure and COVID-19 incidence and mortality in the U.S."	1-Sep-20	JAMA	Estimate that school closure was associated with a 58% decline in COVID-19 mortality and that the effect was largest in states with low cumulative incidence of COVID-19 at the time of school closure. States with the lowest incidence of COVID-19 had a −72% relative change in incidence compared with −49% for those states with the highest cumulative incidence.
Berry et al. (2021); "Evaluating the effects of shelter-in-place policies during the COVID-19 pandemic"	24-Feb-21	PNAS	The estimated effect of SIPO's, an increase in deaths by 0,654 per million after 14 days (significant, cf. Fig. 2), is converted to a relative effect on a state basis based on data from OurWorldInData. For states which did implement SIPO, we calculate the number of deaths without SIPO as the number of official COVID-19 deaths 14 days after SIPO was implemented minus 0,654 extra deaths per million. For states which did not implement SIPO, we calculate the number of deaths with SIPO as the number of official COVID-19 deaths 14 days after March 31 2020 plus 0,654 extra deaths per million. We use March 31 2020 as this was the average date on which SIPO was implemented in the 40 states which did implement SIPO. Using this approximation, the effect of SIPO's in the U.S. is 1,1% more deaths after 14 days. Common standard errors are not available.
Bjørnskov (2021a); "Did Lockdown Work? An Economist's Cross-Country Comparison"	29-Mar-21	CESifo Economic Studies	We use estimates from Table 2 (four weeks). Common estimate is calculated as the average of the effect in Europe and United States, where the effect for each is calculated as (ln(policy stringency) - ln(recommendation stringency)) x estimate.
Blanco et al. (2020); "Do Coronavirus Containment Measures Work? Worldwide Evidence"	1-Dec-20	World Bank Group	The study is not included in the meta-analysis, as it looks at the effect of NPIs on growth rates and does not include an estimate of the effect on total mortality.
Bonardi et al. (2020); "Fast and local: How did lockdown policies affect the spread and severity of the covid-19"	8-Jun-20	0	Find that, world-wide, internal NPIs have prevented about 650,000 deaths (3.11 deaths were prevented for each death that occurred, i.e. 76% effect). However, this effect is for any lockdown including a Swedish lockdown. They do not find an extra effect of stricter lockdowns and state that “our results point to the fact that people might adjust their behaviors quite significantly as partial measures are implemented, which might be enough to stop the spread of the virus.” Hence, whether the baseline is Sweden, which implemented a ban on large gatherings early in the pandemic, or the baseline is “doing nothing” can affect the magnitude of the estimated impacts. Since all Western countries did something and estimates in other reviewed studies are relative to doing less – and, hence not to doing nothing, we report the result from Bonardi et al. as compared to “doing less.” Hence, for Bonardi et al. we use 0% as the common estimate in the meta-analysis for each NPI (SIPO, regional lockdown, partial lockdown, and border closure (stage 1, stage 2 and full) because all NPIs are insignificant (compared to Sweden’s “doing the least”-lockdown).
Bongaerts et al. (2021); "Closed for business: The mortality impact of business closures during the Covid-19 pandemic"	14-May-21	PLOS ONE	Business shutdown saved 9,439 Italian lives by 13th 2020. This corresponds to 32%, as there were 20,465 COVID-19-deaths in Italy by mid April 2020.
Chaudhry et al. (2020); "A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes"	1-Aug-20	EClinacal-Medicine	Finds no effect of partial border closure, complete border closure, partial lockdown (physical distancing measures only), complete lockdown (enhanced containment measures including suspension of all non-essential services), and curfews. In the meta-analysis we use a common estimate of 0%, as estimates and standard errors are not available.
Chernozhukov et al. (2021); "Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S."	1-Jan-21	Journal of Econometrics	The study looks at the effect of NPIs on growth rates but does include an estimate of the effect on total mortality at the end of the study period for employee face masks (-34%), business closure (-29%). and SIPO (-18%), but not for school closures (which we therefore exclude). In reporting the results of their counterfactual, they alter between "fewer deaths with NPI" and "more deaths without NPI.” We have converted the latter to the former as estimate/(1+estimate) so "without business closures deaths would be about 40% higher" corresponds to "with business closures deaths would be about 29% lower.”
Chisadza et al. (2021); "Government Effectiveness and the COVID-19 Pandemic"	10-Mar-21	MDPI	The common estimate is the average effect in Europe and United States respectively calculated as (Actual COVID-19 mortality) / (COVID-19 mortality with recommendation policy) -1, where (COVID-19 mortality with recommendation policy) is calculated as ((Actual COVID-19 mortality) - Estimate x Difference in stringency x population). Stringencies in Europe and United States are equal to the average stringency from March 16th to April 15th 2020 (76 and 74 respectively) and the stringency for the policy based solely on recommendations is 44 following Hale et al. (2020). In the meta-analysis we use the non-linear estimate, but the squared estimate yields similar results.
Dave et al. (2021); "When Do Shelter-in-Place Orders	3-Aug-20	Economic Inpuiry	The study looks at the effect of SIPO's on growth rates but does include an estimate of the effect on total mortality after 20+ days for model 1 and 2 in Table 7. Since model 3, 4 and 5 have estimates

1. Study (Author & title)	2. Date Published	3. Journal	4. Comments regarding meta-analysis
Fight Covid-19 Best? Policy Heterogeneity Across States and Adoption Time"			similar to model 2, we use an average of model 1 to 5, where the estimates of model 3 to 5 are calculated as (common estimate model 2) / (estimate model 2) x estimate model 3/4/5.
Dergiades et al. (2020); "Effectiveness of government policies in response to the COVID-19 outbreak"	28-Aug-20	SSRN	The study is not included in the meta-analysis, as it looks at the effect of NPIs on growth rates and does not include an estimate of the effect on total mortality.
Fakir and Bharati (2021); "Pandemic catch-22: The role of mobility restrictions and institutional inequalities in halting the spread of COVID-19"	28-Jun-21	PLOS ONE	The study is not included in the meta-analysis, as it looks at the effect of NPIs on growth rates and does not include an estimate of the effect on total mortality.
Fowler et al. (2021); "Stay-at-home orders associate with subsequent decreases in COVID-19 cases and fatalities in the United States"	10-Jun-21	PLOS ONE	The study looks at the effect of SIPO's on growth rates but does include an estimate of the effect on total mortality after three weeks (35% reduction in deaths) which is used in the meta-analysis.
Fuller et al. (2021); "Mitigation Policies and COVID-19–Associated Mortality — 37 European Countries, January 23–June 30, 2020"	15-Jan-21	Morbidity and Mortality Weekly Report	For each 1-unit increase in OxCGRT stringency index, the cumulative mortality decreases by 0.55 deaths per 100,000. The common estimate is the average effect in Europe and United States respectively calculated as (Actual COVID-19 mortality) / (COVID-19 mortality with recommendation policy) -1, where (COVID-19 mortality with recommendation policy) is calculated as ((Actual COVID-19 mortality) - Estimate x Difference in stringency x population). Stringencies in Europe and United States are equal to the average stringency from March 16th to April 15th 2020 (76 and 74 respectively) and the stringency for the policy based solely on recommendations is 44 following Hale et al. (2020).
Gibson (2020); "Government mandated lockdowns do not reduce Covid-19 deaths: implications for evaluating the stringent New Zealand response"	18-Aug-20	New Zealand Economic Papers	We use the two graphs to the left in figure 3, where we extract the data from the rightmost datapoint (I.e. % impact of county lockdowns on Covid-19 deaths by 1/06/2020). We then take the average of the estimates found in the two graphs, because it is unclear which estimate the author prefers.
Goldstein et al. (2021); "Lockdown Fatigue: The Diminishing Effects of Quarantines on the Spread of COVID-19 "	4-Feb-21	CID Faculty Working	We convert the effect in Figure 4 after 90 days (log difference -1.16 of a standard deviation change) to deaths per million per stringency following footnote 3 (the footnote says "weekly deaths," but we believe this should be "daily deaths"), so the effect is e^-1.16 − 1 = −0.69 decline in daily deaths per million per SD. We convert to total effect by multiplying with 90 days and "per point" by dividing with SD = 22.3 (corresponding to the SD for the 147 countries with data before March 19, 2020 - using all data yields similar results) yielding -2.77 deaths per million per stringency point. The common estimate is the average effect in Europe and United States respectively calculated as (Actual COVID-19 mortality) / (COVID-19 mortality with recommendation policy) -1, where (COVID-19 mortality with recommendation policy) is calculated as ((Actual COVID-19 mortality) - Estimate x Difference in stringency x population). Stringencies in Europe and United States are equal to the average stringency from March 16th to April 15th 2020 (76 and 74 respectively) and the stringency for the policy based solely on recommendations is 44 following Hale et al. (2020).
Guo et al. (2021); "Mitigation Interventions in the United States: An Exploratory Investigation of Determinants and Impacts"	21-Sep-20	Research on Social Work Practice	We use estimates for "Proportion of Cumulative Deaths Over the Population" (per 10,000) in Table 3. We interpret this number as the change in cumulative deaths over the population in percent and is therefore the same as our common estimate.
Hale et al. (2020); "Global assessment of the relationship between government response measures and COVID-19 deaths"	6-Jul-20	medRxiv	The study is not included in the meta-analysis, as it looks at the effect of NPIs on growth rates and does not include an estimate of the effect on total mortality. They ascertain that "sustained over three months, this would correspond to a cumulative number of deaths 30% lower," however this is not a counterfactual estimate and three months goes beyond the period they have data for.
Hunter et al. (2021); "Impact of non-pharmaceutical interventions against COVID-19 in Europe: A quasi-experimental non-equivalent group and time-series"	15-Jul-21	Eurosurveillance	The study is not included in the meta-analysis, as they report the effect of NPIs in incident risk ratio which are not easily converted to relative effects.

1. Study (Author & title)	2. Date Published	3. Journal	4. Comments regarding meta-analysis
Langeland et al. (2021); "The Effect of State Level COVID-19 Stay-at-Home Orders on Death Rates"	5-Mar-21	Culture & Crisis Conference	The study is not included in the meta-analysis, as it looks at the effect of NPIs on odds-ratios and does not include an estimate of the effect on total mortality.
Leffler et al. (2020); "Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks"	26-Oct-20	ASTMH	Their "mask recommendation" includes some countries, where masks were mandated and may (partially) capture the effect of mask mandates. However, the authors' focus is on recommendation, so we do interpret their result as a voluntary effect - not an effect of mask mandate. Using estimates from Table 2 and assuming NPIs were implemented March 15 (8 weeks in total by end of study period), common estimates are calculated as 8^est-1.
Mccafferty and Ashley (2021); "Covid-19 Social Distancing Interventions by Statutory Mandate and Their Observational Correlation to Mortality in the United States and Europe"	27-Apr-21	Pragmatic and Observational Research	The study is not included in the meta-analysis, as it looks at the effect of NPIs on peak mortality and does not include an estimate of the effect on total mortality.
Pan et al. (2020); "Covid-19: Effectiveness of non-pharmaceutical interventions in the united states before phased removal of social distancing protections varies by region"	20-Aug-20	medRxiv	The study is not included in the meta-analysis, as the cluster the NPIs (e.g. SIPO, mask mandata amd travel restricions are clustered in Level 4).
Pincombe et al. (2021); "The effectiveness of national-level containment and closure policies across income levels during the COVID-19 pandemic: an analysis of 113 countries"	4-May-21	Health Policy and Planning	Policy implementations were assigned according to the first day that a country received a policy stringency rating above 0 in the OxCGRT stay-at-home measure. As the value 1 is a recommendation "recommend not leaving house," we cannot distinguish recommendations from mandates, and, thus, the study is not included in the meta-analysis.
Sears et al. (2020); "Are we #stayinghome to Flatten the Curve?"	6-Aug-20	medRxiv	Find that SIPOs lower mortality by 29-35%. We use the average (32%) as our common estimate. Common standard errors are calculated based on estimates and standard errors from (Table 4) assuming they are linearly related to estimates.
Shiva and Molana (2021); "The Luxury of Lockdown"	9-Apr-21	The European Journal of Developmen t Research	The estimate with 8 weeks lag is insignificant, and preferable given our empirical strategy. However, they use the 4-week lag when elaborating the model to differentiate between high- and low-income countries, so the 4-week lag estimate for rich countries is used in our meta-analysis. Common estimate is calculated as the average of the effect in Europe and United States, where the effect for each is calculated as (policy stringency - recommendation stringency) x estimate.
Spiegel and Tookes (2021); "Business restrictions and Covid-19 fatalities"	18-Jun-21	The Review of Financial Studies	We use weighted average of estimates for Table 4, 6, and 9. Since authors state that they place more weight on the findings in Table 9, Table 9 weights by 50% while Table 4 and 6 weights by 25%. We estimate the effect on total mortality from effect on growth rates based on authors calculation showing that estimates of -0.049 and -0.060 reduces new deaths by 12.5% 15.3% respectively. We use the same relative factor on other estimates.
Stockenhuber (2020); "Did We Respond Quickly Enough? How Policy-Implementation Speed in Response to COVID-19 Affects the Number of Fatal Cases in Europe"	10-Nov-20	World Medical & Health Policy	When calculating arithmetic average / median, the study is included as 0%, because estimates in Table 6 are insignificant and signs of estimates are mixed (higher strictness can cause both fewer and more deaths). We don't calculate common standard errors.
Stokes et al. (2020); "The relative effects of non-pharmaceutical interventions on early Covid-19 mortality: natural experiment in 130 countries"	6-Oct-20	medRxiv	We use estimates from regression on strictness alone (Right panel in Table "Regression results, policy strictness. Baseline is "policy not introduced within policy analysis period" in "Additional file"). We use the average of 24 and 38 days from model 5. There are 23 relevant estimates in total (they analyze all levels within the eight NPI measures in the OxCGRT stringency index). We calculate the effect of each NPI (e.g. closing schools) as the average effect in all of U.S./Europe. This is done by calculating the effect for each state/country based on the maximum level for each measure between Mar 16 and Apr 15 (e.g. if all schools in a state/country are required to close (school closing level 3) the relevant estimate for that state/level is -0.031 (average of -0.464 and 0.402). We assume all NPIs are effective for 54 days (from March 15 to June 1 minus 24 days to reach full effect). Standard errors are converted to common standard errors following the same process (this approach is unique for Stokes, as our general approach is not possible).

1. Study (Author & title)	2. Date Published	3. Journal	4. Comments regarding meta-analysis
Toya and Skidmore (2020); "A Cross-Country Analysis of the Determinants of Covid-19 Fatalities"	1-Apr-20	CESifo Working Papers	It is unclear how they define "lockdown." They write that "many countries [...] imposed lockdowns of varying degrees, some imposing mandatory nationwide lockdowns, restricting economic and social activity deemed to be non-essential," and since all European countries and all states in the U.S. imposed restrictions on economic (closing unessential businesses) and/or social (limiting large gatherings) activity, we interpret this as all European countries and all U.S. states had mandatory nationwide lockdowns. The effect of recommended lockdowns is set to zero in the meta-analysis, as only one country was in this lockdown category (i.e. too few observations, cf. eligibility criteria). The estimate for complete travel closure is -0.226 COVID-deaths per 100,000. Hence, if all of Europe imposed complete travel closure, the total effect would be -0.266 * 748 million (population) * 10 (100,000/1,000,000) equal to 1,690 averted COVID-19 deaths. However, according to OxCGRT-data European countries only had complete travel bans (Level 4: "Ban on all regions or total border closure") in 11% of the time between March 16 and April 15, 2020. So the total effect is 1,690 * 11% = 194 averted deaths. During the first wave 188,000 deaths in Europe was related to COVID-19 (by June 30, 2020), so the total effect is approximated to -0.1% in Europe and, following the same logic, 0% in U.S., where no states closed their borders completely. We use the average, -0.05%, in the meta-analysis. The estimate for mandatory national lockdown is 0.166 (>0) COVID-deaths per 100,000. Since all European countries (and U.S. states) imposed lockdowns, the total effect is 1,241 (553) extra COVID-19 deaths corresponding to 0.7% (0.4%). We use the average of Europe and the U.S., 0.5%, in the meta-analysis. Calculations of the effect of "Mandatory national lockdown" follow the same logic, but we assume 100% of Europe and United States have had "Mandatory national lockdown."
Tsai et al. (2021); "Coronavirus Disease 2019 (COVID-19) Transmission in the United States Before Versus After Relaxation of Statewide Social Distancing Measures"	3-Oct-20	Oxford academic	The study is not included in the meta-analysis, as they report the effect of NPIs on Rt which are not easily converted to relative effects.

8 References

Abadie, Alberto. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature* 59 (2):391–425. https://doi.org/10.1257/jel.20191450.

Alderman, Jillian, and Maretno Harjoto. 2020. "COVID-19: US Shelter-in-Place Orders and Demographic Characteristics Linked to Cases, Mortality, and Recovery Rates." *Transforming Government: People, Process and Policy* ahead-of-print (ahead-of-print). https://doi.org/10.1108/TG-06-2020-0130.

Alemán, Christian, Christopher Busch, Alexander Ludwig, and Raül Santàeulalia-Llopis. 2020. "Evaluating the Effectiveness of Policies Against a Pandemic." ZEW - Centre for European Economic Research Discussion Paper.

Allen, Douglas W. 2021. "Covid-19 Lockdown Cost/Benefits: A Critical Assessment of the Literature." *International Journal of the Economics of Business*, September, 1–32. https://doi.org/10.1080/13571516.2021.1976051.

An, Brian Y., Simon Porcher, Shui-Yan Tang, and Eunji Emily Kim. 2021. "Policy Design for COVID -19: Worldwide Evidence on the Efficacies of Early Mask Mandates and Other Policy Interventions." *Public Administration Review* 81 (6):1157–82. https://doi.org/10.1111/puar.13426.

Aparicio, Ainoa, and Shoshana Grossbard. 2021. "Are COVID Fatalities in the US Higher than in the EU, and If so, Why?" *Review of Economics of the Household* 19 (2):307–26. https://doi.org/10.1007/s11150-020-09532-9.

Arnarson, Björn Thor. 2021. "Breaks and Breakouts: Explaining the Persistence of COVID-19." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3775506.

Ashraf, Badar Nadeem. 2020. "Socioeconomic Conditions, Government Interventions and Health Outcomes during COVID-19," July. http://dx.doi.org/10.13140/RG.2.2.21141.55520.

Atkeson, Andrew. 2021. "A Parsimonious Behavioral SEIR Model of the 2020 COVID Epidemic in the United States and the United Kingdom," February, w28434. https://doi.org/10.3386/w28434.

Atkeson, Andrew, Karen Kopecky, and Tao Zha. 2020. "Four Stylized Facts about COVID-19." *NBER Working Paper*, August, 44. https://doi.org/10.3386/w27719.

Auger, Katherine A., Samir S. Shah, Troy Richardson, David Hartley, Matthew Hall, Amanda Warniment, Kristen Timmons, et al. 2020. "Association Between Statewide School Closure and COVID-19 Incidence and Mortality in the US." *JAMA* 324 (9):1–13. https://doi.org/10.1001/jama.2020.14348.

Bakolis, Ioannis, Robert Stewart, David Baldwin, Jane Beenstock, Paul Bibby, Matthew Broadbent, Rudolf Cardinal, et al. 2021. "Changes in Daily Mental Health Service Use and Mortality at the Commencement and Lifting of COVID-19 'Lockdown' Policy in 10 UK Sites: A Regression Discontinuity in Time Design." *BMJ Open* 11 (5). British Medical Journal Publishing Group:e049721. https://doi.org/10.1136/bmjopen-2021-049721.

Berardi, Chiara, Marcello Antonini, Mesfin G. Genie, Giovanni Cotugno, Alessandro Lanteri, Adrian Melia, and Francesco Paolucci. 2020. "The COVID-19 Pandemic in Italy: Policy and Technology Impact on Health and Non-Health Outcomes." *Health Policy and Technology* 9 (4):454–87. https://doi.org/10.1016/j.hlpt.2020.08.019.

Berry, Christopher R., Anthony Fowler, Tamara Glazer, Samantha Handel-Meyer, and Alec MacMillen. 2021. "Evaluating the Effects of Shelter-in-Place Policies during the COVID-19 Pandemic." *Proceedings of the National Academy of Sciences* 118 (15):e2019706118. https://doi.org/10.1073/pnas.2019706118.

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*, February, ckab017. https://doi.org/10.1093/eurpub/ckab017.

Bjørnskov, Christian. 2021a. "Did Lockdown Work? An Economist's Cross-Country Comparison." *CESifo Economic Studies* 00 (00):14. https://doi.org/10.1093/cesifo/ifab003.

———. 2021b. "Born et al. Om Epidemien i Sverige – Hvad Er Der Galt Og Hvordan Ser Det Ud Nu?" *Punditokraterne* (blog). June 14, 2021. http://punditokraterne.dk/2021/06/14/born-et-al-om-epidemien-i-sverige-hvad-er-der-galt-og-hvordan-ser-det-ud-nu/.

Blanco, Fernando, Drilona Emrullahu, and Raimundo Soto. 2020. "Do Coronavirus Containment Measures Work? Worldwide Evidence," Policy Research Working Papers, , December. https://doi.org/10.1596/1813-9450-9490.

Bonardi, Jean-Philippe, Quentin Gallea, Dimtrija Kalanoski, and Rafael Lalive. 2020. "Fast and Local: How Lockdown Policies Affect the Spread and Severity of Covid-19." *CEPR Covid Economics*, 27.

Bongaerts, Dion, Francesco Mazzola, and Wolf Wagner. 2021. "Closed for Business: The Mortality Impact of Business Closures during the Covid-19 Pandemic." *PLOS ONE* 16 (5). Public Library of Science:e0251373. https://doi.org/10.1371/journal.pone.0251373.

Book, Joakim. 2020. "Oxford's Stringency Index Is Falling Apart – AIER." December 24, 2020. https://www.aier.org/article/oxfords-stringency-index-is-falling-apart/.

Born, Benjamin, Alexander M. Dietrich, and Gernot J. Müller. 2021. "The Lockdown Effect: A Counterfactual for Sweden." *PLOS ONE* 16 (4). Public Library of Science:e0249732. https://doi.org/10.1371/journal.pone.0249732.

Brodeur, Abel, David Gray, Anik Islam, and Suraiya Bhuiyan. 2021. "A Literature Review of the Economics of COVID-19." *Journal of Economic Surveys*, April, joes.12423. https://doi.org/10.1111/joes.12423.

Cerqueti, Roy, Raffaella Coppier, Alessandro Girardi, and Marco Ventura. 2021. "The Sooner the Better: Lives Saved by the Lockdown during the COVID-19 Outbreak. The Case of Italy." *ArXiv:2101.11901 [Econ]*, January. http://arxiv.org/abs/2101.11901.

Chaudhry, Rabail, George Dranitsaris, Talha Mubashir, Justyna Bartoszko, and Sheila Riazi. 2020. "A Country Level Analysis Measuring the Impact of Government Actions, Country Preparedness and Socioeconomic Factors on COVID-19 Mortality and Related Health Outcomes." *EClinicalMedicine* 25 (August):100464. https://doi.org/10.1016/j.eclinm.2020.100464.

Chernozhukov, Victor, Hiroyuki Kasahara, and Paul Schrimpf. 2021. "Causal Impact of Masks, Policies, Behavior on Early Covid-19 Pandemic in the U.S." *Journal of Econometrics*, Pandemic Econometrics, 220 (1):23–62. https://doi.org/10.1016/j.jeconom.2020.09.003.

Chisadza, Carolyn, Matthew Clance, and Rangan Gupta. 2021. "Government Effectiveness and the COVID-19 Pandemic." *Sustainability* 13 (6). Multidisciplinary Digital Publishing Institute:3042. https://doi.org/10.3390/su13063042.

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):323–44. https://doi.org/10.1093/ectj/utaa025.

Coccia, Mario. 2021. “Different Effects of Lockdown on Public Health and Economy of Countries: Results from First Wave of the COVID-19 Pandemic.” *Journal of Economics Library* 8 (1):45–63. https://doi.org/10.1453/jel.v8i1.2183.

Conyon, Martin J., Lerong He, and Steen Thomsen. 2020a. “Lockdowns and COVID-19 Deaths in Scandinavia.” *CEPR Covid Economics*. https://doi.org/10.2139/ssrn.3616969.

———. 2020b. “Lockdowns and COVID-19 Deaths in Scandinavia.” *SSRN Electronic Journal*, June. https://doi.org/10.2139/ssrn.3616969.

Conyon, Martin J., and Steen Thomsen. 2021. “COVID-19 in Scandinavia.” https://doi.org/10.2139/ssrn.3793888.

Dave, Dhaval, Andrew I. Friedson, Kyutaro Matsuzawa, and Joseph J. Sabia. 2021. “When Do Shelter-in-Place Orders Fight Covid-19 Best? Policy Heterogeneity Across States and Adoption Time.” *Economic Inquiry* 59 (1):29–52. https://doi.org/10.1111/ecin.12944.

Dave, Dhaval, Andrew Friedson, Kyutaro Matsuzawa, Drew McNichols, and Joseph Sabia. 2020. “Did the Wisconsin Supreme Court Restart a Covid-19 Epidemic? Evidence from a Natural Experiment.” *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3620628.

Dergiades, Theologos, Costas Milas, Theodore Panagiotidis, and Elias Mossialos. 2020. “Effectiveness of Government Policies in Response to the COVID-19 Outbreak.” *SSRN Electronic Journal*, August. https://doi.org/10.2139/ssrn.3602004.

Doucouliagos, Hristos, and Martin Paldam. 2008. “Aid Effectiveness on Growth: A Meta Study.” *European Journal of Political Economy* 24 (1):1–24. https://doi.org/10.1016/j.ejpoleco.2007.06.002.

Duchemin, Louis, Philippe Veber, and Bastien Boussau. 2020. “Bayesian Investigation of SARS-CoV-2-Related Mortality in France,” June. https://doi.org/10.1101/2020.06.09.20126862.

ECDC. 2020. “COVID-19 in Children and the Role of School Settings in Transmission - First Update.” https://www.ecdc.europa.eu/en/publications-data/children-and-school-settings-covid-19-transmission.

Fakir, Adnan M. S., and Tushar Bharati. 2021. “Pandemic Catch-22: The Role of Mobility Restrictions and Institutional Inequalities in Halting the Spread of COVID-19.” *PLOS ONE* 16 (6). Public Library of Science:e0253348. https://doi.org/10.1371/journal.pone.0253348.

Ferguson, Neil M, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, et al. 2020. “Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID- 19 Mortality and Healthcare Demand,” March, 20.

Flaxman, Seth, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, et al. 2020. “Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe.” *Nature* 584 (7820):257–61. https://doi.org/10.1038/s41586-020-2405-7.

Fowler, James H., Seth J. Hill, Remy Levin, and Nick Obradovich. 2021. “Stay-at-Home Orders Associate with Subsequent Decreases in COVID-19 Cases and Fatalities in the United States.” *PLOS ONE* 16 (6). Public Library of Science:e0248849. https://doi.org/10.1371/journal.pone.0248849.

Frey, Carl Benedikt, Chinchih Chen, and Giorgio Presidente. 2020. “Democracy, Culture, and Contagion: Political Regimes and Countries’ Responsiveness to Covid-19.” *CEPR Covid Economics*. https://cepr.org/sites/default/files/CovidEconomics18.pdf.

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):258–83. https://doi.org/10.1002/pam.22267.

Fuller, James A., Avi Hakim, Kerton R. Victory, Kashmira Date, Michael Lynch, Benjamin Dahl, and Olga Henao. 2021. “Mitigation Policies and COVID-19–Associated Mortality — 37 European Countries, January 23–June 30, 2020.” *Morbidity and Mortality Weekly Report* 70 (2):58–62. https://doi.org/10.15585/mmwr.mm7002e4.

Ghosh, Subhas Kumar, Sachchit Ghosh, and Sai Shanmukha Narumanchi. 2020. “A Study on The Effectiveness of Lock-down Measures to Control The Spread of COVID-19.” *ArXiv:2008.05876 [Physics]*, August. http://arxiv.org/abs/2008.05876.

Gibson, John. 2020. “Government Mandated Lockdowns Do Not Reduce Covid-19 Deaths: Implications for Evaluating the Stringent New Zealand Response.” *New Zealand Economic Papers*, November, 1–12. https://doi.org/10.1080/00779954.2020.1844786.

Goldstein, Patricio, Eduardo Levy Yeyati, and Luca Sartorio. 2021. “Lockdown Fatigue: The Diminishing Effects of Quarantines on the Spread of COVID-19,” June. https://doi.org/10.21203/rs.3.rs-621368/v1.

Goolsbee, Austan, and Chad Syverson. 2021. “Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020.” *Journal of Public Economics* 193 (January):104311. https://doi.org/10.1016/j.jpubeco.2020.104311.

Gordon, Daniel V., R. Quentin Grafton, and Stein Ivar Steinshamn. 2020. “Statistical Analyses of the Public Health and Economic Performance of Nordic Countries in Response to the COVID-19 Pandemic,” November, 2020.11.23.20236711. https://doi.org/10.1101/2020.11.23.20236711.

GRADEpro. 2013. “GRADE Handbook.” October 2013. https://gdt.gradepro.org/app/handbook/handbook.html.

Guallar, María Pilar, Rosa Meiriño, Carolina Donat-Vargas, Octavio Corral, Nicolás Jouvé, and Vicente Soriano. 2020. “Inoculum at the Time of SARS-CoV-2 Exposure and Risk of Disease Severity.” *International Journal of Infectious Diseases* 97 (August):290–92. https://doi.org/10.1016/j.ijid.2020.06.035.

Guo, Shenyang, Ruopeng An, Timothy D. McBride, Danlin Yu, Linyun Fu, and Yuanyuan Yang. 2021. “Mitigation Interventions in the United States: An Exploratory Investigation of Determinants and Impacts.” *Research on Social Work Practice* 31 (1):26–41. https://doi.org/10.1177/1049731520957415.

Gupta, Sumedha, Kosali Simon, and Coady Wing. 2020. “Mandated and Voluntary Social Distancing During The COVID-19 Epidemic: A Review.” *NBER Working Paper Series*, June, w28139. https://doi.org/10.3386/w28139.

Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, et al. 2021. “Variation in Government Responses to COVID-19.” *Nature Human Behaviour* 5 (4):529–38. https://doi.org/10.1038/s41562-021-01079-8.

Hale, Thomas, Andrew J. Hale, Beatriz Kira, Anna Petherick, Toby Phillips, Devi Sridhar, Robin N. Thompson, Samuel Webster, and Noam Angrist. 2020. “Global Assessment of the

Relationship between Government Response Measures and COVID-19 Deaths," July, 2020.07.04.20145334. https://doi.org/10.1101/2020.07.04.20145334.
Helsingen, Lise M., Erle Refsum, Dagrun Kyte Gjøstein, Magnus Løberg, Michael Bretthauer, Mette Kalager, Louise Emilsson, and for the Clinical Effectiveness Research group. 2020. "The COVID-19 Pandemic in Norway and Sweden – Threats, Trust, and Impact on Daily Life: A Comparative Survey." *BMC Public Health* 20 (1):1597. https://doi.org/10.1186/s12889-020-09615-3.
Herby, Jonas. 2021. "A First Literature Review: Lockdowns Only Had a Small Effect on COVID-19." *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.3764553.
Herby, Jonas, Lars Jonung, and Steve H. Hanke. 2021. "Protocol for 'What Does the First XX Studies Tell Us about the Effects of Lockdowns on Mortality? A Systematic Review and Meta-Analysis of COVID-19 Lockdowns.'" *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3872977.
Homburg, Stefan, and Christof Kuhbandner. 2020. "Comment on Flaxman et al. (2020, Nature: The Illusory Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe." *Nature* 584 (7820):257–61.
Hunter, Paul R, Felipe J Colón-González, Julii Brainard, and Steven Rushton. 2021. "Impact of Non-Pharmaceutical Interventions against COVID-19 in Europe in 2020: A Quasi-Experimental Non-Equivalent Group and Time Series Design Study." *Eurosurveillance* 26 (28). https://doi.org/10.2807/1560-7917.ES.2021.26.28.2001401.
Irfan, Omar, Jiang Li, Kun Tang, Zhicheng Wang, and Zulfiqar A Bhutta. 2021. "Risk of Infection and Transmission of SARS-CoV-2 among Children and Adolescents in Households, Communities and Educational Settings: A Systematic Review and Meta-Analysis." *Journal of Global Health* 11 (July):05013. https://doi.org/10.7189/jogh.11.05013.
Jefferson, Tom, Chris B Del Mar, Liz Dooley, Eliana Ferroni, Lubna A Al-Ansary, Ghada A Bawazeer, Mieke L van Driel, et al. 2020. "Physical Interventions to Interrupt or Reduce the Spread of Respiratory Viruses." Edited by Cochrane Acute Respiratory Infections Group. *Cochrane Database of Systematic Reviews* 2020 (11). https://doi.org/10.1002/14651858.CD006207.pub5.
Johanna, Nadya, Henrico Citrawijaya, and Grace Wangge. 2020. "Mass Screening vs Lockdown vs Combination of Both to Control COVID-19: A Systematic Review." *Journal of Public Health Research*, 9. https://dx.doi.org/10.4081%2Fjphr.2020.2011.
Jonung, Lars, and Steve H. Hanke. 2020. "Freedom and Sweden's Constitution." *Wall Street Journal*, May 20, 2020. https://www.wsj.com/articles/freedom-and-swedens-constitution-11589993183.
Jonung, Lars, and Werner Röger. 2006. "The Macroeconomic Effects of a Pandemic in Europe. A Model-Based Assessment", *European Economy*, Economic papers, nr 251, juni, 2006. European Commission. Brussels. https://ec.europa.eu/economy_finance/publications/pages/publication708_en.pdf
Juranek, Steffen, and Floris T. Zoutman. 2021. "The Effect of Non-Pharmaceutical Interventions on the Demand for Health Care and on Mortality: Evidence from COVID-19 in Scandinavia." *Journal of Population Economics* 34 (4):1299–1320. https://doi.org/10.1007/s00148-021-00868-9.

Kapoor, Mudit, and Shamika Ravi. 2020. "Impact of National Lockdown on COVID-19 Deaths in Select European Countries and the US Using a Changes-in-Changes Model." *ArXiv:2006.12251 [Physics, q-Bio, q-Fin]*, June. http://arxiv.org/abs/2006.12251.

Kepp, Kasper Planeta, and Christian Bjørnskov. 2021. "Lockdown Effects on Sars-CoV-2 Transmission – The Evidence from Northern Jutland." *MedRxiv*, January. https://doi.org/10.1101/2020.12.28.20248936.

Laliotis, Ioannis, and Dimitrios Minos. 2020. "Spreading the Disease: The Role of Culture." *CEPR Covid Economics*, June. https://doi.org/10.31235/osf.io/z4ndc.

Langeland, Andy, Jose Marte, and Kyle Connif. 2021. "The Effect of State Level COVID-19 Stay-at-Home Orders on Death Rates," March, 23.

Leffler, Christopher T., Edsel Ing, Joseph D. Lykins, Matthew C. Hogan, Craig A. McKeown, and Andrzej Grzybowski. 2020. "Association of Country-Wide Coronavirus Mortality with Demographics, Testing, Lockdowns, and Public Wearing of Masks." *The American Journal of Tropical Medicine and Hygiene* 103 (6):2400–2411. https://doi.org/10.4269/ajtmh.20-1015.

Lemoine, Philippe. 2020. "Lockdowns, Science and Voodoo Magic." Nec Pluribus Impar. December 4, 2020. https://necpluribusimpar.net/lockdowns-science-and-voodoo-magic/.

Lewis, Nic. 2020. "Did Lockdowns Really Save 3 Million COVID-19 Deaths, as Flaxman et al. Claim?" Climate Etc. June 21, 2020. https://judithcurry.com/2020/06/21/did-lockdowns-really-save-3-million-covid-19-deaths-as-flaxman-et-al-claim/.

Li, Yanni, Mingming Liang, Liang Gao, Mubashir Ayaz Ahmed, John Patrick Uy, Ce Cheng, Qin Zhou, and Chenyu Sun. 2021. "Face Masks to Prevent Transmission of COVID-19: A Systematic Review and Meta-Analysis." *American Journal of Infection Control* 49 (7):900–906. https://doi.org/10.1016/j.ajic.2020.12.007.

Lipp, Allyson, and Peggy Edwards. 2014. "Disposable Surgical Face Masks for Preventing Surgical Wound Infection in Clean Surgery." In *Cochrane Database of Systematic Reviews*, edited by The Cochrane Collaboration, CD002929.pub2. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/14651858.CD002929.pub2.

Liu, Ian T., Vinay Prasad, and Jonathan J. Darrow. 2021. "Evidence for Community Cloth Face Masking to Limit the Spread of SARS-CoV-2: A Critical Review," November. https://www.cato.org/working-paper/evidence-community-cloth-face-masking-limit-spread-sars-cov-2-critical-review.

Mader, Sebastian, and Tobias Rüttenauer. 2021. "The Effects of Non-Pharmaceutical Interventions on COVID-19-Related Mortality: A Generalized Synthetic Control Approach across 169 Countries." SocArXiv. https://doi.org/10.31235/osf.io/v2ef8.

Matzinger, Polly, and Jeff Skinner. 2020. "Strong Impact of Closing Schools, Closing Bars and Wearing Masks during the COVID-19 Pandemic: Results from a Simple and Revealing Analysis." https://doi.org/10.1101/2020.09.26.20202457.

Mccafferty, Sean, and Sean Ashley. 2021. "Covid-19 Social Distancing Interventions by Statutory Mandate and Their Observational Correlation to Mortality in the United States and Europe." *Pragmatic and Observational Research* 12 (April). Dove Press:15–24. https://doi.org/10.2147/POR.S298309.

Moher, D., A. Liberati, J. Tetzlaff, D. G Altman, and for the PRISMA Group. 2009. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *BMJ* 339 (jul21 1):b2535–b2535. https://doi.org/10.1136/bmj.b2535.

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.” *SSRN Electronic Journal*, July. https://doi.org/10.2139/ssrn.3649953.

Nussbaumer-Streit, Barbara, Verena Mayr, Andreea Iulia Dobrescu, Andrea Chapman, Emma Persad, Irma Klerings, Gernot Wagner, et al. 2020. “Quarantine Alone or in Combination with Other Public Health Measures to Control COVID-19: A Rapid Review.” Edited by Cochrane Infectious Diseases Group. *Cochrane Database of Systematic Reviews*, April. https://doi.org/10.1002/14651858.CD013574.

Nuzzo, Jennifer B, Lucia Mullen, Michael Snyder, Anita Cicero, Thomas V Inglesby, Amesh A Adalja, Nancy Connell, et al. 2019. “Preparedness for a High-Impact Respiratory Pathogen Pandemic,” September, 84.

Paldam, Martin. 2015. “Meta-Analysis in a Nutshell: Techniques and General Findings.” *Economics* 9 (1):20150011. https://doi.org/10.5018/economics-ejournal.ja.2015-11.

Pan, William K., Stefanos Tyrovolas, Giné-Vázquez Iago, Rishav Raj Dasgupta, Fernández Daniel, Ben Zaitchik, Paul M. Lantos, and Christopher W. Woods. 2020. “COVID-19: Effectiveness of Non-Pharmaceutical Interventions in the United States before Phased Removal of Social Distancing Protections Varies by Region.” https://doi.org/10.1101/2020.08.18.20177600.

Patel, Urvish, Preeti Malik, Deep Mehta, Dhaivat Shah, Raveena Kelkar, Candida Pinto, Maria Suprun, Mandip Dhamoon, Nils Hennig, and Henry Sacks. 2020. “Early Epidemiological Indicators, Outcomes, and Interventions of COVID-19 Pandemic: A Systematic Review.” *Journal of Global Health* 10 (2):020506. https://doi.org/10.7189/jogh.10.020506.

Perra, Nicola. 2020. “Non-Pharmaceutical Interventions during the COVID-19 Pandemic: A Rapid Review.” *ArXiv:2012.15230 [Physics]*, December. http://arxiv.org/abs/2012.15230.

Pincombe, Morgan, Victoria Reese, and Carrie B Dolan. 2021. “The Effectiveness of National-Level Containment and Closure Policies across Income Levels during the COVID-19 Pandemic: An Analysis of 113 Countries.” *Health Policy and Planning* 36 (7):1152–62. https://doi.org/10.1093/heapol/czab054.

Poeschl, Johannes, and Rasmus Bisgaard Larsen. 2021. “How Do Non- Pharmaceutical Interventions Affect the Spread of COVID-19? A Literature Review.” *Nationalbanken Economic Memo*, no. 4:20. https://www.nationalbanken.dk/en/publications/Documents/2021/04/Economic%20Memo%20nr.%204-2021.pdf.

Pozo-Martin, Francisco, Florin Cristea, Charbel El Bcheraoui, and Robert Koch-Institut. 2020. “Rapid Review Der Wirksamkeit Nicht-Pharmazeutischer Interventionen Bei Der Kontrolle Der COVID-19-Pandemie.” *Robert Koch*, September, 17. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Projekte_RKI/Rapid-Review-NPIs.pdf?__blob=publicationFile.

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* 0 (0). Elsevier. https://doi.org/10.1016/j.ajic.2021.05.011.

Rezapour, Aziz, Aghdas Souresrafil, Mohammad Mehdi Peighambari, Mona Heidarali, and Mahsa Tashakori-Miyanroudi. 2021. “Economic Evaluation of Programs against COVID-19: A Systematic Review.” *International Journal of Surgery* 85 (January):10–18. https://doi.org/10.1016/j.ijsu.2020.11.015.

Robinson, Oliver. 2021. "COVID-19 Lockdown Policies: An Interdisciplinary Review." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3782395.

Sears, James, J. Miguel Villas-Boas, Vasco Villas-Boas, and Sofia Berto Villas-Boas. 2020. "Are We #Stayinghome to Flatten the Curve?," August, 2020.05.23.20111211. https://doi.org/10.1101/2020.05.23.20111211.

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*, August, 202010625. https://doi.org/10.1073/pnas.2010625117.

Shenoy, Ajay, Bhavyaa Sharma, Guanghong Xu, Rolly Kapoor, Haedong Aiden Rho, and Kinpritma Sangha. 2022. "God Is in the Rain: The Impact of Rainfall-Induced Early Social Distancing on COVID-19 Outbreaks." *Journal of Health Economics* 81 (January):102575. https://doi.org/10.1016/j.jhealeco.2021.102575.

Shiva, Mehdi, and Hassan Molana. 2021. "The Luxury of Lockdown." *The European Journal of Development Research*, April. https://doi.org/10.1057/s41287-021-00389-x.

Siedner, Mark J., Guy Harling, Zahra Reynolds, Rebecca F. Gilbert, Sebastien Haneuse, Atheendar S. Venkataramani, and Alexander C. Tsai. 2020. "Social Distancing to Slow the US COVID-19 Epidemic: Longitudinal Pretest–Posttest Comparison Group Study." *PLOS Medicine* 17 (8). Public Library of Science:e1003244. https://doi.org/10.1371/journal.pmed.1003244.

Spiegel, Matthew, and Heather Tookes. 2021. "Business Restrictions and COVID-19 Fatalities." Edited by Itay Goldstein. *The Review of Financial Studies*, June, hhab069. https://doi.org/10.1093/rfs/hhab069.

Stanley, T.D., and Hristos Doucouliagos. 2010. "PICTURE THIS: A SIMPLE GRAPH THAT REVEALS MUCH ADO ABOUT RESEARCH." *Journal of Economic Surveys* 24 (1):170–91. https://doi.org/10.1111/j.1467-6419.2009.00593.x.

Stephens, M., J. Berengueres, S. Venkatapuram, and I. A. Moonesar. 2020. "Does the Timing of Government COVID-19 Policy Interventions Matter? Policy Analysis of an Original Database." https://doi.org/10.1101/2020.11.13.20194761.

Stockenhuber, Reinhold. 2020. "Did We Respond Quickly Enough? How Policy-Implementation Speed in Response to COVID-19 Affects the Number of Fatal Cases in Europe." *World Medical & Health Policy* 12 (4):413–29. https://doi.org/10.1002/wmh3.374.

Stokes, Jonathan, Alex James Turner, Laura Anselmi, Marcello Morciano, and Thomas Hone. 2020. "The Relative Effects of Non-Pharmaceutical Interventions on Early Covid-19 Mortality: Natural Experiment in 130 Countries," October, 2020.10.05.20206888. https://doi.org/10.1101/2020.10.05.20206888.

Subramanian, S. V., and Akhil Kumar. 2021. "Increases in COVID-19 Are Unrelated to Levels of Vaccination across 68 Countries and 2947 Counties in the United States." *European Journal of Epidemiology*, September. https://doi.org/10.1007/s10654-021-00808-7.

Toya, Hideki, and Mark Skidmore. 2020. "A Cross-Country Analysis of the Determinants of Covid-19 Fatalities." *CESifo Working Papers*, April, 14.

Tsai, Alexander C, Guy Harling, Zahra Reynolds, Rebecca F Gilbert, and Mark J Siedner. 2021. "Coronavirus Disease 2019 (COVID-19) Transmission in the United States Before Versus After Relaxation of Statewide Social Distancing Measures." *Clinical Infectious Diseases* 73 (Supplement_2):S120–26. https://doi.org/10.1093/cid/ciaa1502.

Umer, Hamza, and Muhammad Salar Khan. 2020. "Evaluating the Effectiveness of Regional Lockdown Policies in the Containment of Covid-19: Evidence from Pakistan." *ArXiv:2006.02987 [Physics, q-Bio]*, June. http://arxiv.org/abs/2006.02987.

UNICEF. 2021. "In-Person Schooling and COVID-19-Transmission: Review of Evidence 2020." https://www.unicef.org/media/89046/file/In-person-schooling-and-covid-19-transmission-review-of-evidence-2020.pdf.

World Health Organization Writing Group. 2006. "Nonpharmaceutical Interventions for Pandemic Influenza, National and Community Measures." *Emerging Infectious Diseases* 12 (1):88–94. https://doi.org/10.3201/eid1201.051371.

Wu, Samuel X., and Xin Wu. 2020. "Stay-at-Home and Face Mask Policies Intentions Inconsistent with Incidence and Fatality during US COVID-19 Pandemic." https://doi.org/10.1101/2020.10.25.20219279.

Zhang, Mengxi, Siqin Wang, Tao Hu, Xiaokang Fu, Xiaoyue Wang, Yaxin Hu, Briana Halloran, et al. 2021. "Human Mobility and COVID-19 Transmission: A Systematic Review and Future Directions." Preprint. Infectious Diseases (except HIV/AIDS). https://doi.org/10.1101/2021.02.02.21250889.