Studies in Applied Economics
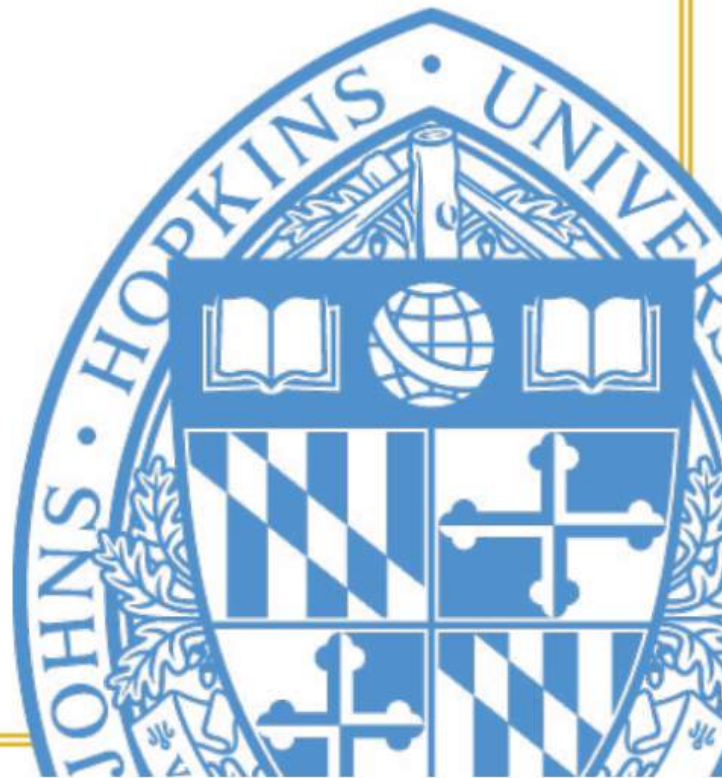
A LITERATURE REVIEW AND META-ANALYSIS OF THE EFFECTS OF LOCKDOWNS ON COVID-19 MORTALITY - II

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 – II

By Jonas Herby, Lars Jonung, and Steve H. Hanke
Published May 20, 2022

About the Series

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

The purpose of this systematic review and meta-analysis is to determine the effect of lockdowns on COVID-19 mortality based on available empirical evidence. Lockdowns are defined as the imposition of at least one compulsory, non-pharmaceutical intervention (NPI). We employ a systematic search and screening procedure in which 19,646 studies are identified that could potentially address the purpose of our study. After three levels of screening, 32 studies qualified. Of those, estimates from 22 studies could be converted to standardized measures for inclusion in the meta-analysis. They are separated into three groups: lockdown stringency index studies, shelter-in-place-order (SIPO) studies, and specific NPI studies. Stringency index studies find that the average lockdown in Europe and the United States in the spring of 2020 only reduced COVID-19 mortality by 3.2%. This translates into approximately 6,000 avoided deaths in Europe and 4,000 in the United States. SIPOs were also relatively ineffective in the spring of 2020, only reducing COVID-19 mortality by 2.0%. This translates into approximately 4,000 avoided deaths in Europe and 3,000 in the United States. Based on specific NPIs, we estimate that the average lockdown in Europe and the United States in the spring of 2020 reduced COVID-19 mortality by 10.7%. This translates into approximately 23,000 avoided deaths in Europe and 16,000 in the United States. In comparison, there are approximately 72,000 flu deaths in Europe and 38,000 flu deaths in the United States each year. When checked for potential biases, our results are robust. Our results are also supported by the natural experiments we have been able to identify. The results of our meta-analysis support the conclusion that lockdowns in the spring of 2020 had little to no effect on COVID-19 mortality. This result is consistent with the view that voluntary changes in behavior, such as social distancing, did play an important role in mitigating the pandemic.

Table of Contents

List of Tables

List of figures

List of acronyms

COVID-19….. Coronavirus Disease 2019
NPI…………..Non-pharmaceutical intervention
OxCGRT…….Oxford COVID-19 Government Response Tracker
PHEIC……….Public Health Emergency of International Concern
PRISMA……..Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PWA…………Precision-weighted average
R_t……………The effective reproductive number. This is the expected number of new infections caused by an infectious individual in a population at time *t*.
SE……………Standard errors
SIPO…………Shelter-in-place order
SIR…………..Susceptible-Infected-Recovered

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

1   Introduction

Social distancing works. If you keep distance from others, your risk of being infected with a communicable disease is reduced. However, the fact that social distancing works does not imply that compulsory non-pharmaceutical interventions (NPIs), commonly known as "lockdowns" – policies that restrict internal movement, close schools and businesses, ban international travel and/or other activities – work. If governments primarily close activities which hardly anyone wants to participate in during an ongoing pandemic, the effect of the lockdown will be modest. If there is too much non-compliance, the effect of the lockdown will be modest. If government only regulates a fraction of the activities where people can become infected, the effect of the lockdown will be modest. If people react strongly to lower infection rates following lockdowns by being much less careful, the effect of the lockdown will be modest. If, if, if…

Although many people perceive lockdowns as extremely effective in reducing Coronavirus Disease 2019 (COVID-19) infections and mortality, it is today – from a research perspective – unknown to what extent lockdowns did in fact reduce COVID-19 infections and COVID-19 mortality. The goal of this study is to answer the following research question: "Were lockdowns effective in reducing COVID-19 mortality?" We also examine if some NPIs were more effective than others.

Definition of "lockdown" and "NPI"
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. During the COVID-19 pandemic, lockdowns have mainly been used to describe two different things. Some use "lockdown" under the definition of "a period of time in which people are not allowed to leave their homes or travel freely". Others use "lockdown" more broadly to describe governments' responses to the pandemic in terms of less or more strict interventions.[1] We follow the latter use and define *lockdown* as any policy consisting of at least one NPI as described above.[2] We use shelter-in-place order (SIPO) to describe the former use of the term "lockdown".

Our focus is on the effect of compulsory NPIs, policies that, for example, restrict internal movement, close schools, and businesses, ban international travel, etc. We do not look at the

---

[1] See https://dictionary.cambridge.org/dictionary/english/lockdown and https://ig.ft.com/coronavirus-lockdowns/ for two different examples of how the term "lockdown" is defined and used.
[2] For example, we will say that the government of Country A introduced the *non-pharmaceutical interventions* of school closures and shelter-in-place-orders as part of the country's *lockdown.*

effect of voluntary behavioral changes (e.g., voluntary mask wearing), the effect of recommendations (e.g., recommended mask wearing), or governmental services (e.g. voluntary mass testing).

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),[3] only Comoros, an island country in the Indian Ocean with a population below 1 million, did not impose at least one NPI (as defined by OxCGRT) before the end of March 2020. Since virtually all countries have implemented some sort of restrictions, we are essentially studying how the degree of lockdowns affects mortality rates.

Figure 1: Percentage of countries with Oxford COVID-19 Government Response Tracker (OxCGRT) stringency index readings above thresholds 65, 70, and 75, respectively
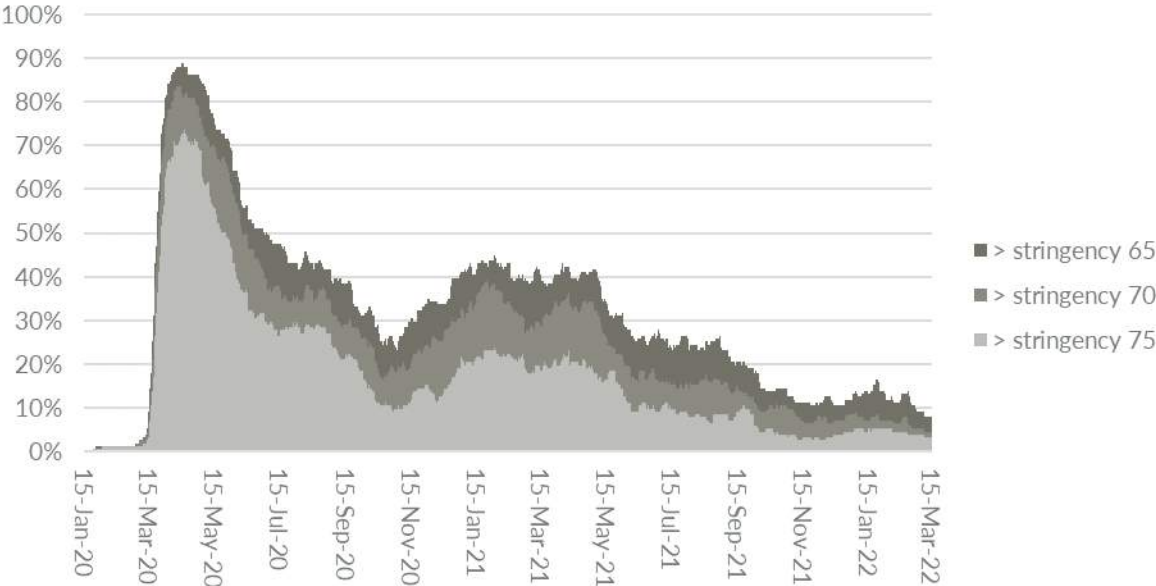
Source: Our World in Data (2022).
Comment: The OxCGRT Stringency Index measures the stringency of lockdowns on a scale from 0 to 100, where a higher value corresponds to stricter lockdowns. The figure shows the share of countries, where the OxCGRT stringency index on a given date surpassed index 65, 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.

---

[3] The Oxford Covid-19 Government Response Tracker (OxCGRT) collects systematic information on policy measures that governments have taken to tackle COVID-19. The different policy responses are tracked since 1 January 2020, cover more than 180 countries, and are coded into 23 indicators, such as school closures, travel restrictions, vaccination policy. These policies are recorded on a scale to reflect the extent of government action, and scores are aggregated into a suite of policy indices.

Do lockdowns work?

One could question the necessity to examine the effectiveness of certain NPIs that have been used for centuries. However, although NPIs such as school and workplace closures were recommended by the World Health Organization (WHO) before the COVID-19 pandemic in the event of an extraordinarily severe pandemic influenza, the evidence of the effectiveness of such measures were, in general, very low (see Table 1).

Table 1: Quality of evidence for selected NPIs as assessed by WHO before the COVID-19 pandemic.

NPI	Quality of evidence as assessed by WHO before COVID-19
School measures and closures	Very low
Workplace measures and closures	Very low
Avoiding crowding	Very low
Entry and exit screening (travelers)	Very low
Internal travel restrictions	Very low
Border closure	Very low

Source: WHO (2019)

Despite the very low quality of evidence (see Table 1), early epidemiological studies predicted that NPIs would have large effects. An often cited model simulation study by researchers at Imperial College London, Ferguson et al. (2020), predicted that a suppression strategy would reduce COVID-19 mortality by up to 99%.[4] Ferguson et al. (2020) state that "it is highly likely that there would be significant spontaneous changes in population behavior even in the absence of government-mandated interventions" but also that "any one intervention in isolation is likely to be limited, requiring multiple interventions to be combined to have a substantial impact on transmission" and "we predict that transmission will quickly rebound if interventions are relaxed" causing many to perceive their projections as a forecast in the case of no lockdown.[5]

---

[4] With $R_0 = 2.0$ and trigger on 60, the number of COVID-19-deaths in Great Britain could be reduced to 5,600 deaths from 410,000 deaths (-99%) with a policy consisting of case isolation + home quarantine + social distancing + school/university closure, see Table 4 in Ferguson et al. (2020). R_0 (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. The lowest effect of lockdowns modelled by Ferguson et al. (2020) was with $R_0 = 2.6$, trigger on 200-400, and case isolation + home quarantine + social distancing. In this case, deaths were predicted to be reduced from 550.000 to 120.000 (-78%).

[5] This perception was supported by statements from the study's authors and Imperial College London. March 17, 2020, Imperial College tweeted that "without more action, the virus would have overwhelmed intensive care units". And, Nial Ferguson is cited for saying that "we're going to have to suppress this virus — frankly, indefinitely — until we have a vaccine.", see https://www.nytimes.com/2020/03/16/us/coronavirus-fatality-rate-white-house.html.

And many hold the results from Ferguson et al. (2020) responsible for the subsequent lockdown in United Kingdom.[6,7]

Already early in the pandemic, there was reason to question whether lockdowns were as effective as promised.[8] First, there was no clear negative correlation between the degree of lockdown and actual outcomes on fatalities in the spring of 2020 (see Figure 2). Given the large perceived effect of lockdowns that studies such as Ferguson et al. (2020) had led to, one would expect to at least observe a simple negative correlation between COVID-19 mortality and the degree to which lockdowns were imposed.

Second, several studies pointed to facts questioning the effect of lockdowns. For example, Atkeson et al. (2020) showed in August 2020 that "across all countries and US 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". Goolsbee and Syverson (2021) found in June 2020 that "legal shutdown orders account for only a modest share of the decline of economic activity […]. While overall consumer traffic fell by 60 percentage points, legal restrictions explain only 7 of that. Individual choices were far more important and seem tied to fears of infection. Traffic started dropping before the legal orders were in place; was highly tied to the number of COVID deaths in the county; and showed a clear shift by consumers away from larger/busier stores toward smaller/less busy ones in the same industry."

---

[6] For example, see https://www.theguardian.com/world/2020/mar/16/new-data-new-policy-why-uks-coronavirus-strategy-has-changed.

[7] Woolhouse (2022), who in January 2020 was appointed to a SAGE sub-committee called the Scientific Pandemic Influenza Group on Modelling (SPI-M) in the United Kingdom, describes how the reaction was due to a misinterpretation of the report from Ferguson et al. (2020): "The report should have been a wake-up call that we needed to invest quickly and heavily in other ways to control novel coronavirus or – according to the model – we'd end up in lockdown. This implication was barely mentioned – lockdown was accepted as a necessity the first time it was proposed. When Report 9 [i.e. Ferguson et al. (2020)]was published the details of the scenarios modelled were quickly forgotten, as were any mentions of the assumptions, caveats and uncertainties of the analysis. Report 9 was condensed to the simple but misleading message that, if the government didn't impose full lockdown immediately, over half a million people would die."

[8] Given Ferguson and his Imperial College team's track record, it's surprising that questions were not raised before jumping on a lockdown bandwagon, see Hanke and Dowd (2022).

Figure 2: The positive correlation between the OxCGRT stringency index and COVID-19 mortality in 44 European countries and the 50 U.S. states (and Washington, D.C.) during the first wave in 2020

Europe

$y = -0.1711x + 135.65$
$R^2 = 0.000$

Deaths per million (April 1 - May 15, 2020)

Average OxCGRT stringency (March 16 - April 30, 2020)

United States

$y = 12.509x - 698.26$
$R^2 = 0.063$

Deaths per million (April 1 - May 15, 2020)

Average OxCGRT stringency (March 16 - April 30, 2020)
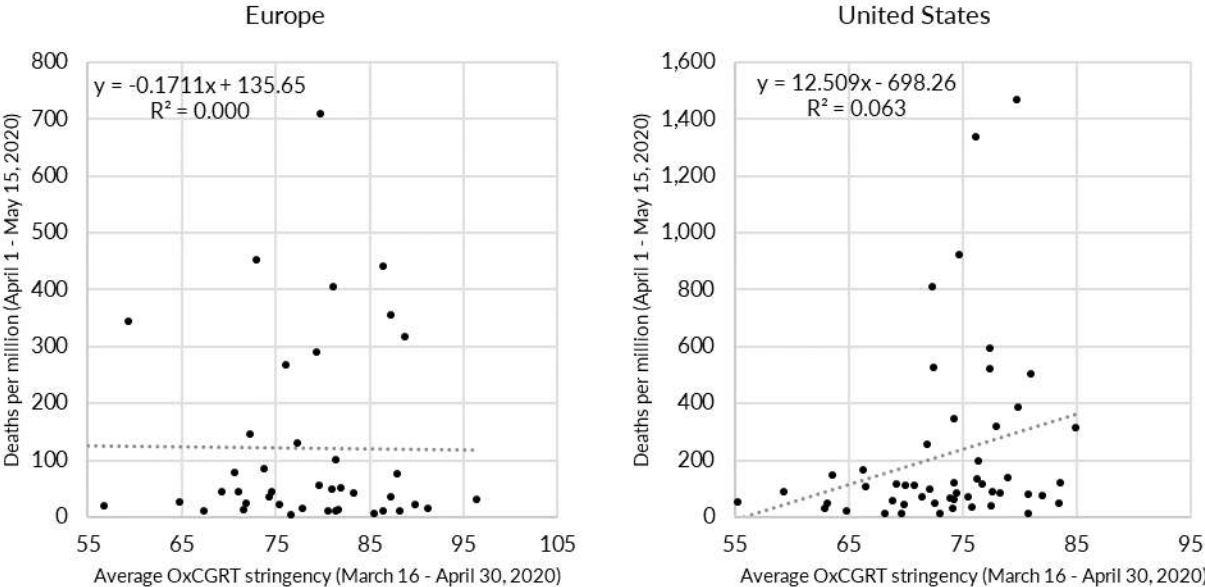
Source: Our World in Data (2022).
Note: The OxCGRT stringency Index measures the stringency of lockdowns on a scale from 0 to 100, where a higher value corresponds to stricter lockdowns. Isle of Man and North Macedonia are not included, because there is no stringency index data for these two countries before April 30, 2020.

Although the externalities associated with a communicable disease like COVID-19 are evident, it is less clear how these externalities can be regulated effectively. [9] Specifically, it remains an open question as to whether lockdowns have had a large, significant effect on COVID-19 mortality.

Our contribution

We address the question "Were lockdowns effective in reducing COVID-19 mortality?" by evaluating the current academic literature on the relationship between lockdowns and COVID-19 mortality rates.[10] Our analysis is based on the evidence found in studies published between January 1, 2020, and February 21, 2022.

We are still in the early phase of the scientific and quantitative evaluation of the effects of lockdowns, and future research will continue to improve our understanding of lockdowns. Still, we find it valuable to summarize in a consistent way the evidence available from the first two years of the pandemic.

Compared to other reviews such as Herby (2021a) and Allen (2021), the main difference in our approach is that we carry out a systematic and comprehensive search strategy to identify all

---

[9] In economics, an externality is an indirect cost or benefit to a third party that arises as an effect of another party's (or parties') activity.

[10] We use "mortality" and "mortality rates" interchangeably to mean COVID-19 deaths as a percentage of population.

papers potentially relevant, and carry out a meta-analysis combining evidence from several existing studies to answer the question we pose. Results need repeated replication to be believable, but replication is unfortunately rare. Mueller-Langer et al. (2019) find that only 0.1% of publications in the top 50 economics journals were replication studies. However, as described by Paldam (2022), the same question is often analyzed in many studies, which use different data samples, estimation models, control variables, etc. The theories of the studies may differ, but the estimation models are often similar. Thus, instead of strict replication, there are often several partial replications. A meta-analysis is a technique developed to analyze if the aggregation of the evidence from studies analyzing the same question leads to a general result. Thus, in our meta-analysis, we aim at replacing replication by presenting results in such a way that they can be systematically assessed and used to derive overall conclusions.

In Figure 3 below, we compare the measured results from our meta-analysis to the forecasts derived from the models used in Ferguson et al. (2020). Overall, the meta-analysis does not support the notion that lockdowns in the spring of 2020 had a large effect on COVID-19 mortality, as many modelling studies had concluded they would.

Figure 3: Divergence between avoided number of deaths in the United States as measured by our meta-results and the forecasted outcome from Imperial College London
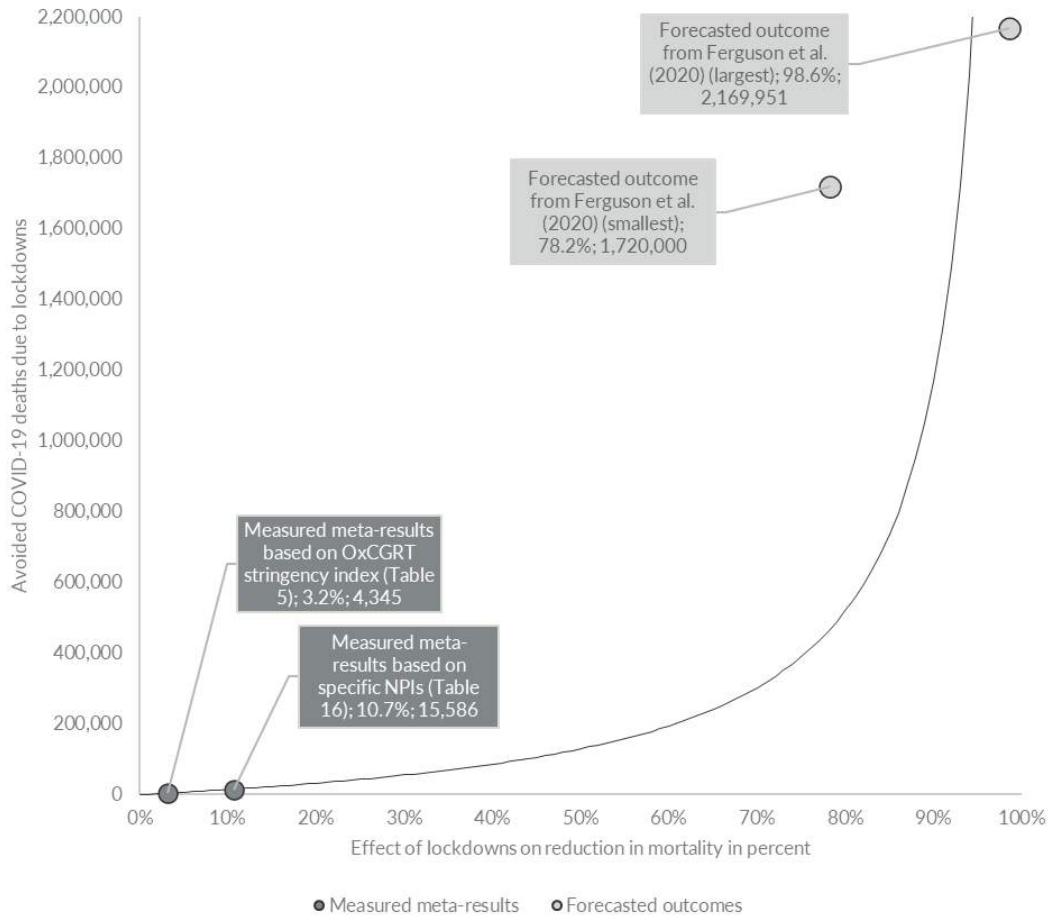
Note: The effect of lockdowns on total mortality based on the meta-study's precision-weighted averages (PWA) is calculated as total COVID-19 deaths by July 1, 2020 (128,063 COVID-19 deaths) x (1/(1-PWA)-1). The relative effect of lockdowns on total mortality based on Ferguson et al. (2020) is calculated as the largest and smallest predicted relative effect multiplied with their mortality estimate of 2.2 million deaths in a "do nothing"-scenario in the United States. For more details see Figure 10 p. 59.

Updates in this version

The first version of this literature review and meta-analysis, Herby et al. (2022), generated countless comments ranging from ad-hominem, wrong, and irrelevant – to of use and of constructive value. For more on the discussion of the first version of this literature review and meta-analysis, see Appendix II.

In this updated version, we have incorporated all constructive comments. Paragraphs have been rewritten for clarity. For example, it was not clear to some commentators that we distinguished between the effect of social distancing and the effect of lockdowns. We have tried to make this distinction more transparent, and the first sentence in the introduction is now "Social distancing works". Our definition of lockdown also created confusion among some critics, so we have elaborated our definition to make it clearer. We have also added more examples to support the understanding of our results and conclusions. For example, we have elaborated section 5.2.3

where we discuss why the effect of lockdowns – as our measured meta-results show – was limited.

Our results have also changed for three reasons. First, we have excluded some studies which we now believe to be ineligible. Second, we have updated our literature search, so more studies are now included. Finally, we have changed some calculations. It is worth noting that we, not the commentators, did identify and correct one computational error that was contained in the first version of this study. Its correction had no significant effect on our overall conclusion.

We have also expanded the section on specific NPIs to give a more in-depth analysis of the results. These updates have changed our estimates, but not the overall conclusion. We believe that one major mistake in our first version was our failure to explain that the overall conclusions do not depend on whether the impact of lockdowns on COVID-19 mortality was 0.2%, 3.0% or 15%. As Figure 3 illustrates, all cases based on actual measurements of saved lives due to lockdowns are much smaller and far removed from the promises made by many epidemiologists, politicians, and the media.

Our literature review and meta-analysis is organized in the following way. In section 2, we describe our identification process for selecting relevant studies. That is, we explain our search strategy and eligibility criteria. In section 3, we present an account of the empirical evidence. Section 4 contains our meta-analysis of the impact of lockdowns on COVID-19 mortality. Section 5 contains our concluding observations and discussion.

2   Identification process: Search strategy and eligibility criteria

Figure 4 presents an overview of our identification process. It uses a flow diagram designed according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines by Moher et al. (2009). Of 19,646 studies identified during our database searches, 1,220 remained after a title-based screening. Then 1,074 studies were excluded because they either did not measure the effect of lockdowns on mortality or did not use an empirical approach. This left 146 studies that were read and inspected carefully. After a more thorough assessment, 114 of the 146 were excluded, leaving 32 eligible studies. The 114 studies excluded in the final step are listed in Table 18 in Appendix I.

Figure 4: The PRISMA flow diagram for the selection of studies

Identification	2021: 18,590 studies identified through extensive database searches (Google Scholar and Scopus). 2022: 1,056 studies identified through Scopus.	→ ←	Search modified to catch references from identified reviews and dedicated COVID-19 portals (e.g. Centre for Economic Policy Research's (CEPR) Covid Economics)
Screening	All 19,646 (18,590 + 1,056) studies screened manually by title (possible related to lockdown and deaths?).	→	18,426 (17,427 + 936 studies excluded because of irrelevant title + 63 doublets excluded)
	1,220 (1,132 + 120) studies possible related to lockdown and deaths screened answering "Measures effect of lockdowns on mortality?" & "Uses empirical approach"?	→	1,074 (1,003 + 71 studies excluded, because the answer is not "yes" to both questions)

Eligibility	146 (129 + 17) full-text studies assessed for eligibility	→	114 studies excluded
			  21 were duplicates with new titles
			  15 only look at timing
			  10 did not look at mortality
			  8 used modelling
			  2 were purely descriptive
			  5 analyzed the effect of social distancing (not lockdowns)
			  14 used time series approach
			  3 were student papers
			  2 did not look at effect of lockdowns
			  10 were not difference-in-difference
			  14 had too few observations
			  10 were synthetic control studies (one obs.)

Included	32 studies included in review	→	Estimates from 10 studies cannot be converted to standardized estimates.
	22 studies included in meta-analysis		

The inclusion rate of 32 eligible studies out of 19,646 identified studies (0.2%) is in the range of other systematic literature reviews on the topic of COVID-19 lockdowns. Our inclusion rate is similar to Talic et al. (2021), who include 72 of 36,729 identified studies (0.2%) and Iezadi et al. (2021) (35 of 12,523, 0.3%), while is relatively low compared to e.g. Rezapour et al. (2021) (26 of 2,176 studies, 1.2%), Zhang et al. (2021) (47 of 1,649 studies, 2.9%), and Johanna et al. (2020) (14 of 623 studies, 2.2%). A major reason for the difference in the inclusion rate is the choice of search strategy. Rezapour et al. (2021), Zhang et al. (2021) and Johanna et al. (2020) identify studies by searching publication databases such as PubMed, Scopus, Web of Science, SAGE, etc., while our search in Google Scholar is more broad. For example, our search also includes presentations and books. Performing our search only with Scopus, instead of both Google Scholar and Scopus, results in an inclusion rate of 0.7%.

Below, we present our search strategy and eligibility criteria, which follow the PRISMA guidelines and are specified in detail in our protocol that we posted in 2021, Herby et al. (2021).[11]

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[12], and a methodology search string.[13] 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 our criteria.[14]

We also included all papers published in *Covid Economics*. Our first search was performed between July 1 and July 5, 2021 and resulted in 18,590 unique studies. 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 materially change the result. Indeed, many papers found using Google Scholar were from these sources. On February 21, 2022, we repeated our search on Scopus resulting in another 1,056 studies.

All 19,646 (18,590 from July 2021, and 1,056 from February 2022) studies were first screened based on the title. Studies clearly not related to our research question were deemed irrelevant.[15]

After screening based on the title, 1,088 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)?

---

[11] The protocol was first published June 23, 2021, and updated the last time October 28, 2021.

[12] 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".

[13] 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".

[14] 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 (2021a); 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).

[15] 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."

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 colleagues.[16]

After the manual screening, 146 studies were retrieved for a full, detailed inspection. These studies were carefully examined, and metadata and empirical results were stored in an Excel spreadsheet. All studies were assessed by at least two researchers. During this process, another 114 papers were excluded because they did not meet our eligibility criteria. A table with all 114 studies excluded in the final step can be found in Appendix I, Table 18. Below we explain the most important of our eligibility criteria. A full list can be found in our protocol Herby et al. (2021).

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. Following our protocol Herby et al. (2021), we exclude studies that use cases, hospitalizations, or other measures.

While we regard analysis based on cases problematic because of large data problems,[17] one could argue that including studies examining the effect of lockdowns on hospitalization could improve the quality of our review and meta-analysis because it would allow us to include more studies. Using the same search strings at Scopus, but replacing ("deaths", "death", and/or "mortality") with ("hospitalization", "intensive care", and/or "ICU"), indicates that including hospitalizations would yield another 1-2 eligible studies.[18]

Although including studies examining the effect of lockdowns on hospitalization would potentially strengthen our results by adding more studies to the review and meta-analysis, we see little reason to believe that doing so would change our results significantly. It is true that a key argument for locking down countries around the world was to protect the healthcare sector and keep hospitalizations down. But one of the arguments for protecting the healthcare sector was that if hospitalizations were high and hospitals were overcrowded, there would be an unusually

---

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

[17] Analyses based on cases 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) find 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, see GRADEpro (2013).

[18] Scopus returns 947 hits on mortality, etc. between January 1, 2020, and June 30, 2021. Searching for hospitalization etc. yields another 35 hits corresponding to 3.7% more studies.

11

high excess mortality rate because COVID-19 patients would not be able to receive treatment.[19] Given this relationship between hospitalizations and deaths, we should see the effect of hospitalizations in our analysis of mortality.

Assessment of actual outcomes (in contrast to modelled outcomes)
There are two different approaches to examine the relationship between mortality rates and lockdown policies. The first approach uses actual, measured mortality data. These are *ex post* studies based on actual mortality outcomes. The second approach uses simulated data on mortality and infection rates generated from models.[20] These are *ex ante* studies based on modelled outcomes.

In this review and meta-analysis, we include all studies from the former group but exclude all *ex ante* studies, as the results from these studies are determined by model assumptions and calibrations and cannot be the basis for solid empirical evidence for policy purposes, before these models have been validated empirically which is exactly the point of our study. This means that we exclude e.g. the much-cited Ferguson et al. (2020) from Imperial College.

Counterfactual difference-in-difference approach
We exclude studies which do not use a counterfactual difference-in-difference approach (called 'controlled before-and-after studies' in some social sciences).[21]

Difference-in-difference is a quasi-experimental design that makes use of longitudinal data from treatment and control groups to obtain an appropriate counterfactual to estimate a causal effect. Difference-in-difference is typically used to estimate the effect of a specific intervention (for example, a non-pharmaceutical intervention) by comparing the changes in outcomes over time between a treatment group (where a specific NPI was in place) and a control group (where the NPI was not in place).

---

[19] E.g. Madsen et al. (2014) find that "high bed occupancy rates were associated with a significant 9 percent increase in rates of in-hospital mortality and thirty-day mortality, compared to low bed occupancy rates. Being admitted to a hospital outside of normal working hours or on a weekend or holiday was also significantly associated with increased mortality."

[20] These simulations are often made in variants of the susceptible-infected-recovered (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 of moving between these states.

[21] See Mailman School of Public Health, Columbia (2007) which has also inspired Figure 5 and other parts of the description of the difference-in-difference method.

Figure 5: Simplified illustration of the difference-in-difference approach compared to interrupted time series approach when trends changes

Panel A: Simplified illustration of the difference-in-difference method
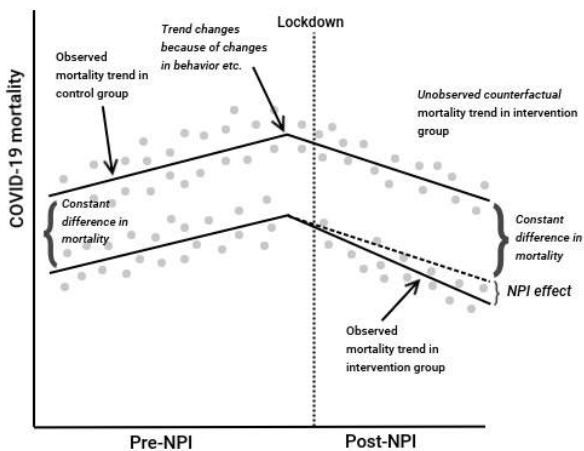
Panel B: Simplified illustration of the interrupted time series method
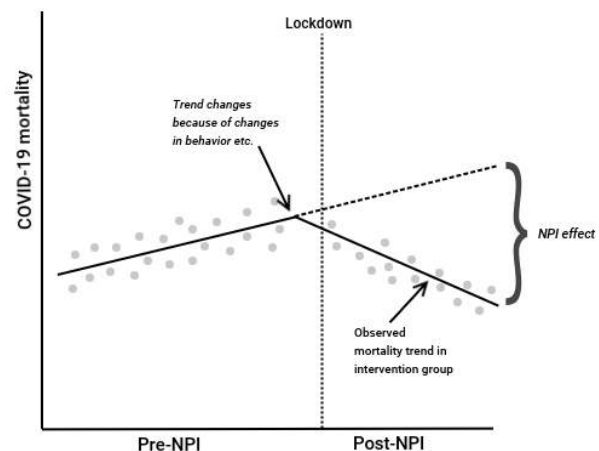
Difference-in-difference is used in observational settings where exchangeability cannot be assumed between the treatment and control groups, as is the case for COVID-19 mortality, where mortality rates differ between countries and over time. The difference-in-difference method relies on a somewhat weaker – although not negligible – exchangeability assumption known as the parallel, or common, trends assumption as illustrated in Figure 5, Panel A.[22]

Difference-in-difference is a useful technique to employ when examining the effect of lockdowns where randomization is not possible. The approach removes biases in post-intervention period comparisons between the treatment and control group that could be the result from permanent differences between those groups (e.g. caused by coincidences early in the pandemic[23]), as well as biases from comparisons over time in the treatment group that could be the result of trends due to other causes of the outcome (e.g. changes in behavior or seasonality).

The exclusion of studies that do not use a counterfactual difference-in-difference approach means that we exclude all studies where the counterfactual is based on forecasting (for example, using a SIR-model calibrated on mortality data). This means that we excludes studies like Duchemin et al. (2020) and Matzinger and Skinner (2020). We also exclude all studies based on interrupted time series designs. Interrupted time series designs are useful when there is a stable long-term period before and after the time of the intervention examined (lockdowns), and where other things are relatively constant and/or can be controlled for. This is not the case with COVID-19 and lockdowns, where the period before (and often after) the intervention is short, where things are far from constant, and where changes in behavior cannot easily be controlled

---

[22] For a more in-depth discussion of the parallel assumption, we refer to McKenzie (2020a), McKenzie (2020b), and McKenzie (2021). Also see McKenzie (2022a) and McKenzie (2022b).

[23] As we describe on page 81, Arnarson (2021) and Björk et al. (2021) show that areas in Europe 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.

for. As illustrated in Figure 5, Panel B, interrupted time series risk overestimating the effect of lockdowns, if, for example, voluntary behavioral changes are important. Excluding interrupted time series studies rule out studies like Bakolis et al. (2021) and Siedner et al. (2020).

Given our criteria, we also exclude the much-cited paper by Flaxman et al. (2020), which claimed that lockdowns saved three million lives in Europe. Flaxman et al. (2020) 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 factor that matters, even if other factors like changes in voluntary behavior, seasonality, etc. caused the observed change in the reproduction rate, R_t. Figure 6 illustrates how problematic Flaxman's assumption is. The figure shows Flaxman et al. (2020)'s estimate of the effect of various NPIs on the effective reproduction number, R_t,[24] in Denmark and Sweden. According to the results, banning public events had close-to-zero effect in Denmark (Panel A) but huge effects in Sweden (Panel B).

Figure 6 The assumptions used by Flaxman et al. lead to two contradictory conclusions: That banning public events had no effect in Denmark but were extremely effective in Sweden in March 2020

Panel A: R_t in Denmark
R_t in Denmark is assumed to drop after the implementation of "Complete lockdown" (i.e., caused by the closure of businesses).

Panel B: R_t in Sweden
R_t in Sweden is assumed to drop after "Public events banned". An intervention that had close-to-zero effect in Denmark.
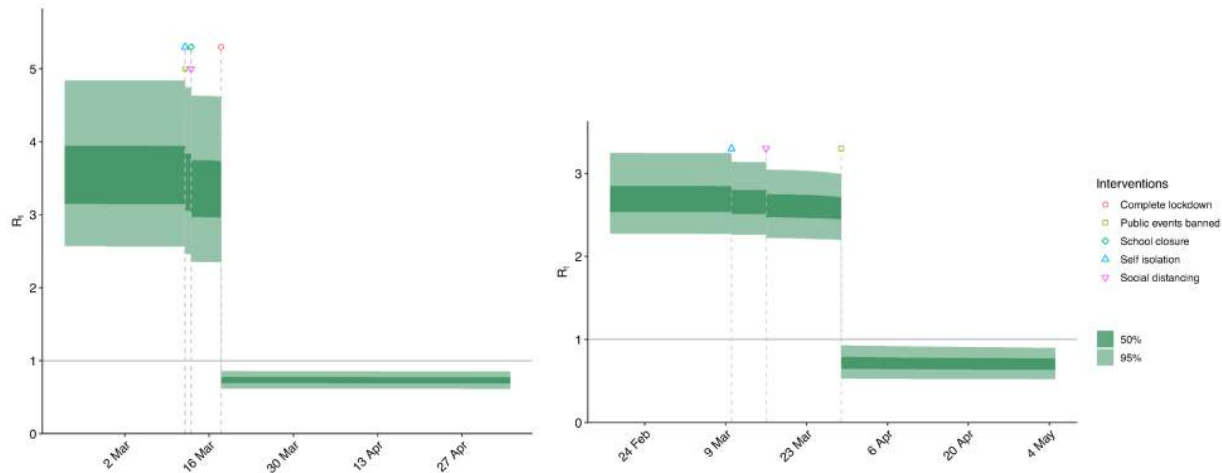
Source: Flaxman et al. (2020), Extended Data Fig. 1 & Fig. 2.

Flaxman et al. (2020) are aware of this problem and state that "our parametric form of R_t assumes that changes in R_t are an immediate response to interventions rather than gradual changes in behavior". Despite stating that the results cannot be interpreted as an effect of

---

[24] The effective reproductive number, denoted R_t, is the expected number of new infections caused by an infectious individual.

lockdowns, media around the globe – supported by statements by the study's authors[25] – reported these findings as "proof" that lockdowns had saved millions of lives.[26] Similar interpretation problems are typical and are found in Brauner et al. (2021)[27] and Hsiang et al. (2020) for example.

In our view, Flaxman et al. (2020), Brauner et al. (2021), Hsiang et al. (2020), etc. illustrate how problematic it is to force data to fit a certain model and certain assumptions if you want to infer the effect of lockdowns on COVID-19 mortality, and how these assumptions – while not being academically incorrect, as they are readily available in their paper – can lead to misguided perceptions in the media.[28] Including the estimates from these studies and interpreting them as the effect of lockdowns would without any doubt greatly overstate the effectiveness of lockdowns.

Jurisdictional variance - few observations
We also exclude studies with little jurisdictional variance.[29] For example, we exclude Conyon et al. (2020) 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 (As of matter of fact, research has shown that Sweden was particularly unlucky in the spring of 2020. Arnarson (2021) and Björk et al. (2021) show that areas in Europe – like Sweden – 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). Another example is Wu and Wu (2020), who use all U.S. states, but pool groups of states so they

---

[25] For example, Dr. Seth Flaxman said "Lockdown averted millions of deaths", see e.g. https://www.bbc.com/news/health-52968523, and Samir Bhatt said "Our estimates show that lockdowns had a really dramatic effect in reducing transmission," and ""Without them [lockdowns] we believe the toll would have been huge,", see e.g. https://news.wgcu.org/2020-06-09/modelers-suggest-pandemic-lockdowns-saved-millions-from-dying-of-covid-19.

[26] For example, see https://www.reuters.com/article/us-health-coronavirus-lockdowns-idUSKBN23F1G3, https://www.bbc.com/news/health-52968523, https://www.imperial.ac.uk/news/198074/lockdown-school-closures-europe-have-prevented/, https://www.france24.com/en/20200609-covid-19-lockdowns-saved-millions-of-lives-and-easing-curbs-risky-studies-find, and https://www.washingtonpost.com/health/2020/06/08/shutdowns-prevented-60-million-coronavirus-infections-us-study-finds/.

[27] Brauner et al. (2021) states that "if increasing availability of information reduces infection growth rates, it would cause us to overstate the effectiveness of anti-contagion policies" and Hsiang et al. (2020) write that "our approach cannot distinguish direct effects on transmission in schools and universities from indirect effects, such as the general population behaving more cautiously after school closures signaled the gravity of the pandemic".

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

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

end up with basically three observations.[30] None of the excluded studies cover more than 10 jurisdictional areas.

Synthetic control studies

The synthetic control method is a special case of the difference-in-difference 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 counterfactual 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 derived through the synthetic control. We exclude synthetic control studies because of too little jurisdictional variance, as these studies examine the effects of lockdowns based on one country/state compared to a synthetic counterfactual.

But – as discussed by Bjørnskov (2021b) – synthetic control studies also have empirical problems in relation to studying the effect of lockdowns. Bjørnskov (2021b) 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 (2021b) estimates that *actual Sweden* experienced approximately 500 fewer deaths the first 11 weeks of 2020 and 4,500 fewer deaths in 2019 compared to *synthetic Sweden*. Such empirical problems are inherent to all synthetic control studies of COVID-19 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.[31] 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 well suited for studying the effects of lockdowns.

The role of optimal timing

One important thesis on the effect of lockdowns is that timing is important for a lockdown to be effective. The rationale behind this thesis is straightforward (assuming lockdowns are effective): If an epidemic is growing exponentially, the benefit of intervening sooner rather than later is disproportionately large. For example, if the doubling-time is one week, then locking down one week earlier will more than halve the total number of deaths, assuming that the pre-lockdown reproduction number is larger than two and *if* the lockdown brings the reproduction number below one. On the other hand, locking down too strongly and too early can result in a resurgence when restrictions are lifted if there is a failure to completely eliminate the virus, with potentially

---

[30] One can discuss whether our criteria should in fact exclude Wu and Wu (2020), when their data include all U.S. states. However, by pooling the data, they also lose much of the variation in data which was our primary reason for requiring a minimum of jurisdictional variance. Excluding Wu and Wu (2020) is unlikely to change our results as they conclude that "there were no significant differences in total positive cases, average daily new cases and average daily fatality among the 3 groups [SIPO and mask mandates, SIPO but no mask mandate, and neither SIPO nor mask mandate] during the pandemic period. This study suggested that SIPO and mask mandate orders in the general public alone, probably have limited effects in lowering transmission and fatality."

[31] Also see Abadie (2021).

higher deaths than if it was permitted to spread to a small extent prior to the lockdown. Hence, the argument goes that there is an optimal timing of lockdowns.[32]

We, however, evaluate the general effect of lockdowns, i.e., whether lockdowns on average have been effective in reducing COVID-19 mortality. We therefore exclude studies which solely analyze the effect of optimally timed lockdowns in contrast to less well-timed lockdowns. There are several reasons for this exclusion.[33]

First, studies searching for the optimal timing of lockdowns will by design find inflated effects of the average lockdown, because they – if optimal timing is important – will neglect all the less well-timed lockdowns implemented around the world. And hence, these studies will not result in an unbiased estimate of the average effect of lockdowns.

Secondly, it is inherently 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. In fact, it is difficult for a democratic country's political leaders to impose and enforce a lockdown, unless there is a widespread belief that a danger is imminent.

Björk et al. (2021) illustrate the difficulties in analyzing the effect of timing in Europe. They find that a 10-stringency-points-stricter lockdown would reduce COVID-19 mortality by a total of 200 deaths per million[34] if done in week 11, 2020, but would only have approximately 1/3 of the effect if implemented one week earlier or later and close to 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 mandatory restrictions 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 correct, what Björk et al. (2021) find is that information and signaling are far more important than the strictness of the lockdown. There may be other interpretations, but our point is that studies focusing on timing cannot differentiate between these two conflicting interpretations.

This view is also supported by Figure 7, which illustrates that all European countries and US states which were hit hard and early by COVID-19 in the spring of 2020 experienced high overall mortality rates, whereas all countries hit relatively late experienced low mortality rates. The figure shows that there is no doubt that being prepared for a pandemic and knowing when it arrives at your doorstep is vital. But to what degree this can be attributed to well-timed lockdowns or simply to alert citizens is a question, which is not easily answered and may previously have been misunderstood or neglected in prior research on e.g. the 1918 Spanish Flu pandemic (we will get back to this issue in section 5.2.4 on p. 71).

---

[32] For more on this thesis see e.g. Abernethy and Glass (2022) and Oraby et al. (2021).

[33] This exclusion criteria was mistakenly not made public in our protocol Herby et al. (2021), but "only looks at timing" was decided upon as an exclusion criteria mid-September 2021 (documentation can be provided on request). Also see Herby (2021d).

[34] 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 7: **All countries and states that were hit late by the pandemic experienced lower COVID-19 mortality rates**
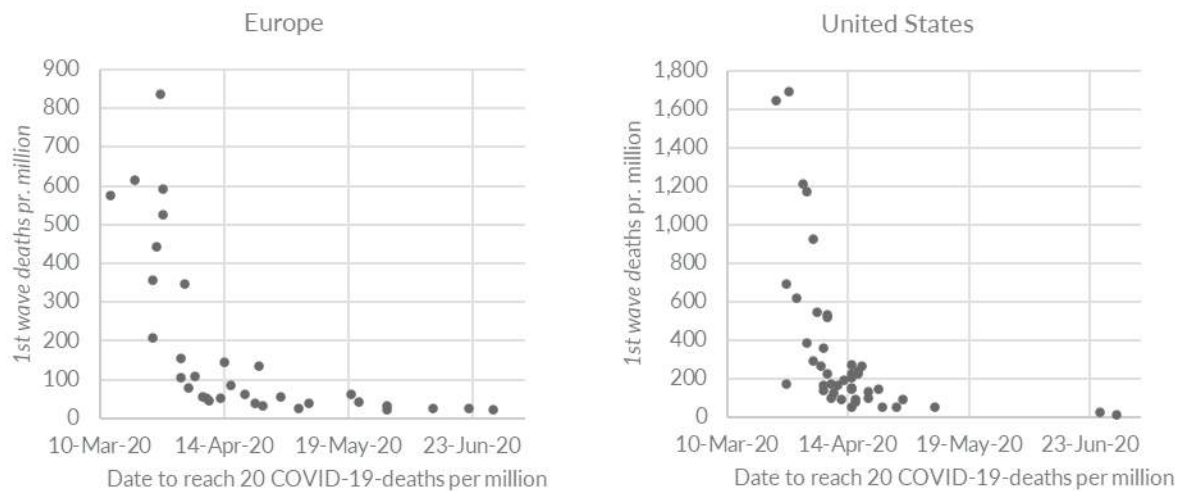
Comment: The figure shows the relationship between early pandemic strength and total 1st wave of COVID-19 mortality. On the X-axis is "Days to reach 20 COVID-19-deaths per million (measured from February 15, 2020)." The Y-axis shows mortality (deaths per million) by June 30, 2020.
Source: Reported COVID-19 deaths and OxCGRT stringency for European countries and US states with more than one million citizens. Data from Our World in Data (2022).

We are aware of three reviews (Iezadi et al. (2021), Perra (2020), and Stephens et al. (2020)), which opine on the importance of timing. Stephens et al. (2020) 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. (2020) 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." Iezadi et al. (2021) write that "it is very important to contain the spread of the infection at the very early stage of the outbreak. At later stages, no NPIs, even if implemented harshly, might be very effective", while Perra (2020) writes that "countries that acted early, with respect to the local spread, were most successful in controlling the spread and reported markedly lower death tolls". But these three reviews do not distinguish between the effect of information (which is the effect of being hit late by the pandemic) and the effect of lockdowns.

Although the verdict on optimal timing is still out, we would like to stress the importance of alternative interpretations here. As Figure 7 illustrates, one can easily interpret the lower mortality rates as an effect of early lockdowns even when it is caused by changes in voluntary behavior or – not unlikely – by a combination of both. One should be careful concluding that early lockdown is important when alternative conclusions – such as changing information and voluntary behavior – may explain outcomes equally well.

We further note that even if future research finds that the timing of lockdowns *is* crucial, such knowledge may not be useful for future policymakers. First, it is not easy to know when the right timing is. When COVID-19 hit Europe and the United States, it was virtually impossible to determine the right timing. The World Health Organization declared the COVID-19 pandemic a

Public Health Emergency of International Concern (PHEIC) on January 30, 2020, WHO (2020a). However, this was the sixth PHEIC in just 11 years, and could not reasonably justify a lockdown.[35] March 11, 2020, was the first time the WHO characterized COVID-19 as a pandemic, WHO (2020b). 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 date that a lockdown response to the WHO's announcement could potentially have a large effect due to the lag between infection and death, the mortality rate in Italy was a staggering 178 COVID-19 deaths per million, with an additional 13 per million dying each day.[36]

Second, we already pointed to the fact that policymakers (at least in democratic countries) need support from the electorate to impose and enforce lockdowns. So even if a few people know the right time to impose a lockdown, this information is only useful if citizens and politicians agree that there is a dangerous and threatening infectious disease, and act upon that threat.[37] But under these conditions, they are also likely to respond significantly (and voluntarily) to recommendations, making a lockdown less necessary. In fact, data from the influenza surveillance program in Denmark from Statens Serum Institut (2020) show that the influenza vanished *before* lockdowns were implemented but possibly *coinciding with* the announcement of coming lockdowns which spurred significant voluntary behavioral changes. However, the influenza vanished at the exact same time in Norway and Sweden, as illustrated in Figure 8, suggesting that *if* lockdowns spurred significant voluntary behavioral changes, the Swedish 500-person limit on public gatherings effective by March 12, 2020, may have been sufficient to spur these changes.

---

[35] Wilder-Smith and Osman (2020) state that "six events were declared PHEIC between 2007 and 2020: the 2009 H1N1 influenza pandemic, Ebola (West African outbreak 2013-2015, outbreak in Democratic Republic of Congo 2018-2020), poliomyelitis (2014 to present), Zika (2016) and COVID-19 (2020 to present)."

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

[37] In his book, "The Premonition: A Pandemic Story", M. Lewis (2021) describes how White House experts wanted to close schools in United States in the beginning of the 2009 swine flu, but did not have the necessary support to implement what the group thought was the right decision. It turned out that closing schools would in fact have been a mistake. Later, in early 2020, the same group again wanted to close schools, etc. Again, they did not have the necessary political support, but this time it remains unknown if it could potentially have saved a significant number of lives.

Figure 8: **Influenza disappeared at the same time in Denmark, Norway, and Sweden in March 2020 despite radical differences in lockdown policies**
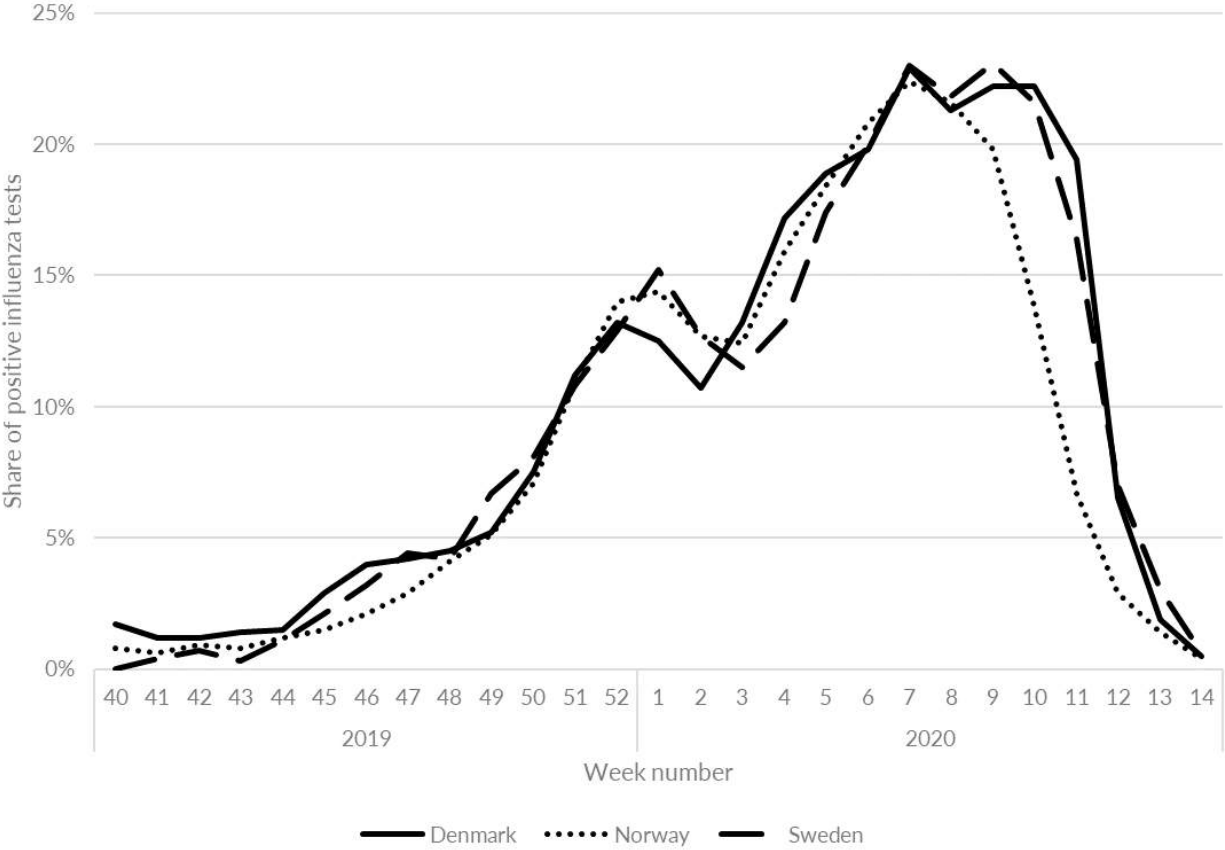
Source: Data from Emborg et al. (2021)
Note: In Sweden, gatherings were limited to 500 persons from Thursday, March 12, 2020 (week 11), while high schools and higher education was closed from Wednesday, March 18, 2020 (week 12).

To conclude: We believe that most – if not all – studies focusing on timing fails to distinguish between the effects of lockdowns and the effect of voluntary behavioral changes.

3   The empirical evidence

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

3.1   Preliminary considerations

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

The empirical interpretation is based on results – not the studies' authors' conclusions
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 the 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%.[38]

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. This means that for lockdowns with a stringency of at least 124, the lockdown theoretically reduces mortality. However, the stringency index is limited to values between 0 and 100 by design, so for all possible values of the stringency index, the total effect of lockdowns on mortality is positive (more deaths), so Chisadza et al.'s conclusion is infeasible.

This is illustrated in Figure 9 below. The figure describes the total policy effect based on Chisadza et al. (2021) estimates for their squared specification. Starting from a lockdown with a stringency of 0 (no lockdown) and increasing stringency from there, a stricter lockdown increases mortality. But, at stringency 62.4, a stricter lockdown reduces mortality at the margin. However, the total effect is still an increase in mortality for stringency values below 124. And because stringency values are capped at 100, there are no such lockdowns that decrease mortality overall.

---

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

Figure 9: The total policy effect, including infeasible values outside the range of the OxCGRT stringency index, as estimated by Chisadza et al. (2021)
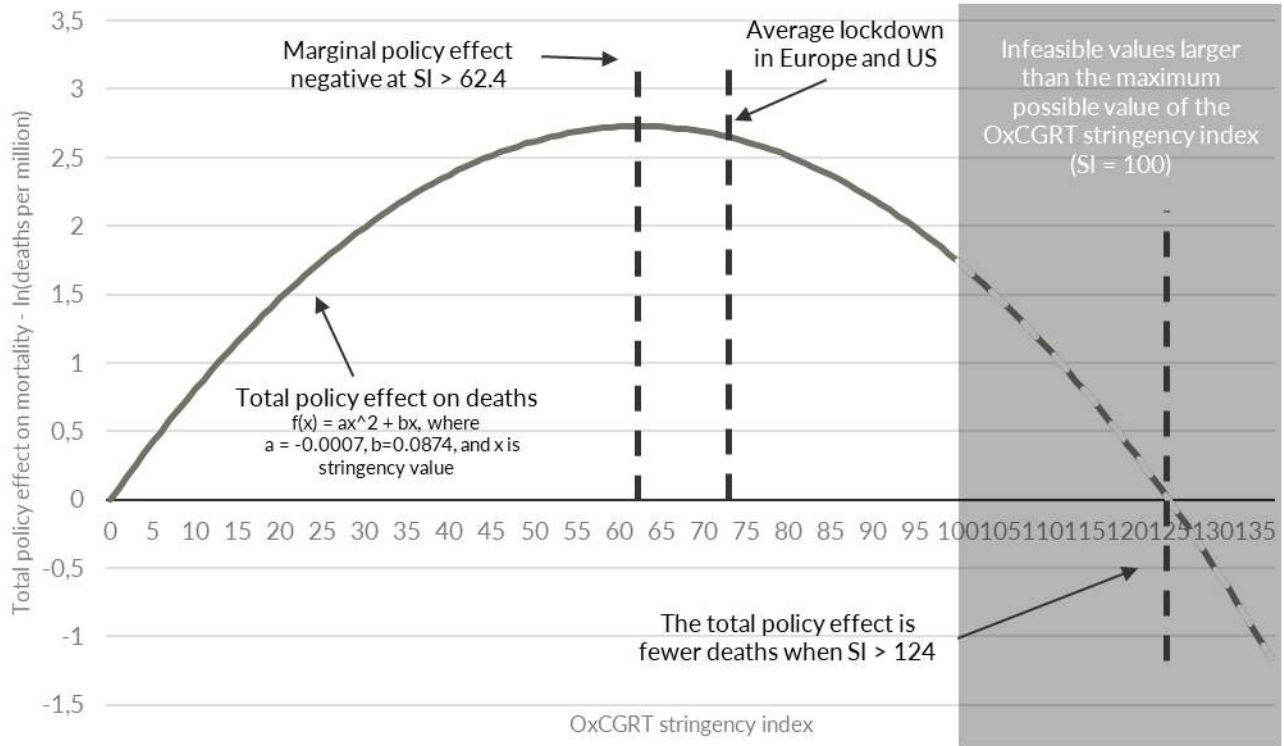
Source: Chisadza et al. (2021)
Note: SI = OxCGRT stringency Index. The OxCGRT stringency Index measures the stringency of lockdowns on a scale from 0 to 100, where a higher value means stricter lockdowns.

Again, 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 substantially reduces the efficacy of lockdowns on mortality. The explanation could be that people responded to information about national conditions. For these studies, we present both estimates in Table 2, but – following Doucouliagos and Paldam (2008) – we use an average of the estimates in our meta-analysis to avoid giving more weight to a study with multiple models

relative to studies with just one principal model (note that for Chernozhukov et al. (2021), we can only include the model estimating large effects of lockdowns, because the report the counterfactual effect for this model only – see Table 19).

For studies which look at different classes of countries (e.g., rich and poor), we report both estimates in Table 2 but use the estimate for rich Western countries in our meta-analysis, where we derive standardized estimates for Europe and the United States.

Effects are measured relative to "doing the least in the spring of 2020"
Virtually all countries in the world implemented some kind of lockdown 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 did during the first wave, when Sweden, due to constitutional constraints,[39] implemented very few restrictions compared to other western countries. However, some studies state that they *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) 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. 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 "doing the least", 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 the least" when available. Hence, for Bonardi et al. (2020), we state that the effect of lockdowns is zero (compared to "doing the least").

Note that this also means that our results cannot say much about the importance of *signaling*. One could imagine that lockdown may serve as a signal to citizens that now is the time to be careful. If signaling is thought to be important, future research should focus on finding the least costly signals.

---

[39] See Jonung and Hanke (2020).

3.2 Overview of the findings of eligible studies

Table 2 covers the 32 studies eligible for our review.[40] Out of these 32 studies, 19 were peer-reviewed and 13 were working papers. The studies analyze lockdowns during the first wave. Most of the studies (25) use data collected before September 1st, 2020, and 6 use data collected before May 1st, 2020. Only two studies use data collected after January 1st, 2021. All studies are cross-sectional, ranging across jurisdictions. Geographically, 15 studies cover countries worldwide, two cover European countries, 13 cover the United States, one covers Europe and the United States, and one covers the OECD member countries. Seven studies analyze the effect of SIPOs, 11 studies analyze the effect of stricter lockdowns (measured by the OxCGRT stringency index), 13 studies analyze specific NPIs independently, and one study analyzes other measures.[41]

Several studies find no statistically significant effect of lockdowns on mortality. This includes Bjørnskov (2021a) and P. Goldstein et al. (2021), 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 An et al. (2021) and Guo et al. (2021) who find no significant negative (fewer deaths) effect of any of the analyzed NPIs, including business closures, school closures, and border closures.

Other studies find a significant *negative* relationship between lockdowns and mortality. Fowler et al. (2021) conclude that SIPOs reduce COVID-19 mortality by 35%, while Chernozhukov et al. (2021) state that employee mask mandates reduces mortality by 34% and closing businesses and bars reduces mortality by 29%.

A few 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 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 we believe to be the best (albeit still imperfect) measure, as it overcomes the measurement problems related to proper reporting of COVID-19 deaths.

Table 2: 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	COVID-19 mortality	The study uses state-level data from all U.S. states published by the COVID-19 Tracking Project, and a multivariate regression analysis to empirically investigate the impacts of the duration of shelter-in-place orders on mortality.	The study finds that shelter-in-place orders are - for the average duration - associated with 1% (insignificant) fewer deaths per capita.	The study does not use a stringent difference-in-difference approach; but this approach is included as the study uses panel data with a time dimension.

---

[40] The following numbers are based on data presented in Table 4.

[41] Yue Li et al. (2021) use "data on the stringency of state social distancing measures from wallethub.com".

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
cases, mortality, and recovery rates"				
An et al. (2021); "Policy Design for COVID-19: Worldwide Evidence on the Efficacies of Early Mask Mandates and Other Policy Interventions"	COVID-19 deaths	The authors use 164 countries to estimate the long run efficacy of early mandate adoption. They use both a country fixed-effects model and a country random effects model. They also use several other methods in order to estimate the specific NPI´s effect on infection rates and mortality, and they look specifically at timing for each NPI.	With different modeling approaches and alternative measurements (for both the focal independent and dependent variables), the analysis shows that domestic lockdowns and restaurant closures do not display any consistent associations with new infection and mortality rates in the short term. Mass gathering bans and school closures need more time to manifest their short-term efficacies. Both cross-sectional and longitudinal analyses provide consistent evidence that only mask mandates demonstrate persistent long-run efficacy from early adoption.	The study uses several different methods to estimate how effective timing and specific NPIs are and finds that the only NPI that shows consistent reduction of infection rates and mortality both in the short and long term is mask mandates.
Ashraf (2020); "Socioeconomic conditions, government interventions and health outcomes during COVID-19"	COVID-19 mortality	The study's focus is on the effectiveness of policies targeted to diminish the effect of socioeconomic inequalities (economic support) on COVID-19 deaths. The study uses data from 80 countries worldwide and includes the OxCGRT stringency as a control variable in its models. The paper finds a significant negative (fewer deaths) effect of stricter lockdowns. The effect of lockdowns becomes insignificant when the author includes an interaction term between the socioeconomic conditions index and the economic support index in the study's 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 becomes insignificant when including an interaction term between the socioeconomic conditions index and the economic support index.	
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 Hopkins University and SIPO data from the University of Washington to estimate the effect of SIPOs. They find no detectable effects of SIPOs 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.	A 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. The paper looks at the effect 14 days after SIPOs 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	The study uses excess mortality and OxCGRT stringency data from 24 European countries to estimate the effect of lockdowns on the number of deaths in the subsequent one, two, three and four weeks.	A stricter lockdown (OxCGRT stringency) does not have a significant effect on excess mortality. Specifications using instrument variables yields similar results.	The study finds a positive (more deaths) effect after one and two weeks, which could indicate that other factors (omitted variables) affect the results.

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
Blanco et al. (2020); "Do Coronavirus Containment Measures Work? Worldwide Evidence"	COVID-19 mortality	The study uses 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 orders, bans on gatherings, bans on public events, school closures, lockdowns of workplaces, interruption of public transportation services, and international border closures. They address the possible endogeneity of the NPIs by using instrumental variables.	When using the naïve dummy variable approach, all parameters are statistically 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.	The study runs the same model four times for each of the different NPIs (stay at home-orders, bans on meetings, bans on public events, and mobility restrictions). These NPIs were often introduced almost simultaneously so there is a high risk of multicollinearity with each run capturing the same underlying effect. Indeed, the size and standard errors of the estimates are worryingly similar. The study looks at the effect 14 days after NPIs are implemented, which is a fairly short lag given that the time between infection and deaths is 2-3 weeks; see 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	The study uses NPI data collected from news headlines from LexisNexis and death data from Johns Hopkins University up to April 1st, 2020 in a panel structure with 184 countries. The study controls for country fixed-effects, day fixed-effects, and within-country evolution of the disease.	The study finds that certain interventions (SIPO, regional lockdown and partial lockdown) work (in developed countries), but that stricter interventions (SIPOs) do not have a larger effect than less strict interventions (e.g., restrictions on gatherings). It finds no effect of border closures.	The study finds 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
Chernozhukov et al. (2021); "Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S."	Growth rates	The study uses COVID-deaths from the New York Times, Johns Hopkins University, and data on U.S. states' policies from Raifman et al. (2020) to estimate the effect of SIPOs, closed nonessential businesses, closed K-12 schools, closed restaurants except takeout, closed movie theaters, and face mask mandates for employees in public facing businesses.	The study 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) have no statistically significant effect. The effect of school closures is highly sensitive to the inclusion of national case and death data.	The authors state that "our regression specification for case and death growths is explicitly guided by an SIR model although our causal approach does not hinge on the validity of an SIR model." We are uncertain if this means that data are managed to fit an SIR-model (and thus should fail our eligibility criteria).
Chisadza et al. (2021); "Government Effectiveness and the COVID-19 Pandemic"	COVID-19 mortality	The study uses COVID-19-deaths and OxCGRT stringency from 144 countries to estimate the effect of lockdowns on the number of COVID-19-deaths. It finds a significant positive (more deaths) non-linear association between government response indices and the number of deaths.	In the authors' linear model, an increase by 1 on "stringency index" increases the number of log deaths by 0.0130 per million (corresponding to 1.3%). In their non-linear model, 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 authors state 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).
Clyde et al. (2021); "A Study of the Effectiveness of Governmental Strategies	COVID-19 mortality	The study uses data for NPIs in 37 OECD countries from the Oxford COVID-19 Government Response Tracker (OxCGRT) to examine the impact of policy variable	Results indicate that only school closings and public transportation closings have a persistently significant	

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
for Managing Mortality from COVID-19"		changes between March 15 and October 31, 2020.	impact. Stay-at-home policies only show a significant impact after 70 days. Workplace closings, restrictions on the size of gatherings, and restrictions on internal travel show no significant impact on mortality rates. Moreover, stricter measures are not significantly associated with lower growth rates in mortality.	
Dave et al. (2021); "When Do Shelter-in-Place Orders Fight Covid-19 Best? Policy Heterogeneity Across States and Adoption Time"	COVID-19 mortality	The study 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 SIPOs. The authors find that a 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 a significant effect on mortality after 20+ days (see Figure 4). They indicate that the lacking significance may be due to long term estimates being identified of a few early adopting states.	The study finds no overall significant effect of a SIPO on deaths but does find a negative effect (fewer deaths) in early adopting states.	The study finds large effects of a SIPO on deaths after 6-14 days in early adopting states (see Table 8), which is before a SIPO-related effect would be seen. This could indicate that other factors rather than SIPOs drive the results.
Dergiades et al. (2020); "Effectiveness of government policies in response to the COVID-19 outbreak"	COVID-19 mortality	The study uses daily deaths from the European Centre for Disease Prevention and Control and OxCGRT stringency from 32 countries worldwide (including U.S.) to estimate the effect of lockdowns on the number of deaths.	The study 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.	The 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.
Ertem et al. (2021); "The impact of school opening model on SARS-CoV-2 community incidence and mortality"	COVID-19 deaths	The study uses 459 U.S. counties to estimate the impact of school opening on mortality. It uses multivariate Poisson regressions with robust standard errors. The authors define 3 school modes, as they focus on the difference between traditional and virtual.	"There was no impact of school opening mode on subsequent COVID-19-related deaths during the entire 12-week period after school opening in any region (Fig. 4 and Extended Data Fig. 2)." The authors also conclude that there are major limitations because the underlying reason for regional differences could not be delineated.	
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	The study 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 Johns Hopkins University. The study instruments stringency using day-to-day changes in the stringency of the restrictions in the rest of the world.	The study finds large causal effects of stricter restrictions on the weekly growth rate of recorded deaths attributed to COVID-19. It shows that more stringent interventions help more in richer, more educated, more democratic, and less corrupt countries with older, healthier populations and more effective governments.	The authors find 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.

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
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	The study 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. It finds a large and early (after few days) effect of SIPOs 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 might have reduced confirmed cases by 390,000 (the 95% confidence interval spans from 170,000 to 680,000) and fatalities by 41,000 (from 27,000 to 59,000) within the first three weeks in localities that implemented stay-at-home orders.	The study finds the largest effect of SIPOs 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.
Fuller et al. (2021); "Mitigation Policies and COVID-19–Associated Mortality — 37 European Countries, January 23–June 30, 2020"	COVID-19 mortality	The study uses COVID 19-deaths and OxCGRT stringency in 37 European countries to estimate the effect of lockdowns on the number of COVID-19 deaths. It finds a significant negative (fewer deaths) effect of stricter lockdowns after the 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	The study uses data for every county in the United States from March through June 1, 2020, to estimate the effect of SIPOs (called "lockdowns") on COVID-19 mortality. Policy data are acquired from American Red Cross reporting on emergency regulations. The author's 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. The author handles causality problems using instrument variables (IV).	The author finds no statistically significant effect of SIPOs.	The author uses the word "lockdown" as synonym for SIPO (writes "technically, government-ordered community quarantine")
P. Goldstein et al. (2021); "Lockdown Fatigue: The Diminishing Effects of Quarantines on the Spread of COVID-19 "	COVID-19 mortality	The study uses panel data from 152 countries with data from the onset of the pandemic until December 31, 2020. It 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 days, the effect of lockdowns is smaller in the short run, but after 90 days the effect is almost the same as during the first lockdown (only approximately 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	The study uses policy data of 1,470 executive orders from the state–government websites for all 50 U.S. states and Washington DC and COVID-19 deaths from Johns Hopkins University in a random-effect spatial error panel model to estimate the effect on COVID-19 deaths of nine NPIs: SIPO, strengthened	Two mitigation strategies (all school closure and mandatory self-quarantine of travelers) showed a positive (more deaths) impact on COVID-19-deaths per 10,000. Six mitigation strategies (SIPO,	The study only concludes 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

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
		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.	public school closure, large gathering bans [>10], any gathering ban, restaurant/bar limit to dining out only, and nonessential business closure) did not show any impact (Table 3, "Proportion of Cumulative Deaths Over the Population).	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.
Hale et al. (2020); "Global assessment of the relationship between government response measures and COVID-19 deaths"	COVID-19 mortality	The study uses the OxCGRT stringency and COVID-19-deaths from the European Centre for Disease Prevention and Control for 170 countries. It 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).	The study 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.	
Hale, Angrist, Hale, et al. (2021); "Government responses and COVID-19 deaths: Global evidence across multiple pandemic waves"	COVID-19 mortality	The study uses the OxCGRT stringency to analyze the effect of stricter lockdowns in 113 countries through 1-3 waves.	The study finds that stricter lockdowns reduce mortality and that the effect was particularly large – even during the first wave – in countries which experienced three waves.	The authors' results on three wave countries are based on just 10 countries. However, since there is no obvious reason why lockdowns in countries, which later experienced a third wave, should be particularly effective, this could indicate that some model misspecification is driving the results.
Leffler et al. (2020); "Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks"	COVID-19 mortality	The study uses 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.	The study 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). It finds no significant effect of the number of weeks with internal lockdowns and international travel restrictions (Table 2).	The authors' "mask recommendation" category includes some countries where masks were mandated (see Supplemental Table A1) and may (partially) capture the effect of mask mandates. The study looks at duration, which may cause a causality problem, because politicians may be less likely to ease restrictions when there are many cases or deaths.
Yue Li et al. (2021); "Association of state social distancing restrictions with nursing home COVID-19 and non-COVID-19 outcomes"	COVID-19 mortality	The study uses data on the stringency of state social distancing measures from wallethub.com (based on 17 state COVID-19 policy metrics) to categorize US states into either "high" or "low" stringency. The study then uses linear regression to estimate cumulative numbers of deaths among residents in long-term care in low and high stringency states based on mortality data from 14,046 nursing homes.	The study finds that stricter lockdowns reduced COVID-19 mortality and across-facility disparities in COVID-19 outcomes, but also caused more deaths due to non–COVID reasons among nursing home residents.	
Mccafferty and Ashley (2021); "Covid-19 Social Distancing Interventions by Statutory Mandate and Their Observational Correlation to Mortality in the United States and Europe"	Other	The study uses data from 27 U.S. states and 12 European countries to analyze the effect of NPIs on peak mortality rate using general linear mixed-effects modelling.	The study finds that no mandate (school closures, prohibition on mass gatherings, business closures, stay at home orders, severe travel restrictions, and closure of non-essential businesses)	

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
			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	The study uses county-level data for all U.S. states. Mortality data are obtained from Johns Hopkins University, while policy data are obtained from official governmental websites. It 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 or stay-at-home, public mask requirements, or travel restrictions. It uses a stepped-wedge cluster randomized trial (SW-CRT) for clustering and negative binomial mixed model regression.	The study 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.	The authors 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	The study 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 (the authors use one variable covering both recommendations and orders).	The study finds that shelter-in-place recommendations or 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	The study uses cellular location data from all 50 states and the District of Columbia to investigate mobility patterns during the pandemic across states and time. The authors estimate the effect of stay-at-home policies on COVID-19 mortality by adding COVID-19 death tolls and the timing of SIPOs for each state. "We obtain travel activity and social distancing data from the analytics company Unacast." "To denote periods before or after a state implemented a 'stay at home order,' we obtain the date each statewide policy was issued [45] for all fifty states and the District of Columbia" – from the New York Times.	The study finds 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. The 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 a SIPO is a reduction in deaths by 29%-35% compared to a situation without a SIPO but with pre-mandate social distancing. These estimates are insignificant at a 95% confidence interval.
Shiva and Molana (2021); "The Luxury of Lockdown"	COVID-19 mortality	The study uses COVID-19-deaths and OxCGRT stringency from 169 countries to estimate the effect of a lockdown on the number of deaths 1-8 weeks later. It finds that stricter lockdowns reduce COVID-19-deaths 4 weeks later (but insignificantly after 8 weeks) and have the greatest effect in high income countries. It	A stricter lockdown (1 stringency point) reduces deaths by 0,1% after 4 weeks. After 8 weeks the effect is insignificant.	

1. Study (Author & title)	2. Measure	3. Description	4. Results	5. Comments
		finds no effect of workplace closures in low-income countries.		
Spiegel and Tookes (2021a); "All or Nothing? Partial Business Shutdowns and COVID-19 Fatality Growth"	Growth rates	The study uses hand-collected policy data for all US counties from March through December 2020 to estimate the effect of capacity limits on spas, bars, restaurants, and gyms.	The study finds mixed results, where partial capacity restrictions on restaurants and bars are more effective than full shutdowns, and full shutdowns of gyms reduce fatality growth rate, while partial capacity restrictions are counterproductive.	
Spiegel and Tookes (2021b); "Business restrictions and Covid-19 fatalities"	COVID-19 mortality	The study uses 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.	The study 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 the authors 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 not readily intuitive; e.g., mask recommendations increase deaths by 48% while mask mandates reduce deaths by 12%, and closing restaurants and bars reduces deaths by 50%, while closing bars but not restaurants only reduces deaths by 5%, although the latter could potentially be explained by more crowding in open venues. They handle early effect (within 1-2 weeks) in Table 5.
Stokes et al. (2020); "The relative effects of non-pharmaceutical interventions on early Covid-19 mortality: natural experiment in 130 countries"	COVID-19 mortality	The study 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). It 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, and 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) have a large significant negative (fewer deaths) effect, while the effect of stricter restrictions on gatherings are insignificant. The authors recommend that the closing of schools (level 1) has a very large (in absolute terms it is twice the effect of border quarantines) positive 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 (the authors' 0–24-	The authors' results are counter intuitive and somewhat inconclusive. Why does limiting very large gatherings (>1,000) work, while stricter limits do not? Why does recommending school closures cause more deaths? Why is the effect of border closures before the first death insignificant, while the effect of closing borders after the first death is significant (and large)? 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
			day model returns mostly insignificant results).	
Yang et al. (2021); "What is the relationship between government response and COVID-19 pandemics? Global evidence of 118 countries"	COVID-19 deaths	The study uses OxCGRT stringency covering 118 countries from January 1 to April 13, 2021 to estimate the effect of lockdowns on mortality.	The study finds a statistical negative (fewer deaths) effect of stricter lockdowns on COVID-19 mortality rates 21 days later.	The study finds a (relatively large) effect after just seven days before any policy could be effective, 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 draw any firm conclusions based on the overview presented in Table 2. For example, 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 Imperial College London (Ferguson et al. (2020))? This question is the subject for our meta-analysis in the next section. Here, it turns out that Ashraf's (2020) -0.073 to -0.326 deaths/million per stringency point represents a relatively modest effect—one that corresponds to only 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 22 studies in which we can derive a *standardized estimate,* which is the relative effect of lockdowns on COVID-19 mortality. That is, we obtain an estimate of how many percent fewer deaths there were due to lockdowns. For some studies, the authors state the relative effect, and our standardized estimate is thus readily available. For other studies, their estimate must be converted to our standardized estimate. Doing so is fairly straightforward for most studies and our calculations are explained in Table 19. However, the estimates from 10 studies cannot be converted to our standardized estimates. This include studies estimating the effect of lockdowns on mortality growth rates, unless the authors, like e.g. Chernozhukov et al. (2021), calculate a counterfactual scenario.[42] Also, Mccafferty and Ashley (2021) estimate the effect of lockdowns on peak mortality, but an estimate of peak mortality cannot meaningfully be converted to our standardized estimate, which measures the relative impact on COVID-19 mortality.

---

[42] One reason that we cannot calculate a counterfactual based on the estimated effect on growth rates is that we do not know the effect on the distribution and vertex of the death curve. If lower growth rates simply flatten the curve, the effect on total mortality can be limited even if the effect on growth rates is substantial.

The conclusions in the excluded studies are, overall, in correspondence with the conclusions in the included studies, see Table 3. Five excluded studies find lockdowns reduce mortality, two find mixed effects of NPIs with some NPIs working and others not working, and three find no or positive (more deaths) effects of NPIs. In comparison, eight of the included studies find lockdowns reduce mortality, 11 find mixed or insignificant effects of NPIs, and three studies find no or positive (more deaths) effects of lockdowns. Our reasons for not including a study are described in Table 19 in Appendix I.

Table 3: Conclusions from included and excluded studies in the meta-analysis are similar

Conclusion	Number (share) of studies included in meta-analysis	Number (share) of studies *not* included in meta-analysis
Find that lockdowns reduce mortality	8 (36%)	5 (50%)
Find mixed effects of NPIs	11 (50%)	3 (30%)
Find that lockdowns have no effect (or increases mortality)	3 (14%)	2 (20%)
Total	22	10

Note: Mixed effects means that some of the examined NPIs reduce mortality while others increase mortality. The category "Find that lockdowns reduce mortality" includes studies which find at least one significant negative (fewer deaths) estimate ($p < 0.05$) and no significant positive estimates (more deaths). The category "Find mixed effects of NPIs" includes studies which find both significantly negative and significantly positive estimates. The category "Find that lockdowns have no effect (or increases mortality)" includes studies which find no significant estimates or find at least one significant positive (more deaths) estimate ($p < 0.05$) and no significant negative (fewer deaths) estimates.

The studies we examine are placed in three categories. Eight studies analyze the effect of stricter lockdowns based on the OxCGRT stringency indices, 12 studies analyze the effect of SIPOs (six studies only analyze SIPOs and six analyze SIPOs together with other interventions), and eight studies analyze the effect of specific NPIs independently.[43] Each of these categories is handled so that comparable estimates can be made across categories.

Bias dimensions

Not all eligible studies are of the same quality. One way to handle this problem is to evaluate the quality of each study and use this evaluation to weigh or group the studies. However, there is currently no consensus as to best practices and/or an established scientific framework for evaluating the effectiveness of NPIs and lockdowns.[44] As a result, such evaluation risk being subjective. Instead, we investigate whether there are any biases in the reviewed studies that can affect the studies' conclusions. We do this by dividing them into different "bias dimensions". Below, we describe the dimensions, as well as our reasons to believe why they could describe important biases. We also describe which group we perceive as "better" (meaning "probably less biased"). However, it should be noted that the primary objective of these dimensions is to identify and understand any biases in the studies, which can affect our overall results.

- *Peer-reviewed vs. working papers*: We distinguish between peer-reviewed studies and working papers. All else being equal, we perceive peer-reviewed studies as better than working papers.[45]

---

[43] The total is larger than 22 because the 12 SIPO studies include six studies which look at multiple measures.
[44] See Banholzer et al. (2022).
[45] Vetted papers from CEPR Covid Economics are considered as working papers in this regard.

- *Long vs. short data periods*: 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 omit the last part of the first wave and may yield biased results if lockdowns only "flatten the curve" and do not prevent deaths. All else being equal, we perceive studies based on long periods as better than studies based on short periods.

- *No early effect vs. early effect on mortality*: On average, it takes approximately three to four weeks from infection to death.[46] However, several studies find effects of lockdown on mortality almost immediately, which is clearly inconsistent with the standard view of COVID-19 transmission. 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. Thus, we distinguish between studies which find an effect on mortality sooner than 14 days after lockdown and those that do not.[47] Note that many studies do not look at the short term and thus fall into the latter category by default. All else being equal, we perceive no early effect as better than an early effect (or no info).

- *Lagged vs. no lag of policy measures*: On average, it takes approximately three to four weeks from COVID-19 infection to death. Hence, the effect of a lockdown policy on mortality should be seen about three weeks after the policy measure is implemented, and a specification with a 2-4 weeks lagged policy variable is expected to be better than a specification with no lag, because the latter also captures the development around the policy decision which is not influenced by the policy.[48] All else being equal, we perceive a lagged effect to be better than no lag.

- *Panel vs. no panel estimation:* The development of a pandemic in a country is affected by several factors and therefore may be inherently different from country to country. One way to handle these intrinsic differences is to exploit the panel data structure using fixed or random effects regression. Thus, we distinguish between studies which use fixed or random effects and those that do not. All else being equal, we perceive fixed/random effects regression as better than those that do not use fixed/random effects.

- *Verified vs. unverified data:* We distinguish between studies using verified data (e.g., from OxCGRT or New York Times) and unverified data (e.g., data collected by researchers but

---

[46] Leffler et al. (2020) write "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) state 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."

[47] 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."

[48] See Chernozhukov et al. (2022) for a discussion of why this may affect estimates.

not readily publicly available and not updated on a continuous basis). All else being equal, we perceive studies using verified data as better than studies using unverified data.[49]

- *Address vs. do not address causality*: Not all studies address the causality/endogeneity question. We distinguish between studies addressing the causality question and studies which do not address the question. We consider the question as addressed if the authors handle causality technically (e.g., using instrument variable or lagged dependents) or argue for the causality of their results. All else being equal, we perceive studies that address causality as better than studies which do not address causality.

- *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. 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", the research field is registered as "Economics"). Where no corresponding author was available, the affiliation of the first author has been used. Afterwards, all research fields have been classified as either from the "Social Science" or "Other."[50] All else being equal, we perceive social sciences as better suited for examining the effect of policies than other sciences.

We also considered including a bias 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 accurate 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). Second, a major goal with lockdowns was to save lives. To the extent 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 of the 32 studies, Bjørnskov (2021a),is based on excess mortality, we have to disregard this bias dimension.

---

[49] Eight studies are based on stringency data from OxCGRT, two are based on specific NPI data from OxCGRT, three are based on data from New York Times, one on data from COVID-19 Tracking Project, and one on data from Response2covid19. These 15 studies are considered being based on verified data. Three studies collect their own data, two studies rely on data from other studies, one study uses purchased data from Burbio.com, one study use American Red Cross reporting on emergency regulation. These six data sources are considered as being unverified (the latter two because they require login credentials, and thus are not easily verified by external researchers).

[50] Research fields classified as social sciences are 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.

Meta-data used for our bias dimensions as well as other relevant information are shown in Table 4.

Table 4: Bias dimension data for the studies included in the meta-analysis

1. Study (Author & title)	2. Publication status	3. End of data period	4. Earliest effect	5. Policy/ death lag	6. Fixed/random effects	7. Verified data	8. Handling of causality	9. Field of research	10. Quality index
Alderman and Harjoto (2020); "COVID-19: US shelter-in-place orders and demographic characteristics linked to cases, mortality, and recovery rates"	Peer-review	11-Jun-20	n/a	No	No	Other official source (Verified)	Does not handle causality	Economics (Social science)	0.25
An et al. (2021); "Policy Design for COVID-19: Worldwide Evidence on the Efficacies of Early Mask Mandates and Other Policy Interventions"	Peer-review	15-Jul-20	n/a	Yes	Yes	Other official source (Verified)	Does not handle causality	Public policy (Social science)	0.14
Ashraf (2020); "Socioeconomic conditions, government interventions and health outcomes during COVID-19"	WP	20-May-20	n/a	No	Yes	OxCGRT stringency index (Verified)	Other model handling	Economics (Social science)	0.25
Berry et al. (2020); "Evaluating the effects of shelter-in-place policies during the COVID-19 pandemic"	Peer-review	30-May-20	8-14 days	No	Yes	Other research data (Unverified)	Lag dependent (>2 weeks)	Public policy (Social science)	0.00
Bjørnskov (2021); "Did Lockdown Work? An Economist's Cross-Country Comparison"	Peer-review	30-Jun-20	<8 days	Yes	Yes	OxCGRT stringency index (Verified)	Instrument variable	Economics (Social science)	0.77
Bonardi et al. (2020); "Fast and local: How did lockdown policies affect the spread and severity of the covid-19"	WP	13-Apr-20	<8 days	Yes	Yes	Own data (Unverified)	Argumentation	Economics (Social science)	0.39
Chernozhukov et al. (2021); "Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S."	Peer-review	03-Jun-20	n/a	Yes	Yes	Other research data (Unverified)	Other model handling	Economics (Social science)	0.14
Chisadza et al. (2021); "Government Effectiveness and the COVID-19 Pandemic"	Peer-review	01-Sep-20	n/a	No	no	OxCGRT stringency index (Verified)	Does not handle causality	Economics (Social science)	0.25
Dave et al. (2020); "When Do Shelter-in-Place Orders Fight Covid-19 Best? Policy Heterogeneity Across States and Adoption Time"	Peer-review	20-Apr-20	Finds no effect	Yes	Yes	New York Times (Verified)	Argumentation	Economics (Social science)	0.77
Ertem et al. (2022); "The impact of school opening model on SARS-CoV-2 community incidence and mortality"	Peer-review	21-Dec-20	>21 days	Yes	Yes	Other data (Unverified)	Does not handle causality	Engineering (Other)	0.77
Fowler et al. (2021); "Stay-at-home orders associate with subsequent decreases in COVID-19 cases and fatalities in the United States"	Peer-review	07-May-20	<8 days	No	Yes	New York Times (Verified)	Other model handling	Public Health (Social science)	0.39

Study		Date				Data	Causality	Field	
Fuller et al. (2021); "Mitigation Policies and COVID-19–Associated Mortality — 37 European Countries, January 23–June 30, 2020"	WP	30-Jun-20	n/a	Yes	No	OxCGRT stringency index (Verified)	Does not handle causality	Epidemiology (Other)	0.14
Gibson (2020); "Government mandated lockdowns do not reduce Covid-19 deaths: implications for evaluating the stringent New Zealand response"	Peer-review	01-Jun-20	Finds no effect	Yes	Yes	Other data (Unverified)	Instrument variable	Economics (Social science)	0.77
Goldstein et al. (2021); "Lockdown Fatigue: The Diminishing Effects of Quarantines on the Spread of COVID-19 "	WP	31-Dec-20	<8 days	Yes	Yes	OxCGRT stringency index (Verified)	Lag dependent (>2 weeks)	International Development (Social science)	0.56
Guo et al. (2021); "Mitigation Interventions in the United States: An Exploratory Investigation of Determinants and Impacts"	Peer-review	07-Apr-20	n/a	No	Yes	Own data (Unverified)	Does not handle causality	Social work (Social science)	0.06
Hale et al. (2021); "Government responses and COVID-19 deaths: Global evidence across multiple pandemic waves"	Peer-review	11-Mar-21	n/a	Yes	Yes	OxCGRT stringency index (Verified)	Other model handling	Government (Social science)	0.77
Leffler et al. (2020); "Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks"	Peer-review	09-May-20	n/a	No	No	OxCGRT data (Verified)	Does not handle causality	Ophthalmology (Other)	0.39
Sears et al. (2020); "Are we #stayinghome to Flatten the Curve?"	WP	29-Apr-20	Finds no effect	No	Yes	New York Times (Verified)	Argumentation	Economics (Social science)	0.39
Shiva and Molana (2021); "The Luxury of Lockdown"	Peer-review	08-Jun-20	15-21 days	Yes	Yes	OxCGRT stringency index (Verified)	Does not handle causality	Government (Social science)	0.77
Spiegel and Tookes (2021); "Business restrictions and Covid-19 fatalities"	Peer-review	31-Dec-20	15-21 days	Yes	No	Own data (Unverified)	Other model handling	Management (Social science)	0.25
Stokes et al. (2020); "The relative effects of non-pharmaceutical interventions on early Covid-19 mortality: natural experiment in 130 countries"	WP	01-Jun-20	n/a	Yes	Yes	OxCGRT data (Verified)	Other model handling	Economics (Social science)	0.06
Yang et al. (2021); "What is the relationship between government response and COVID-19 pandemics? Global evidence of 118 countries"	Peer-review	03-Dec-20	<8 days	Yes	No	OxCGRT stringency index (Verified)	Does not handle causality	Economics (Social science)	0.39

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 4. In Table 19 in Appendix I, we describe for each paper how we interpret the estimates

and how they are converted to a standardized 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 (SE)[51] and use the precision of each estimate (defined as 1/SE) to calculate the precision-weighted average of all estimates. The precision-weighted average (PWA) is our primary indicator of the efficacy of lockdowns, but we also report arithmetic averages and medians in the meta-analysis.

Sensitivity analyses
Given the relatively low number of studies in the meta-analysis, a study with an outlier estimate or outlier weight may influence our primary indicator of the efficacy of lockdowns. One way to deal with this uncertainty and illustrate the robustness of our estimates is to cap the estimates and weights at the end of the tails.

We therefore carry out four sensitivity analyses, where we replace the outlier (min/max) estimate/weight with the nearest estimate/weight and recalculate the PWA. For instance, the conclusion of Chisadza et al. (2021) is an outlier which finds that the average lockdown *increases* COVID-19 mortality. In one sensitivity analysis, we replace the estimate from Chisadza et al. (2021) with the nearest estimate from Bjørnskov (2021a) and recalculate the PWA. We report the result of our four sensitivity analyses as a span (from/to) at the bottom of each table.

Quality-adjusted precision-weighted average
As a supplement to the PWA and the sensitivity analyses, we also calculate a quality-adjusted PWA based on our bias dimensions displayed in Table 4. The quality-adjusted PWA is calculated as the PWA weighted by a quality index, where the score on the quality index for each study is the squared number of bias dimensions, where the study is of "better" quality. Hence, each study can score between 0 and 64 on the index (because it includes eight bias dimensions). Finally, the index is normalized to 0-1 by dividing by 64.

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

Eight eligible studies examine the link between lockdown stringency and COVID-19 mortality.[52] The results from these studies, converted to standardized estimates, are presented in Table 5

---

[51] Standard errors are converted such that the t-value, calculated based on standardized 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).

[52] An earlier version of our meta-study also included Stockenhuber (2020). However, Stockenhuber (2020) does not use a difference-in-difference approach and is excluded in this version.

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.[53] Many countries followed similar patterns and almost all countries closed schools, 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), see Hale, Angrist, Goldszmidt, et al. (2021b), 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 far from 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 across two countries. In any case, there is no better available measure to adopt for cross-country comparisons.

Since the studies examined use different units of estimates, we have created standardized estimates for Europe and United States to make them comparable. The standardized 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)[54] compared to the most lenient lockdown, which we define as a COVID-19 policy based solely on recommendations (stringency 44).[55]

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. 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 the United States, it was 30 points. Hence, the total effect of the lockdowns compared to the recommendation policy was (using rounded numbers) -6.37 deaths/million in Europe (32 x -0.200) and -5.91 (30 x -0.200) deaths/million in United States. With populations of 748 million

---

[53] 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 Public information campaigns," is not an intervention following our lockdown 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 our conclusions.

[54] Unless otherwise noted, we use these values in our calculations. The average stringency index is relatively stable during the first wave until end of June 2020. For instance, the average stringencies are 73 and 72, respectively, between March 16th and June 30th, 2020.

[55] In reality, the most lenient lockdown varies from study to study depending on the group of countries and/or states included in the study. However, for practical purposes our definition is sufficient to calculate standardized estimates.

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 standardized estimate is thus -2.4%, see Table 5. Our approach is not unproblematic. First of all, the level of stringency varies over time for all countries. Secondly, OxCGRT has changed the index over time, and a 10-point difference today may not be the exact same as a 10-point difference when the studies were finalized. However, we believe these problems are small and unlikely to significantly alter our results.

Table 5 demonstrates that the studies find that lockdowns, on average, have reduced COVID-19 mortality rates by 3.2% (precision-weighted average) and the sensitivity analysis shows a span from 4.4% to 3.0%. The results yield an arithmetic average of 8.9% and a median of 5.8%. To put the estimate in perspective, there were 188,542 registered COVID-19 deaths in Europe and 128,063 COVID-19 deaths in United States by June 30, 2020. Thus, the 3.2% PWA (8.9% arithmetic average, 5.8% median) corresponds to 6,000 (18,000, 12,000) avoided deaths in Europe and 4,000 (13,000, 8,000) avoided deaths in the United States.[56] In comparison, there are approximately 72,000 flu deaths in Europe and 38,000 flu deaths in the United States each year.[57]

Hence, based on the stringency index studies, we find that mandated lockdowns in Europe and the United States had little to no effect on COVID-19 mortality rates.

---

[56] The estimate from Fuller et al. (2021), the only study finding a substantial effect of lockdowns (-35%), corresponds to 103,000 avoided deaths in Europe and 70,000 avoided deaths in United States.

[57] The average estimated flu deaths in United States the five years prior to COVID-19 was 38,400 according to the CDC (2022), and the WHO (2022) states that there are 72,000 flu deaths in Europe each year.

Table 5: Estimates of the effect on COVID-19 mortality of the average lockdown in Europe and in the United States from studies based on the OxCGRT stringency index

Effect on COVID-19 mortality	Standardized estimate (Estimated Averted Deaths / Total Deaths)	Standard error	Weight (1/SE)
Bjørnskov (2021)	-0.3%	0.822%	122
Shiva and Molana (2021)	-4.0%**	0.395%	253
Chisadza et al. (2021)	11.7%**	1.442%	69
Goldstein et al. (2021)	-7.5%	1.964%	51
Fuller et al. (2021)	-35.3%**	9.085%	11
Ashraf (2020)	-2.4%**	0.391%	256
Yang et al. (2021)	-16.3%**	4.523%	22
Hale et al. (2021)	-16.9%**	2.812%	36
Precision-weighted average (arithmetic average / median)	**-3.2% (-8.9%/-5.8%)**		
Sensitivity analysis (quality-adjusted PWA)	**-4.4% to -3.0% (-3.8%)**		

*Note: ** (*) denote significance at $p < 0.01$ ($p < 0.05$). The table shows the estimates for each study converted to a standardized 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% lower COVID-19 mortality. For details on how the estimates are converted to standardized estimates see Table 19 in Appendix I. The quality-adjusted PWA is calculated as the PWA weighted by a quality index, where the score on the quality index for each study is the number of bias dimensions squared (except "peer-reviewed" and "social sciences"), where the study is of "better" quality, see Table 4.*

We now turn to the bias dimensions. Table 6 presents the results differentiated by the bias dimensions. We find no evidence of significant biases. Although the effect is generally of a larger magnitude (more negative, i.e., fewer deaths) for studies we – all else being equal – perceive as better, the difference is marginal. Only one bias dimension, social sciences vs. other sciences, shows a large difference. This is because Fuller et al. (2021) find that lockdowns reduced COVID-19 mortality rates by 35.3%, and is the only study from the non-social sciences. Fuller et al. (2021) do not exploit the panel structure of their data by using fixed or random effects, nor do they address the causality question. There are three studies which both lag the policy implementation, exploit panel estimation, and address the causality question. These three studies find that lockdowns reduced COVID-19 mortality rates by 4.9% (compared to 2.6% for the other studies).

Table 6: Estimates of the effect on COVID-19 mortality of the average lockdown in Europe and in the United States from studies based on the OxCGRT stringency index classified according to the bias dimensions

Values show effect on COVID-19 mortality	Precision-weighted average	Arithmetic average	Median
Peer-reviewed vs. working papers			
Peer-reviewed [5]	-2.4%	-5.2%	-4.0%
Working paper [3]	-5.0%	-15.1%	-7.5%
Long vs. short data period			
Data period ends after 31 May 2020 [7]	-3.5%	-9.8%	-7.5%
Data period ends before 31 May 2020 [1]	-2.4%	-2.4%	-2.4%
No early effect on mortality			
Does not find an effect within the first 14 days [1]	-4.0%	-4.0%	-4.0%
Finds effect within the first 14 days (including n/a) [7]	-2.8%	-9.6%	-7.5%
Lag vs. no lag of policy measures			
Lag policy implementation [6]	-5.6%	-13.4%	-11.9%
No lag of policy implementation [2]	1.5%	4.7%	4.7%
Panel vs. no panel estimation			
Panel estimation [5]	-3.8%	-6.2%	-4.0%
No panel estimation [3]	0.6%	-13.3%	-16.3%
Verified vs. unverified data			
Verified data [8]	-3.2%	-8.9%	-5.8%
Unverified data [0]	n/a	n/a	n/a
Address vs. do not address causality			
Address causality [4]	-3.7%	-6.8%	-5.0%
Do not address causality [4]	-2.7%	-11.0%	-10.2%
Social sciences vs. other sciences			
Social sciences [7]	-2.8%	-5.1%	-4.0%
Other sciences [1]	-35.3%	-35.3%	-35.3%

Note: The table shows the standardized estimate as described in Table 5 for each bias dimension. The number of studies in each category is in square brackets. A negative number corresponds to fewer deaths, so -5% means 5% lower COVID-19 mortality.

Overall conclusion on stringency index studies

Compared to a policy based solely on recommendations, we find little evidence that *stricter* lockdowns had a noticeable impact on COVID-19 mortality. Only one study, Fuller et al. (2021), finds a substantial effect of stricter lockdowns compared to the most lenient lockdown, while the remaining studies find little to no effect. Indeed, according to the stringency index studies, the average lockdown in Europe and the United States only reduced COVID-19 mortality by 3.2% using the precision-weighted average. The sensitivity analysis ranges from 4.4% and 3.0%, and overall, our bias dimensions do not suggest that biases are important.

We stress that this result does not imply that lockdowns do not work. It simply indicates that the most lenient lockdowns had virtually the same effect on mortality as stricter lockdowns. Since no

country did nothing, we cannot reject the thesis that some NPI would be required, e.g., to spur voluntary behavioral changes.[58]

It should also be noted that the eight stringency studies are all based on the same index (OxCGRT stringency index). Although OxCGRT is widely recognized as the best index recording the strictness of 'lockdown style' policies that restrict people's behavior and tracks and compares policy responses around the world, rigorously and consistently, we cannot rule out the possibility that the lack of evidence of the efficacy of lockdowns is caused by the limitations of the index.[59] In the following section, we will look at the effect of one of the strictest NPIs used during the COVID-19 pandemic, SIPOs, following the same structure as the current section.

4.2 Shelter-in-place order (SIPO) studies

We have identified 12 eligible studies that estimate the effect of Shelter-In-Place Orders (SIPOs) on COVID-19 mortality, see Table 7.[60] Six 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 United States issued SIPOs in the spring of 2020. 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 US-data as the marginal effect of SIPOs on top of other restrictions, although we cannot rule out 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 7. This table demonstrates that the studies generally find that SIPOs have reduced COVID-19 mortality by 2.0% (precision-weighted average) and the sensitivity analysis shows a span from 4.1% to 1.4%. The arithmetic average estimate is 7.8% and the median is 0.5%. To put these numbers into perspective, there were 188,542 registered COVID-19 deaths in Europe and 128,063 COVID-19 deaths in the United States by June 30, 2020. Thus, the reduction of 2.0% PWA (7.8% arithmetic average, 0.5% median) corresponds to approximately 4,000 (16,000, 1,000) avoided deaths in Europe and 3,000 (11,000, 1,000) avoided deaths in the United States had all countries and states

---

[58] As noted earlier, Sweden limited public gatherings to 500 people on March 12, 2020, and – at the same time as Denmark and Norway – eliminated influenza, see Figure 8. Based on this experience, limiting gatherings to 500 people could be sufficient, if some NPI is required to spur voluntary behavioral changes.

[59] Morgenstern (1963) describes in detail how such indices can be "fuzzy" metrics containing lots of errors in mismeasurement.

[60] An earlier version of our meta-analysis also included Aparicio and Grossbard (2021) and Chaudhry et al. (2020). However, these studies do not use a difference-in-difference approach and have been excluded in this version.

implemented SIPOs. In comparison, there are approximately 72,000 flu deaths in Europe and 38,000 flu deaths in the United States each year.[61]

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 by 6.0%, whereas the latter finds that SIPOs *decrease* COVID-19 mortality by 5.1%. As we will see below, this difference may – at least partly – be explained by the data period covered by each study.

Table 7: Estimates of the effect on COVID-19 mortality of shelter-in-place orders (SIPO)

Values show effect on COVID-19 mortality	Standardized estimate (Estimated Averted Deaths / Total Deaths)	Standard error	Weight (1/SE)
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
Stokes et al. (2020)	4.9%	2.8%	36
Spiegel and Tookes (2021b)	13.1%*	6.6%	15
Bonardi et al. (2020)*	0.0%	n/a	n/a
Guo et al. (2021)	4.6%	14.8%	7
An et al. (2021)	15.6%	7.8%	13
Precision-weighted average (arithmetic average / median) where SIPO is one of several variables	6.0% (3.4%/4.7%)		
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
Alderman and Harjoto (2020)	-1.0%	0.6%	169
Berry et al. (2020)*	1.1%	n/a	n/a
Fowler et al. (2021)	-35.0%**	7.0%	14
Gibson (2020)	-6.0%	24.3%	4
Dave et al. (2020)	-40.8%	36.1%	3
Precision-weighted average (arithmetic average / median) where SIPO is the only variable	-5.1% (-19.0%/-19.1%)		
Precision-weighted average (arithmetic average / median) for all studies	**-2.0% (-7.8%/-0.5%)**		
Sensitivity analysis (quality-adjusted PWA)	**-4.1% to -1.4% (-1.8%)**		

*Note: ** (*) denote significance at p < 0.01 (p < 0.05). A negative number corresponds to fewer deaths, so -5% means 5% lower COVID-19 mortality. The quality-adjusted PWA is calculated as the PWA weighted by a quality index, where the score on the quality index for each study is the number of bias dimensions squared (except "peer-reviewed" and "social sciences"), where the study is of "better" quality, see Table 4.*
* Bonardi et al. (2020) and Berry et al. (2021) do not affect the precision-weighted average, as we do not know the standard error.
** An et al. (2021) report estimates for SIPO introduced both early and late.[62] For simplicity we only report the average estimate and standard error (detailed estimates are +13% and +18%). Stokes et al. (2020) report estimates for SIPO implemented before first death and within 14 days after first death. Again, for simplicity we only report the average estimate and standard error (detailed estimates are +1.8% and +8.0%%, average SE is calculated as sqrt((SE$_1^2$+SE$_2^2$+…+SE$_n^2$)/k), where k is the number of estimates).

---

[61] The average estimated flu deaths in United States the five years prior to COVID-19 was 38,400 according to CDC (2022), and WHO (2022) writes that there are 72,000 flu deaths in Europe each year.

[62] An et al. (2021) define early mandate adoption as being taken within 14 days (the median in their dataset) after the first reported infection in each country.

Table 8 presents the results differentiated by bias dimensions. One bias dimension, long vs. short data period, shows a large potential bias driven by relatively short data periods. The four studies with relatively short data periods find a very large effect of SIPOs (25.9% reduction in mortality rates), while studies based on longer data periods find a modest increase in mortality rates of 1.0%. The last data points for the three studies which find the – by far – largest effects of SIPOs (Sears et al. (2020), Fowler et al. (2021), and Dave et al. (2021)) are April 29, May 7, and April 20 in 2020, respectively. These findings may indicate that SIPOs can delay deaths but not eliminate them completely. However, these studies were done very early in the pandemic and could not – as do the other studies – "stand on the shoulder of giants". The bias dimensions "Lag vs. no lag of policy implementation", "Panel vs. no panel estimation", and "Address vs. do not address causality" also find some potential bias, although of less magnitude.

Table 8: Estimates of the effect on COVID-19 mortality of shelter-in-place orders (SIPO) classified according to the bias dimensions

Values show effect on COVID-19 mortality	Precision-weighted average	Arithmetic average	Median
Peer-reviewed vs. working papers			
Peer-review [8]	-2.3%	-8.4%	-3.5%
Working paper [2]	-0.2%	-13.6%	0.0%
Long vs. short data period			
Data period ends after 31 May 2020 [6]	1.0%	1.5%	1.9%
Data period ends before 31 May 2020 [4]	-25.9%	-25.8%	0.0%
No early effect on mortality			
Finds effect within the first 14 days [4]	-4.3%	-16.5%	-19.1%
Does not find an effect within the first 14 days (including n/a) [6]	-1.7%	-4.8%	0.0%
Lag vs. no lag of policy measures			
Lag policy implementation [6]	3.8%	-5.2%	-0.6%
No lag of policy implementation [4]	-4.2%	-15.9%	0.0%
Panel vs. no panel estimation			
Panel estimation [8]	-6.4%	-13.3%	-11.9%
No panel estimation [2]	0.2%	6.0%	0.0%
Verified vs. unverified data			
Verified data [6]	-2.6%	-14.7%	-16.6%
Unverified data [4]	2.5%	-1.5%	0.0%
Address vs. do not address causality			
Address causality [7]	-6.7%	-16.2%	-17.7%
Do not address causality [3]	0.2%	6.4%	0.6%
Social sciences vs. other sciences			
Social sciences [10]	-2.0%	-9.4%	-3.5%
Other sciences [0]	n/a	n/a	n/a

Note: The table shows the standardized estimate as described in Table 7 for each bias dimension. The number of studies in each category is in square brackets (note that the numbers do not include Bonardi et al. (2020) and Berry et al. (2021), because they do not affect the precision-weighted average, as we do not know the standard error). A negative number corresponds to fewer deaths, so -5% means 5% lower COVID-19 mortality.

There are three studies that both use long data series, lag the policy implementation, exploit panel estimation, and address the causality question. These studies find that SIPOs *increased* COVID-19 mortality rates by 0.6% (compared to an average reduction in mortality by 2.5% in the other studies).

Overall conclusion on SIPO studies
We find that SIPOs had little to no effect on COVID-19 mortality. On average, countries in Europe and states in the United States which used SIPOs only reduced COVID-19 mortality by 2.0% (precision-weighted average). The sensitivity analysis ranges from 1.4% to 4.1%, and our bias dimensions suggest that using long data series is important and that this will further reduce the estimated effect of SIPOs, possibly making the effect of a SIPO on COVID-19 mortality positive (more deaths).

Multiple studies find a small *positive* effect of SIPOs on COVID-19 mortality. Although such a result might appear to be counterintuitive, it could, for, example be the result of an (asymptomatic) infected person being isolated at home under a SIPO who can infect family members with a higher viral load causing more severe illness.[63][64] Our result is in line with Nuzzo et al. (2019),who state that "in the context of a high-impact 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),which concludes that "forced isolation and quarantine are ineffective and impractical."[65]

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

4.3   Studies of other specific NPIs

A total of eight eligible studies examine the effect of specific NPIs.[66] The definition of these specific NPIs varies from study to study which makes comparison difficult. The variety of 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), Spiegel and Tookes (2021b) examine bar and/or restaurant closure but not business closure, and Guo et al. (2021) look at both business closures and bar/restaurant closures independently.

---

[63] See Guallar et al. (2020), who conclude, "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."

[64] One could imagine that SIPOs also affected other types of deaths, including deaths by despair, impacts of deferred diagnoses, etc. However, all studies in Table 7 examine the effect of SIPOs on COVID-19 deaths – not overall mortality – so these are not likely explanations.

[65] 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.

[66] 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.

Some studies include several NPIs (e.g. Stokes et al. (2020) and Spiegel and Tookes (2021b)), while others cover very few. For example, Leffler et al. (2020) look at internal lockdowns of any type, mask recommendations, and international travel restrictions). Too few NPIs in a model are potentially a problem because they can capture the effect of excluded NPIs.[67] On the other hand, several NPIs in a model increase the risk of multiple test bias. Also, looking at one NPI at a time may be problematic, as behavioral spillover effects may not be fully captured. For example, if we show that closing bars work because people who went to bars were more likely to be infected than people not going to bars, then this finding does not automatically imply that closing bars will have a significant impact on the overall number of infections, if people adjust their behavior according to official case numbers and are more careful when case numbers rise.

The differences in the choice of NPIs and in the number of NPIs generally make it challenging to create an overview of the results. In the following, we go through the evidence for the effectiveness of specific NPIs. First, we cover business closures, then school closures, limiting gatherings, border closures and face masks, as these NPIs are all covered by at least four studies. Last, we cover NPIs covered by one or two studies (cancellation of public events, closing public transport, and restrictions on internal movement).

Business closures
Five studies examine the effect of business closure on COVID-19 mortality.[68] Table 9 presents an overview of the estimates in these studies. Closing businesses reduced COVID-19 mortality rates by 7.5% (precision-weighted average), and an arithmetic average of 10.5% and a median of 5.5%. The sensitivity analysis shows a span from 6.6% to 9.3% reduction in mortality rates.

Three studies find little to no effect, one study (Spiegel and Tookes (2021b)) finds some effect, and another study (Chernozhukov et al. (2021)) finds a relatively large effect. It should be noted that the estimate from Chernozhukov et al. (2021) is based on their model without the national death-variable (see their Table 7). They also run a model with national deaths, where they find that business closures *increase* mortality (see their Table 9). However, since they do not calculate a counterfactual for this model, this estimate is not included in Table 9. If there is a large effect, it seems related to closing bars and restaurants. Indeed, the "close business" category in Chernozhukov et al. (2021) is an average of closed businesses, restaurants, and movie theaters. And, the "closing bars and restaurants"-submeasure (in grey in Table 9) in Spiegel and Tookes (2021b) delivers largest relative effect. Note that overall estimate of business closures for Spiegel and Tookes (2021b) is much smaller than the estimate for just its "closing bars and restaurants"-submeasure.

---

[67] Say two studies, A and B, examine the effect of lockdowns. Study A examines school closure and business closure, whereas study B examines business closure and SIPO. Then, the estimates from study A could capture the effect of the omitted variable SIPO, and the estimates from study B could capture the effect of school closures. Based on study A and B, we would report precision-weighted averages on three estimates, but since they all potentially capture the effect of omitted variables, our precision-weighted average would be biased towards larger effects.

[68] An earlier version of our meta-analysis included Bongaerts et al. (2021), which looked at business closures in Italy and thus did not meet our eligibility criteria of sufficient jurisdictional variance.

Table 9: Estimates of the effect on COVID-19 mortality of business closures

Effect on COVID-19 mortality	Description	Standardized estimate (Estimated Averted Deaths / Total Deaths)	Standard error	Weight (1/SE)
Chernozhukov et al. (2021)[*]	Measure is average of business and bar/restaurant closure	-28.6%	20.842%	5
Spiegel and Tookes (2021b)	Business closure (average)	-13.3%[*]	5.812%	17
	Bars & restaurants	*-50.2%[**]*	*8.735%*	*11*
	Bars but not restaurants	*-4.9%*	*4.208%*	*24*
	Gyms closed	*-13.8%[**]*	*4.272%*	*23*
	Spas closed	*15.9%[**]*	*4.782%*	*21*
Stokes et al. (2020)[*]	Workplace closing	-4.9%[**]	1.822%	55
	Implemented before 1st death	*-4.8%[*]*	*1.955%*	*51*
	Implemented 14 days after 1st death	*-5.0%[**]*	*1.678%*	*60*
Guo et al. (2021)	Business closure (average)	-0.4%	15.616%	6
	Business closure	*4.7%*	*13.600%*	*7*
	Bars & restaurants	*-5.5%*	*17.400%*	*6*
An et al. (2021)	Business closure (average)	-5.5%	13.062%	8
	Early business closure	*-7.5%*	*5.760%*	*17*
	Late business closure	*-3.6%*	*17.551%*	*6*
Precision-weighted average (arithmetic average / median)		**-7.5% (-10.5%/-5.5%)**		
Sensitivity analysis (quality-adjusted PWA)		**-9.3% to -6.6% (-11.9%)**		

*Note: ** (*) denote significance at p < 0.01 (p < 0.05). For studies with several estimates related to the measure, we report all submeasures in grey but focus on the average of these measures and use the average when calculating the PWA. The average standard error is calculated as $sqrt((SE_1^2 + SE_2^2 + ... + SE_n^2)/k)$, where k is the number of estimates. The quality-adjusted PWA is calculated as the PWA weighted by a quality index, where the score on the quality index for each study is the number of bias dimensions squared (except "peer-reviewed" and "social sciences"), where the study is of "better" quality, see Table 4. Values in grey are not included in the calculations of the precision-weighted average, the arithmetic average, and the median. A negative number corresponds to fewer deaths, so -5% means 5% lower COVID-19 mortality.*
[*] The estimate from Chernozhukov et al. (2021) is based on their model without the national deaths-variable (see their Table 7). They also run a model with national deaths, where they find that business closures increase mortality slightly (see their Table 9). However, since they do not calculate a counter factual for this model, this estimate is not included in the Table 9.
[**] Stokes et al. (2020) consider two distinct time periods of 24 days to account for the changing magnitude of effects of an exponentially transmitted virus: i) 0-24 days and ii) 14-38 after the first COVID-19 death. We report both of them and use the average in the precision-weighted average.

School closures

Four studies examine the effect of school closure on COVID-19 mortality.[69] Table 10 presents an overview of the estimates in the studies. Closing schools reduced COVID-19 mortality rates by 5.9% (precision-weighted average) with an arithmetic average of 0.2% and a median of 0.0%. The sensitivity analysis shows a span from 2.5% to 6.2%.

Since schools in United States were closed almost simultaneously, the estimate from Guo et al. (2021) suffers from lack of variation in school closures, but has little impact on the precision-

---

[69] An earlier version of our meta-analysis included Auger et al. (2020), which looked at school closures in the United States. However, Auger et al. (2020) represents an interrupted time series analysis and, thus, did not meet our eligibility criteria.

weighted average due to the low precision/weight.[70] Ertem et al. (2021) look at school re-openings and find a (small) increase in mortality rates following school reopenings. Assuming that the effect of closing and reopening is identical, this corresponds to a (small) decrease in mortality rates following school closures.

Table 10: Estimates of the effect on COVID-19 mortality of school closures

Effect on COVID-19 mortality	Description	Standardized estimate (Estimated Averted Deaths / Total Deaths)	Standard error	Weight (1/SE)
Stokes et al. (2020)[*]	School closures (average)	-10.9%[**]	1.567%	64
	School closing implemented before 1st death	*-2.2%*	*1.681%*	*59*
	School closing implemented after 1st death	*-19.7%[**]*	*1.443%*	*69*
Guo et al. (2021)[**]	School closures (average)	10.1%	20.662%	5
	Public school closure	*-0.2%*	*23.400%*	*4*
	All school closure	*20.4%*	*17.500%*	*6*
An et al. (2021)	School closures (average)	3.8%	7.374%	14
	Early school closure	*2.8%*	*7.590%*	*13*
	Late school closure	*4.7%*	*7.152%*	*14*
Ertem et al. (2022)[***]	Opening schools (virtual vs. traditional)	-3.7%[**]	1.982%	50
Precision-weighted average (arithmetic average / median)		**-5.9% (-0.2%/0.0%)**		
Sensitivity analysis (quality-adjusted PWA)		**-6.2% to -2.5% (-1.2%)**		

*Note: ** (*) denote significance at p < 0.01 (p < 0.05). For studies with several estimates related to the measure, we report all submeasures in grey but focus on the average of these measures and use the average when calculating the PWA. The average standard error is calculated as sqrt((SE$_1^2$ + SE$_2^2$ + … + SE$_n^2$) / k), where k is the number of estimates. Chernozhukov et al. (2021) also examine the effect of school closures but does not report a counterfactual estimate. Based on a back-of-the envelope sensitivity analysis it is unlikely that including Chernozhukov et al. (2021) would have a considerable effect on the result. The quality-adjusted PWA is calculated as the PWA weighted by a quality index, where the score on the quality index for each study is the number of bias dimensions squared (except "peer-reviewed" and "social sciences"), where the study is of "better" quality, see Table 4. Values in grey are not included in the calculations of the precision-weighted average, the arithmetic average, and the median. A negative number corresponds to fewer deaths, so -5% means 5% lower COVID-19 mortality.*
[*] Stokes et al. (2020) consider two distinct time periods of 24 days to account for the changing magnitude of effects of an exponentially transmitted virus: i) 0-24 days and ii) 14-38 after the first COVID-19 death. We report both of them and use the average in the precision-weighted average.
[**] The estimate from Guo et al. (2021) is based on school closures in United States where all 50 states closed schools between March 13, 2020, and March 23, 2020, and 44 states closed schools in just four school days (March 15, 2020 (Sunday) to March 19, 2020 (Friday)), see Table 1 in Auger et al. (2020). Hence, the difference-in-difference effect is based on a maximum of seven school days. As noted by Chernozhukov et al. (2021), this lack of cross-sectional variation can lead to considerable uncertainty over the effects of school closures, and the results from Guo et al. (2021) should be treated with this in mind.
[***] The standard error for Ertem et al. (2021) is underestimated and, thus, the weight, 1/SE, is overestimated. See the description for Ertem et al. (2021) in Table 19 in Appendix I for further details.

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

---

[70] According to Auger et al. (2020), all 50 states closed schools between March 13, 2020, and March 23, 2020, which means that all difference-in-difference is based on maximum seven school days (44 states closed schools in just four school days (March 15, 2020 (Sunday) to March 19, 2020 (Friday)), see Table 1 in Auger et al. (2020)).

of age) might facilitate transmission." UNICEF (2020) and ECDC (2020) reach similar conclusions. UNICEF (2020) 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) concludes, "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 concludes "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." [71]

Even though UNICEF (2020) and ECDC (2020) published their reviews in December 2020, there were still at least 160 countries who closed schools during 2021, according to Oxford COVID-19 Government Response Tracker.[72]

---

[71] Isphording et al. (2021) apply a quasi-experimental approach where they use the staggered timing of summer breaks across federal states in Germany to estimate the causal impact of school re-openings after the summer break in 2020. They find no evidence of a positive effect of school re-openings on case numbers. Rather, they find that the end of summer breaks had a negative but insignificant effect on the number of new confirmed cases.

[72] See Hale, Angrist, Goldszmidt, et al. (2021a).

Limiting gatherings

Four studies examine the effect of limiting gatherings. Table 11 presents an overview of the estimates in the studies. Limiting gatherings *increased* COVID-19 mortality rates by 5.9% (precision-weighted average) and the sensitivity analysis shows a span from a 4.9% to an 8.9% *increase* in mortality rates. The arithmetic average is 8.5% and the median is 7.0%, while the quality-adjusted PWA is 9.8%.

It is worth noting that no studies have estimates showing that limiting gatherings reduced COVID-19 mortality. Indeed, all four studies find positive – and sometimes rather large – effects of limiting gatherings on mortality.

Table 11: Estimates of the effect on COVID-19 mortality of limiting gatherings

Effect on COVID-19 mortality	Description	Standardized estimate (Estimated Averted Deaths / Total Deaths)	Standard error	Weight (1/SE)
Stokes et al. (2020)	Limiting gatherings (average)	3.8%**	1.307%	77
	Restrictions implemented before 1st death	*2.4%*	*1.403%*	*71*
	Restrictions implemented after 1st death	*5.2%***	*1.204%*	*83*
Spiegel and Tookes (2021b)	Limiting gatherings (average)	16.0%*	7.014%	14
	Gatherings <=10	*4.5%*	*8.034%*	*12*
	Gatherings 11-100	*19.1%***	*6.249%*	*16*
	Gatherings (limit > 100)	*24.4%***	*6.631%*	*15*
Guo et al. (2021)	Limiting gatherings (average)	5.7%	13.100%	8
	Gatherings <=10	*8.7%*	*13.050%*	*8*
	Gatherings 11-100	*2.7%*	*13.150%*	*8*
An et al. (2021)	Limiting gatherings (average)	8.4%	12.442%	8
	Early gathering limits	*1.1%*	*8.149%*	*12*
	Late gathering limits	*15.7%*	*15.595%*	*6*
Precision-weighted average (arithmetic average / median)		**5.9% (8.5%/7.0%)**		
Sensitivity analysis (quality-adjusted PWA)		**4.9% to 8.9% (9.8%)**		

*Note: ** (*) denote significance at p < 0.01 (p < 0.05). For studies with several estimates related to the measure, we report all submeasures in grey but focus on the average of these measures and use the average when calculating the PWA. The average standard error is calculated as sqrt((SE₁² + SE₂² + … + SEₙ²)/k), where k is the number of estimates. A negative number corresponds to fewer deaths, so -5% means 5% lower COVID-19 mortality. The quality-adjusted PWA is calculated as the PWA weighted by a quality index, where the score on the quality index for each study is the number of bias dimensions squared (except "peer-reviewed" and "social sciences"), where the study is of "better" quality, see Table 4. Values in grey are not included in the calculations of the precision-weighted average, the arithmetic average, and the median. Stokes et al. (2020) consider two distinct time periods of 24 days to account for the changing magnitude of an exponentially transmitted virus: i) 0-24 days and ii) 14-38 after the first COVID-19 death. We report both of them and use the average in the precision-weighted average.*

Travel restrictions

Five studies examine the effect of travel restrictions.[73] Table 12 presents an overview of the estimates in these studies. Travel restrictions reduced COVID-19 mortality rates by 3.4% (precision-weighted average) and the sensitivity analysis shows a span from 0.4% to 4.7%. The arithmetic average is an 5.3% *increase* in mortality and the median is 0.0%, while the quality-adjusted PWA shows an *increase* of 2.1%. Note that the description of the NPI varies greatly between studies and may not be comparable. This may partly explain the large span of estimates (from a *reduction* of 15.6% to an *increase* of 36.3%).

Table 12: Estimates of the effect on COVID-19 mortality of travel restrictions

Effect on COVID-19 mortality	Description	Standardized estimate (Estimated Averted Deaths / Total Deaths)	Standard error	Weight (1/SE)
Leffler et al. (2020)	International travel restrictions	-15.6%	5.737%	17
Stokes et al. (2020)	Travel restrictions (average)	-6.1%**	1.320%	76
	Travel restrictions implemented before 1st death	*-2.1%*	*1.417%*	*71*
	Travel restrictions implemented after 1st death	*-10.0%***	*1.216%*	*82*
Bonardi et al. (2020)	Border closures	0.0%	n/a	n/a
Guo et al. (2021)	Quarantine of travelers	36.3%	16.950%	6
An et al. (2021)	Travel restrictions (average)	12.1%	8.440%	12
	Early travel restrictions	*5.5%*	*5.366%*	*19*
	Late travel restrictions	*18.7%*	*10.662%*	*9*
Precision-weighted average (arithmetic average / median)		-3.4% (5.3%/0.0%)		
Sensitivity analysis (quality-adjusted PWA)		-4.7% to -0.4% (1.1%)		

*Note: ** (*) denote significance at p < 0.01 (p < 0.05). For studies with several estimates related to the measure, we report all submeasures in grey but focus on the average of these measures and use the average when calculating the PWA. The average standard error is calculated as $sqrt((SE_1^2 + SE_2^2 + … + SE_n^2)/k)$, where k is the number of estimates. A negative number corresponds to fewer deaths, so -5% means 5% lower COVID-19 mortality. The quality-adjusted PWA is calculated as the PWA weighted by a quality index, where the score on the quality index for each study is the number of bias dimensions squared (except "peer-reviewed" and "social sciences"), where the study is of "better" quality, see Table 4. Values in grey are not included in the calculations of the precision-weighted average, the arithmetic average, and the median. Stokes et al. (2020) consider two distinct time periods of 24 days to account for the changing magnitude of effects of an exponentially transmitted virus: i) 0-24 days and ii) 14-38 after the first COVID-19 death. We report both of them and use the average in the precision-weighted average.*

---

[73] An earlier version of our meta-analysis also included Toya and Skidmore (2021). However, Toya and Skidmore (2021) does not use a difference-in-difference approach and is excluded in this version.

Mask mandates

The three studies examining the effect of mask mandates – an intervention that was not widely used in the spring of 2020, and in many countries was even discouraged – on average find that masks mandates reduced COVID-19 mortality by 18.7% (precision-weighted average), see Table 13. The sensitivity analysis shows a span from 12.5% to 19.9%, and the arithmetic average and median are 18.7% and 13.5%, respectively.

The description of the NPI varies greatly between studies and may not be comparable. Chernozhukov et al. (2021) find that "employee face masks" reduces mortality by 34%, and, thus, do not – like An et al. (2021) – look at a general mask mandate. Spiegel and Tookes (2021b) look at both "employee face masks" and "mandatory face masks," finding similar effects. We do not include the estimate on "mask recommendation" from Leffler et al. (2020), as it is not a mandated NPI, but interestingly Leffler et al. (2020) find that mask recommendations reduced mortality by 23%, which is in the same order of magnitude we find for mask mandates.

Table 13: Estimates of the effect on COVID-19 mortality of mask mandates

Effect on COVID-19 mortality	Description	Standardized estimate (Estimated Averted Deaths / Total Deaths)	Standard error	Weight (1/SE)
Chernozhukov et al. (2021)	Employee face masks	-34.0%	8.511%	12
Spiegel and Tookes (2021b)	Average of below measures	-13.5%	5.131%	19
	Employee face masks	*-12.2%*	*4.272%*	*23*
	Mandatory face masks	*-14.9%*	*5.866%*	*17*
An et al. (2021)	Mask mandates	-8.5%	12.697%	8
	Early mandates	*-8.0%*	*17.246%*	*5.8*
	Late mandates	*-9.1%*	*5.000%*	*20*
Precision-weighted average (arithmetic average / median)		**-18.7% (-18.7%/-13.5%)**		
Sensitivity analysis (quality-adjusted PWA)		**-19.9% to -12.5% (-18.2%)**		

*Note: ** (*) denote significance at p < 0.01 (p < 0.05). For studies with several estimates related to the measure, we report all submeasures in grey but focus on the average of these measures and use the average when calculating the PWA. The average standard error is calculated as $sqrt((SE_1{}^2 + SE_2{}^2 + ... + SE_n{}^2)/k)$, where k is the number of estimates. A negative number corresponds to fewer deaths, so -5% means 5% lower COVID-19 mortality. The quality-adjusted PWA is calculated as the PWA weighted by a quality index, where the score on the quality index for each study is the number of bias dimensions squared (except "peer-reviewed" and "social sciences"), where the study is of "better" quality, see Table 4. Values in grey are not included in the calculations of the precision-weighted average, the arithmetic average, and the median.*

Our findings are in contrast to several other reports and studies. WHO (2019) concludes that "ten RCTs were included in meta-analysis, and there was no evidence that face masks are effective in reducing transmission of laboratory-confirmed influenza" and in their Influenza Pandemic Preparedness Strategy UK Department of Health, Social Services, and Public Safety (2011) states that "in line with the scientific evidence, the Government will not stockpile facemasks for general use in the community". Liu et al. (2021) conclude in a review 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, Jefferson et al. (2020), concludes, "There is low certainty evidence from nine trials (3,507 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; 3,005 participants)".[74]

However, it should be noted that even if no effect is found in controlled settings, this does not necessarily imply that mask mandates do 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 socialize, or masks may function as a constant reminder of the presence of the pandemic). In a cluster-randomized trial of community-level mask promotion in Bangladesh, Abaluck et al. (2022) found that the intervention (which included free masks, information on the importance of masking, role modeling by community leaders, and in-person reminders for 8 weeks) reduced symptomatic seroprevalence by 9.3%. Another possible explanation is that masks reduce the viral inoculum and that this affects mortality. For example, Bielecki et al. (2021) – in a sort of natural experiment – find that in two groups of Swiss soldiers, soldiers in the one group – those who were *not* physically distancing or wearing surgical masks before being exposed to COVID-19 – had more symptoms (47%) than the other group who *were* physically distancing and wearing surgical masks before being exposed to COVID-19.

---

[74] 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, Yanni 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%).

Other NPIs (cancelling public events, closing public transportation, restrictions on internal movement, "lockdown vs. no lockdown")

Table 14 presents the estimates for NPIs that are only covered by one study, as well as Leffler et al. (2020)'s estimate on lockdown vs. no lockdown. The estimates are all close-to-zero, but – needless to say – the uncertainty is large. We do note, however, that Leffler et al. (2020) support the other results in the meta-analysis, suggesting the effect of lockdowns of any type is limited.

Table 14: Estimates of the effect on COVID-19 mortality of other NPIs

Effect on COVID-19 mortality	Description	Standardized estimate (Estimated Averted Deaths / Total Deaths)	Standard error	Weight (1/SE)
Stokes et al. (2020)	Cancelling public events (average)	2.0%	2.165%	46
	Cancelling public events before 1st death	*-1.5%*	*2.324%*	*43*
	Cancelling public events after 1st death	*5.5%***	*1.994%*	*50*
Stokes et al. (2020)	Closing public transport (average)	0.1%	2.478%	40
	Closing public transport before 1st death	*2.0%*	*2.659%*	*38*
	Closing public transport after 1st death	*-1.7%*	*2.282%*	*44*
Stokes et al. (2020)	Restricting internal movement (average)	1.4%	2.412%	41
	Restricting internal movement before 1st death	*0.5%*	*2.588%*	*39*
	Restricting internal movement after 1st death	*2.3%*	*2.221%*	*45*
Leffler et al. (2020)	Lockdown of any type	1.7%	9.015%	11

Note: ** (*) denote significance at p < 0.01 (p < 0.05). For studies with several estimates related to the measure, we report all submeasures in grey but focus on the average of these measures and use the average when calculating the PWA. The average standard error is calculated as sqrt((SE₁² + SE₂² + … + SEₙ²)/k), where k is the number of estimates. A negative number corresponds to fewer deaths, so -5% means 5% lower COVID-19 mortality. Stokes et al. (2020) consider two distinct time periods of 24 days to account for the changing magnitude of effects of an exponentially transmitted virus: i) 0-24 days and ii) 14-38 after the first COVID-19 death. We report both of them and use the average in the precision-weighted average.

4.4 The overall effect of lockdown policies based on SIPO and specific NPIs

Overview of specific NPIs

Table 15 below summarizes our results on SIPO and the other specific NPIs. The central precision-weighted average in column 2 is small for most NPIs and even positive for limiting gatherings. Only masks mandates seem to have a notable effect on mortality rates but note that the estimate is based on just three studies (column 3). Column 4 presents the results of the sensitivity analyses. The precision-weighted averages are generally robust to the sensitivity analyses, and the quality-adjusted PWA generally find a less promising effect than the precision-weighted average (business closures is the only NPI where the quality-adjusted PWA is 'preferable' to the precision-weighted average).

Table 15: Summary of estimates of specific non-pharmaceutical interventions (NPIs)

1. NPI (number of studies)	2. Precision-weighted average (PWA)	3. Sensitivity analysis	4. Quality-adjusted PWA
SIPO (12)	-2.0%	-4.1% to -1.4%	-1.8%
Business closure (5)	-7.5%	-9.3% to -6.6%	-11.9%
School closures (4)	-5.9%	-6.2% to -2.5%	-1.2%
Limiting gatherings (4)	5.9%	4.9% to 8.9%	9.8%
Travel restrictions (5)	-3.4%	-4.7% to -0.4%	1.1%
Mask mandates (3)	-18.7%	-19.9% to -12.5%	-18.2%
Public events cancellation (1)	2.0%	n/a	2.0%
Public transportation closures (1)	0.1%	n/a	0.1%
Internal movement restrictions (1)	1.4%	n/a	1.4%

Note: The table summarizes the precision-weighted averages from Table 7, Table 9, Table 10, Table 11, Table 12, Table 13, and Table 14.

The average effect of lockdowns during spring of 2020

The overview in Table 15 allows us to estimate the effect of the average lockdown policies in the spring of 2020. First, we use OxCGRT data to calculate the share of the population which faced each of the NPIs from Table 15 in the spring of 2020. We focus on NPIs in the period between March 16 and April 15, 2020, which would have the greatest impact on deaths, since death rates flattened after this period. We only look at whether each NPI was mandated or not, and not whether it was strict or more lenient (e.g., we code both "2 - Require closing (only some levels or categories, e.g., just high school, or just public schools)" and "3 - Require closing all levels" as "closed schools"). This means that we overestimate the effect if stricter NPIs are more effective than more lenient NPIs. Also, as mentioned earlier, each precision-weighted average risk to be biased towards larger effect, since the estimate in each study may capture the effect of multiple (omitted) NPIs (also see footnote 67).

Based on this approach and with the bias towards overestimating the effect of lockdowns in mind, Table 16 presents the effect of the average lockdown in the spring of 2020. Our

calculations suggest that the average lockdown in Europe and the United States – based on estimates for specific NPIs – reduced COVID-19 mortality rates by 10.7% (precision-weighted average) with a span in the sensitivity analysis from 0.7% (worst case) to 16.0% (best case). The quality-adjusted PWA is a 3.2% reduction in mortality rates. This precision-weighted average of 10.7% is larger than the effect found in the studies based on the OxCGRT stringency index (3.2% reduction), but still relatively small and far from the large effects promised by many epidemiological models early in the pandemic, such as Ferguson et al. (2020).[75] To put the estimate in perspective, there were 188,542 registered COVID-19 deaths in Europe and 128,063 COVID-19 deaths in the United States by June 30, 2020. Thus, the 10.7% corresponds to 23,000 avoided COVID-19 deaths in Europe (best case: 26,000 avoided deaths, worst case: 1,000 avoided deaths) and 16,000 avoided COVID-19 deaths in the United States (best case: 25,000 avoided deaths, worst case: 1,000 avoided deaths). In comparison, there are approximately 72,000 flu deaths in Europe and 38,000 flu deaths in the United States each year.[76]

Table 16: Estimates of the effect on COVID-19 mortality of the average lockdown in Europe and in the United States from studies based on specific NPIs

1. NPI (number of studies)	2. Share of time with mandate (population weighted)	3. Impact on mortality (PWA · share)	4. Sensitivity analysis (best case to worst case)	5. Quality-adjusted PWA
SIPO (12)	70%	-1.4%	-2.9% to -1.0%	-1.2%
Business closure (5)	92%	-6.9%	-8.6% to -6.1%	-10.9%
School closures (4)	97%	-5.7%	-6.0% to -2.4%	-1.1%
Limiting gatherings (4)	95%	5.6%	4.7% to 8.4%	9.3%
Travel restrictions (5)	93%	-3.1%	-4.3% to -0.4%	1.0%
Mask mandates (3)	10%	-1.9%	-2.1% to -1.3%	-1.9%
Public events cancellation (1)	95%	1.9%	1.9% to 1.9%	1.9%
Public transportation closures (1)	14%	0.0%	0.0% to 0.0%	0.0%
Internal movement restrictions (1)	64%	0.9%	0.9% to 0.9%	0.9%
Total impact of the average lockdown policy		**-10.7%**	**-16.0% to -0.7%**	**-3.2%**

Note: Column 2 shows the share of the time between March 16 and April 15, 2020, where each NPI was implemented. Column 3 shows the impact of the NPI given the precision-weighted average (PWA) in Table 15 and the share of the time the NPI was implemented, see Column 2. Column 4 shows the best case (where all PWAs are in the lower end of the sensitivity analysis) and the worst case (where all PWAs are in the upper end of the sensitivity analysis), and column 5 shows the quality-adjusted PWA. The total impact of the average lockdown policy is calculated as the product of (1 – (estimates in column 3)) - 1. Implementing all NPIs has an impact of -26.4%.

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

---

[75] Note that the stringency index does not include face mask mandates.

[76] The average estimated flu deaths in the United States the five years prior to COVID-19 was 38,400 according to the CDC (2022), and the WHO (2022) writes that there are 72,000 flu deaths in Europe each year.

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 1,220 studies, 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 1,220 studies, 32 met our eligibility criteria, and standardized estimates for our meta-analysis could be calculated for 22 of the eligible studies.

5.1   Conclusions

Overall, our meta-analysis fails to confirm the notion that lockdowns – at least in the spring of 2020 – 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 3.2% compared to the most lenient COVID-19 policy. Shelter-in-place orders (SIPOs) were also ineffective. They only reduced COVID-19 mortality by 2.0%. Based on nine specific NPIs, we estimate that the average lockdown in Europe and the United States in the spring of 2020 reduced mortality by 10.7%. The 3.2% to 10.7% correspond 6,000-23,000 avoided deaths in Europe and 4,000-16,000 avoided deaths in the United States. In comparison, there are approximately 72,000 flu deaths in Europe and 38,000 flu deaths in the United States each year.[77] Thus, lockdowns in Europe and the United States on average saved lives corresponding to 9%-35% of an average flu season.

Of the specific NPIs, we find that mask mandates had the largest effect (reducing COVID-19 mortality by 18.7%), but the estimate is based on just three studies with heterogeneity in the definition of the mandate. Limiting gatherings were counterproductive and *increased* mortality by 5.9%.

Our measured meta-results are supported by the natural experiments we have been able to identify through our work and by searches in the abstract and citation database Scopus (see Table 17).

Overall, our meta-analysis supports the conclusion that lockdowns – at least in the spring of 2020 – had little to no effect on COVID-19 mortality.

Throughout the meta-analysis, we have focused on the precision-weighted average as our primary indicator of the efficacy of lockdowns. However, as shown Figure 10, the overall conclusion holds regardless of which studies or measures one chooses to emphasize. The figure presents the effect on mortality in the United States based on the measured estimates from all stringency studies as well as our two central measured estimates for the effect of lockdowns in the spring of 2020 (the precision-weighted average from the stringency studies in Table 5 and the

---

[77] The average estimated flu deaths in the United States the five years prior to COVID-19 was 38,400 according to the CDC (2022), and the WHO (2022) writes that there are 72,000 flu deaths in Europe each year.

estimate based on specific NPIs in Table 16). We have added the max and min forecasted estimates from Ferguson et al. (2020) for comparison. Even if we cherry-pick the most preferable empirical estimate from Fuller et al. (2021), the effect on mortality is far from the effects promised based on Ferguson et al. (2020) and corresponds to the mortality from less than two influenza seasons.[78]

Figure 10: Divergence between avoided number of deaths in the United States as measured by the meta-results, studies based on the OxCGRT stringency index, and the forecasted outcome from Imperial College London
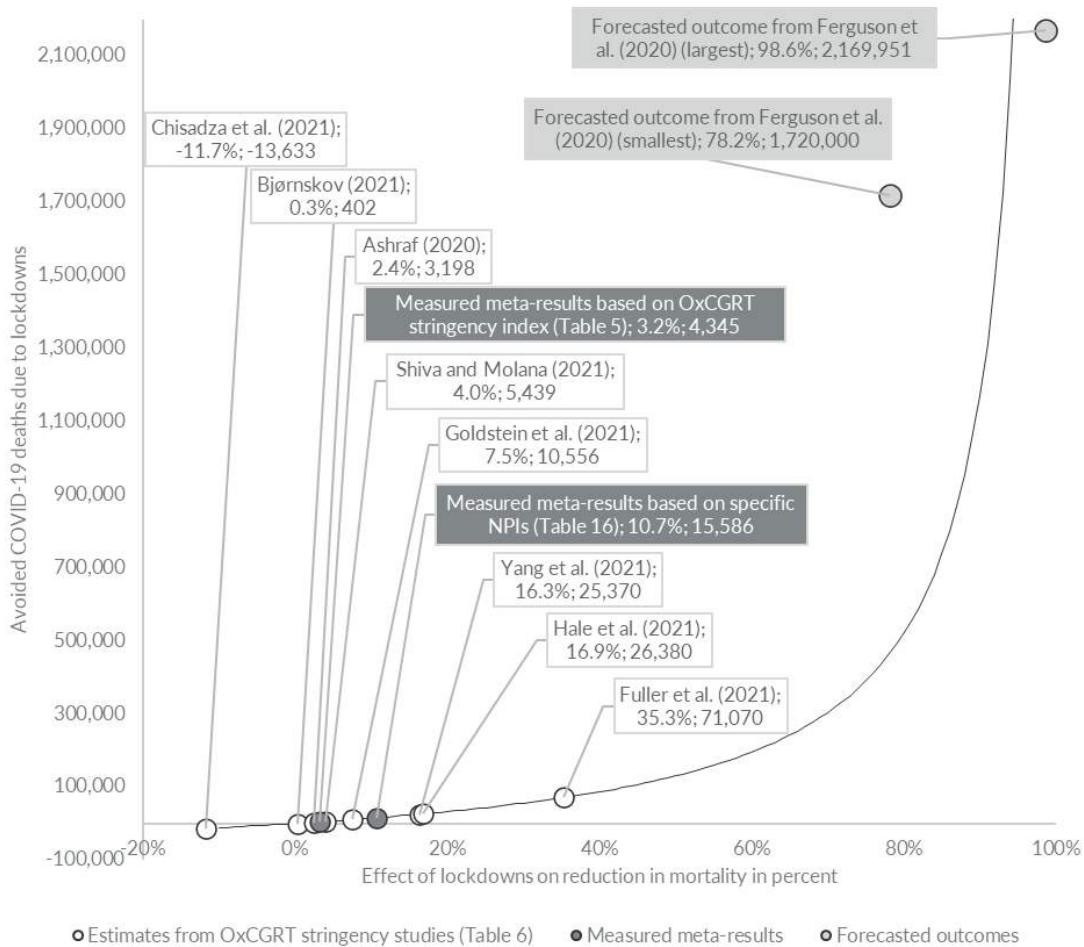
Note: The estimates in the group "Estimates from stringency studies" are from Table 5. For the group "Meta-study PWA" the estimate for "PWA stringency studies" is from Table 5, and "average lockdown spring 2020" is from Table 16. Both estimates illustrate the effect of the average lockdown in Europe and United States in the spring of 2020. The effect of lockdowns on total mortality based on the meta-study's precision-weighted averages (PWA) is calculated as total COVID-19 deaths by July 1, 2020 (128,063 COVID-19 deaths) x (1/(1-PWA)-1). The relative effect of lockdowns on total mortality based on Ferguson et al. (2020) is calculated as the largest and smallest predicted relative effect multiplied with their mortality estimate of 2.2 million deaths in a "do nothing"-scenario in United States. The estimates from Ferguson et al. (2020) is for a two year period, but the relative effect is largest early in the pandemic.

---

[78] The five years prior to COVID-19, the flu caused 38,400 deaths annually on average according to CDC (2022).

Limitations

Our study has a number of limitations. Most importantly, we cannot say whether NPIs are crucial as a means of signaling the danger of the pandemic. It is possible that – even if stricter lockdowns and each individual NPI are ineffective –the government 'doing something' is necessary to spur voluntary behavioral changes. The question is whether people had understood how serious the situation was if governments had not resorted to measures that went beyond the usual. If the answer to this question is that some sort of lockdown was indeed necessary (and there is some evidence suggesting this)[79], the next obvious question is: How little intervention does it take to send the necessary signal? Our results suggest that the answer to this question is "relatively little", but how to send the best signal with as little cost to society as possible, is an obvious area for future research.

Another limitation is the limited number of studies. This is especially true for the specific NPIs where our estimates are often based on 3-5 studies. We hope that our work will inspire more researchers to think about ways to examine the effect of specific NPIs so future meta-analyses have more studies to rely on.

Also, we do not analyze the role of timing. As pointed out in the section 2.2, p. 16, and in section 5.2.4, p. 71, we believe that many studies examining the role of timing are fundamentally flawed because they do not distinguish between voluntary behavioral changes and lockdowns (i.e. mandatory behavioral changes). We also point out that even if there *is* an optimal timing, we cannot be sure that democratic governments will ever be able to react in time on this information and implement the NPIs accordingly. However, these are problems that may be addressed by future research. In any case researchers should find ways to distinguish between the effects of voluntary behavioral changes (possibly spurred by signaling) and those of lockdowns.

Our meta-analysis shows that some NPIs may have a measurable effect on mortality. Mask mandates especially look relatively effective, although our estimate is based on just three studies and the 18.7% is still a relatively low effect compared to the effects promised by many epidemiological models early in the pandemic.[80] However, that an NPI has a measurable effect on mortality is a necessary – but not sufficient – requirement to make the policy beneficial and desirable. Also, there is some evidence that mask recommendations can be sufficient to reap much of the effect of mask mandates. Thus, future research is needed to estimate the broader costs of mask mandates – including effects on welfare, trust etc. – before one can conduct an actual cost-benefit analysis, which can answer whether mask mandates are a desirable policy.

---

[79] For example, using survey data immediately (same day) before and after Boris Johnson announced the UK lockdown, Eggers and Harding (2021) find that "the lockdown announcement made people more supportive of the government's response to the crisis but also (perhaps surprisingly) more concerned about the pandemic". E.g., people responding after the announcement were more likely to respond "I fear for my future" and "I have started not going out at all".

[80] Some have argued that mask mandates may be efficient regulation during future influenza seasons. However, an effect of 18.7% will only reduce the number of deaths by approximately 25.000 in Europe and the United States combined during an average flu season. Although this *is* a large benefit, it should be compared to the economic cost of mandating more than 1 billion people to wear facemasks.

We only look at mortality. It is possible that there are other benefits related to lockdowns which are not captured in the studies looking at mortality rates. For example, Banholzer et al. (2022) believe that "interventions that reduce the number of new infections can have downstream effects on various outcomes, including disease-related deaths, cases of severe illness and hospitalizations, cases with long-term health effects after infection, the efficiency of testing and contact tracing" etc. While this may be true, we believe it is unlikely that the effect of lockdowns on infections has been so different from the effect of lockdowns on mortality that it changes the overall conclusion. Even if the effect on infections is two or three times larger than the effect on deaths, the overall effect is limited and far from the effect promised based on model studies, see Figure 10. However, only future research can tell whether this immediate assessment holds true.

We also restrict our search strategy to studies using a "counterfactual difference-in-difference approach". We believe difference-in-difference studies are better suited than other widely used empirical methods to examine the true effect of lockdowns because they allow us to leave out the effect of voluntary behavior changes. There is, however, no doubt that the results from other study methods are of great interest because they can give us insights with regard to e.g., the importance of voluntary behavior changes. We welcome future research on these other methodologies.

5.2   Discussion

5.2.1   Conclusions are in line with other reviews

Overall, we conclude that stricter 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), stating, "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 the conclusion in Allen (2021): "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[81], suggesting that the impact of lockdowns is

---

[81] We are uncertain if these numbers are correct, as Poeschl and Larsen (2021) list only one study examining bar closures in their overview table, although they review two studies. Also, Poeschl and Larsen (2021) do not look at business closures, and at least one of their studies examine this. Weber (2020) writes that "In an estimation without the non-positivity constraints, the sum of all sector closure effects is insignificant at the one percent level".

limited and not that far from zero, which contradicts their conclusion.[82] Based on two interrupted time series studies, Iezadi et al. (2021) find that overall NPIs reduced daily ICU admissions by 16.5%. Mendez-Brito et al. (2021) find that school closing is the most effective measure, although only 14 out of 24 studies (58%) found an association between school closures and number of cases, suggesting a limited effect.[83] Herby (2021a) concludes that "mandated behavior changes accounts for only 9% (median: 0%) of the total effect on the growth of the pandemic stemming from behavioral changes. The remaining 91% (median: 100%) of the effect was due to voluntary behavior changes."

The findings contained in Johanna et al. (2020) are in contrast to our results. 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. (2020) 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.

5.2.2   Causality or correlation?

As pointed out by Bjørnskov (2021a), there is a potential endogeneity problem (also referred to as reverse causality) "which derives from the nature of political reactions to the virus that could rely on the reported number of infections. If an increase in the reported infection rate leads government to introduce lockdown policies, and if a declining reported infection rate subsequently leads them to ease the lockdown, the estimated association between policy stringency and mortality is biased."

Several studies explicitly claim that they examine the actual causal relationship between lockdowns and COVID-19 mortality. Some studies use instrumental variables[84], lagged dependents[85], or other techniques to establish a causal relationship, while others make causality probable using arguments.[86] Sebhatu et al. (2020) show that government policies are strongly

---

[82] If the true estimate was far from zero, we would expect to see relatively few estimates which are zero or positive (more deaths). If, on the other hand, the true value is around 0, we would expect to see that approximately half of the "guesses" are greater than zero, while half are lower than zero. Such an outcome is illustrated in the figure below.

[83] Talic et al. (2021) is another systematic review and meta-analysis which looks at the effectiveness of public health measures in reducing the incidence of COVID-19. However, their focus is on voluntary measures. They state that "the findings of this review suggest that personal and social measures, including handwashing, mask wearing, and physical distancing are effective at reducing the incidence of covid-19. More stringent measures, such as lockdowns and closures of borders, schools, and workplaces need to be carefully assessed by weighing the potential negative effects of these measures on general populations Further research is needed to assess the effectiveness of public health measures after adequate vaccination coverage".

[84] E.g., Bjørnskov (2021a).

[85] E.g., P. Goldstein et al. (2021).

[86] 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."

driven by the policies initiated in neighboring countries rather than by the severity of the pandemic in their own countries.

In short, Sebhatu et al. (2020) show that 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. Sebhatu et al. (2020) also find that the death rate predicts the stringency of countries' policy adoptions, but the effect is small explaining only 2.1 stringency points on average (in comparison, the gap between the strictest and most lenient lockdown in Europe was between 67 to 92 stringency points in the period from March 16 to April 15, 2020).[87] The very low mortality rates on the day of lockdown (defined as the day when the OxCGRT stringency index exceeds 60) is illustrated in Figure 11.

Note that decisions were often taken days in advance of implementation (e.g. Denmark announced the lockdown March 11, 2020, but schools closed 5 days later (March 16) and businesses seven days later (March 18)), where mortality rates were even lower (and knowledge even more sparse) and the reporting of deaths was in some cases delayed. This lag between decision and implementation is non-negligible. On average, mortality rates were about 50% lower three days prior to lockdown and 75% lower five days prior to lockdown.

---

[87] Sebhatu et al. (2020) estimate that the impact of "death rate (/100,000)" on stringency is 11.706 (Table S3) and highly significant. On average, the death rate (/100,000) was 0.18 (Table S4) meaning that the average explanatory power was 2.1 stringency points. According to OxCGRT, the average death rate (/100,000) in Europe was approximately 0.08 and 0.19 when reaching stringency 60 and 70, respectively. Only San Marino had a substantial death rate and – given the estimate from Sebhatu et al. (2020) – a substantial impact on stringency, when reaching stringency 60 (and Netherlands and Spain when reaching stringency 70). In San Marino, Netherlands, and Spain the total impact measured as death rate estimate was 34, 5, and 2 respectively when reaching stringency index 60 and 172, 15, and 13 respectively when reaching stringency index 60. No other country had an impact above 5 points.

Figure 11: Mortality rates in European countries were very low prior to lockdown-decisions
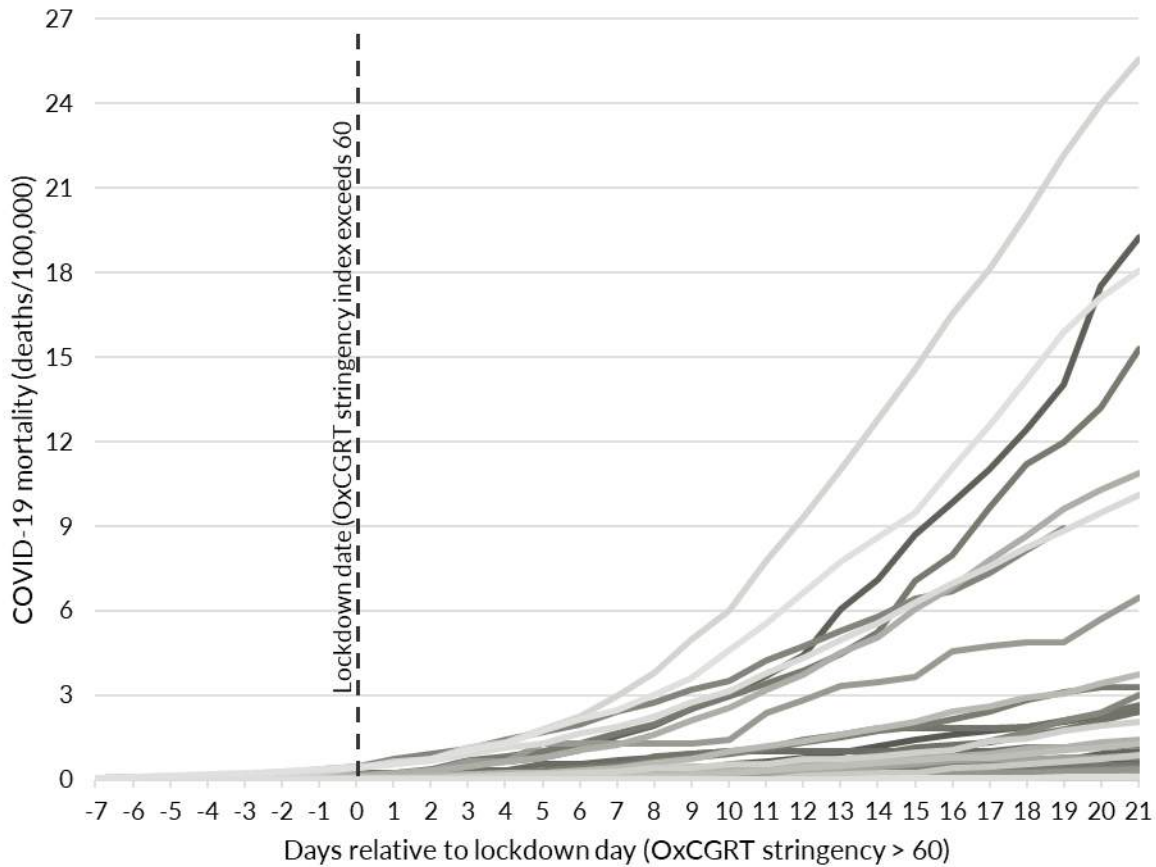
Source: Our World in Data (2022).
Note: The figure shows the development in mortality in 37 European countries with more than 100,000 inhabitants. Day 0 is the day where the OxCGRT stringency index crosses 60

Also, most eligible studies examine the first wave, where most countries – due to limited testing capabilities – had little information about the progress of the pandemic, and thus, the policy response in a given country is unlikely to be greatly caused by the severity of the pandemic. Bjørnskov (2021a) points to governments' ability to react quickly and note that "although one might think that policy making reacts quickly to changing mortality during an emergency, exploring the determinants of changes in the stringency indices reveals that an increase in the contemporaneous mortality or an increase in the reported number of Sars-CoV-2 cases was not associated with stricter lockdown measures" and conclude that "it is highly unlikely that there is a substantial endogeneity problem in the following as mortality changes only affect policy changes with a three-week lag, and as policy changes cannot affect the mortality rate before another two to three weeks have passed. As such, any bias is likely to be small and practically negligible." Finally, 11 of the 22 studies in the meta-analysis handles the causality question, but their results are not much different from the other 11 studies, implying that causality is not a major problem.

Hence, we believe there is a strong case for a causal relationship in our results and that what the studies examine is the effect (of the strictness) of lockdowns on mortality and not the opposite (mortality rates' effect on (the strictness of) lockdowns, although the issue can never be finally settled with observational studies.[88]

5.2.3 Why are the effects of lockdowns limited?

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 (e.g., Ferguson et al. (2020)). But it is evident that modelers around the globe failed to accurately forecast the development of the pandemic.

One example is the projections of COVID-19 inpatients that were published during the Danish negotiations of a reopening in the spring of 2021. Figure 12 below compares the projections to actual data following the reopening. Not only did the modelers fail to project the number of COVID-19 inpatients following the reopening of the economy, the actual outcome was below even the most optimistic lockdown scenario (the lower bound of the grey shaded area). It should be noted that this forecasting failure was in no way unique to the Danish health authorities. Most health authorities and expert modelers failed to correctly project the development of the pandemic.[89]

---

[88] Note that the "RCT like" studies, we have identified, support our conclusions based on observational studies. See Table 17.

[89] See, for example, https://www.spectator.co.uk/article/how-did-sage-scenarios-compare-to-reality-an-update and https://www.svd.se/hundratusen-skulle-do--modellen-slog-fel.

Figure 12: Model forecasts of COVID-19 inpatients with and without reopening compared to actual outcomes (Denmark, 2021)
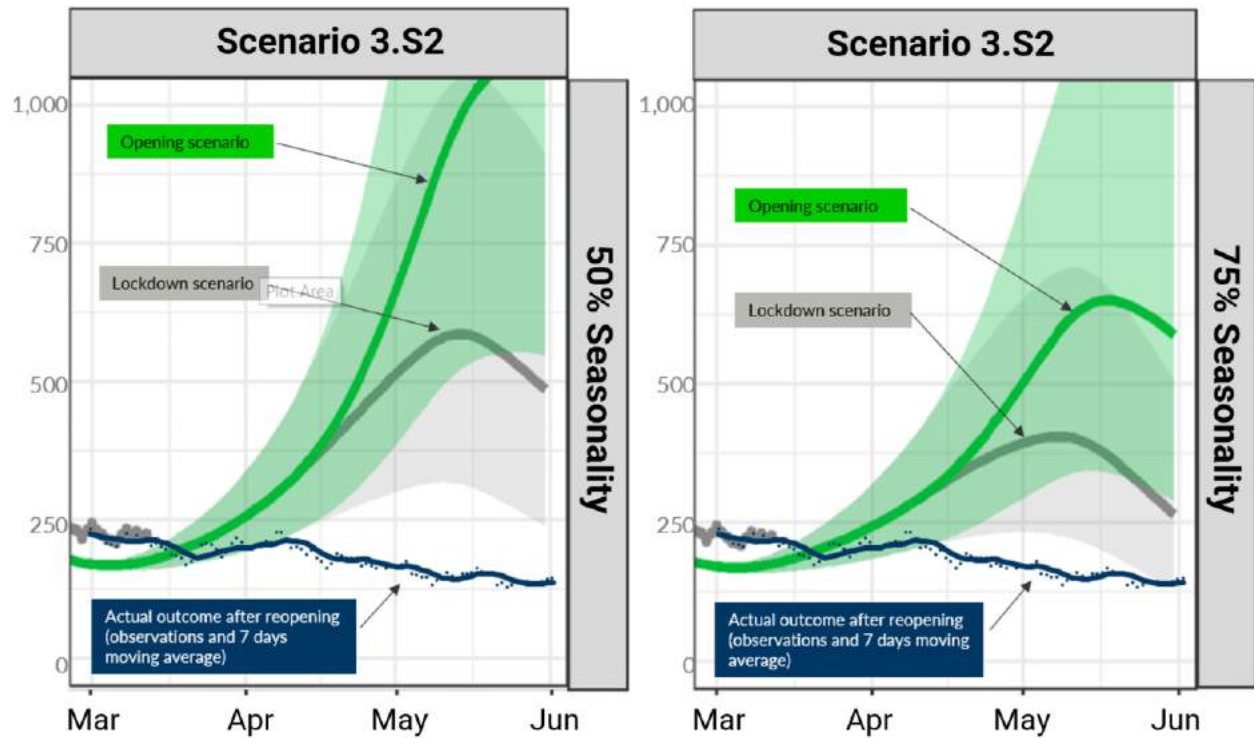
Source: Statens Serum Institut (2021) and Danmarks Statistik (2022).
Note: The green line shows the projections if the economy was reopened, while the grey line shows the projections with continued lockdown. The blue dots and line show the actual development after the economy was reopened. "50% sæson" and "75% sæson" denotes how much seasonal variation is included. Text has been translated from Danish to English.

We propose four factors that might explain the difference between our conclusions and the views embraced by some epidemiologists.

People respond voluntarily to dangers

First, people respond to dangers outside their door when they are aware of it. When a pandemic rages, people engage in social distancing regardless of what the government mandates. 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.[90] On the 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 (2021a) reviews studies that 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. Andersen et al. (2020) find that consumer spending fell almost as much in Sweden as in Denmark despite Sweden having a very

---

[90] In a randomized control trial, Helsingen, Løberg, et al. (2020) find that "provided good hygiene and social distancing measures, there was no increased COVID-19 spread at training facilities." Their study shows that many activities can be safe with focus on hygiene and social distancing.

limited lockdown.[91] Interestingly, the response in Sweden was especially large among 70+year-olds and even larger than in Denmark – possibly due to the larger outbreak in Sweden. Chetty et al. (2020) show that high-income individuals reduced spending sharply in mid-March 2020, particularly in areas with high rates of COVID-19 infection and in sectors that require physical interaction. By comparing counties with and without restrictions, Goolsbee and Syverson (2021) conclude that only 7%-points of the 60%-point decline in business activity could be attributed to legal restrictions, and that the shift was highly tied to the number of COVID deaths in the county. Most of the decline resulted from consumers voluntarily choosing to avoid stores and restaurants. The point from Andersen et al. (2020), Chetty et al. (2020) and Goolsbee and Syverson (2021) is illustrated in below figure from Maas (2020).

Figure 13: Lockdown-policy differences and consumer activity in Iowa and Illinois
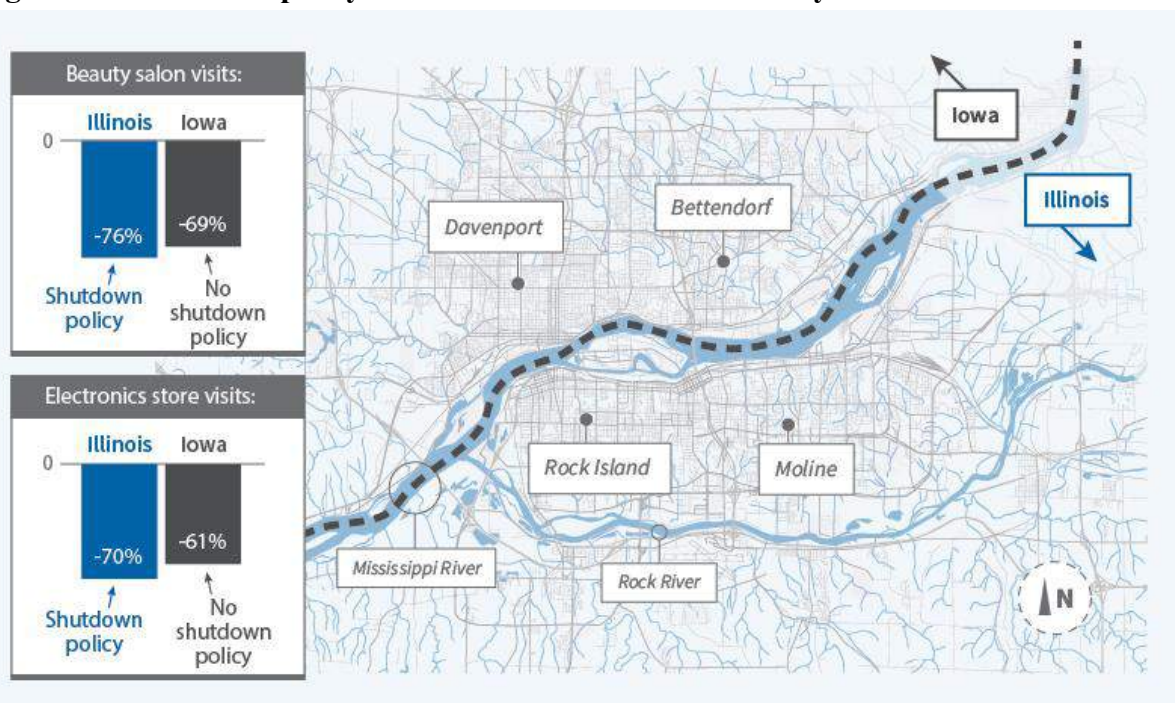
Source: Maas (2020) based on data from SafeGraph.

Gupta et al. (2020) find that "information-based policies and events such as first cases had the largest effects". They also find that SIPOs did not affect social distancing (as measured by a mixing index which – based on cell phone data – measured the exposure of a smart device to other devices). At a first glance this seems to be in conflict with the findings by Joshi and Musalem (2021) who find that SIPOs increases time spent at home, but one obvious explanation is that people stop mixing voluntarily before they are mandated to stay at home, so the SIPOs do not increase social distancing (and thus, reduce infections) but only increase time spent at home. After all, there are several ways you can leave your house without mixing with others. Bor et al.

---

[91] Andersen et al. (2020) analyze transaction data for Denmark and Sweden from a large bank in Scandinavia to reach this conclusion.

(2021) find that intrinsic motivations related to the severity of the pandemic (as measured by national case numbers) play a significant role when citizens increase their attention to the health authorities' advice during an epidemic.

These voluntary behavioral changes may also explain why epidemiological model simulations such as Ferguson et al. (2020) – which do not model behavior endogenously[92] – fail to forecast the effect of lockdowns. Voluntary behavioral changes also explain why the flu almost disappeared in Denmark in March 2020 before a single restriction was implemented. In Denmark, schools closed by March 16th and businesses by March 18th. But at these dates the share of positive influenza tests in Denmark had dropped from about 25% – a level which had been more or less constant for two months – before March 11th, when the Danish prime minister held a press conference, to 5%-10% (Statens Serum Institut (2020)). As we showed in Figure 8, p. 20, the same pattern was seen in Norway and Sweden.

Similar drops in the share of positive influenza tests were seen in Norway and Sweden (Emborg et al. (2021), also see Figure 8 on p. 72). In the United States, Ziedan et al. (2020) find that "aggregate trends in outpatient visits show a 40% decline after the first week of March 2020, only a portion of which is attributed to state policy." Tsai and Tzu-Ting (2021) find similar results for Taiwan. Overall, we believe that Allen (2021) is correct, when he concludes, "The ineffectiveness [of lockdowns] stemmed from individual changes in behavior: either non-compliance or behavior that mimicked lockdowns."

---

[92] In fact, Ferguson et al. (2020) describe their results as "unlikely", as they are based on the assumption of the "absence of any […] spontaneous changes in individual behavior".

Mandates only regulate a fraction of our potential contagious contacts
Second, mandates only regulate a fraction of our potential contagious contacts. Figure 14 illustrates infection locations in Germany during the early pandemic. The figure shows that most of the infections in Germany assigned to an outbreak (defined as at least two cases) occurred in homes (including homes for the elderly), hospitals, and workplaces that were not subject to general restrictions applied throughout the society and where potentially effective interventions, such as handwashing, coughing etiquette, ventilation, distancing etc. could neither be regulated nor enforced but relied solely on voluntary behavioral changes. In total, 77% of infections occurred in homes, hospitals, and workplaces, and the share was – despite the first weeks of the pandemic when awareness was low – large (above 60%) despite variations in the use of NPIs.

Figure 14: Homes, hospitals and workplaces were the main drivers of infections in Germany and the location for 77% of all infections
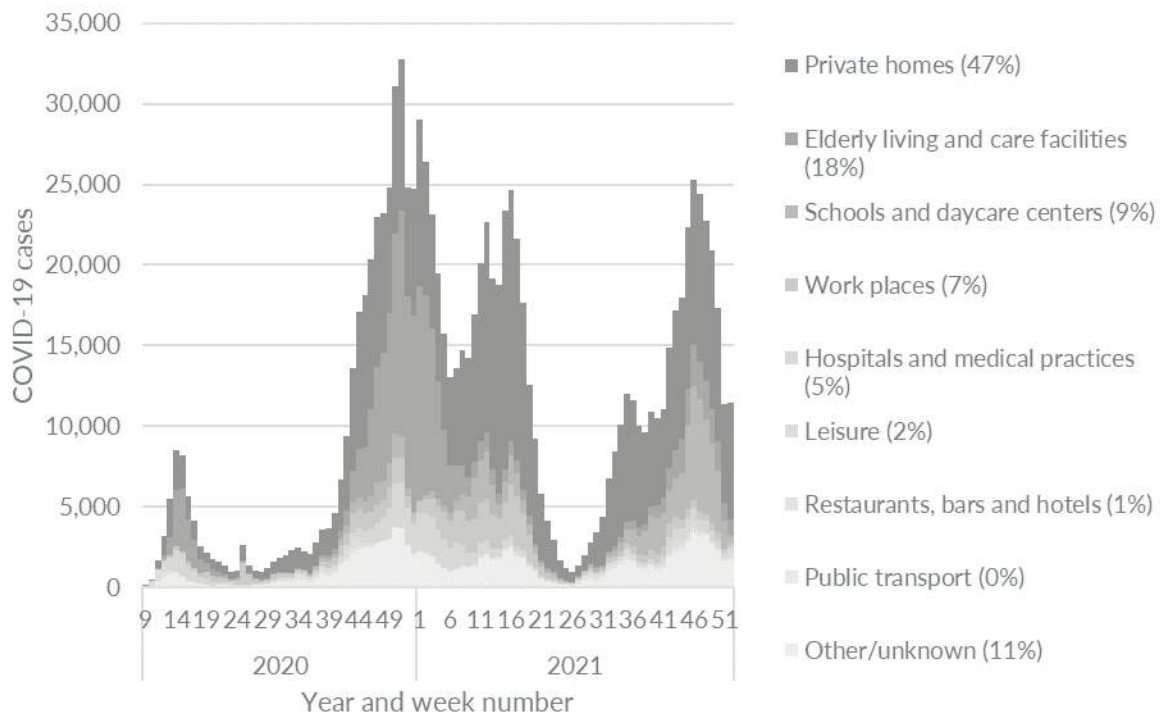
Source: Robert Koch Institut (2022)
Note: Laboratory-confirmed COVID-19 cases assigned to an outbreak by infection setting and reporting week. Number in parentheses is the share of total cases during the covered period. The data covers COVID-19 outbreaks with two or more cases which includes about 15% of all cases of infection in Germany.

The data in Figure 14 only covers COVID-19 outbreaks with two or more cases, which includes about 15% of all cases of infection in Germany. But data from other countries show similar patterns. Lee et al. (2020) write that "early contact tracing studies and a large study of more than 59,000 case contacts in South Korea found household contacts to be greater than six times more likely to be infected with SARS-CoV-2 than other close contacts" and Zhao et al. (2020) find that "69.2% of total cases were clustered in a home, apartment or residential estate."

In a Danish matched case-control study based on data from November 2020, Munch et al. (2022) find that contact with an infected person at home or at work was a substantial risk factor.[93] They find no infection risk associated with community exposures such as shopping at supermarkets, travel by public transport, dining at restaurants, and private social events with few participants. The results were confirmed by Lendorf (2021) who found that 69% were infected at places with no general restrictions.[94] Both Munch et al. (2022) and Lendorf (2021) are based on data from a period with relatively few infections and some restrictions (gatherings limited to 10 persons, restaurants and bars etc. had to close at 10 pm, and face masks were mandatory indoors, except when seated). Nevertheless, their results illustrate that people in countries like Denmark, Finland, and Norway – countries that allowed people to go to work, use public transport, and meet privately at home during the pandemic – had ample opportunities to legally meet with – and get infected by – others. Still, these countries experienced relatively low COVID-19 mortality rates.

Behavioral responses may counteract any initial effect of lockdowns
Third, even if lockdowns are successful in initially reducing the spread of COVID-19, the behavioral responses may counteract the effect 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.[95] As pointed out by Deaton and Cartwright (2018), randomization "does not relieve us of the need to think about (observed or unobserved) covariates". Also, this kind of second-order behavior response may also explain why closing down non-essential businesses simply reallocates consumer visits away from "nonessential" to "essential" businesses, as shown by Goolsbee and Syverson (2021), with limited impact on the total number of contacts.[96]And this probable behavior response to changes in infection levels limits the knowledge we can obtain from randomized control trials examining specific NPIs (if for example, masking children in schools reduces the infections among children and teachers, this does not necessarily imply that masking children reduces infection rates overall).

Also, Joshi and Musalem (2021) find that the effect of SIPOs on mobility decreases as time passes and infection rates drop.

Some NPIs may have led to unintended consequences
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

---

[93] Munch et al. (2022) write "contact (OR 4.9, 95% CI 2.4–10) and close contact (OR 13, 95% CI 6.7–25) with a person with a known SARS-CoV-2 infection were main determinants. Contact most often took place in the household or workplace".

[94] 23% were infected in the household, 27% were infected at work, and 19% were infected by close acquaintances.

[95] 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.

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

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 evidence that limiting gatherings was counterproductive and increased COVID-19 mortality by 5.9%, see Table 11.

5.2.4 Objections to the results of the meta-analysis

Our results and conclusions go against the conventional wisdom that lockdowns were effective in reducing COVID-19 mortality, and, indeed, the first version of our study did spawn a wide range of objections to our measured meta-results. We address the most important of these in this section.

The "timing of lockdowns is crucial"-objection

One objection to our conclusions is 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 first note that this objection does not necessarily contradict 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 when compared to doing very little (see section 3.1 p. 23 for further discussion). However, there is little solid evidence supporting the timing thesis, because it is inherently difficult to analyze (see Section 2.2 p. 16 for further discussion).

In Figure 7 we showed that all countries and states that were hit late by the pandemic experienced low COVID-19 mortality rates. This pattern – where areas hit late in the pandemic also have lower deaths rates – was also found during the Spanish Flu in 1918. Figure 15 below shows how cities in the United States that were hit early by the Spanish Flu in the autumn of 1918 also experienced large excess mortality. The data is from Hatchett et al. (2007) and Markel et al. (2007), who both conclude that the low excess mortality was due to lockdowns early in the pandemic. But, as the figure clearly shows, cities which implemented lockdowns early in the pandemic were also hit relatively late compared to other cities making it difficult to assess whether the measured effect of early lockdowns is related to lockdowns or is related to voluntary behavior changes instead.

Hatchett et al. (2007) touch on the subject by noting that "in addition, cities whose epidemics began later tended to intervene at an earlier stage of their epidemics, presumably because local officials in these cities observed the effects of the epidemic along the Eastern seaboard and resolved to act quickly". But if local officials observed the effects on the east coast, many ordinary people probably did too, spurring – or at least laying the groundwork for – voluntary behavioral changes. Indeed, if we exclude cities which were hit early, the average excess

mortality in Markel et al. (2007) is similar across response times indicating that information and voluntary behavioral changes could be driving their results.[97]

Figure 15: **Cities hit late by the Spanish Flu in 1918 experienced lower excess mortality**

Panel A: Relationship between early pandemic strength, response times, and total 1st wave mortality - data from Hatchett et al. (2007)

Panel B: Relationship between early pandemic strength, response times, and excess mortality – data from Markel et al. (2007)
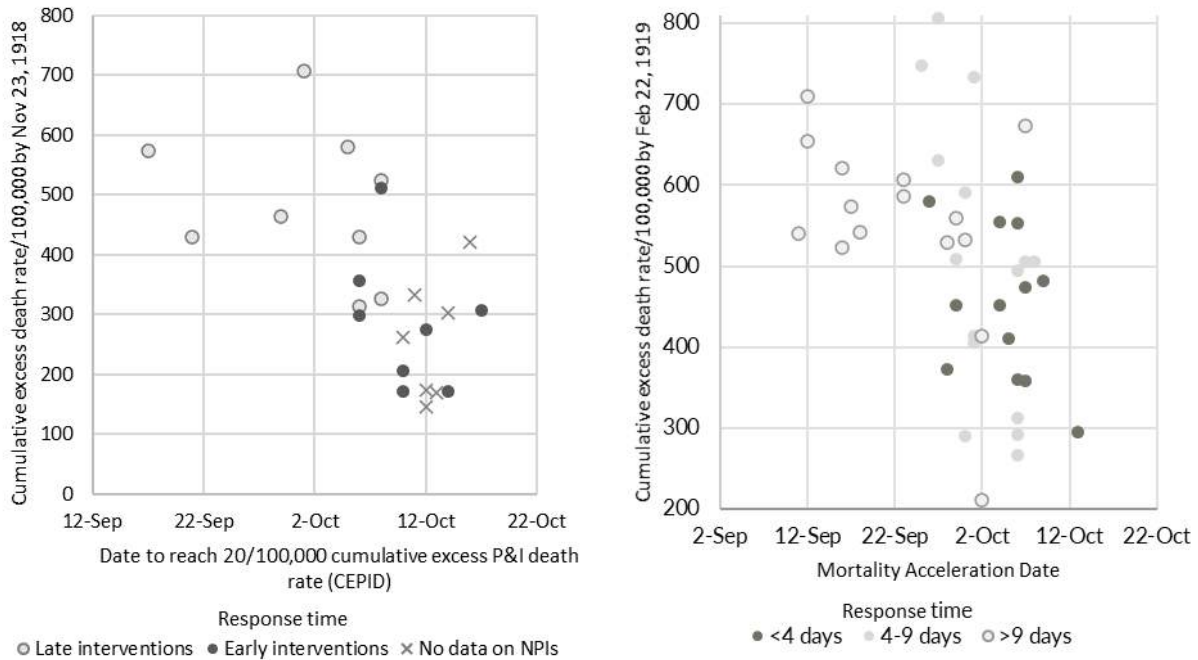
Source: Hatchett et al. (2007) and Markel et al. (2007)
Note: The figure illustrates that cities that were hit late by the 1918 Spanish Flu pandemic generally experienced lower excess death rates despite response times.

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. If lockdowns are effective as long as they are well timed, our results – which show the average effect of lockdowns in Europe and the United States – show that governments, on average. were unable to time lockdowns properly to obtain a substantial effect on mortality.[98] The problem of

---

[97] For cities with a response time <4 days, the average excess morality (unweighted) is 458 compared to 500 (4-9 days) and 486 (>9 days) if we only include cities with a "mortality acceleration date" after September 24, 1918. Choosing later cut-off dates does not change the picture. For Hatchett et al. (2007), this comparison is more difficult, because there is little overlap between early and late intervention cities.

[98] As an example of how difficult timing is, Herby (2021b) and Herby (2021c) show how the Danish authorities – responding to local outbreaks in the fall of 2021 – in several occasions closed schools *after* case numbers had regressed to the level before the outbreak. The figures below illustrate this fact. The black line is test-corrected case numbers, the red vertical line is the day the schools closed, and the pink vertical line is the earliest day the effect of the school closure can be seen in cases (in Denmark in the fall of 2021, people were advised to wait four

proper timing is well known from the debate about discretionary economic stabilization policies. Selecting the proper timing for such measures have proved a disappointment.[99] Thus, discretionary approaches have largely been abandoned for rule-based stabilization policies.

*The "your results do not apply to later waves and variants"-objection*
Another objection has been that the studies included in the meta-analysis cover, for the most part, the first wave, and that one can imagine that lockdowns may be effective against later waves or variants. This objection is a hypothesis, and only future research can show if this hypothesis is true or false. We note that even if future research falsifies this hypothesis, yet another hypothesis can then be proposed: that – again – the historical evidence is not relevant for future waves and variants. In the end, it will be a political discussion how heavy such speculations should weigh when making decisions, when historical data find limited effects of lockdowns.

The "there are too few studies to know for sure"-objection
Several commentators have pointed out that our conclusion is based on relatively few studies. While more studies are always better, we have included all existing empirical evidence and covered far more studies than needed to e.g., bring a new drug to the market.[100] It is worth noting that optimal sample sizes can be and are surprisingly small in number (see Hanke and Mehrez (1979)). Communicable diseases can be handled using pharmaceutical interventions and/or non-pharmaceutical interventions. From a scientific and political perspective, the evidence required to implement either of these interventions should not differ. Hence, it is inconsistent to have a political regime where pharmaceutical interventions *may* only be used if one can prove they are *effective* and that negative side effects are *small*, while *non*-pharmaceutical interventions *will* be used unless one can prove they are *ineffective* and that negative side effects are *large*.

days between close contact to an infected person and being tested).

---
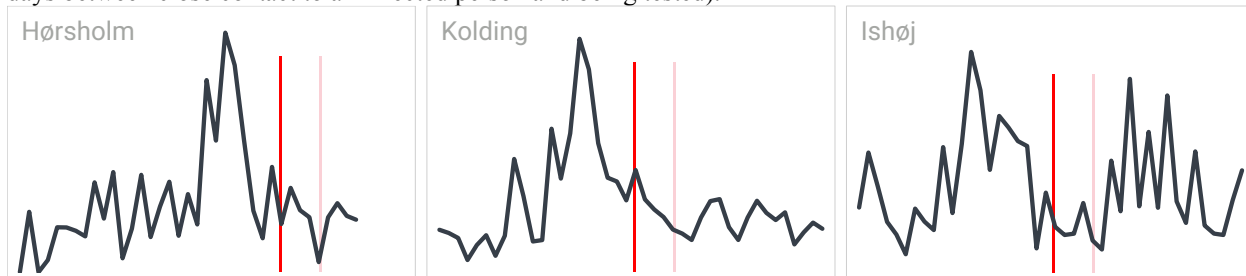
[99] This is partly due to policy lags, i.e. the lag between the time an epidemic problem arises, and the effect of a policy intended to counteract it. In monetary and fiscal policies, the four lags are the *recognition lag*, the *decision lag*, the *implementation lag* and the *effectiveness lag*. These lags are likely to be relevant for pandemic policies too. As described on p. 19, the first time WHO characterized COVID-19 as a pandemic was by March 11, 2020, 2½ month after the first case in China.

[100] In general, 3-4 clinical trials are required before a drug will be approved. Usually, a Phase 1 for safety, a Phase 2 to find the best dose (most effective while being safe), and 1-2 randomized Phase 3 trials to confirm the benefit seen in the Phase 2 trial.

The "we cannot be sure without randomized control trials"-objection
Another objection is that our findings are not based on randomized control trials (RCTs). The obvious response is that a RCT has never been conducted for lockdowns. Therefore, we are limited to observational studies.

We agree that, preferably, the effect of lockdowns should be tested using randomized control trials (RCT), although Deaton and Cartwright (2018) argue that "the lay public, and sometimes researchers, put too much trust in RCTs over other methods of investigation". Unfortunately, there are very few classic RCT studies except for some little relevant mask studies (see Hirt et al. (2022)). Given the lack of RCTs, observational studies are our best way to know if lockdowns work. We only have one source of evidence: history, with all its covariates and missing/bad data. If we reject that source, we have nothing to rely on. RCT is not the only kind of research that can improve policy, and, as Caplan (2022) puts it: "What's the empirical evidence that RCTs actually improve policy?"

However, we have – after we, among other things, conducted a search on Scopus[101] – knowledge of a few "RCT like" studies which use natural experiments to examine the effects of lockdowns, mask mandates, school closures and SIPOs. The studies are described in Table 17 and overall find conclusions very similar to the measured meta-results.

Table 17: The results in the identified natural experiments are similar to our measured meta-results

Study	Conclusion	NPI and design	Description
Kepp and Bjørnskov (2021); "Lockdown Effects on Sars-CoV-2 Transmission – The evidence from Northern Jutland"	"Efficient infection surveillance and voluntary compliance make full lockdowns unnecessary at least in some circumstances." *This result is similar to the measured meta-results, see section 4.*	Lockdown Quasi-random policy change	Kepp and Bjørnskov (2021) use evidence from a quasi-natural experiment when seven of the 11 municipalities in Northern Jutland in Denmark went into extreme lockdown after the discovery of mutations of Sars-CoV-2 in mink (and *not* because of general level of infections).
Dave et al. (2020); "Did the Wisconsin Supreme Court Restart a Covid-19 Epidemic? Evidence from a Natural Experiment"	"We find no evidence that the repeal of the state SIPO impacted social distancing, COVID-19 cases, or COVID-19-related mortality during the fortnight following enactment. Estimated effects were economically small and nowhere near statistically different from zero." *This result is similar to the measured meta-results, see section 4.*	SIPO Quasi-random policy change	Dave et al. (2020) use the Wisconsin Supreme Court abolishment of Wisconsin's "Safer at Home" order (a SIPO) as a natural experiment.

---

[101] The search was based on the same methodology as in our search strategy described in section 2.1, but we replace the methodology search string with "natural experiment" and "regression discontinuity".

Study	Conclusion	NPI and design	Description
Wang (2022); "Stay at home to stay safe: Effectiveness of stay-at-home orders in containing the COVID-19 pandemic"	"We find that although residents in both groups were staying at home even before the implementation of any order, these orders reduced the number of new COVID-19 cases by [5.4%][102]." *This result is similar to the measured meta-results, see section 4.*	SIPO Regression discontinuity design	Wang (2022) compares counties close to the border between states with SIPOs and states without SIPOs.
Hansen and Mano (2021); "Mask Mandates Save Lives"	"Our results imply that mandates saved 87,000 lives [in the United States] through December 19, 2020, while a nationwide mandate could have saved 58,000 additional lives." This corresponds to an impact of 36% fewer deaths.[103] *This result is relatively large compared to the measured meta-results, see section 4.*	Mask mandates Regression discontinuity design	Hansen and Mano (2021) rely on the variation between counties across "mask borders", i.e. a state border that separates two counties, in which one county is in a state with a mask mandate at a given time and the other county is in a state without a mask mandate at the same time. Interestingly, they find mask mandates are four times more effective in counties which are positively inclined to wearing masks which may indicate an important voluntary effect.[104]
Abaluck et al. (2022); "Impact of community masking on COVID-19: A cluster-randomized trial in Bangladesh"	"The intervention reduced symptomatic seroprevalence by 9.3% (adjusted prevalence ratio (aPR) = 0.91 [0.82, 1.00]; control prevalence 0.76%; treatment prevalence 0.68%). In villages randomized to surgical masks (n = 200), the relative reduction was 11.2% overall (aPR = 0.89 [0.78, 1.00])". *This result is relatively small compared to the measured meta-results, see section 4.*	Mask mandates Cluster-randomized trial of community-level mask promotion	Abaluck et al. (2022) carried out an experiment in Bangladesh with cross-randomized mask promotion strategies at the village and household level, including cloth vs. surgical masks. All intervention arms received free masks, information on the importance of masking, role modeling by community leaders, and in-person reminders for 8 weeks. The control group did not receive any interventions. Neither participants nor field staff were blinded to intervention assignment.

---

[102] In their abstract they write 7.6%, but that result is from their difference-in-differences model. The 5.4% is from their regression discontinuity design (see Supporting Information Table A14).

[103] By December 19, 2020, 321,035 had died with COVID-19 in the United States. The results from Hansen and Mano (2021) imply that 263,035 would have died with a nationwide mandate and 408,035 without any mandates. Hence, the effect of masks is 36%.

[104] Hansen and Mano (2021) writes: "Specifically, mask mandates reduce COVID-19 cases and deaths by -78.03 and -1.45, respectively, in the median county more positively inclined to wearing masks in our sample. While the same numbers for the median county more negatively inclined towards to wearing masks are -44.54 and -0.37, also respectively".

Study	Conclusion	NPI and design	Description
Fukumoto et al. (2021); "No causal effect of school closures in Japan on the spread of COVID-19 in spring 2020"	"We do not find any evidence that school closures in Japan reduced the spread of COVID-19. Our null results suggest that policies on school closures should be reexamined given the potential negative consequences for children and parents." *This result is similar to the measured meta-results, see section 4.*	School closures Regression discontinuity design	Matches each municipality with open schools to a municipality with closed schools that is the most similar in terms of potential confounders to estimate the causal impact of closing schools.

Note: We found Kepp and Bjørnskov (2021) and Dave et al. (2020) during our initial search on Google Scholar. One of the authors pointed us toward Hansen and Mano (2021), and we knew Abaluck et al. (2022) from the media. The other studies were identified in a search on Scopus using the same disease search string and government response search string, but replacing the methodology search string with first "natural experiment" then "regression discontinuity design", see The Royal Swedish Academy of Sciences (2021). We did not find any additional relevant natural experiments in the SCOPUS search (several studies claim to be natural experiments, but as Digitale et al. (2021) note, "the term 'natural experiment' is somewhat of a misnomer. Policy responses being studied are not naturally occurring, but are decisions driven by the pandemic's trajectory and social and political will"). The Kepp and Bjørnskov (2021), Fukumoto et al. (2021), and Wang (2022) are not included in our review and meta-analysis, because they focus on cases and not on COVID-19 mortality. Dave et al. (2020) is not included in our review, because it is a synthetic control method and lack jurisdictional variance. Hansen and Mano (2021) is not included, because their working paper was published after our search on Google Scholar.

Overall, the results from the natural experiments in Table 17 are similar to our own conclusions. They, for the most part, only find marginal effects of lockdowns and NPIs, except for Hansen and Mano (2021) who examine mask mandates. These studies do not live up to our eligibility criteria and are not included in the meta-analysis (see notes to Table 17) but should be noteworthy given their credibility.

The "every time a country has locked down, the mortality rate has dropped"-objection
We agree that the general pattern has been that mortality rates usually – but not always – drop after lockdowns are imposed. However, this does not imply causality, as people voluntarily change behavior when responding to information. Also, while this has been the general pattern, there are examples which, as a minimum, question the causality in the argument. Figure 15 shows daily deaths in Slovenia and Slovakia during the 2020/21 winter. Both Slovenia and Slovakia introduced strict lockdowns in late October 2020. On October 24th, Slovakia had issued SIPOs, limited gatherings, banned indoor eating and drinking at restaurants, closed schools, and issued mask mandates.[105] Nevertheless, the death toll continued to rise, and the death rate stayed above pre-lockdown levels for at least six months. Slovenia experienced a more classic wave, but the daily death rate did not peak until 6 weeks after the lockdown, making it unlikely to be caused by the lockdown (usually, it takes 3-4 weeks from infection to death.)

Figure 16: **Lockdowns in Slovakia and Slovenia did not make mortality rates drop**
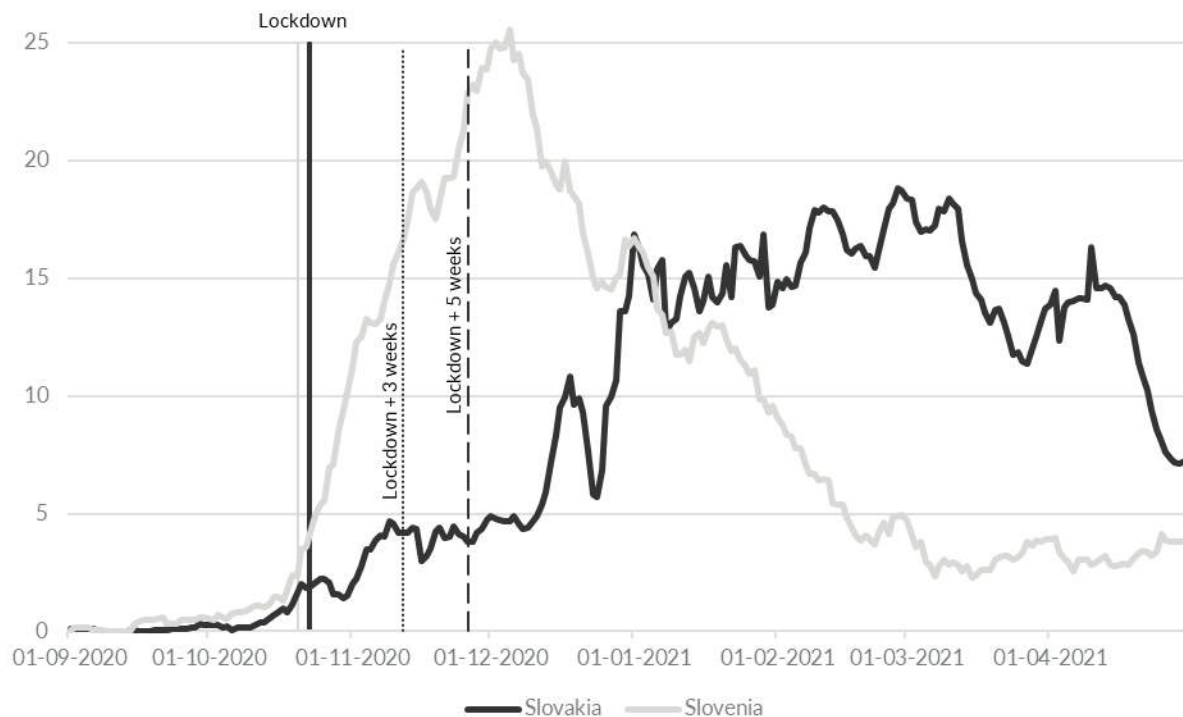
Source: Our World in Data (2022)
Note: Daily deaths are shown as seven days average. The black (Slovakia) and grey (Slovenia) vertical lines illustrates the day when the OxCGRT stringency index surpassed 70. The dotted and dashed lines show approximately three and five weeks after lockdowns were introduced.
Source: Our World in Data and Folkhälsomyndigheten.

---

[105] Source: https://crisis24.garda.com/alerts/2020/10/slovakia-authorities-to-introduce-partial-nationwide-lockdown-from-october-24-update-13.

The "what about Zero-Covid-countries"-objection?

Some have pointed to countries like Australia and New Zealand, which have followed a strategy with very strict lockdowns as a response to even relatively few infections (known by many as a "Zero-Covid"-strategy). For example, Melbourne's SIPO in response to the Delta strain lasted 262 days.[106] Comparing COVID-19 mortality rates in Australia to mortality rates in Europe and the United States, this Zero-Covid strategy appears to be effective when measured by COVID-19 mortality rates. But as illustrated in Figure 17, the immediate effectiveness is less obvious compared to other island countries of which at least some have used more lenient COVID-19 policies (e.g. by December 31, 2021, Iceland had never issued a SIPO and only closed schools for 82 days in total, whereas New Zealand (/Australia) had issued a SIPO for 191 (/363) days and closed schools for 200 (/396 days).[107] We therefore caution against attributing the low mortality rates in New Zealand and Australia to strict lockdowns when more obvious explanations – such as being island countries – may explain the differences.

Figure 17: COVID-19 mortality rates have been relatively low in several island countries despite significant differences in their lockdown policies (2020-2021)
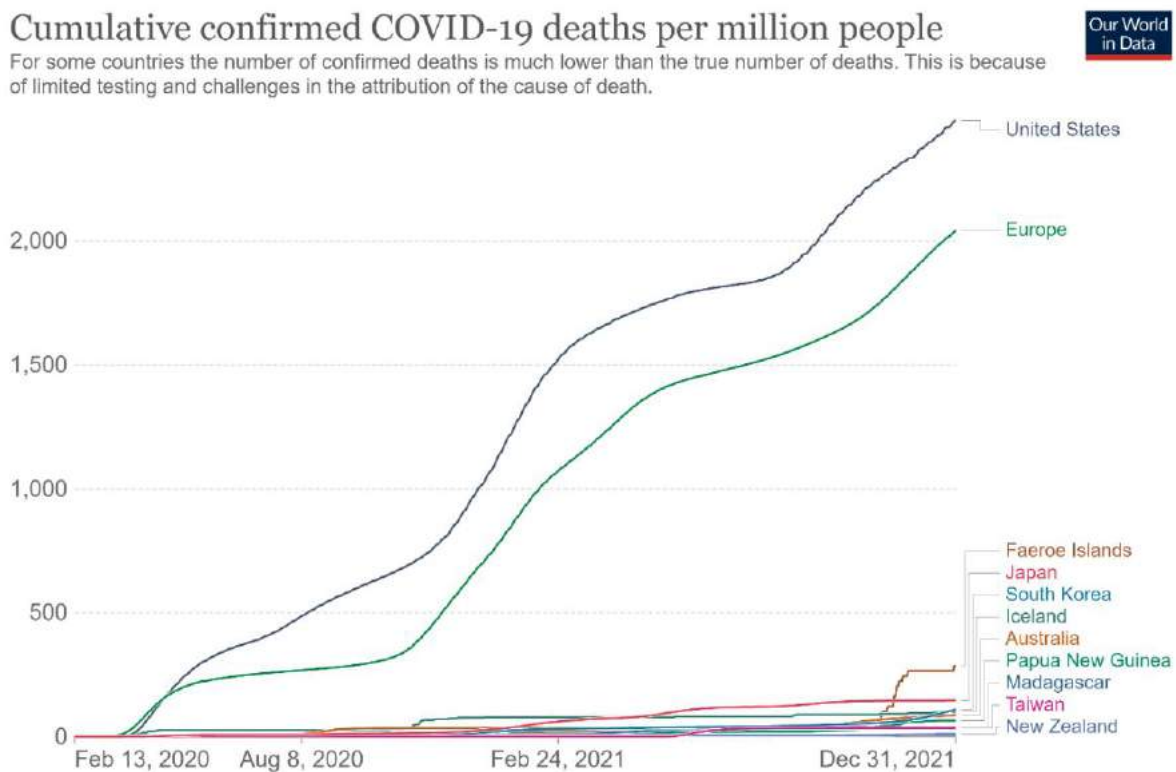
Source: Our World in Data (2022).
Note: South Korea is included as it is de facto an island.

---

[106] See https://www.reuters.com/world/asia-pacific/melbourne-readies-exit-worlds-longest-covid-19-lockdowns-2021-10-20/.
[107] Source: Hale, Angrist, Goldszmidt, et al. (2021a). Note that Hale, Angrist, Goldszmidt, et al. (2021a) only report one the extent of lockdowns in countries with several jurisdictional regions may be over.

5.2.5 Other factors are more likely to explain the differences in COVID-19 mortality between countries than differences in lockdowns

But what else 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 counterparts. Using data from Germany, Laliotis and Minos (2020) show that the spread of COVID-19 and the resulting deaths in predominantly Catholic regions with stronger social and family ties were much higher when compared to non-Catholic ones at the local NUTS 3 level.[108] Albæk (2021) notes that trust in others seems to be an important factor, see Figure 18.[109] And Bollyky et al. (2022) find that "measures of trust in the government and interpersonal trust, as well as less government corruption, had larger, statistically significant associations with lower standardized infection rates". Thornton (2022) finds that "if all societies had trust in government at least as high as Denmark, which is in the 75th percentile, the world would have experienced 13% fewer infections. If social trust—trust in other people—reached the same level, the effect would be even larger, with 40% fewer infections globally." Similar results are found in several other studies.

---

[108] 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.

[109] It is remarkable that the five countries above the confidence interval in Figure 18 – Belgium, Italy, the United Kingdom, Spain, and Sweden – all can be found among the countries hit early by the pandemic, see Figure 7, which may indicate that the mortality rates in these countries could have been much lower, if they had not been among the first countries to be hit by the COVID-19 pandemic. (There are no country labels in Figure 7, but the five countries all experienced more than 500 deaths per million during the first wave. The mortality rates by June 30, 2020, were 838 per million in Belgium, 576 per million in Italy, 593 per million in United Kingdom, 615 per million in Spain, and 525 per million in Sweden).

Figure 18: Countries with more trust in others experienced lower COVID-19 mortality rates
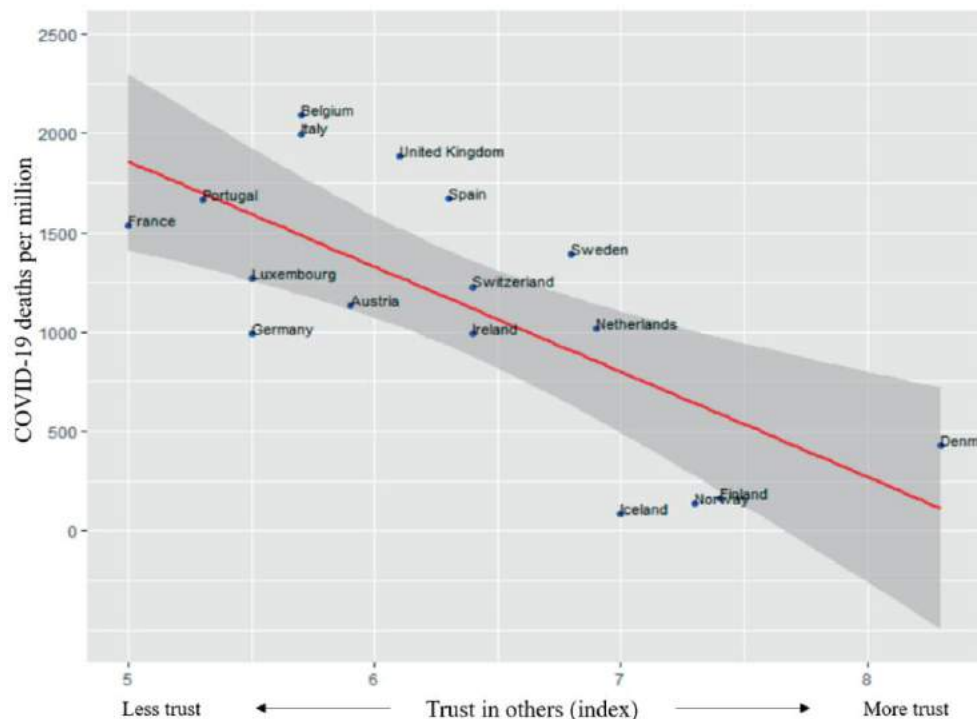
Government communication may also have played a large role. Compared to its Scandinavian neighbors, the communication from Swedish health authorities was far more subdued and embraced the idea of public health vs. economic trade-offs. An illustration of the differences in perspective on the coming pandemic was visible on March 7th, 2020, when the Danish final in the European Song Contest – based on national health authorities' strong recommendations – was held *without* audience in Denmark but – again based on national health authorities' recommendations – *with* audience in Sweden.[110]

This clearly illustrates differences in the health authorities' assessment of the risk, and this difference may explain why Helsingen, Refsum, et 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.

---

[110] Another example is the Danish Prime Minister, Mette Frederiksen, who on several occasions said that every single death is a tragedy. A word choice that was very distinct from that of the Swedish state epidemiologist Anders Tegnell's.

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 in Europe 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 persisted in later waves. The importance of the timing of the winter holiday is illustrated in Figure 19 borrowed from Andersson (2022), which illustrate 1) excess mortality in Swedish regions (län) where the winter holiday was in week 9, 2) excess mortality in Swedish regions where the winter holiday was in other weeks, and 3) excess mortality in other Nordic countries. The figure illustrates how excess deaths in Sweden in the spring of 2020 were primarily driven by regions with winter holidays in week 9, while excess mortality in other regions were comparable to the excess mortality in other Nordic countries.

Figure 19: The excess mortality in Sweden in the spring of 2020 emerged primarily in regions with winter holidays in week 9, when ski tourists were unknowingly exposed to a COVID-19 virus outbreak in the Alps
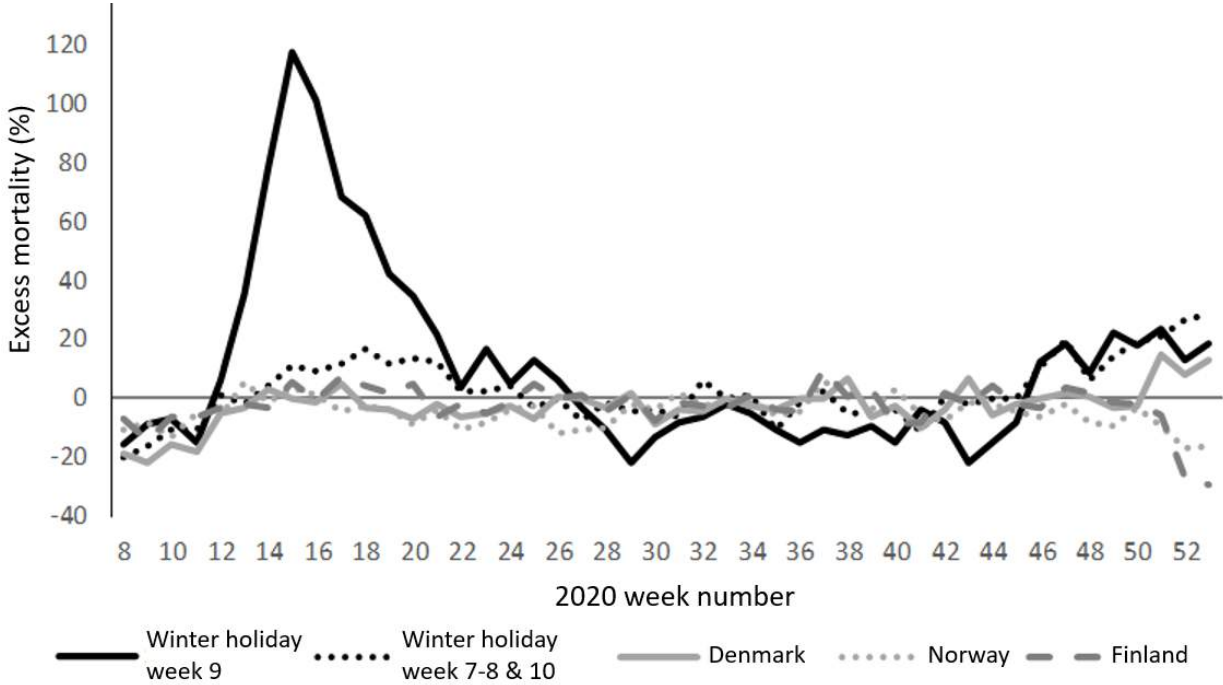
Source: Andersson (2022)
Note: Axis titles and legends have been translated from Swedish to English.

Also, Sweden had more frail elders due to very mild flu seasons in 2018/19 and 2019/20 as well as very few deaths during the 2019 summer compared to earlier years and compared to other Nordic countries (see Herby (2020)) which affected mortality in Sweden (see Juul et al. (2022); Zahran et al. (2022)). Had the winter holiday in Sweden been in week 7 or week 8 as in Denmark

and had mortality rates in Sweden in 2018/19 and 2019/20 been comparable to mortality rates in Denmark, the Swedish COVID-19 situation could have turned out very differently.[111]

5.3 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 finds little to no positive health 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 etc. so that citizens can voluntarily respond to the pandemic in a way that mitigates their exposure.[112]

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: Our study finds that lockdowns had little to no effect in reducing COVID-19 mortality.

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.[113] They have contributed to reducing economic activity, raising unemployment, reducing schooling[114], causing

---

[111] 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.

[112] As noted by Thaler and Sunstein (2009), it might be fruitful to consider how nudging can influence citizens' behavior without coercion.

[113] In a review, Onyeaka et al. (2021) conclude that "the impact of the lockdown has had far-reaching effects in different strata of life, including; changes in the accessibility and structure of education delivery to students, food insecurity as a result of unavailability and fluctuation in prices, the depression of the global economy, increase in mental health challenges, wellbeing and quality of life amongst others."

[114] Engzell et al. (2021) find that children, on average, learn almost nothing in the weeks they received virtual education. The effect was particularly pronounced in less educated homes, where the negative effects for children with low-educated parents were 60 percent greater than those from well-educated homes. Rose et al. (2021) find that school closures in the UK in the spring of 2020 had put six- and seven-year-old students about two months behind in reading and seven months in math. Lindberg (2021) find that half of all students between 5th and 8th grade experienced got significantly less out of online teaching. Lindberg (2021) notes that school closures especially have caused academic and well-being problems among students from socially disadvantaged homes. Rajmil et al. (2021) indicate that children's health has suffered in most of the world. Of particular concern is new research showing that even very young children's development has suffered tremendously during the lockdowns. Some of the clearest evidence for the problems comes from Rhode Island, where Brown University's Resonance project, Deoni et al. (2021), has followed children's developments for more than ten years. Children's development is measured with so-called Mullen scales for early learning, which measure their fine motor skills, language and visual reception, and which are considered a counterpart to intelligence goals for young children. Deoni et al. (2021) find that children's scores on these goals have dropped markedly between 27 to 37 points compared to the normal of about 100 points before the shutdowns. The study also finds, like several other studies, that the losses in early development are significantly greater among poor and poorly educated families. Halloran et al. (2021) find that passing rates declined 10.1% in districts *without* in-person teaching relative to districts *with* in-person teaching

political unrest[115], contributing to domestic violence, loss of life quality[116], and the undermining of liberal democracy. These costs to society must be compared to the benefits of lockdowns, which our meta-analysis has shown are little to none.[117]

Such a standard benefit-cost calculation leads to a strong conclusion: until future research based on credible empirical evidence can prove that lockdowns have large and significant reductions in mortality, lockdowns should be rejected out of hand as a pandemic policy instrument.

---

and that the effect was larger in districts with larger populations of students who are Black, Hispanic or eligible for free and reduced price lunch. Goldhaber et al. (2022) conclude that "remote instruction was a primary driver of widening achievement gaps. Math gaps did not widen in areas that remained in-person (although there was some widening in reading gaps in those areas)."

[115] See Wood et al. (2021).

[116] Patrick et al. (2022) find that deviations from expected historical trends in alcohol use that occurred during the pandemic among U.S. young and middle adults included increases in alcohol use frequency and increases in the use of alcohol to relax/relieve tension and because of boredom. These shifts were likely due, in part, to drinking while alone and at home.

[117] A crucial research object here is to determine not only how many deaths lockdowns avoid gross (e.g., avoided COVID-19 deaths), but also the net effect on excess mortality over a longer period. Also crucially, it matters which lives are saved and what lives are lost due to lockdowns. Lockdowns have an impact on the Health-Related Quality-of-Life (HRQoL) of the population, a key concept in health economics. Deaths results in a health loss. Loss of health can also be a consequence of a loss in the quality of life. Quality-adjusted life years (QALY) is a generally world-wide accepted measure of health loss which include both loss of health due to loss of life expectancy and loss of quality of life. In a recent study by Persson et al. (2021), the HRQoL was measured by a web based survey sent to randomized samples of the adult Swedish population before the outbreak of the pandemic in Sweden in February 2020 (n=1,016) and during the outbreak of the pandemic. The first wave pandemic data was collected in April 2020 (n=1,003), one-month after the outbreak and, the second wave data was collected in January 2021 (n=1,013), after 10-months living under the pandemic. The number of quality-adjusted life-years (QALYs) lost per month was calculated for both pandemic surveys. Loss of health for the Swedish population for one month in April 2020 was 29 000 QALYs and for the other month January 2021 was estimated to 44 000 QALYs. These monthly losses of health are in the same magnitude as the total loss of health due to excess mortality for the entire year 2020 which is 42 802 QALYs. Similar estimates of the NPIs and the related loss of quality of life are also available from the US, see Hay et al. (2021). It is fair to conclude that any analysis of the costs to society of NPIs should include measures of the HRQoL. However, such measures are so far only available for a few countries like Sweden and the United States.

Appendix I. Supplementary information

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 18: Studies excluded in the meta-study 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 et al. (2021); "Timing of social distancing policies and COVID-19 mortality: county-level evidence from the U.S."	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	Duplicate
Amuedo-Dorantes, Kaushal and Muchow (2021); "Timing of social distancing policies and COVID-19 mortality: county-level evidence from the US"	Only looks at timing
Aparicio and Grossbard (2021); "Are Covid Fatalities in the US Higher than in the EU, and If so, Why?"	Not difference-in-
Arruda et al. (2021); "ASSESSING THE IMPACT OF SOCIAL DISTANCING ON COVID-19 CASES AND DEATHS IN BRAZIL: AN INSTRUMENTED	Social distancing (not
Auger et al. (2020); "Association between statewide school closure and COVID-19 incidence and mortality in the US"	Uses a time series
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	Uses a time series
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
Bharati and Fakir (2020); "Pandemic Catch-22: How effective are mobility restrictions in halting the spread of COVID-19 in developing countries"	Duplicate
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 et al. (2021); "Closed for business: The mortality impact of business closures during the Covid-19 pandemic"	Too few observations
Bongaerts et al. (2021); "Closed for business: The mortality impact of business closures during the Covid-19 pandemic"	Duplicate
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
Borri et al. (2020); "The "Great Lockdown": Inactive workers and mortality by Covid-19"	Too few observations
Bushman et al. (2020); "Effectiveness and compliance to social distancing during COVID-19"	Social distancing (not
Canatay et al. (2021); "Critical country-level determinants of death rate during Covid-19 pandemic"	Not difference-in-
Caselli et al. (2020); "From the lockdown to the new normal: An analysis of the limitations to individual mobility in Italy following the Covid-19 crisis"	Do not look at mortality
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
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
Chaudhry et al. (2020); "A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on	Not difference-in-
Chernozhukov, Kasahara and Schrimpf (2021); "Mask mandates and other lockdown policies reduced the spread of COVID-19 in the US"	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
Ciminelli and Garcia-Mandico (2020); "When and how do business shutdowns work? Evidence from Italy's first COVID-19 wave"	Too few observations
Ciminelli and Garcia-Mandico (2021); "Business shutdowns and covid-19 mortality"	Duplicate
Coccia (2020); "The effect of lockdown on public health and economic system: findings from first wave of the COVID-19 pandemic for designing	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
Dey et al. (2021); "Lag time between state-level policy interventions and change points in COVID-19 outcomes in the United States"	Uses a time series
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	Uses modelling
Filias (2020); "The impact of government policies effectiveness on the officially reported deaths attributed to covid-19."	Student paper
Fountoulakis et al. (2020); "Factors determining different death rates because of the COVID-19 outbreak among countries"	Not difference-in-
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	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	Too few observations
Guo et al. (2020); "Social distancing interventions in the United States: An exploratory investigation of determinants and impacts"	Duplicate
Hale et al. (2020); "Global Assessment of the Relationship between Government Response Measures and COVID-19 Deaths"	Duplicate
Huber and Langen (2020); "The impact of response measures on COVID-19-related hospitalization and death rates in Germany and Switzerland"	Duplicate

1. Study (Author & title)	2. Reason for exclusion
Huber and Langen (2020); "Timing matters: the impact of response measures on COVID-19-related hospitalization and death rates in Germany and	Only looks at timing
Hunter et al. (2021); "Impact of non-pharmaceutical interventions against COVID-19 in Europe: A quasi-experimental non-equivalent group and time-	Not difference-in-
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	Too few observations
Juranek and Zoutman (2021); "The Effect of Non-Pharmaceutical Interventions on the Demand for Health Care and Mortality: Evidence on COVID-19 in	Duplicate
Juranek and Zoutman (2021); "The effect of non-pharmaceutical interventions on the demand for health care and on mortality: evidence from COVID-	Duplicate
Kakpo and Nuhu (2020); "Effects of Social Distancing on COVID-19 Infections and Mortality in the US"	Social distancing (not
Kapitsinis (2021); "The underlying factors of excess mortality in 2020: a cross-country analysis of pre-pandemic healthcare conditions and strategies to	Not difference-in-
Kapoor and Ravi (2020); "Impact of national lockdown on COVID-19 deaths in select European countries and the US using a Changes-in-Changes	Too few observations
Khan et al. (2021); "Assessing the Impact of Policy Measures in Reducing the COVID-19 Pandemic: A Case Study of South Asia"	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 US 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	Do not look at mortality
Langeland et al. (2021); "The Effect of State Level COVID-19 Stay-at-Home Orders on Death Rates"	Not difference-in-
Le et al. (2020); "Impact of government-imposed social distancing measures on COVID-19 morbidity and mortality around the world"	Uses a time series
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	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	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	Only looks at timing
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	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	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
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
Pan et al. (2021); "Heterogeneity in the Effectiveness of Non-pharmaceutical Interventions During the First SARS-CoV2 Wave in the United States"	Duplicate
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	Only looks at timing
Porto et al. (2022); "Lockdown, essential sectors, and Covid-19: Lessons from Italy"	Too few observations
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
Shanmugam et al. (2021); "A report card on prevention efforts of covid-19 deaths in US"	Not difference-in-
Siedner et al. (2020); "Social distancing to slow the US COVID-19 epidemic: Longitudinal pretest–posttest comparison group study"	Duplicate
Siedner et al. (2020); "Social distancing to slow the US COVID-19 epidemic: Longitudinal pretest–posttest comparison group study"	Uses a time series
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
Stamam et al. (2020); "IMPACT OF LOCKDOWN MEASURE ON COVID-19 INCIDENCE AND MORTALITY IN THE TOP 31 COUNTRIES OF THE	Uses a time series
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
Stockenhuber (2020); "Did We Respond Quickly Enough? How Policy-Implementation Speed in Response to COVID-19 Affects the Number of Fatal	Not difference-in-
Supino et al. (2020); "The effects of containment measures in the Italian outbreak of COVID-19"	Uses a time series
Thayer et al. (2021); "An interrupted time series analysis of the lockdown policies in India: a national-level analysis of COVID-19 incidence."	Uses a time series
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
Toya and Skidmore (2021); "A Cross-Country Analysis of the Determinants of Covid-19 Fatalities"	Not difference-in-
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
Tsai et al. (2021); "Coronavirus Disease 2019 (COVID-19) Transmission in the United States Before Versus After Relaxation of Statewide Social	Uses a time series
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 US"	Do not look at mortality
Wu and Wu (2020); "Stay-at-home and face mask policies intentions inconsistent with incidence and fatality during US 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-	Do not look at mortality
Yehya et al. (2021); "Statewide Interventions and Coronavirus Disease 2019 Mortality in the United States: An Observational Study"	Is purely descriptive
Yehya, Venkataramani and Harhay (2020); "Statewide Interventions and Coronavirus Disease 2019 Mortality in the United States: An Observational	Only looks at timing
Yilmazkuday (2021); "Stay-at-home works to fight against COVID-19: international evidence from Google mobility data"	Social distancing (not
Ylli et al. (2020); "The lower COVID-19 related mortality and incidence rates in Eastern European countries are associated with delayed start of	Not effect of lockdowns
Zimmerman and Anderson (2021); "Association of the timing of school closings and behavioral changes with the evolution of the coronavirus disease	Uses a time series

Interpretation of estimates and conversion to standardized estimates

In Table 19, we describe for each study how we interpret their results and convert the estimates to our standardized estimate. For studies not included in the meta-analysis, we describe why. Standard errors are converted such that the t-value, calculated based on standardized 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 19: Notes concerning the standardization of results of the studies included in the meta-analysis

1. Study (Author & title)	2. Date Published	3. Journal	4. Notes concerning the calculation of standardized estimates
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.
An et al. (2021), 19; "Policy Design for COVID-19: Worldwide Evidence on the Efficacies of Early Mask Mandates and Other Policy Interventions"	06-sep-21	Public Administration Review	We use the country fixed-effects models as the authors state "To capture the dynamic nature of the relationships, namely, how policy adoption relates to infection and mortality rates over time, we turn to panel data using within-country variations (i.e., country fixed-effects)."
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 standardized 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)
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, see Fig. 2), is converted to a relative effect on a state basis based on data from Our World in Data. 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 US is 1,1% more deaths after 14 days. Standardized 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). Bjørnskov (2021a) uses a log-log specification which means that the standardized estimate can be calculated as the average of the effect in Europe and United States, where the effect for each is calculated as exp((ln(policy stringency) x estimate) – exp(ln(recommendation stringency)) x estimate).
Bonardi et al. (2020); "Fast and local: How did lockdown policies affect the spread and severity of the covid-19"	8-Jun-20	CEPR Covid Economics	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 standardized 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).
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." They have two model specifications. One *excluding* national case numbers (their Table 7) and one *including* national case numbers (their Table 9). The latter find much smaller effects of some NPIs on mortality, but since they only calculate the counterfactual for the specification excluding national case numbers, we do not include the estimates from their Table 9 in our analysis.
Chisadza et al. (2021); "Government Effectiveness and the COVID-19 Pandemic"	10-Mar-21	MDPI	They use a Poisson model, so a one-unit change in the predictor variable will change the log of the dependent variable (mortality) by respective regression coefficient. With the estimates from Table 2 (model 4), the change in ln(mortality) is equal to SI x 0.0874 - SI^2 x -0.0007, where SI is the value of the stringency index. Stringencies in Europe and United States are equal to the average stringency

1. Study (Author & title)	2. Date Published	3. Journal	4. Notes concerning the calculation of standardized estimates
			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). Hence, the standardized estimate for Europe can be calculated as exp((76-44) x 0.0874 - (76^2-44^2) x 0.0007)-1 = 9%. The effect for United States is calculated similarly (14%). The standardized standard error is calculated as (standardized estimate / estimate x standard error). However, since Chisadza et al. (2021) use a quadratic term, we cannot calculate the standard error. Instead, we use the standard error from their linear model as a proxy for the standard error for the average effect. Note: In an earlier version of this study, we used the linear estimate without taking the exponential.
Dave et al. (2021); "When Do Shelter-in-Place Orders Fight Covid-19 Best? Policy Heterogeneity Across States and Adoption Time"	3-Aug-20	Economic Inquiry	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 similar to model 2, we use an average of model 1 to 5, where the estimates of model 3 to 5 are calculated as (standardized estimate model 2) / (estimate model 2) x estimate model 3/4/5.
Ertem et al. (2021); "The impact of school opening model on SARS-CoV-2 community incidence and mortality"	27-okt-21	Nature Medicine	Include OxCGRT policy variables as covariates in their regression models, but do not present estimates for these variables. Since we do not have access to covariation between coefficients, the coefficients for different weeks are assumed independent. This results in an underestimation of standard error for our standardized estimate. We report the results from these models from the school mode–week interaction terms as marginal effects that are interpreted as the adjusted absolute effect of school mode per week on the outcome.
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 standardized 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, Angrist, Goldszmidt, et al. (2021b).
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.
P. Goldstein et al. (2021); "Lockdown Fatigue: The Diminishing Effects of Quarantines on the Spread of COVID-19 "	4-Feb-21	CID Faculty Working Paper	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, so the effect is e^-1.16 – 1 = −0.69 decline in weekly deaths per million per standard deviation. 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 151 countries with data before March 19, 2020 - using all data yields similar results) yielding -0.03 deaths per week per million per stringency point. The standardized 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). Actual COVID-19 mortality is cumulative mortality by June 30, 2020. Hence, the standardized estimate is calculated for the first wave. If we instead calculate the standardized estimate for the full data period used by P. Goldstein et al. (2021), the average stringency is 58 in Europe and 50 in United States and cumulative mortality is much larger. Both effects cause the estimated effect to be (much) lower.
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 standardized estimate.
Hale, Angrist, Hale, et al. (2021); "Government responses and COVID-19	09-jul-21	PLOS ONE	We use the estimate from Table 1 (1). Standardized estimate is calculated as the average of the effect in Europe and United States, where the effect for each is calculated as exp((policy stringency - recommendation stringency) x estimate)-1. Stringencies in Europe

1. Study (Author & title)	2. Date Published	3. Journal	4. Notes concerning the calculation of standardized estimates
deaths: Global evidence across multiple pandemic waves"			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).
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), standardized estimates are calculated as $8^{est}-1$.
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 standardized estimate. Standardized 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 Development 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. Standardized estimate is calculated as the average of the effect in Europe and United States, where the effect for each is calculated as $\exp((\text{policy stringency} - \text{recommendation stringency}) \times \text{estimate})-1$. 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).
Spiegel and Tookes (2021b); "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.
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 Table "Regression results, mean policy strictness (combination of timing and strictness)" in "Additional file"). We use the average of their 24-day and 38-day specification from model 5. We calculate the effect of each NPI as the average effect in all of US/Europe. First, mortality rates without interventions are calculated for each country/state as the number of days each intervention was in effect before May 8, 2020 (24 days before end of study period). Based on this, we calculate the effect on mortality for each intervention in each country/state as "days in in effect" x estimate x population (/mio). E.g., in Austria, workplaces were closed for 54 days in this period (workplaces closed on March 16, 2020). The total effect of workplace closures on mortality is then 54 days x -0.286 (average of -0.258 (24-day spec.) and -0.313 (38-day spec.) x 9 mio = 139 avoided deaths. Doing the same for other NPIs, the total effect of the Austrian lockdown was 336 avoided deaths by June 1st, so 1,004 would have died *without* lockdown while 668 died *with* lockdown. This is done for all countries in Europe and all states in United States. In total, 204,709 would have died in Europe *without* lockdowns (compared to 172,714 *with* lockdowns) and 122,469 (compared to 108,293) in United States. The effect of each intervention in Europe and United States is then calculated assuming the intervention is in place from March 15th, 2020 (54 days). Hence, for Europe, the effect of workplace closure is 54 days x -0.286 (est.) x 748 million = 11,529 avoided deaths out of 204,709 potential deaths which is equal to -5.6%. For United States the effect is 5,060 avoided deaths out of 122,469 potential deaths (-4.1%), and the average effect of workplace closure is thus 4.9%.
Yang et al. (2021); "What is the relationship between government response and COVID-19 pandemics? Global evidence of 118 countries"	28-aug-21	Structural Change and Economic Dynamics	The standardized 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 + Avoided mortality). Avoided mortality is calculated as Estimate x Difference in stringency x number of days 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). Actual COVID-19 mortality is cumulative mortality by June 30, 2020, and the number of days is 107 (from March 16 to June 30, 2020). Hence, the standardized estimate is calculated for the first wave. If we instead calculate the standardized estimate for the full data period used by Yang et al. (2021), the average stringency is 57 in Europe and 51 in United States and cumulative mortality is much larger. Both effects cause the estimated effect for the full period to be (much) lower.

Appendix II: The Anatomy of the Negative Spin-Meisters

The Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise, which one of us (Hanke) founded and co-directs, published "A Systematic Literature Review and Meta-Analysis of the Effects of Lockdowns on COVID-19 Mortality" in its *Studies in Applied Economics* working paper series on January 21, 2022. The working paper's findings – that lockdowns had little to no public health effects measured by mortality – and its policy conclusions – that lockdown policies are ill-founded and should be rejected out of hand – attracted considerable attention in the media, in the White House and halls of the U.S. Congress, among public health experts, and within the chattering classes around the world.

But, it was the strong endorsement of Dr. Marty Makary, a distinguished Professor of Medicine at the Johns Hopkins School of Medicine, during his February 2 appearance on *Tucker Carlson Tonight* that set off a media firestorm. Indeed, on February 3, the Science Media Centre in London issued a press release, Science Media Centre (2022), with statements by Prof. Neil Ferguson, Dr. Seth Flaxman, Prof. Samir Bhatt – all affiliated with Imperial College London and authors of two of the studies (Ferguson et al. (2020) and Flaxman et al. (2020)), we implicitly criticized – and Prof. David Paton (Nottingham University Business School). The release contained several criticisms of our working paper from the Imperial College team, many of which were unscientific[118] or clearly flawed.[119,120] Those were authored by Prof. Ferguson, Dr. Flaxman, and Prof. Bhatt. The press release also contained positive comments by Prof. Paton.

The accompanying Figure 20 denotes five of the most scientific criticisms raised in Science Media Centre (2022) as well as five criticisms raised by follow-up "fact checkers". Although we believe that the criticisms are with little or no merit, we are not going to engage in a critique of them in this Appendix, as the relevant criticisms are dealt with, either directly or indirectly, in the text of this working paper.[121] Our purpose is to illustrate how biased and politicized the media was when it came to its reportage of our working paper.

---

[118] For example, Dr. Seth Flaxman wrote that "smoking causes cancer, the earth is round, and ordering people to stay at home (the correct definition of lockdown) decreases disease transmission". Needless to say, this is in no way a scientific comment. The effect of lockdowns is – unlike smoking and the shape of the Earth – still an open scientific question. And despite Flaxman's attempt to monopolize the term "lockdown," our definition is widely used (see footnote 1 on p. 7 for an example).

[119] For example, Dr. Seth Flaxman wrote that "[the only studies looked at] do not include key facts about disease transmission such as… lockdowns do not immediately save lives, because there's a lag from infection to death, so to see the effect of lockdowns on Covid deaths we need to wait about two or three weeks." We were very aware of this point and mentioned this lag several times in our meta-analysis. For example, we wrote "we distinguish between studies which find an effect on mortality sooner than 14 days after lockdown and those that do not."

[120] Prof. Samir Bhatt wrote that "two years in, it seems still to focus on the first wave of SARS-COV2 and in a very limited number of countries." Dr. Bhatt made this statement even though we clearly described how we had gathered all available studies which covered Europe, the United States, and worldwide.

[121] Referring to numbers in Figure 20: 1) The term "lockdown" has mainly been used to describe two different things. We define our use of lockdown on p. 1. 2) We examine the average lockdown – not the optimal lockdown.

Figure 20: Flow chart for criticisms raised in Science Media Centre
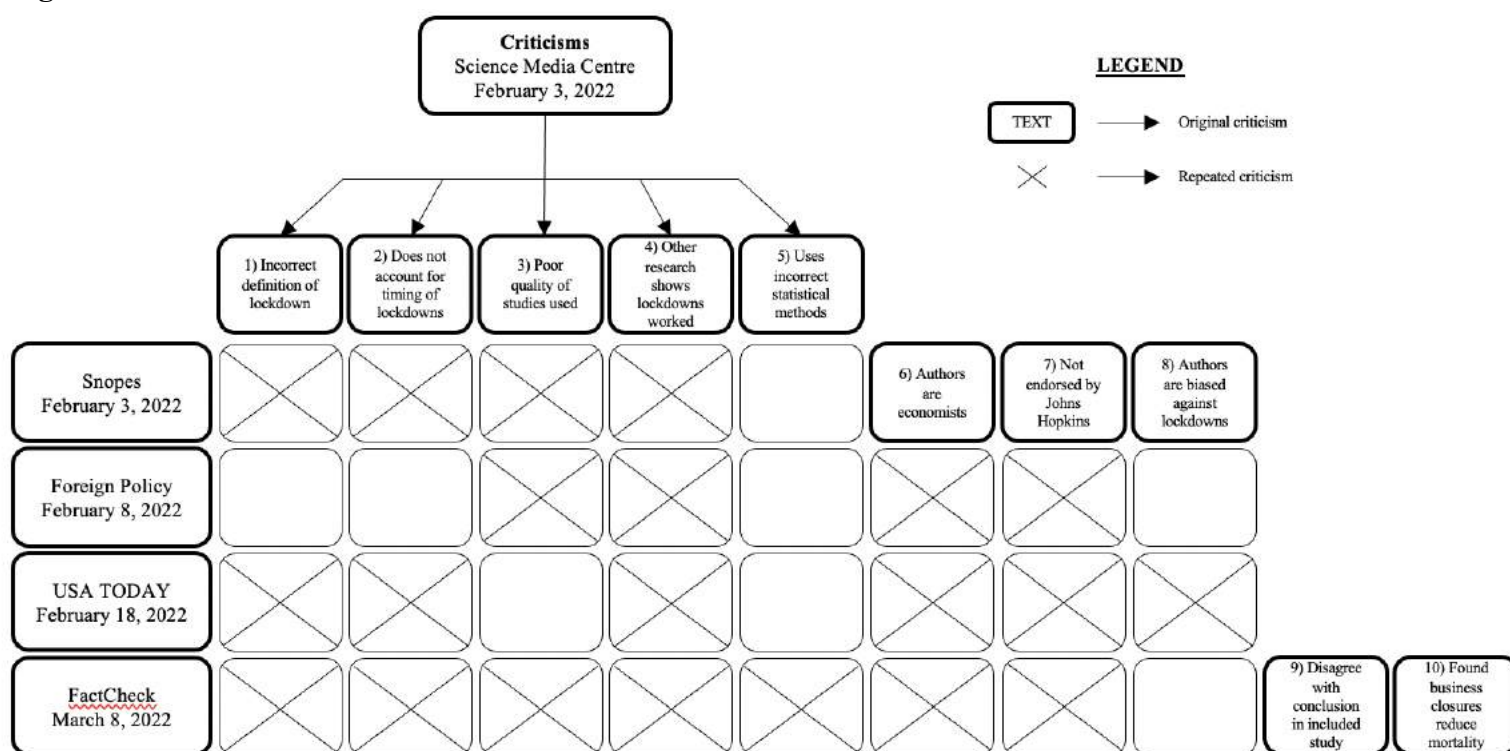
Snopes, which advertises itself as a "fact-checking" website, was quick to jump on the bandwagon. A few hours after the Science Media Centre press release, *Snopes* published a report that contained eight criticisms, Evon (2022). Of those, five were lifted from the Science Media Centre press release. And three new but irrelevant "criticisms" were added, see Figure 20 and footnote 121.

Hence, there are good reasons to exclude studies focusing on timing. We discuss timing on p. 16 and p. 71. 3) The quality of the included studies is handled using bias dimensions (see p. 33). 4) We clearly describe which studies are included and – to handle the "Other research shows lockdowns prevent deaths"-critique –we now explain in more detail why some of the prominent studies such as Flaxman et al. (2020) are both very problematic and ineligible, see section 2.2 on p. 11. We also relate our results to the conclusions in other reviews in section 5.2.1 on p. 61. 5) We use more bias-dimensions to handle the "Used incorrect statistical methods"-critique (see p. 33). 6) The "authors are economists" is an unscientific and irrelevant comment which we do not handle, but simply note that economists are skilled in handling meta-analysis in a wide range of contexts. 7) The "not endorsed by Johns Hopkins" critique is an unscientific and irrelevant comment. As a matter of fact, the Johns Hopkins University does not endorse specific research projects published by its faculty and staff. That said, our study was published by the Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise, a research institute located at the Johns Hopkins University. 8) The "authors are biased against lockdowns" critique is an unscientific comment which was handled in the working paper where our search strategy and eligibility criteria were clearly described (see section 2 on p. 8 in this updated version). 9) We disagree with the conclusion in Chisadza et al. (2021). We explain why in section 3.1, p. 21. 10) We handle this critique on p. 8, where we write: "We believe that one major mistake in our first version was our failure to explain that the overall conclusions do not depend on whether the impact of lockdowns on COVID-19 mortality was 0.2%, 3.0% or 15%."

On February 8, *Foreign Policy* published a critique, Garrett (2022). It copied three of the Science Media Centre's criticisms and two of the new ones added by *Snopes*. In short, *Foreign Policy*'s article presented nothing new.

Then, *USA Today* entered the picture on February 18, Sadeghi (2022). The fact checkers at *USA Today* offered seven criticisms—four copied from the Science Media Centre and three from *Snopes*.

On March 8, *FactCheck* presented ten criticisms, Robertson (2022). They included six from the Science Media Centre (including all five that we identified in Figure 20), two from *Snopes*, and two new confused and misleading criticisms, see Figure 20 and footnote 121.

There was, of course, a great deal of reportage about our working paper that appeared in the early February-April 2022 period. This material was highly repetitive, echoing material presented on February 3 in either the Science Media Centre press release or the *Snopes* report. There were several similar reports, but to avoid repetition ourselves, we limit our review to the five reports contained in Figure 20.

The striking feature of the media flow is its unoriginality. Indeed, there is little evidence that the post-Science Media Centre press release authors even engaged in basic "primary" reading of our text, let alone any "critical" reading of the text. What is evident is what Martin Heidegger identified in his 1927 magnum opus *Being and Time* as "idle talk" ("Gerede"). This is inauthentic discourse, simply adopting and circulating others' opinions about something without ever engaging in even reading a primary text.

If that was not bad enough, the post-Science Media Centre critiques are clearly biased as they all exclude any mention of Prof. Paton's favorable appraisals of our working paper. For example, Prof. Paton made the following four points in Science Media Centre (2022):

1. "Both parts of the paper (systematic review and the meta analysis) make a significant contribution to our understanding of lockdown effects."
2. "Key to a systematic review like this are the sets of search & exclusion criteria. The paper is very transparent about this which is good. They focus on difference-in-difference empirical studies. i.e. they look at papers which compare the impact of an intervention on mortality by looking before & after, but relative to other areas which did not have the intervention. As a result, modelling studies (like the well-known Flaxman Nature paper) are excluded. That is not controversial."
3. "[The result] is pretty consistent with other, non-systematic reviews (e.g. Herby & Allen) which is reassuring. It is also consistent with the (few) studies which look at the impact on overall excess mortality."
4. "More marginal in my view is their exclusion of synthetic control method (SCM) papers. Some of these papers find a significant impact of NPIs on mortality so including them might have led to somewhat higher average mortality effects. The paper gives a robust defence of their exclusion, but I think you would get people on both sides of that debate."

91

As illustrated in Figure 21, none of these positive comments saw the light of day in the "fact checking" that followed the Science Media Centre's press release on February 3. That's because the new cottage industry called "fact checking" has arguably become highly politicized. As a result, there is not much fact checking, but rather opinions about whether the so-called fact checkers agreed or disagreed with the policy implications or conclusions of what they are supposed to be fact checking. So, for the most part, fact checkers were not engaged in fact checking, but were engaged in publishing opinion and narrative. By hiding behind the shroud of "facts" and "fact-checking," they have attempted to cast doubt, in our case, via innuendo.

Figure 21: Flow chart for __appraisals__ raised in Science Media Centre
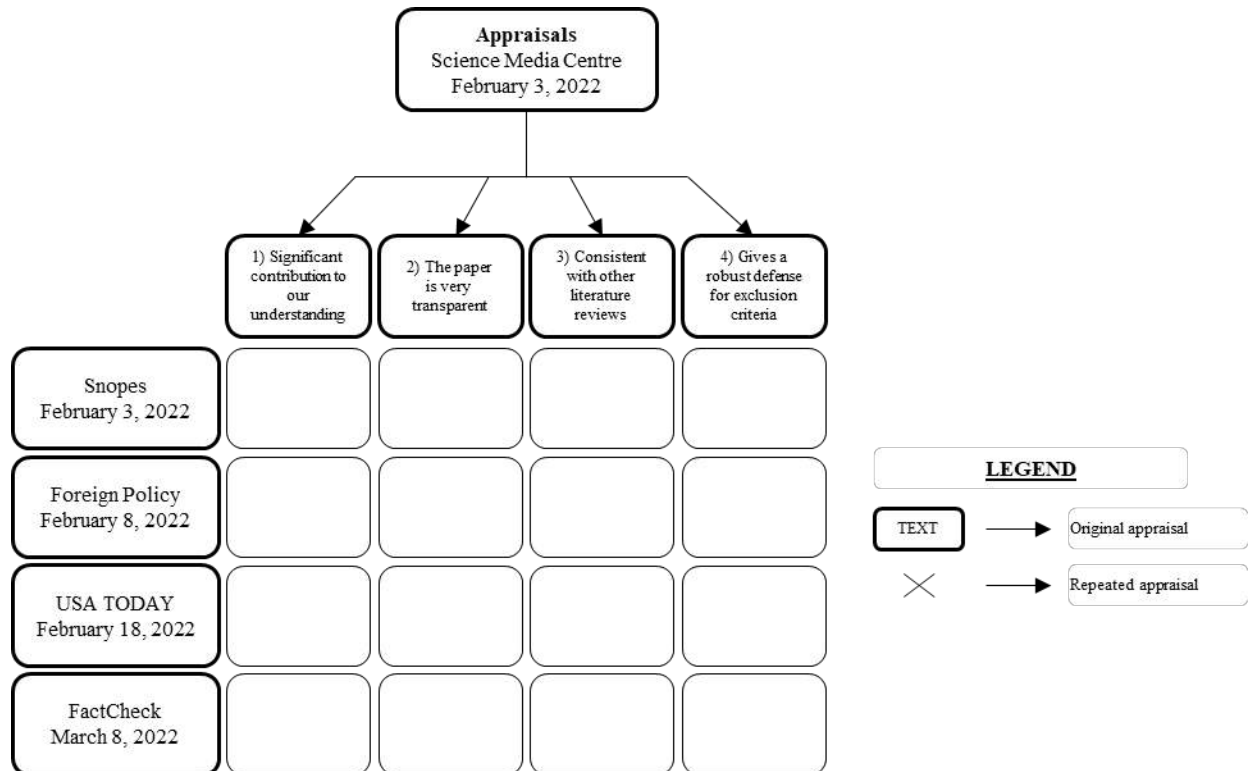
In closing this Appendix, we would like to indicate that we received extensive private reviews and comments on our working paper. After all, that's the purpose of publishing working papers. Our professional correspondents did engage in a serious primary reading of our working paper and made many useful comments and suggestions. Most of their names appear in our thank-you note following the abstract.

We engaged in a thorough review and revision of our January 21, 2022 working paper. We can report that one error of commission was found in the original. It was not detected by any fact checkers or by those we corresponded with, but by us. The error was a computational error that involved logarithms. It was "small" and did not materially affect our results.

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.

Abaluck, Jason, Laura H. Kwong, Ashley Styczynski, Ashraful Haque, Md. Alamgir Kabir, Ellen Bates-Jefferys, Emily Crawford, et al. 2022. "Impact of Community Masking on COVID-19: A Cluster-Randomized Trial in Bangladesh." *Science* 375 (6577):eabi9069. https://doi.org/10.1126/science.abi9069.

Abernethy, Gavin M., and David H. Glass. 2022. "Optimal COVID-19 Lockdown Strategies in an Age-Structured SEIR Model of Northern Ireland." *Journal of The Royal Society Interface* 19 (188):20210896. https://doi.org/10.1098/rsif.2021.0896.

Albæk, Karsten. 2021. "Covid-19 epidemien i Vesteuropa: betydningen af befolkningstæthed, andel ældre og tillid for dødelighed og økonomisk aktivitet," 17.

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.

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.

Andersen, Asger Lau, Emil Toft Hansen, Niels Johannesen, and Adam Sheridan. 2020. "Pandemic, Shutdown and Consumer Spending: Lessons from Scandinavian Policy Responses to COVID-19." *Https://Arxiv.Org/Abs/2005.04630*, May. http://arxiv.org/abs/2005.04630.

Andersson, Fredrik N. G. 2022. "Frivillig Anpassning Eller Tvingande Restriktioner? Överdödligheten, Pandemibekämpningen Och Den Ekonomiska Utvecklingen i Europa under 2020." *Forthcoming in Statsvetenskaplig Tidskrift*.

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. "How a School Holiday Led to Persistent COVID-19 Outbreaks in Europe." *Scientific Reports* 11 (1):24390. https://doi.org/10.1038/s41598-021-03927-z.

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.

Banholzer, Nicolas, Adrian Lison, and Werner Vach. 2022. "Comment on 'A Literature Review and Meta-Analysis of the Effects of Lockdowns on COVID-19 Mortality,'" March, 9.

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.

Bielecki, Michel, Roland Züst, Denise Siegrist, Daniele Meyerhofer, Giovanni Andrea Gerardo Crameri, Zeno Stanga, Andreas Stettbacher, Thomas Werner Buehrer, and Jeremy Werner Deuel. 2021. "Social Distancing Alters the Clinical Course of COVID-19 in Young Adults: A Comparative Cohort Study." *Clinical Infectious Diseases* 72 (4):598–603. https://doi.org/10.1093/cid/ciaa889.

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.

Bollyky, Thomas J, Erin N Hulland, Ryan M Barber, James K Collins, Samantha Kiernan, Mark Moses, David M Pigott, et al. 2022. "Pandemic Preparedness and COVID-19: An Exploratory Analysis of Infection and Fatality Rates, and Contextual Factors Associated with Preparedness in 177 Countries, from Jan 1, 2020, to Sept 30, 2021." *The Lancet*, February, S0140673622001726. https://doi.org/10.1016/S0140-6736(22)00172-6.

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. https://cepr.org/sites/default/files/CovidEconomics23.pdf.

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/.

Bor, Alexander, Frederik Jørgensen, and Michael Bang Petersen. 2021. "Estimating the Municipality Level Attention to Physical Distancing During the Second Half of 2020," August, 6. https://t.co/ytnTInJr89.

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.

Brauner, Jan M., Sören Mindermann, Mrinank Sharma, David Johnston, John Salvatier, Tomáš Gavenčiak, Anna B. Stephenson, et al. 2021. "Inferring the Effectiveness of Government Interventions against COVID-19." *Science* 371 (6531):eabd9338. https://doi.org/10.1126/science.abd9338.

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.

Caplan, Bryan. 2022. "The Efficacy of RCTs: Actual Empirics." Econlib. February 24, 2022. https://www.econlib.org/observational-evidence-on-the-efficacy-of-rcts/.

CDC. 2022. "Burden of Influenza." Centers for Disease Control and Prevention. January 7, 2022. https://www.cdc.gov/flu/about/burden/index.html.

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.
———. 2022. "Comments on the 'John Hopkins' Meta Study (Herby et al., 2022) and Chisadza et al. (2021)," March. https://cdn.factcheck.org/UploadedFiles/Chernozhukov-Kasahara-Schrimpf-comments-on-meta-analysis-and-Chisadza-et-al.pdf.

Chetty, Raj, John Friedman, Nathaniel Hendren, Michael Stepner, and The Opportunity Insights Team. 2020. "How Did COVID-19 and Stabilization Policies Affect Spending and Employment? A New Real-Time Economic Tracker Based on Private Sector Data." *NBER Working Paper Series*, June. https://doi.org/10.3386/w27431.

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.

Clyde, William, Andreas Kakolyris, and Georgios Koimisis. 2021. "A Study of the Effectiveness of Governmental Strategies for Managing Mortality from COVID-19." *Eastern Economic Journal* 47 (4):487–505. https://doi.org/10.1057/s41302-021-00202-x.

Conyon, Martin J., Lerong He, and Steen Thomsen. 2020. "Lockdowns and COVID-19 Deaths in Scandinavia." *SSRN Electronic Journal*, June. https://doi.org/10.2139/ssrn.3616969.

Danmarks Statistik. 2022. "SMIT1: COVID-19 Smittespredning Mv. Pr. Dag (Eksperimentel Statistik) Efter Nøgletal Og Tid." 2022. https://www.statistikbanken.dk/pxs/225353.

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.

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science & Medicine* 210 (August):2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.

Deoni, Sean C. L., Jennifer Beauchemin, Alexandra Volpe, and Viren D'Sa. 2021. "Impact of the COVID-19 Pandemic on Early Child Cognitive Development: Initial Findings in a Longitudinal Observational Study of Child Health." Preprint. Pediatrics. https://doi.org/10.1101/2021.08.10.21261846.

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.

Digitale, Jean C., Kristefer Stojanovski, Charles E. McCulloch, and Margaret A. Handley. 2021. "Study Designs to Assess Real-World Interventions to Prevent COVID-19." *Frontiers in Public Health* 9 (July):657976. https://doi.org/10.3389/fpubh.2021.657976.

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.

Eggers, Andrew C., and Robin Harding. 2021. "Rallying in Fear? Estimating the Effect of the UK COVID-19 Lockdown with a Natural Experiment." *European Journal of Political Research*, June, 1475-6765.12467. https://doi.org/10.1111/1475-6765.12467.

Emborg, Hanne-Dorthe, AnnaSara Carnahan, Karoline Bragstad, Ramona Trebbien, Mia Brytting, Olav Hungnes, Emma Byström, and Lasse S Vestergaard. 2021. "Abrupt Termination of the 2019/20 Influenza Season Following Preventive Measures against COVID-19 in Denmark, Norway and Sweden." *Eurosurveillance* 26 (22). https://doi.org/10.2807/1560-7917.ES.2021.26.22.2001160.

Engzell, Per, Arun Frey, and Mark D. Verhagen. 2021. "Learning Loss Due to School Closures during the COVID-19 Pandemic." *Proceedings of the National Academy of Sciences* 118 (17):e2022376118. https://doi.org/10.1073/pnas.2022376118.

Ertem, Zeynep, Elissa M. Schechter-Perkins, Emily Oster, Polly van den Berg, Isabella Epshtein, Nathorn Chaiyakunapruk, Fernando A. Wilson, et al. 2021. "The Impact of School Opening Model on SARS-CoV-2 Community Incidence and Mortality." *Nature Medicine* 27 (12):2120–26. https://doi.org/10.1038/s41591-021-01563-8.

Evon, Dan. 2022. "Here's What We Know About 'Johns Hopkins Study' on Lockdowns." Snopes.Com. February 3, 2022. https://www.snopes.com/news/2022/02/03/johns-hopkins-study-on-lockdowns/.

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. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf.

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.

Fukumoto, Kentaro, Charles T. McClean, and Kuninori Nakagawa. 2021. "No Causal Effect of School Closures in Japan on the Spread of COVID-19 in Spring 2020." *Nature Medicine* 27 (12):2111–19. https://doi.org/10.1038/s41591-021-01571-8.

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.

Garrett, Laurie. 2022. "Economists Are Fueling the War Against Public Health." Foreign Policy. February 8, 2022. https://foreignpolicy.com/2022/02/08/economists-are-fueling-the-war-against-public-health/.

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.

Goldhaber, Dan, Thomas J Kane, Andrew McEachin, Emily Morton, Tyler Patterson, and Douglas O Staiger. 2022. "The Consequences of Remote and Hybrid Instruction During the Pandemic," May, 36.

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.

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. 2021a. "A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker - OxCGRT)." *Nature Human Behaviour* 5 (4):529–38. https://doi.org/10.1038/s41562-021-01079-8.

———. 2021b. "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, Noam Angrist, Andrew J. Hale, Beatriz Kira, Saptarshi Majumdar, Anna Petherick, Toby Phillips, et al. 2021. "Government Responses and COVID-19 Deaths: Global Evidence across Multiple Pandemic Waves." Edited by Holly Seale. *PLOS ONE* 16 (7):e0253116. https://doi.org/10.1371/journal.pone.0253116.

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.

Halloran, Clare, Rebecca Jack, James Okun, and Emily Oster. 2021. "Pandemic Schooling Mode and Student Test Scores: Evidence from US States." w29497. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w29497.

Hanke, Steve H., and Kevin Dowd. 2022. "Imperial College's Fear Machine." *National Review*, March 30, 2022. https://www.nationalreview.com/2022/03/imperial-colleges-fear-machine/.

Hanke, Steve H., and Abraham Mehrez. 1979. "An Optimal Sampling Procedure for the Collection of Residential Water Use Data." *Water Resources Research* 15 (6):1343–48. https://doi.org/10.1029/WR015i006p01343.

Hansen, Niels-Jakob, and Rui C. Mano. 2021. "Mask Mandates Save Lives." *IMF Working Papers* 2021 (205):1. https://doi.org/10.5089/9781513577616.001.

Hatchett, Richard J., Carter E. Mecher, and Marc Lipsitch. 2007. "Public Health Interventions and Epidemic Intensity during the 1918 Influenza Pandemic." *Proceedings of the National Academy of Sciences* 104 (18):7582–87. https://doi.org/10.1073/pnas.0610941104.

Hay, Joel W., Cynthia L. Gong, Xiayu Jiao, Nadine K. Zawadzki, Roy S. Zawadzki, A. Simon Pickard, Feng Xie, Samuel A. Crawford, and Ning Yan Gu. 2021. "A US Population Health Survey on the Impact of COVID-19 Using the EQ-5D-5L." *Journal of General Internal Medicine* 36 (5):1292–1301. https://doi.org/10.1007/s11606-021-06674-z.

Helsingen, Lise M., Magnus Løberg, Erle Refsum, Dagrun Kyte Gjøstein, Paulina Wieszczy, Ørjan Olsvik, Frederik E. Juul, et al. 2020. "Randomized Re-Opening of Training Facilities during the COVID-19 Pandemic." Preprint. Public and Global Health. https://doi.org/10.1101/2020.06.24.20138768.

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. 2020. "Working Paper: Exceptionally Many Vulnerable – 'Dry Tinder' – in Sweden Prior to COVID-19." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3702595.

———. 2021a. "A First Literature Review: Lockdowns Only Had a Small Effect on COVID-19." *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.3764553.

———. 2021b. "Hvorfor Lukker Myndighederne Skolerne?" *Punditokraterne* (blog). March 31, 2021. https://punditokraterne.dk/2021/03/31/hvorfor-lukker-myndighederne-skolerne/.

———. 2021c. "Nedlukningen Af Hørsholm Kommune Var Unødvendig." *Punditokraterne* (blog). May 21, 2021. https://punditokraterne.dk/2021/05/21/nedlukningen-af-hoersholm-kommune-var-unoedvendig/.

———. 2021d. "Hvad Betyder Timingen Af Nedlukninger? Virker Det, Hvis Man Lukker Ned Tidligt?" *Punditokraterne* (blog). September 16, 2021. https://punditokraterne.dk/2021/09/16/hvad-betyder-timingen-af-nedlukninger-virker-det-hvis-man-lukker-ned-tidligt/.

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.

———. 2022. "A Literature Review and Meta-Analysis of the Effects of Lockdowns on COVID-19 Mortality." *Studies in Applied Economics*, The Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise, , no. 200 (January):62. https://sites.krieger.jhu.edu/iae/files/2022/01/A-Literature-Review-and-Meta-Analysis-of-the-Effects-of-Lockdowns-on-COVID-19-Mortality.pdf.

Hirt, Julian, Perrine Janiaud, and Lars G. Hemkens. 2022. "Randomized Trials on Non-Pharmaceutical Interventions for COVID-19: A Scoping Review." *BMJ Evidence-Based Medicine*, January, bmjebm-2021-111825. https://doi.org/10.1136/bmjebm-2021-111825.

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. https://www.researchgate.net/publication/342253045_Comment_on_Flaxman_et_al_2020_The_illusory_effects_of_non-pharmaceutical_interventions_on_COVID-19_in_Europe.

Hsiang, Solomon, Daniel Allen, Sébastien Annan-Phan, Kendon Bell, Ian Bolliger, Trinetta Chong, Hannah Druckenmiller, et al. 2020. "The Effect of Large-Scale Anti-Contagion Policies on the COVID-19 Pandemic." *Nature* 584 (7820):262–67. https://doi.org/10.1038/s41586-020-2404-8.

Iezadi, Shabnam, Kamal Gholipour, Saber Azami-Aghdash, Akbar Ghiasi, Aziz Rezapour, Hamid Pourasghari, and Fariba Pashazadeh. 2021. "Effectiveness of Non-Pharmaceutical Public Health Interventions against COVID-19: A Systematic Review and Meta-Analysis." Edited by Prasenjit Mitra. *PLOS ONE* 16 (11):e0260371. https://doi.org/10.1371/journal.pone.0260371.

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.

Isphording, Ingo E., Marc Lipfert, and Nico Pestel. 2021. "Does Re-Opening Schools Contribute to the Spread of SARS-CoV-2? Evidence from Staggered Summer Breaks in Germany." *Journal of Public Economics* 198 (June):104426. https://doi.org/10.1016/j.jpubeco.2021.104426.

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.

Joshi, Yogesh V., and Andres Musalem. 2021. "Lockdowns Lose One Third of Their Impact on Mobility in a Month." *Scientific Reports* 11 (1):22658. https://doi.org/10.1038/s41598-021-02133-1.

Juul, Frederik E., Henriette C. Jodal, Ishita Barua, Erle Refsum, Ørjan Olsvik, Lise M. Helsingen, Magnus Løberg, Michael Bretthauer, Mette Kalager, and Louise Emilsson. 2022. "Mortality in Norway and Sweden during the COVID-19 Pandemic." *Scandinavian Journal of Public Health* 50 (1):38–45. https://doi.org/10.1177/14034948211047137.

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.

Lee, Elizabeth C., Nikolas I. Wada, M. Kate Grabowski, Emily S. Gurley, and Justin Lessler. 2020. "The Engines of SARS-CoV-2 Spread." *Science* 370 (6515):406–7. https://doi.org/10.1126/science.abd8755.

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/.

Lendorf, Maria. 2021. "Projekt COVID-19 Mapping: Afdækning Af Smittespredningsmønsteret i Region Hovedstaden," March. https://www.nordsjaellandshospital.dk/nyheder/Documents/Projekt%20COVID-19-Mapping%20Data%20rapport_kort_09032021.pdf.

Lewis, Michael. 2021. *The Premonition: A Pandemic Story*. First Edition. New York: W.W.
    Norton & Company.

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.

Li, Yue, Zijing Cheng, Xueya Cai, Yunjiao Mao, and Helena Temkin-Greener. 2021.
    "Association of State Social Distancing Restrictions with Nursing Home COVID-19 and
    Non-COVID-19 Outcomes." Preprint. Health Policy.
    https://doi.org/10.1101/2021.02.08.21251382.

Lindberg, Anette. 2021. "Grundskolers erfaringer med nødundervisning under covid-19-
    pandemien," 83. https://www.eva.dk/sites/eva/files/2021-
    03/Rapport%20GRUNDSKOLE.pdf.

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.

Maas, Steve. 2020. "Consumers' Fear of Virus Outweighs Lockdown's Impact on Business."
    NBER (blog). August 2020. https://www.nber.org/digest/aug20/consumers-fear-virus-
    outweighs-lockdowns-impact-business.

Madsen, Flemming, Steen Ladelund, and Allan Linneberg. 2014. "High Levels Of Bed
    Occupancy Associated With Increased Inpatient And Thirty-Day Hospital Mortality In
    Denmark." *Health Affairs* 33 (7):1236–44. https://doi.org/10.1377/hlthaff.2013.1303.

Mailman School of Public Health, Columbia. 2007. "Difference-in-Difference Estimation."
    2007. https://www.publichealth.columbia.edu/research/population-health-
    methods/difference-difference-estimation.

Markel, Howard, Harvey B. Lipman, J. Alexander Navarro, Alexandra Sloan, Joseph R.
    Michalsen, Alexandra Minna Stern, and Martin S. Cetron. 2007. "Nonpharmaceutical
    Interventions Implemented by US Cities During the 1918-1919 Influenza Pandemic."
    JAMA 298 (6):644. https://doi.org/10.1001/jama.298.6.644.

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.

McKenzie, David. 2020a. "Revisiting the Difference-in-Differences Parallel Trends Assumption:
    Part I Pre-Trend Testing." *World Bank* (blog). January 21, 2020.

https://blogs.worldbank.org/impactevaluations/revisiting-difference-differences-parallel-trends-assumption-part-i-pre-trend.

———. 2020b. "Revisiting the Difference-in-Differences Parallel Trends Assumption: Part II What Happens If the Parallel Trends Assumption Is (Might Be) Violated?" *World Bank* (blog). February 3, 2020. https://blogs.worldbank.org/impactevaluations/revisiting-difference-differences-parallel-trends-assumption-part-ii-what-happens.

———. 2021. "An Adversarial or 'Long and Squiggly' Test of the Plausibility of Parallel Trends in Difference-in-Differences Analysis." *World Bank* (blog). March 10, 2021. https://blogs.worldbank.org/impactevaluations/adversarial-or-long-and-squiggly-test-plausibility-parallel-trends-difference.

———. 2022a. "A New Synthesis and Key Lessons from the Recent Difference-in-Differences Literature." *World Bank* (blog). January 10, 2022. https://blogs.worldbank.org/impactevaluations/new-synthesis-and-key-lessons-recent-difference-differences-literature.

———. 2022b. "Explaining Why We Should Believe Your DiD Assumptions." *World Bank* (blog). January 24, 2022. https://blogs.worldbank.org/impactevaluations/explaining-why-we-should-believe-your-did-assumptions.

Mendez-Brito, Alba, Charbel El Bcheraoui, and Francisco Pozo-Martin. 2021. "Systematic Review of Empirical Studies Comparing the Effectiveness of Non-Pharmaceutical Interventions against COVID-19." *Journal of Infection*, June, S0163445321003169. https://doi.org/10.1016/j.jinf.2021.06.018.

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.

Morgenstern, Oskar. 1963. *On the Accuracy of Economic Observations*. 2d ed., Completely rev. Princeton, N.J: Princeton University Press.

Mueller-Langer, Frank, Benedikt Fecher, Dietmar Harhoff, and Gert G. Wagner. 2019. "Replication Studies in Economics—How Many and Which Papers Are Chosen for Replication, and Why?" *Research Policy* 48 (1):62–83. https://doi.org/10.1016/j.respol.2018.07.019.

Munch, Pernille Kold, Laura Espenhain, Christian Holm Hansen, Luise Müller, Tyra Grove Krause, and Steen Ethelberg. 2022. "Societal Activities Associated with SARS-CoV-2 Infection: A Case-Control Study in Denmark, November 2020." *Epidemiology and Infection* 150:e9. https://doi.org/10.1017/S0950268821002478.

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.

Onyeaka, Helen, Christian K Anumudu, Zainab T Al-Sharify, Esther Egele-Godswill, and Paul Mbaegbu. 2021. "COVID-19 Pandemic: A Review of the Global Lockdown and Its Far-Reaching Effects." *Science Progress* 104 (2):003685042110198. https://doi.org/10.1177/00368504211019854.

Oraby, Tamer, Michael G. Tyshenko, Jose Campo Maldonado, Kristina Vatcheva, Susie
      Elsaadany, Walid Q. Alali, Joseph C. Longenecker, and Mustafa Al-Zoughool. 2021.
      "Modeling the Effect of Lockdown Timing as a COVID-19 Control Measure in Countries
      with Differing Social Contacts." *Scientific Reports* 11 (1):3354.
      https://doi.org/10.1038/s41598-021-82873-2.

Our World in Data. 2022. "COVID-19 Data Explorer." Our World in Data. 2022.
      https://ourworldindata.org/coronavirus-data-explorer.

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.

———. 2022. "The Political Economy of Meta-Analysis," February, 24.
      http://www.martin.paldam.dk/Papers/Meta-method/12-Meta-mining.pdf.

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.

Patrick, Megan E., Yvonne M. Terry-McElrath, Richard A. Miech, Katherine M. Keyes, Justin
      Jager, and John E. Schulenberg. 2022. "Alcohol Use and the COVID-19 Pandemic:
      Historical Trends in Drinking, Contexts, and Reasons for Use among U.S. Adults."
      Social Science & Medicine 301 (May):114887.
      https://doi.org/10.1016/j.socscimed.2022.114887.

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.

Persson, Ulf, Sara Olofsson, Ning Yan Gu, Cynthia L Gong, Xiayu Jiao, and Joel W Hay. 2021.
      "Health-Related Quality-of-Life in the Swedish General Population during COVID-19
      Based on Pre- and Post- Pandemic Outbreak Measurements." *Nordic Journal of Health
      Economics*, September, 18.

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%20Mem
      o%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.

Rajmil, Luis, Anders Hjern, Perran Boran, Geir Gunnlaugsson, Olaf Kraus de Camargo, and Shanti Raman. 2021. "Impact of Lockdown and School Closure on Children's Health and Well-Being during the First Wave of COVID-19: A Narrative Review." *BMJ Paediatrics Open* 5 (1):e001043. https://doi.org/10.1136/bmjpo-2021-001043.

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.

Robert Koch Institut. 2020. "Epidemiologisches Bulletin 38/2020." . . *September*, 10.

Robertson, Lori. 2022. "What We've Learned About So-Called 'Lockdowns' and the COVID-19 Pandemic." FactCheck.Org. March 8, 2022. https://www.factcheck.org/2022/03/scicheck-what-weve-learned-about-so-called-lockdowns-and-the-covid-19-pandemic/.

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

Rose, Susan, Liz Twist, Pippa Lord, Simon Rutt, Karim Badr, Chris Hope, and Ben Styles. 2021. "Impact of School Closures and Subsequent Support Strategies on Attainment and Socio-Emotional Wellbeing in Key Stage 1: Interim Paper 1," January. https://educationendowmentfoundation.org.uk/public/files/Publications/Covid-19_Resources/Impact_of_school_closures_KS1_interim_findings_paper_-_Jan_2021.pdf.

Sadeghi, McKenzie. 2022. "Fact Check: Working Paper Isn't Proof COVID-19 Restrictions Don't Work, Experts Say." USA Today. 2022. https://www.usatoday.com/story/news/factcheck/2022/02/18/fact-check-working-paper-isnt-proof-lockdown-dont-work-experts-say/6749032001/.

Science Media Centre. 2022. "Expert Reaction to a Preprint Looking at the Impact of Lockdowns, as Posted on the Johns Hopkins Krieger School of Arts and Sciences Website." February 3, 2022. https://www.sciencemediacentre.org/expert-reaction-to-a-preprint-looking-at-the-impact-of-lockdowns-as-posted-on-the-john-hopkins-krieger-school-of-arts-and-sciences-website/.

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. 2021a. "All or Nothing? Partial Business Shutdowns and COVID-19 Fatality Growth." *SSRN Electronic Journal*.
https://doi.org/10.2139/ssrn.3835510.

———. 2021b. "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.

Statens Serum Institut. 2020. "Uge 13-2020 - Markant fald i influenza i de seneste to uger." March 29, 2020. https://www.ssi.dk/aktuelt/nyhedsbreve/influenza-nyt/2019-2020/13-2020.

———. 2021. "Tillægsnotat af den 21. marts 2021 til "Notat af d. 16. marts 2021: Prognoser for genåbningen d. 1. marts og d. 15. marts, samt genåbningsscenarier d. 6. april","" March, 16.

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.

Talic, Stella, Shivangi Shah, Holly Wild, Danijela Gasevic, Ashika Maharaj, Zanfina Ademi, Xue Li, et al. 2021. "Effectiveness of Public Health Measures in Reducing the Incidence of Covid-19, SARS-CoV-2 Transmission, and Covid-19 Mortality: Systematic Review and Meta-Analysis," 15. https://doi.org/doi.org/10.1136/bmj-2021-068302.

Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Rev. and Expanded ed. New York: Penguin Books.

The Royal Swedish Academy of Sciences. 2021. "Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 202," 48. https://www.nobelprize.org/uploads/2021/10/advanced-economicsciencesprize2021.pdf.

Thornton, Jacqui. 2022. "Covid-19: Trust in Government and Other People Linked with Lower Infection Rate and Higher Vaccination Uptake." *BMJ*, February, o292. https://doi.org/10.1136/bmj.o292.

Toya, Hideki, and Mark L. Skidmore. 2021. "A Cross-Country Analysis of the Determinants of COVID-19 Fatalities." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3832483.

Tsai, Yung-Yu, and Yang Tzu-Ting. 2021. "Measuring Voluntary Responses in Healthcare Utilization During and After COVID-19 Pandemic: Evidence from Taiwan." Preprint. In Review. https://doi.org/10.21203/rs.3.rs-681310/v1.

UK Department of Health, Social Services, and Public Safety. 2011. "UK Influenza Pandemic Preparedness Strategy."

UNICEF. 2020. "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.

Wang, Guihua. 2022. "Stay at Home to Stay Safe: Effectiveness of Stay-at-home Orders in Containing the COVID-19 Pandemic." *Production and Operations Management*, February, poms.13685. https://doi.org/10.1111/poms.13685.

Weber, Enzo. 2020. "Which Measures Flattened the Curve in Germany?" *COVID ECONOMICS*, no. 24 (June):205–17. https://cepr.org/sites/default/files/news/CovidEconomics24.pdf.

WHO. 2019. *Non-Pharmaceutical Public Health Measures for Mitigating the Risk and Impact of Epidemic and Pandemic Influenza*. https://apps.who.int/iris/rest/bitstreams/1257621/retrieve.

———. 2020a. "Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-NCoV)." January 30, 2020. https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov).

———. 2020b. "WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020." March 11, 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

———. 2022. "Burden of Influenza." 2022. https://www.euro.who.int/en/health-topics/communicable-diseases/influenza/seasonal-influenza/burden-of-influenza.

Wilder-Smith, Annelies, and Sarah Osman. 2020. "Public Health Emergencies of International Concern: A Historic Overview." *Journal of Travel Medicine* 27 (8):taaa227. https://doi.org/10.1093/jtm/taaa227.

Wood, Reed M., Gina Yannitell Reinhardt, Babak Rezaeedaryakenari, and Leah Windsor. 2021. "Resisting Lockdown: The Influence of COVID-19 Restrictions on Social Unrest." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3951980.

Woolhouse, Mark Edward John. 2022. *The Year the World Went Mad*.

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.

Yang, Qi-Cheng, Xia Chen, Chun-Ping Chang, Di Chen, and Yu Hao. 2021. "What Is the Relationship between Government Response and COVID-19 Pandemics? Global Evidence of 118 Countries." *Structural Change and Economic Dynamics* 59 (December):98–107. https://doi.org/10.1016/j.strueco.2021.08.007.

Zahran, Sammy, Levi Altringer, and Ashok Prasad. 2022. "The Longevity-Frailty Hypothesis: Evidence from COVID-19 Death Rates in Europe." *International Journal of Environmental Research and Public Health* 19 (4):2434. https://doi.org/10.3390/ijerph19042434.

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.

Zhao, Pengcheng, Nan Zhang, and Yuguo Li. 2020. "A Comparison of Infection Venues of COVID-19 Case Clusters in Northeast China." *International Journal of Environmental Research and Public Health* 17 (11):3955. https://doi.org/10.3390/ijerph17113955.

Ziedan, Engy, Kosali Simon, and Coady Wing. 2020. "Effects of State COVID-19 Closure Policy on NON-COVID-19 Health Care Utilization." w27621. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w27621.