Public Choice
https://doi.org/10.1007/s11127-024-01216-7

Were COVID-19 lockdowns worth it? A meta-analysis

Jonas Herby[1] · Lars Jonung[2] · Steve H. Hanke[3]

Received: 27 October 2023 / Accepted: 11 October 2024

Abstract
Following the onset of the COVID-19 pandemic, an unprecedented use of mandatory lockdowns—defined as the imposition of at least one compulsory, non-pharmaceutical intervention—took place. We conduct a meta-analysis to determine the effect of these lockdowns on COVID-19 mortality. Our meta-analysis finds that lockdowns in the spring of 2020 had a relatively small effect on COVID-19 mortality and is consistent with the view that voluntary changes in behavior, such as social distancing, played an important role in mitigating the pandemic. Given the enormous economic costs associated with lockdowns and our findings of the relatively small health benefits, the efficacy of lockdowns during the COVID-19 pandemic is called into question.

Keywords COVID-19 · Restrictions · Lockdown · Non-pharmaceutical interventions · Mortality · Meta-analysis

JEL Classification I18 · I38 · D19

1 Introduction[1]

In April of 1932, Joseph Schumpeter delivered a lecture at the University of Bonn at a conference on social and economic development. It was then that Schumpeter said, 'When something fundamentally new occurs in the world, we are confronted by an enigma' (Stolper, 1994, p. 110). Such an enigma is what confronted policymakers with the outbreak of COVID-19 in early 2020. Indeed, policymakers were faced with a serious information deficit, with the factors relevant for decision-making wrapped in a fog of uncertainty. This setting was ripe for an 'availability cascade' in which decision-makers, knowing little about COVID-19, took their cues from the first movers (Kuran and Sunstein, 2007). Not

[1] This paper is based on a longer manuscript (Herby et al., 2022).

✉ Lars Jonung
lars.jonung@nek.lu.se

Jonas Herby
jherby@gmail.com

1 Copenhagen, Denmark

2 Department of Economics at, Lund University, Lund, Sweden

3 The Johns Hopkins University, Baltimore, USA

Published online: 28 November 2024

surprisingly, 'rational herding' and 'a reputational cascade' followed, with most decision-makers going along with the crowd and jumping on the 'lockdown bandwagon.' In consequence, the public choices associated with COVID-19 lockdowns were governed by a simple copycat phenomenon. When lockdowns were imposed in one country, neighboring countries quickly followed suit (Herby et al., 2022, p. 3). The regulatory response to the COVID-19 pandemic harks back to the swine flu episode of 1976. It was then that, at the behest of the U.S. Centers for Disease Control (CDC), a vaccine program was hurriedly put in place. Fortunately, a comprehensive *ex-post* analysis of the swine flu episode was conducted by Richard Neustadt and Harvey Fineberg (Neustadt and Fineberg, 1978). The authors found that theories were "spun" by medical professionals from "meager evidence," with conclusions "fueled by conjectures." And that the head of the CDC, David Sencer, "knew every facet of the empire," and that his lobbying for total swine flu vaccine inoculations was driven largely by "empire building," not an assessment of the benefits and costs of the mandates. So-called "national emergencies" often follow this same pattern of rushed decisions made with a paucity of data and without analyses of their benefits and costs. The results lead to what Robert Higgs diagnosed as the ratchet effect: whenever a national emergency is proclaimed, the size and scope of the government bureaus and agencies responsible for the emergency response permanently ratchet up (Higgs, 1987). The list of major public policy decisions made in such a way is long, and includes the U.S. "War on Terror" (Niskanen, 2006).

Amidst the COVID-19 pandemic, at least 99 governments declared a state of emergency (Bjørnskov and Voigt, 2022a) while many others were endowed with powers akin to those granted under constitutional emergency provisions. These powers facilitated and permitted the use of non-pharmaceutical interventions (NPIs), commonly known as "lockdowns." Such policies typically entailed mandatory restrictions on internal movement, closures of schools and businesses, and bans on international travel, among others.

Lockdowns were frequently instituted based on the advice from selected COVID-19-focused health experts (Easterly, 2013; Koppl, 2023). However, the empirical evidence supporting their guidance was limited while both official pandemic plans and reviews of available evidence warned against lockdowns (WHO, 2019), hinting at a pronounced action bias.[2] Furthermore, these experts' perspectives might have been influenced by prevailing public policies, curtailed media freedom (Bjørnskov and Voigt, 2022a), and public opinion, potentially compromising their objectivity (Murphy et al., 2021).

Despite the prevalent popular belief that lockdowns were effective during the COVID-19 pandemic, recent studies indicate that the utilization of emergency powers correlates positively with the number of deaths in the aftermath of natural disasters (Bjørnskov and Voigt, 2022b). A plausible explanation for this might be the lack of robust systemic mechanisms that assist policymakers in discerning effective from less-effective policy responses during disasters. This same knowledge gap might have been substantial during the pandemic (Pennington, 2021). Thus, the prevailing belief in the efficacy of lockdowns could be misplaced, potentially leading to their extended use (Dur, 2001; Pennington, 2023).

While lockdowns were implemented to manage pandemic externalities, infectious disease externalities in the absence of government intervention can be less significant and less costly to society than commonly believed (Leeson and Rouanet, 2021). Economic theory

[2] The concept of "action bias" refers to the tendency of individuals to prefer taking action rather than remaining inactive or maintaining the status quo, especially in situations of uncertainty. This bias can manifest in decision-making contexts, where the pressure to do something, even if it is not the most efficient or effective response, can lead to suboptimal outcomes (Patt and Zeckhauser 2000).

suggests that lockdowns are far from an efficient means of regulating pandemic-related externalities, raising the possibility that allowing civil society to self-regulate might have been a more effective approach (Herby, 2020; Mulligan, 2023). In addition, research conducted before the pandemic suggests that public health regulations are often influenced by private interests rather than public ones. The allocation of public health resources frequently reflects these private interests, leading to policies that may have unintended, adverse effects. Instead of promoting health and consumer welfare, such policies can sometimes undermine them (Leeson and Thompson, 2023).

These considerations indicate that lockdowns may represent a significant instance of government failure.

The purpose of our meta-study is therefore to determine whether lockdowns were, in fact, effective in reducing COVID-19 mortality.[3,4] As such, we aim to estimate the benefit side of the cost–benefit analysis necessary to evaluate the desirability of locking down society in an epidemic emergency.

We use "NPI" to describe *any government mandate which directly restricts peoples' possibilities*. Our definition does *not* include governmental recommendations, governmental information campaigns, access to mass testing, voluntary social distancing, etc., but *does* include mandated interventions such as closing schools or businesses and mandated face masks. We define a *lockdown* as any policy consisting of at least one NPI.[5,6]

Compared to other reviews (Allen, 2021; Herby, 2021a), the main difference in our approach is that we carry out a systematic and comprehensive search strategy to identify all potentially relevant papers and carry out a meta-analysis combining evidence from several existing studies to answer the question we pose. In our meta-analysis, we aim to present results capable of being systematically assessed and used to derive overall conclusions.

2 Methodology

The methodology follows our protocol posted in 2021 (Herby et al., 2021).[7]

[3] Physical distancing is effective. 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 lockdowns work (Herby et al., 2022, p. 1).

[4] 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 influenza pandemic, the evidence of the effectiveness of such measures was, in general, very low (WHO 2019).

[5] 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. We use shelter-in-place order to describe the former use of the term "lockdown" (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).

[6] 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*. We note that our results apply to the average lockdowns in Europe and the United States in the spring of 2020 when most countries implemented several NPIs.

[7] The protocol was first published June 23, 2021, and last updated on October 28, 2021 (see Supplementary Material A).

2.1 Eligibility criteria

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 of infections, hospitalizations, or other similar measures.[8]

We exclude studies which do not use a counterfactual difference-in-difference approach (DD).[9] DD removes biases in post-intervention period comparisons between the treatment and control group that could be the result of permanent differences between those groups (e.g. caused by coincidences early in the pandemic[10]), 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). A potentially important bias is the reverse causality problem.[11] However, for a number of reasons we believe that this bias is relatively unimportant:

First, several studies explicitly claim that they examine the actual causal relationship between lockdowns and COVID-19 mortality. Some studies use instrumental variables (Bjørnskov, 2021a), lagged dependents (Goldstein et al., 2021; Yang et al., 2021), or other techniques to establish a causal relationship, while others make causality probable using arguments.[12]

Second, several studies find that government pandemic policies were strongly driven by the policies initiated in neighboring countries rather than by the severity of the pandemic

[8] 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 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 some studies find that SIPOs reduce the number of cases but have no significant effect on the number of COVID-19 deaths (Chernozhukov et al., 2021). 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 (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). Finally, mortality is hierarchically the most important outcome (GRADEpro 2013).

[9] For a more in-depth discussion of the difference-in-difference approach, we refer to David McKenzie at the World Bank (McKenzie 2020, 2021, 2022a, 2022).

[10] 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; Björk et al., 2021; Herby et al., 2022, p. 81).

[11] If increasing infection rates lead governments to introduce lockdown policies, and declining infection rates subsequently lead them to ease lockdowns, the estimated association between policy stringency and mortality is biased (Bjørnskov 2021a). It is essential for the DD estimates that the treatment is the only important factor that differs between treatment and control group in the post-treatment period (Chabé-Ferret 2017; Goodman-Bacon and Marcus 2020). This may not be satisfied if there is a great degree of reverse causality.

[12] An argument could be "estimated case reductions accelerate over time, becoming largest after 20 days following enactment of a SIPO. These findings are consistent with a causal interpretation" (Dave et al., 2021) or "prior to first Covid-19 death (when policymakers could not possibly be reacting to deaths in their own country)" (Stokes et al., 2020).

in their own countries (Engler et al., 2021; Mistur et al., 2023; Sebhatu et al., 2020). In other words, an availability cascade was driving public policy (Kuran & Sunstein, 2007). Hence, it was not the severity of the pandemic that drove the adoption of lockdowns, but rather the propensity to copy policies initiated by neighboring countries.

Third, there were very few deaths (and cases) on the date countries locked down. Of 37 European countries with more than 100,000 inhabitants, no country had more than one registered COVID-19 deaths per million on the day the stringency index crossed 60. And the decision to lock down was often taken several days earlier where mortalities were even lower (Herby et al., 2022, p. 64).

One study shows that the death rate *does* predict the stringency of policies adopted in various countries (Sebhatu et al., 2020). But the effect is negligible, explaining only 2.1 stringency points on average (in comparison, the gap between the strictest and most lenient lockdowns in Europe was between 67 and 92 stringency points in the period from 16 March to 15 April 2020) (Herby et al., 2022, sec. 5.2.2).

While DD might not remove all of the biases, it "provides important advantages over methods like before-and-after comparisons and interrupted time-series designs" (Goodman-Bacon and Marcus, 2020).

The exclusion of studies that do not use a counterfactual DD approach means that we exclude all studies where the counterfactual is based on *forecasting* (for example, using a SIR-model calibrated on mortality data).[13] Hence, we exclude studies like the much-cited, but often criticized (Hanke and Dowd, 2022), study from Imperial College London (Ferguson et al., 2020) which predicted that a suppression strategy would reduce COVID-19 mortality by up to 99%.[14] We also exclude all studies based on interrupted time series designs.[15] Our criteria also exclude a much-cited paper (Flaxman et al., 2020), which – based on the very problematic (implicit) assumption that voluntary social distancing had zero effect – claimed that lockdowns saved three million lives in Europe.[16]

Our protocol also excludes studies with little jurisdictional variance.[17] These excluded studies focused on high profile places such as Italy, Sweden, and New York, that – possibly because they were hit early and were surprised by the pandemic – experienced very high

[13] A common problem with epidemiological models is that they don't take spontaneous behavior changes into account and therefore produce erroneous forecasts (Coccia 2023). And even when they do, these behavior changes and consequently the results are based on the authors' assumptions rather than empirical evidence (Atkeson 2023; Toxvaerd 2020).

[14] 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 (Ferguson et al., 2020, p. 13). 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 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%).

[15] 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 for (Herby et al., 2022, p. 13).

[16] The authors assumed 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.

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

death tolls during the first wave, and cannot – even when combined – be assumed to produce unbiased knowledge on the effects of lockdown measures due to selection bias (see Supplementary Material A). Our protocol also excludes *synthetic control studies* because of too little jurisdictional variance, as these studies examine the effect of lockdowns based on a single country/state compared to a synthetic counterfactual. Empirical problems inherent to all synthetic control studies of COVID-19 also encourage exclusion: that the synthetic control should be fitted based on a long period before the intervention or the event one is studying the consequences of – i.e., the lockdown (Abadie, 2021; Bjørnskov, 2021b). It is not possible to address this problem for the coronavirus pandemic, as there clearly is no long period with coronavirus before the lockdown. Hence, the synthetic control method is, by design, not well suited for studying the effects of lockdowns. Retrospectively, excluding these studies in our protocol may have been unnecessary and one to two of the excluded studies could have been included in our meta-study without imposing selection bias (see Supplementary Material A). However, the inclusion of these studies would not have altered our meta-results substantially; in fact, our conclusions would be stronger. That said, we think it prudent to avoid changing our research protocol *ex post facto*.

Since we evaluate the general effect of lockdowns, i.e., whether lockdowns on average have been effective in reducing COVID-19 mortality, we also exclude studies which solely analyze the effect of optimally timed lockdowns in contrast to less well-timed lockdowns, as they will by design find inflated effects of the average lockdown. *If* optimal timing is important, these studies will neglect all the less well-timed lockdowns implemented around the world.[18,19] Also, 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).[20]

2.2 Empirical considerations

Several studies adopt multiple models to understand the effect of lockdowns. For example, one study (Bjørnskov, 2021a) estimated 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.

[18] 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 criterion mid-September 2021 (documentation can be provided on request).

[19] The rationale behind the optimal timing-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 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 (Abernethy and Glass 2022; Oraby et al., 2021).

[20] We find indications from the COVID-19 pandemic and studies on the 1918 influenza pandemic (Hatchett et al., 2007; Herby 2021b, 2022; Markel et al., 2007) that the alleged importance of the timing of lockdowns may be ill-founded. COVID-19 mortality data show that all European countries and U.S. states which were hit hard and early by COVID-19 in the spring of 2020 experienced high overall mortality rates; whereas, *all* countries and U.S. states hit relatively late experienced low mortality rates regardless of the timing of lockdowns(Herby 2021b; Herby et al., 2022, Fig. 7). We find the same pattern for the Spanish Flu in 1918 (Herby et al., 2022, Fig. 15).

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. In one interesting example (Chernozhukov et al., 2021), the authors estimated two models: one with and one 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 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 (Doucouliagos & Paldam, 2008).[21]

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

3 Results

3.1 Identification process

Figure. 1 presents an overview of our identification process which followed our protocol posted in 2021 (Herby et al., 2021).[22] It uses a flow diagram designed according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). 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. Of 19,646 studies identified during our database searches,[23] 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

[21] For the mentioned study (Chernozhukov et al., 2021) we can only include the model estimating large effects of lockdowns, as the authors reported the counterfactual effect for this model only (Herby et al., 2022, tbl. 19).

[22] The protocol was first published June 23, 2021, and last updated on October 28, 2021 (see Supplementary Material A).

[23] The studies 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, and a methodology search string. We also required mentions of “deaths”, “death”, and/or “mortality”.

of the 146 were excluded, leaving 32[24] eligible studies of which the estimates from 22[25] could be converted to standardized estimates for the meta-analysis.[26] 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.[27,28]

3.2 *Meta*-analysis

In the meta-analysis, we include the 22 studies for which we can derive a *standardized estimate,*[29] which is the *relative effect* of lockdowns on COVID-19 mortality (how many fewer deaths there were due to lockdowns compared to a situation without lockdown). For some studies, the authors state the relative effect, and our standardized estimate is thus readily available.[30] For other studies, their estimates are converted to our standardized estimate using a hypothetical framework where either all jurisdictions or none would be subject to measures going beyond the minimum. For example, if a study finds that lockdowns reduce

[24] (Alderman and Harjoto 2020; An et al., 2021; Ashraf 2020; Berry et al., 2021; Bjørnskov 2021a; Blanco et al., 2020; Bonardi et al., 2020; Chernozhukov et al., 2021; Chisadza et al., 2021; Clyde et al., 2021; Dave et al., 2021; Dergiades et al., 2020; Ertem et al., 2021; Fakir and Bharati 2021; Fowler et al., 2021; Fuller et al., 2021; Gibson 2020; Goldstein et al., 2021; Guo et al., 2021; Hale et al., 2020; Hale, Angrist, Hale, et al., 2021a, 2021b; Leffler et al., 2020; Li et al., 2021; Mccafferty and Ashley 2021; Pan et al., 2020; Pincombe et al., 2021; Sears et al., 2020; Shiva and Molana 2021; Spiegel and Tookes 2021a, 2021b; Stokes et al., 2020; Yang et al., 2021).

[25] (Alderman and Harjoto 2020; An et al., 2021; Ashraf 2020; Berry et al., 2021; Bjørnskov 2021a; Bonardi et al., 2020; Chernozhukov et al., 2021; Chisadza et al., 2021; Dave et al., 2021; Ertem et al., 2021; Fowler et al., 2021; Fuller et al., 2021; Gibson 2020; Goldstein et al., 2021; Guo et al., 2021; Hale, Angrist, Hale, et al., 2021a, 2021b; Leffler et al., 2020; Sears et al., 2020; Shiva and Molana 2021; Spiegel and Tookes 2021a; Stokes et al., 2020; Yang et al., 2021).

[26] For more details on our search strategy we refer to the longer manuscript (Herby et al., 2022, sec. 2.1).

[27] The inclusion rate in related reviews are: 72 of 36,729 identified studies (0.2%) (Talic et al., 2021), 35 of 12,523 studies (0.3%) (Iezadi et al., 2021), 26 of 2,176 studies (1.2%) (Rezapour et al., 2021), 47 of 1,649 studies (2.9%) (Zhang et al., 2021), and 14 of 623 studies (2.2%) (Johanna et al., 2020). A major reason for the difference in our inclusion rate is the choice of search strategy. Three of the reviews (Johanna et al., 2020; Rezapour et al., 2021; Zhang et al., 2021) identified studies by searching publication databases such as PubMed, Scopus, Web of Science, SAGE, etc., while our search in Google Scholar is broader. 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%.

[28] Our meta-analysis would have been more efficacious if there would have been more studies that qualified for inclusion. This is particularly true for specific NPIs. Even though we used all of the studies that qualified, the number was relatively small.

[29] The estimates from 10 of the 32 eligible studies cannot be converted to our standardized estimates and are excluded from the meta-analysis (Herby et al., 2022, sec. 3.2). The conclusions in these 10 excluded studies are, overall, consistent with the conclusions in the included studies (Herby et al., 2022, tbl. 3).

[30] This includes studies estimating the effect of lockdowns on mortality growth rates, unless the authors calculated a counterfactual scenario. 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.

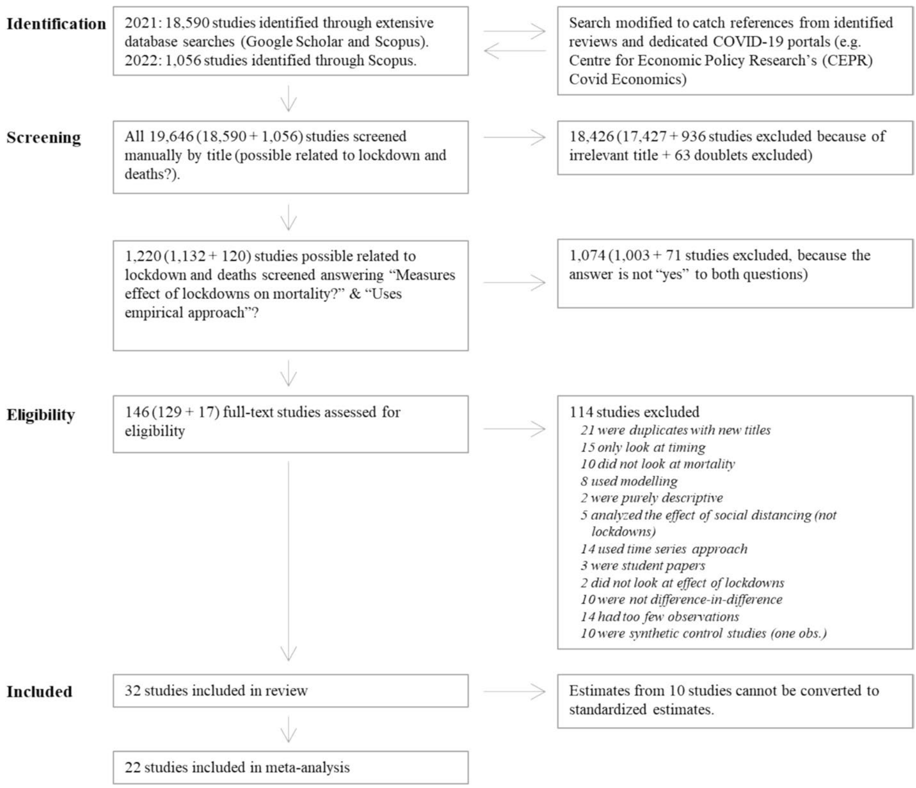

Fig. 1 The PRISMA flow diagram for the selection of studies

COVID-19 mortality by six deaths per million and there were 294 reported deaths per million at the end of the study period, our standardized estimateis two percent.[31,32]

We also convert standard errors (SE) and use the precision of each estimate (defined as 1/SE) to calculate the *precision-weighted average* (PWA) of all estimates (Paldam, 2015; Stanley and Doucouliagos, 2010).[33] The PWA is our primary indicator of the efficacy of lockdowns, but we also report arithmetic averages and medians in the meta-analysis, as well as the alternative weighted averages using the inverse of each estimate's variance (1/SE^2).[34]

[31] We have described for each paper how we interpret the estimates and how they are converted to a standardized estimate (Herby et al., 2022, tbl. 19). All calculations are available in the supplementary spreadsheet.

[32] Note that this approach implicitly assumes that the results also hold for units outside the sample used in the difference-in-difference analysis and thus the assumptions required are stronger.

[33] 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).

[34] The use of 1/SE rather than 1/SE^2 can mitigate the impact of outliers – studies with exceptionally small or large effect sizes coupled with small standard errors. Given the controversies about the topic at hand and

3.3 Handling biases and uncertainty

3.3.1 *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, since there is currently no consensus as to best practices and/or an established scientific framework for evaluating the effectiveness of lockdowns, such an evaluation risks being subjective. Instead, we investigate the importance of any biases by dividing the studies into eight "bias dimensions."[35] In general, our bias dimensions do not show notable biases and are not reported here (Herby et al., 2022, tbls. 6–8), although shelter-in-place order-studies using shorter data periods seem to find much larger effects on mortality.[36]

3.3.2 Selective reporting

Selective reporting is the tendency of editors, referees, or authors themselves to prefer empirical estimates that are conclusive, have a particular sign supported by theory or intuition, or both (Havranek et al., 2015). We investigate two aspects of selective reporting.

First, as our analysis focuses on authors' main specifications, we examine whether authors have a bias in their preferences, causing them to favor estimates that are particularly large or particularly small. We define this as 'within bias,' as this bias exists within each study.

Second, we explore whether researchers, as a group, tend to avoid publishing (including working papers) certain types of results (e.g., analyses that counter-intuitively find that lockdowns led to more deaths). We refer to this as 'between bias,' as it reflects a collective tendency across different studies to avoid publishing results that are contrary to prevailing theories or expectations. We avoid using the term 'publication bias' because we include working papers in our meta-analysis, but 'between bias' is comparable to the general notion of publication bias.

Our analyses of selective reporting can be found in Supplementary Material B. We find no signs of 'within bias.' The estimates from the main specification are within the range of estimates for each study. Also, the estimates from the main specifications are very close to the precision-weighted estimates within each study.

We do find some evidence of 'between bias.' Our funnel plots suggest that estimates finding that lockdowns increase COVID-19 mortality are less likely to be reported than those indicating a reduction in mortality. However, our meta-regression analyses reveal that this bias is not statistically significant and that the PET meta-average is very close to the PWA's.

Footnote 34 (continued)

the relatively few available studies, we believe this approach offers advantages in this specific context, particularly because we do not know the source of potential biases (Paldam 2015; Stanley and Doucouliagos 2012). We note that the weights yield similar results which is normally the case (Paldam 2015), although the alternative weights, $1/SE^2$, generally result in smaller estimated effects of lockdowns.

[35] The eight bias dimensions are: 1) Peer-reviewed vs. working papers, 2) long vs. short data period, 3) no early effect on mortality, 4) lag vs. no lag of policy measures, 5) panel vs. no panel estimation, 6) verified vs. unverified data, 7) addresses vs. does not address causality, and 8) social sciences vs. other sciences (Herby et al., 2022, p. 33).

[36] A study is categorized as using a short data period, if the data period ends before May 31, 2020.

3.3.3 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.[37] We report the result of our four sensitivity analyses as a span (from/to) at the bottom of each table.

3.3.4 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 (Herby et al., 2022, tbl. 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, shelter-in-place order-studies, and studies analyzing specific NPIs.

3.4 The effect of lockdowns based on stringency index studies

The results from the eight studies examining the link between lockdown stringency and COVID-19 mortality are presented in Table 1 below. All studies are based on the COVID-19 Government Response Tracker's stringency index of Oxford University's Blavatnik School of Government (OxCGRT) (Hale et al., 2020).[38]

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.[39] Many countries followed similar patterns and almost all countries closed schools, while only a few countries issued shelter-in-place orders without closing businesses too. 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 working from home)) (Hale et al.,

[37] For instance, in one sensitivity analysis, we replace the estimate from an outlier which found that the average lockdown increased COVID-19 mortality (Chisadza et al., 2021) with the nearest estimate (Bjørnskov 2021a) and recalculate the PWA.

[38] It is important to note that the index is far from perfect. It is certainly possible to identify errors and omissions in the index (Book 2020). 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 are few better available measures to adopt for cross-country comparisons.

[39] One parameter, "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 states in the U.S. in the OxCGRT database, only Andorra, Belarus, Bosnia and Herzegovina, Faeroe Islands, and Moldova – less than 1.6% of the population – did not receive the maximum score by March 20, 2020, so the parameter simply shifts the index upward in parallel and should not have a notable impact on our conclusions.

2021a, 2021b). However, the effect of including recommendations in the index is primarily to make an upward, parallel shift in the index and should not alter the results relative to our focus on mandated NPIs.

Since the stringency index includes recommendations, the standardized estimates show the effect of the average lockdown in Europe and the United States (with average stringencies of 76 and 74, respectively, between March 16th and April 15th, 2020)[40] compared to a policy based solely on recommendations (stringency 44).[41,42]

Table 1 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%, a median of 5.8%, and 2.8% when weighted by inverse variance.

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.[43]

Hence, based on the stringency index studies, we find that mandated lockdowns in Europe and the United States had a relatively small effect on COVID-19 mortality rates.

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

[41] For example, one study (Ashraf 2020) estimated that the effect of stricter lockdowns was -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 based solely on recommendations (76 vs. 44). In the United States, it was 30 points stricter. Hence, we calculate the total effect of the average lockdown compared to a policy based solely on non-mandated recommendations to be (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 and 333 million, respectively, the total effect as estimated 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, 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%, c.f. Table 1.

[42] 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 might not be precisely the 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.

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

Table 1 Standardized 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 (2021a, 2021b)	−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%**	0.553%	181
Yang et al. (2021)	−16.3%**	4.523%	22
Hale et al., (2021a, 2021b)	−16.9%**	2.812%	36
Precision-weighted average (arithmetic average / median / weighted by inverse variance)	−3.2% (−8.9%/−5.8%/−2.8%)		
Sensitivity analysis (quality-adjusted PWA)	−4.4% to −3.0% (−3.8%)		

** (*) denote significance at $p < 0.01$ ($p < 0.05$). A negative number corresponds to fewer deaths, so −5% means 5% lower COVID-19 mortality

Since almost all countries instituted some pandemic response, we cannot reject the hypothesis that some NPIs would be required to spur voluntary behavioral changes.[44,45] Thus, Table 1 may simply indicate that the most lenient lockdowns had virtually the same effect on mortality as stricter lockdowns. Thus, the incremental effect of stricter lockdowns is small.

In the following section, we will look at the effect of lockdowns based on studies which examine specific NPIs.

[44] It is crucial to spread information rapidly during a pandemic. It is worth noting that the employment of ordinary information channels to spread information might have been insufficient in early 2020. In this case, lockdowns – because they are highly unusual – most probably worked as information instruments that signaled to the population that the public health situation was serious. As a result, their announcement effect was pronounced, and peoples' voluntary behavior was adjusted accordingly. This announcement effect has been observed for emergency declarations, which appear early and separate from other policies (Gupta et al., 2020). Using survey data immediately (same day) before and after Boris Johnson announced the U.K. lockdown, a study (Eggers and Harding 2021) found 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." People who responded after the announcement were more likely to respond by stating "I fear for my future" and "I have started not going out at all." These results are consistent with observations from the Nordic countries (Herby et al., 2022, Fig. 8) and may indicate that even minor restrictions are sufficient to spur significant voluntary behavioral changes.

[45] 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. It has been described in detail how such indices can be "fuzzy" metrics that are often replete with mismeasurements (Morgenstern 1963).

3.5 The effect of lockdowns based on studies of specific NPIs

A total of 14 studies examined the effect of specific NPIs.[46] The definition of these specific NPIs varies from study to study which makes comparisons difficult. The variety of definitions can be seen in the analysis of non-essential business closures and bar/restaurant closures, which has been examined as a combined parameter (the average of business closures and bar/restaurant closures in each state) (Chernozhukov et al., 2021), as bar and/or restaurant closures but not business closures (Spiegel and Tookes, 2021a), and as both business closures and bar/restaurant closures independently (Guo et al., 2021).

Some studies include several NPIs (Spiegel and Tookes, 2021a; Stokes et al., 2020), while others cover very few. For example, one study looked at internal lockdowns of any type, mask recommendations, and international travel restrictions (Leffler et al., 2020), and six studies only look at shelter-in-place orders (Alderman and Harjoto, 2020; Berry et al., 2021; Dave et al., 2021; Fowler et al., 2021; Gibson, 2020; Sears et al., 2020). Too few NPIs in a model are potentially a problem because they can capture the effect of excluded NPIs.[47] 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 deaths if people adjust their behavior according to official case numbers and are less careful if case numbers fall following bar closures.

Table 2 below summarizes our results on specific NPIs.[48] The central precision-weighted average in column 2 is small for most NPIs and even positive for limiting gatherings. Only mask 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 and column 5 presents the results when estimates are weighted by inverse variance. The precision-weighted averages are generally robust to the sensitivity analyses, and both the quality-adjusted PWA and the weighting by inverse variance generally finds 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, while limiting gatherings is the only ‘preferable’ NPI using inverse variance weights).

The overview in Table 2 allows us to estimate the effect of the average lockdown policy in the spring of 2020, just as we did for the stringency studies in Sect. 3.4. First, we use OxCGRT data to calculate the share of the population that faced each of the NPIs from Table 2 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 typically flattened after this period. We only look at whether each NPI was mandated or not, and not

[46] 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.

[47] Say two studies, A and B, examine the effect of lockdowns. Study A examines school closures and business closures, whereas study B examines business closures 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.

[48] Detailed results are available for each of the NPIs (Herby et al., 2022) (see column 6 for a reference to the relevant table for each specific NPI).

Table 2 Summary of standardized estimates of specific non-pharmaceutical interventions (NPIs)

1. NPI (number of studies)*	2. Precision-weighted average (PWA)	3. Sensitivity analysis	4. Quality-adjusted PWA	5. Weighted by inverse variance	6. Reference to detailed results in longer manuscript(Herby et al., 2022)
Shelter-in-place order (12)	−2.0%	−4.1% to −1.4%	−1.8%	−0.9%	Table 7
Business closure (5)	−7.5%	−9.3% to −6.6%	−7.8%	−5.7%	Table 9
School closures (4)	−5.9%	−6.2% to −2.5%	−6.7%	−7.8%	Table 10
Limiting gatherings (4)	5.9%	4.9% to 8.9%	5.9%	4.3%	Table 11
Travel restrictions (5)	−3.4%	−4.7% to −0.4%	−3.2%	−5.9%	Table 12
Mask mandates (3)	−18.7%	−19.9% to −12.5%	−18.7%	−17.9%	Table 13
Public events cancellation (1)	2.0%	n/a	2.0%	2.0%	Table 14
Public transportation closures (1)	0.1%	n/a	0.1%	0.1%	Table 14
Internal movement restrictions (1)	1.4%	n/a	1.4%	1.4%	Table14

The table summarizes the results from Table 7 and Table 9, Table 10 , Table 11, Table 12, Table 13, and Table 14 in the longer manuscript (Herby et al., 2022). A negative number corresponds to fewer deaths, so −5% means 5% lower COVID-19 mortality

* The number in parentheses in column 1 indicates the number of studies covering the specific NPI. The total is larger than 14 because several studies look at multiple NPIs

whether it was strict or more lenient.[49] 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 risks being biased towards the larger effect, since the estimate in each study may capture the effect of multiple (omitted) NPIs (see also footnote 46).

Based on this approach and with the bias towards overestimating the effect of lockdowns in mind, Table 3 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 0.5% (worst case) to 15.8% (best case). The quality-adjusted PWA is a 11.5% reduction in mortality rates, while the result using inverse variance weights is 6.5%. The 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 (Ferguson et al., 2020). To put the estimate in perspective, the 10.7% corresponds to 23,000 avoided COVID-19 deaths in Europe and 16,000 avoided COVID-19 deaths in the United States.

4 Discussion

Throughout our meta-analysis, we have focused on the precision-weighted average as our primary indicator of the efficacy of lockdowns. However, as shown in Fig. 2, the overall conclusion holds regardless of which of the studies or measures included in our meta-analysis is emphasized. Figure. 2 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 1 and the estimate based on specific NPIs in Table 3). For comparison, we have added the max and min forecasted estimates from the epidemiological modeling exercises conducted by Imperial College London (Ferguson et al., 2020). Even if we pick the most extreme empirical estimate of the effects of lockdowns on mortality (Fuller et al., 2021), the measured effect of lockdowns is much less than those generated by Imperial College London (Ferguson et al., 2020) and their epidemiological modeling exercises.

Just why do the findings of our meta-analysis about the efficacy of lockdowns seem to differ from the general views embraced by most epidemiologists? We offer four potential explanations (Herby et al., 2022, sec. 5.2.3).

First, people respond voluntarily to new dangers. 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 disease prevention efforts, like social distancing and increased focus on hygiene, becomes elevated when infection rates are high.

Second, mandates only regulate a fraction of our potential contagious contacts. Data from the Robert Koch Institute in Germany (Robert Koch Institut, 2022) showed that 77% 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 society and where potentially

[49] That is, 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.”.

Table 3 Standardized 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) (PWA ·share)	3. Impact on mortality	4. Sensitivity analysis (best case to worst case)	5. Quality-adjusted PWA	6. Weighted by inverse variance
Shelter-in-place order (12)	70%	−1.4%	−2.9% to −1.0%	−1.2%	−0.6%
Business closure (5)a	92%	−6.9%	−8.6% to −6.1%	−7.2%	−5.3%
School closures (4)	97%	−5.7%	−6.0% to −2.4%	−6.5%	0.0%
Limiting gatherings (4)	95%	5.6%	4.7% to 8.4%	5.6%	4.1%
Travel restrictions (5)	93%	−3.1%	−4.3% to −0.4%	−3.0%	−5.5%
Mask mandates (3)b	10%	−1.9%	−2.1% to −1.3%	−1.9%	−1.9%
Public events cancellation (1)	95%	1.9%	1.9% to 1.9%	1.9%	1.9%
Public transportation closures (1)	14%	0.0%	0.0% to 0.0%	0.0%	0.0%
Internal movement restrictions (1)	64%	0.9%	0.9% to 0.9%	0.9%	0.9%
Total impact of the average lockdown policy		−10.7%	−15.8% to −0.5%	−11.5%	−6.5%

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 2 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). Column 5 shows the quality-adjusted PWA, and column 6 the estimated effect when weighted på inverse variance. The total impact of the average lockdown policy is calculated as the product of (1 – (estimates in column 3)) − 1. Implementing all NPIs 100% of the time has an impact of − 26.4%

[a]The effect from closing non-essential businesses is possibly related to the closure of bars (Herby et al., 2022, tbl. 9)

[b]Our result on mask mandates is somewhat in contrast to several reviews of randomized control trials (RCTs) on mask usage (Jefferson et al., 2023; Liu et al., 2021; UK Department of Health, Social Services, and Public Safety 2011; WHO, 2019). 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 mask when they socialize, or masks may function as a constant reminder of the presence of the pandemic)

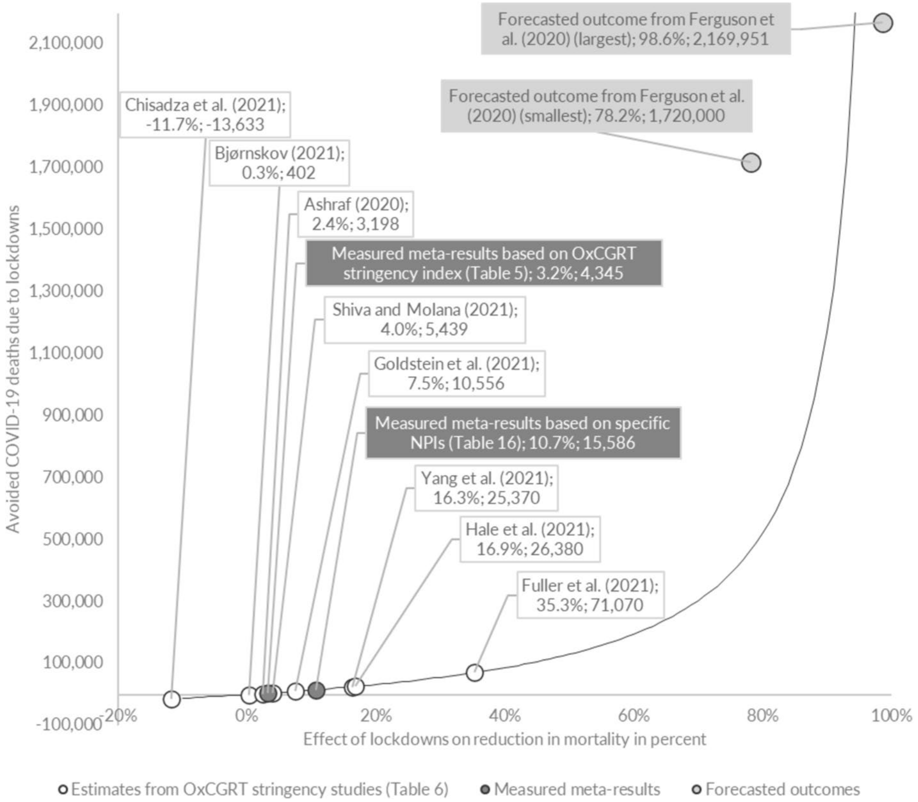

Fig. 2 Divergence between avoided number of deaths in the United States as measured by the meta-results, studies based on the OxCGRT stringency index, and the forecasted outcome from Imperial College London

effective interventions, such as handwashing, coughing etiquette, ventilation, distancing, etc. could neither be regulated nor enforced but relied solely on voluntary behavioral changes.[50]

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. The economic intuition is straightforward. If closing bars and restaurants causes the prevalence of the disease to fall toward zero, the demand for disease prevention efforts like social distancing and increased focus on hygiene also falls towards zero, and the disease will return. In economic terms, lockdowns are substitutes for – not complements to – voluntary behavioral changes (Atkeson, 2021).[51]

Fourth, unintended consequences may play a larger role than recognized. We already pointed to the possible unintended consequence of shelter-in-place orders, which may isolate an infected person at home with his/her family where he/she risks infecting family

[50] Data from other countries show similar patterns (Lee et al., 2020; Lendorf 2021; Munch et al., 2022a, 2022b; Zhao et al., 2020).

[51] This kind of behavior response may also explain why increases in COVID-19 cases were unrelated to levels of vaccination across 68 countries and 2,947 counties in the United States (Subramanian and Kumar 2021). When people are vaccinated and protected against severe disease, they have less reason to be careful.

members with a higher viral load, causing more severe illness. But often, lockdowns have limited peoples' access to safe (outdoor) places such as beaches, parks, and zoos, or included outdoor mask mandates or strict outdoor gathering restrictions, pushing people to meet at less safe (indoor) places. Indeed, we do find evidence that limiting gatherings was counterproductive and increased COVID-19 mortality by 5.9%, see Table 2.

5 Conclusions

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. Indeed, we find that lockdowns had a relatively small effect on COVID-19 mortality. Studies based on the OxCGRT stringency index found that the average lockdown in Europe and the United States only reduced COVID-19 mortality by 3.2%. 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% "range"[52] corresponds to a range of 6,000–23,000 avoided deaths in Europe and 4,000–16,000 avoided deaths in the United States. For context, there are approximately 72,000 flu deaths in Europe and 38,000 flu deaths in the United States each year.[53]

The results of our meta-analysis are consistent with several non-meta reviews, which have concluded that "the most recent research has shown that lockdowns have had, at best, a marginal effect on the number of Covid-19 deaths" (Allen, 2021), and 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" (Herby, 2021a). In addition, a WHO report (World Health Organization Writing Group, 2006) stated that "reports from the 1918 influenza pandemic indicate that …[NPIs]… did not stop or appear to dramatically reduce transmission."[54] Our measured meta-results are also supported by the natural experiments we have been able to identify through our work and by searches in the abstract and citation database Scopus (Herby et al., 2022, tbl. 17). Finally, our results are consistent with new research showing that the use of emergency powers are positively correlated with the number of deaths following natural disasters (Bjørnskov and Voigt, 2022b).

All policy decisions should be based on a comparison of benefits and costs. When it comes to the policy decision associated with the imposition of lockdowns, our meta-analysis provides a considerable amount of new evidence about the benefits of lockdowns. Given our findings that lockdowns had a negligible effect on COVID-19 mortality, it is clear that their benefits were much more limited than had been previously assumed. The costs to society associated with lockdowns must be compared to the health benefits in full cost–benefit calculus when evaluating the use of lockdowns. Given the enormous economic, social,[55] and political costs of lockdowns to society and the findings from our

[52] We use the word "range" to indicate the range of values observed, but we must stress that the 3.2% and 10.7% estimates are derived from two different methods to estimate the overall effect of the average lockdown in Europe and the United States.

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

[54] To avoid confusion, we have replaced the words "social-distancing measures" with "…NPI…" because the WHO, in its cited report, defined "social-distancing measures" as we define NPIs.

[55] For instance, one study finds mask mandates make people more selfish (Cardella et al., 2024).

meta-analysis, lockdowns should only be considered with an utmost degree of caution.[56] Lockdowns are not likely to emerge as a constructive and efficient tool.[57]

Given that lockdown measures were predominantly instituted based on wide-reaching political consensus, our findings offer a possible explanation as to why so few governments have undertaken rigorous assessments of the efficacy of their lockdown strategies, as governments are likely to exhibit hesitancy in evaluating policies when there's a potential to uncover their inefficacy or detrimental effects (Mavrot and Pattyn, 2022).

When it comes to COVID-19, officialdom has gone beyond the failure to conduct *ex-post* evaluations. Indeed, it has actively engaged in the censorship of findings that question the efficacy of COVID-19 lockdown policies. Our experience fosters a healthy dose of public choice cynicism about leaders in the public health field (Bhattacharya and Hanke, 2023).

Appendix 1: Analysis of selective reporting (publication *bias*)

Background

Selective reporting is defined as the tendency of editors, referees, or researchers themselves to prefer empirical estimates that are conclusive, have a particular sign supported by theory or intuition, or support authors' political preferences (Havranek et al., 2015).

A large literature finds selective reporting, e.g. (Doucouliagos & Paldam, 2008; Havranek et al., 2015), and reports how such bias should be modelled (Paldam, 2015, 2022; Stanley & Doucouliagos, 2010).

Given the highly politicized nature of the policy response to the COVID-19 pandemic, such biases are also possible in the literature examining the effects of lockdowns. We therefore investigate two aspects of selective reporting in this context.

In this supplementary section, we investigate two types of biases:

1) *Within bias*: This occurs when researchers exhibit a preference for estimates that are either unusually large or small within a given study.
2) *Between bias*: This refers to a collective tendency among researchers to avoid publishing certain types of results, such as findings that contradict prevailing theories (e.g., analyses suggesting that lockdowns led to more deaths). We use this term instead of 'publication bias' because our meta-analysis includes working papers, though 'between bias' aligns with the general concept of publication bias.

Within bias.

[56] A thorough cost–benefit analysis is beyond the scope of this paper. We stress that it is important to take voluntary behavior changes into account when examining the costs of lockdowns. See pp. 141–148 in the book, "Did Lockdowns Work? The verdict on Covid Restrictions" where Table 18 contains a stylized overview of the costs (Herby et al., 2023, tbl. 18).

[57] See for example a study covering 25 European countries showing that countries with a high level of stringency of lockdowns experienced the sharpest decline in economic growth without displaying lower excess mortality. Lockdowns emerge in this cross-country analysis as highly costly without any significant effect on excess mortality (Andersson and Jonung 2024).

To examine the degree of within bias, we have collected and standardized all estimates and standard errors from each of the studies in the meta-analysis as reported in our main text. Table 4 shows the resulting standardized estimates and standard errors as well as the internal precision-weighted average (internal PWA) for each study.[58] In the column, "Specification", we mark whether the estimate is from the specification which the researchers regard as their main specification (marked with "Main") or from multiple specifications ("Average") (see Sect. 2.2 in main paper).

Figure 3 shows all standardized estimates from studies in the meta-analysis emphasizing the estimate used in the meta-analysis. Although Fig. 3 illustrates variation between studies, there is no visual indication that researchers favor particularly large or particularly small estimates. There is only one study where the estimate from the main specification diverts largely from the other specifications (Yang et al., 2021). Yang et al., (2021) provide four specifications which only differ in the lag between changes in lockdown stringency and death. When the lag is short (0, 7, or 14 days), they find large positive effects (more deaths) of stricter lockdowns. When the lag is long (21 days), they find a small negative effect (fewer deaths). Our selection of the latter is justified by the reverse causality problem (see Sect. 2.1 in the main paper).

In Fig. 4 we compare the estimates from the main specification with the internal PWA within each study. The internal PWA is calculated as the product of each estimate and its corresponding precision (one divided by the estimate's standard error) divided by the sum of precisions. The figure shows that the main specifications are very close to the internal precision-weighted averages except for Yang et al., (2021).

Based on * MERGEFORMAT Fig. 3 and * MERGEFORMAT Fig. 4, we conclude that the main specifications do not suffer from *within* bias.

Between bias

* MERGEFORMAT Fig. 5 presents two funnel plots of the estimates from the main specifications. Funnel plots are essential graphical tools in meta-analyses for assessing publication bias and other potential biases. They display the treatment effects from individual studies against a measure of study precision, such as the precision (one divided by the standard error). In the absence of bias, the plot typically resembles an inverted funnel, with points symmetrically distributed around the overall effect size. Asymmetry in the funnel plot may suggest the presence of biases, including selective reporting or small-study effects. Therefore, funnel plots provide a visual method for evaluating the robustness and reliability of aggregated results in a meta-analysis.

The funnel plot in Panel A corresponds to estimates from the stringency studies, while the funnel plot in Panel B corresponds to estimates from SIPO (Shelter-in-Place Orders) studies. Both scatter plots resemble an inverted funnel. However, in both Panel A and Panel B the right-hand side appears more sparsely populated with studies compared to the left-hand side, indicating that fewer studies report positive effects (more deaths) of lockdowns. Numerically smaller and positive estimates typically show more precision. Large point estimates are usually associated with substantial uncertainty (low precision).

It is noteworthy that the funnels have a similar shape despite the estimates being of two different characters (stringency studies estimate the effect of the average lockdown while

[58] The internal PWA is calculated as the product of each estimate and its corresponding precision (one divided by the estimate's standard error) divided by the sum of precisions.

SIPO studies estimate the effect of shelter-in-place orders. To test for between bias, we run FAT-PET meta-regression analysis. The FAT-PET meta-regression serves to explore potential sources of bias, particularly to investigate the relationship between standard errors and effect estimates, which might indicate between bias. In meta-regression, the focus is on detecting patterns or biases across studies rather than merely summarizing the estimates.

By simple graphical transformation the funnel graph becomes the meta-regression model in Eq. (1) below (Paldam, 2015; Stanley & Doucouliagos, 2010):

$$b_i = \beta_F s_i + \beta_M + u_i \tag{1}$$

where b_i is the studies' point estimates, s_i is the studies' standard errors, β_F is the FAT, funnel asymmetry test, and β_M is the PET, precision estimate test. The noise term is u_i. When the funnel is asymmetric (indicating between bias), the FAT is non-zero and the PET differs from the mean.

Equation (1) can be estimated using weighted least squares (WLS) regression weighting the squared errors by the inverse of the estimates' individual variances (i.e.,$1/s_i^2$), or by dividing Eq. (1) through by s_i to account for obvious heteroskedasticity which gives:

$$b_i/s_i = \beta_F + \beta_M \cdot 1/s_i + u_i/s_i \tag{2}$$

Which again can be written as:

$$t_i = \beta_F + \beta_M \cdot p_i + v_i \tag{3}$$

where t_i is the t-value, p_i is the precision, and v_i is the noise term.

Equation (3) can be estimated using OLS. The results of the FAT-PET are presented in* MERGEFORMAT Table 5, which shows that the FAT, β_F, is non-zero but statistically insignificant for both stringency and SIPO studies. The PET is in both groups similar to the PWA and statistically insignificant.[59]

This implies that our results are not significantly affected by 'between bias'.

Conclusions

In this supplementary material, we examine potential biases in our meta-analysis. Specifically, we investigate: 1) whether researchers exhibit a 'within bias,' favoring estimates that are particularly large or small within individual studies; and 2) whether there is a 'between bias,' where researchers collectively avoid publishing certain types of results, such as analyses showing counterintuitive findings (e.g., lockdowns leading to more deaths).

Our analysis reveals no evidence of 'within bias'. Our funnel plots indicate some 'between bias', but our meta-regression analyses find no significant effect thereoff.

[59] We also estimated the squared version termed the PEESE MRA as described by Paldam (2015). This approach yields similar conclusions. Estimates are available in the supplementary spreadsheet.

Table 4 Overview of all estimates

Study	Estimated effect	Standard error	Specification	Internal PWA
Stringency studies				
Bjørnskov (2021)	−0.3%	0.822%	Main	−0.2%
	26.4%	18.455%		
	−31.1%	23.511%		
	3.1%	51.458%		
Shiva and Molana (2021)	−4.0%	0.395%	Main	−4.0%
	−2.5%	0.460%		
	−2.5%	0.314%		
	−3.6%	0.339%		
Chisadza et al. (2021)	11.7%	1.442%	Main	12.8%
	49.3%	6.062%		
	8.0%	0.984%		
	6.3%	0.781%		
	26.9%	3.308%		
	23.4%	2.885%		
Goldstein et al. (2021)	−7.5%	1.972%	Main	−7.6%
	−7.8%	1.595%		
	−7.6%	1.807%		
Fuller et al. (2021)	−35.3%	9.085%	Main	−35.3%
Ashraf (2020)	−2.4%	0.553%	Average	−3.8%
	−0.9%	0.795%		
	−3.9%	0.006%*		
Yang et al., (2021)	−16.3%	4.523%	Main	−16.3%
	28.6%	7.106%		
	53.8%	8.890%		
	26.8%	6.975%		
Hale et al., (2021a, 2021b)	−16.9%	2.812%	Main	−20.1%
	−14.3%	2.854%		
	−11.6%	2.897%		
	−47.6%	4.534%		
SIPO studies				
Chernozhukov et al. (2021)	−17.7%	14.259%	Main	−17.7%
Stokes et al. (2020)	4.9%	2.803%	Main	1.1%
	−1.9%	0.719%		
	2.0%	3.680%		
	1.9%	3.619%		
	0.8%	6.882%		
	2.3%	3.900%		
	−3.5%	1.798%		
	7.4%	4.930%		
	7.5%	4.799%		
	6.9%	6.156%		
	7.7%	6.854%		

Table 4 (continued)

Study	Estimated effect	Standard error	Specification	Internal PWA
Spiegel and Tookes (2021a, 2021b)	13.1%	6.568%	Average	8.8%
	1.0%	4.846%		
	3.3%	5.611%		
	24.0%	7.907%		
Bonardi et al. (2020)	0.0%	n/a		
Guo et al. (2021)	4.6%	14.800%	Main	4.6%
An et al. (2021)	15.6%	8.721%	Main	15.9%
	16.2%	9.086%		
Sears et al., (2020)*	−32.2%	17.576%	Main	−32.2%
Alderman and Harjoto (2020)	−1.0%	0.592%	Main	−1.0%
Berry et al. (2020)	1.1%	n/a	Main	1.1%
Fowler et al. (2021)	−35.0%	6.969%	Main	−35.0%
Gibson (2020)	−6.0%	24.515%	Average	−5.8%
	−8.1%	27.500%		
	−3.9%	21.111%		
Dave et al. (2020)	−40.4%	36.130%	Average	−40.0%
	−50.2%	43.188%		
	−36.1%	34.820%		
	−38.8%	34.532%		
	−39.3%	34.326%		
	-37.6%	31.900%		

(Ashraf, 2020) report the SE to 0.000. We use 0.00049 in the analysis to avoid dividing by zero. (Sears et al., 2020) provides three estimates, but they only standardized one estimate

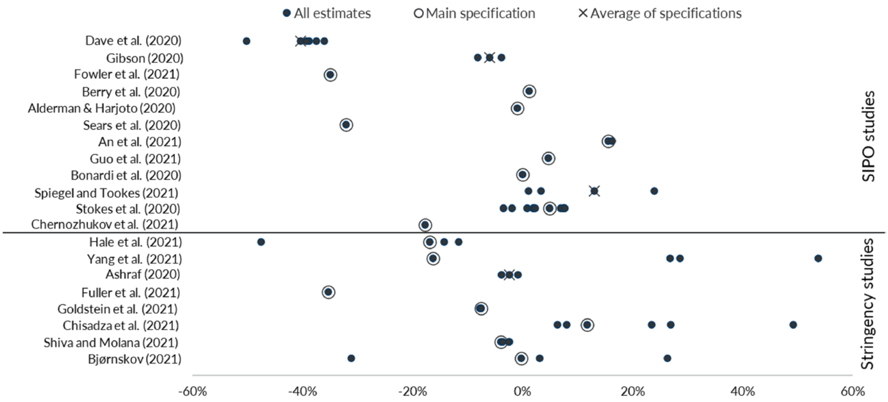

Fig. 3 Overview of all estimates *Note*: Crosses mark where the average of multiple specifications is used in our meta-analysis because there is no obvious main specification

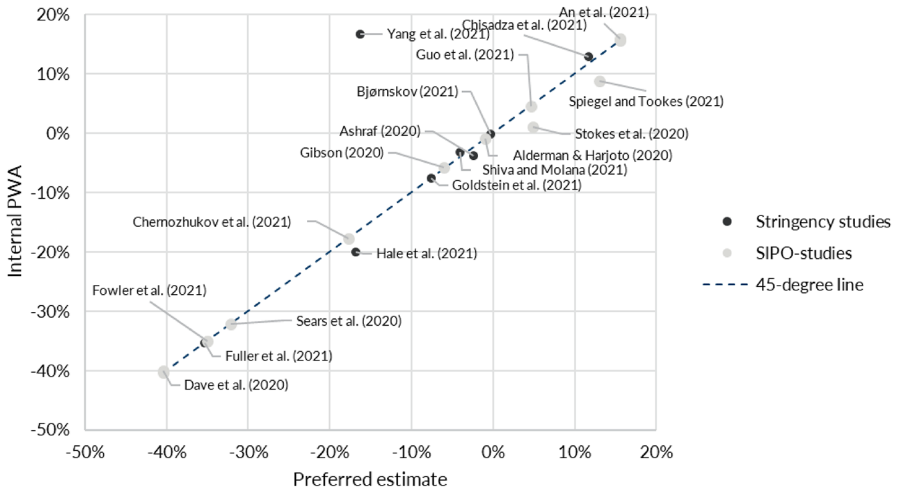

Fig. 4 Comparison of estimates from main specifications to precision-weighted average within each study *Source*: * MERGEFORMAT Table 4. *Note*: The internal precision-weighted average (PWA) is calculated for each study as the product of each single estimate and its corresponding precision (one divided by the standard error) divided by the sum of precisions

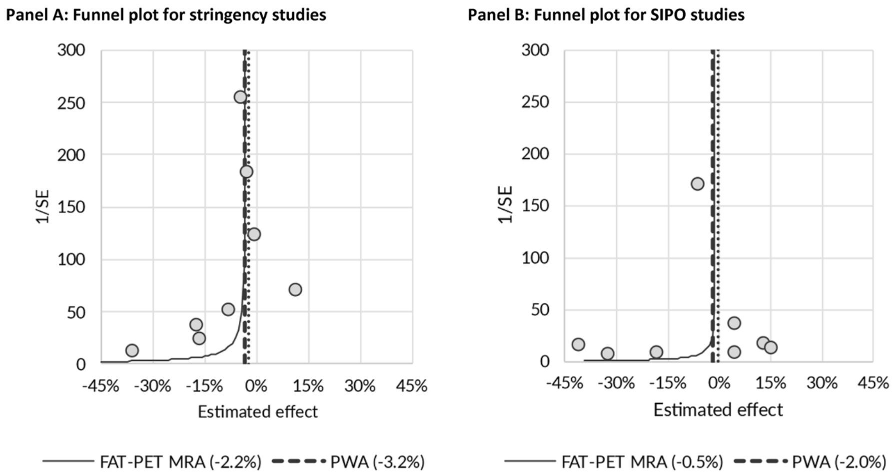

Fig. 5 Funnel plots *Note*: The grey dots mark the estimates used in the meta-analysis

Table 5 FAT-PET results

		Stringency studies	SIPO studies
Standard error (β_F)	*Est*	−0.972	-0.388
	SE	(2.8797)	(0.8238)
	95% CI	[−8.018; 6.075]	[−2.288; 1.512]
	P-value	0.747	0.650
Constant (β_M)	*Est*	−0.022	−0.005
	SE	(0.0234)	(0.0149)
	95% CI	[−0.079; 0.035]	[−0.040; 0.029]
	P-value	0.386	0.734

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11127-024-01216-7.

Acknowledgements We have received helpful comments from Douglas Allen, Torben M. Andersen, Fredrik N. G. Andersson, Andreas Bergh, Jonas Björk, Anders Björkman, Christian Bjørnskov, Joakim Book, Gunnar Brådvik, Kristoffer Torbjørn Bæk, Dave Campbell, Bernard Casey, Kevin Dowd, James Forder, Ulf Gerdtham, Nicholas Hanlon, Caleb Hofmann, Olga B. Jonas, Daniel B. Klein, Fredrik Charpentier Ljungqvist, Christian Heebøl-Nielsen, Martin Paldam, Jonas Ranstam, Spencer Ryan, John Strezewski, Roger Svensson, Ulf Persson, Anders Waldenström, and Joakim Westerlund. We also thank several proofreaders who reviewed the first version of this study and helped us find ways in which we could clarify our methodology and facilitate the understanding of our results. We thank Line Andersen, Troels Sabroe Ebbesen, Anders Lund Mortensen, and Gabriel Kjær-Westermann for excellent research assistance. Needless to say, the usual disclaimer holds: All remaining errors are our own.

Funding Open access funding provided by Lund University. No funding was received to assist with the preparation of this manuscript.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Abadie, A. (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

Abernethy, G. M., & Glass, D. H. (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

Alderman, J., & Harjoto, M. (2020). COVID-19: US shelter-in-place orders and demographic characteristics linked to cases, mortality, and recovery rates. *Transforming Government: People, Process and Policy, 15*(4), 627–644. https://doi.org/10.1108/TG-06-2020-0130

Allen, D. W. (2021). Covid-19 lockdown Cost/Benefits: A critical assessment of the literature. *International Journal of the Economics of Business, 22*(1), 32. https://doi.org/10.1080/13571516.2021.1976051

An, B. Y., Porcher, S., Tang, S., & Kim, E. E. (2021). Policy design for covid-19: worldwide evidence on the efficacies of early mask mandates and other policy interventions. *Public Administration Review, 81*(6), 1157–1182. https://doi.org/10.1111/puar.13426

Andersson, F. N. G., & Jonung, L. (2024). The Covid-19 lesson from Sweden: Don't lockdown. *Economic Affairs*. https://doi.org/10.1111/ecaf.12611

Arnarson, B. T. (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, B. N. (2020). Socioeconomic conditions, government interventions and health outcomes during COVID-19. *Covid Economics., 37*(14), 141–62. https://doi.org/10.13140/RG.2.2.21141.55520

Atkeson, A. (2021). A Parsimonious behavioral SEIR model of the 2020 COVID Epidemic in the United States and the United Kingdom, w28434. https://doi.org/10.3386/w28434

Atkeson, A. (2023). COVID-19: Epidemiological models. *Annual Review of Financial Economics, 15*(1), 7–27. https://doi.org/10.1146/annurev-financial-110821-020639

Bell, D., Nicoll, A., Fukuda, K., Horby, P., Monto, A., Hayden, F., Wylks, C., Sanders, L., & Van Tam, J. (2006). Non-pharmaceutical interventions for pandemic influenza, national and community measures. *Emerging infectious diseases., 12*(1), 88–94.

Berry, C. R., Fowler, A., Glazer, T., Handel-Meyer, S., & MacMillen, A. (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

Bhattacharya, J., & Hanke, S. H. (2023). SSRN and medRxiv censor counter-narrative science. *Econ Journal Watch, 20*(2), 450–459.

Björk, J., Mattisson, K., & Ahlbom, A. (2021). Impact of winter holiday and government responses on mortality in Europe during the first wave of the COVID-19 pandemic. *European Journal of Public Health, 31*(2), 272–277. https://doi.org/10.1093/eurpub/ckab017

Bjørnskov, C. (2021b, June 14). Born et al. om epidemien i Sverige – hvad er der galt og hvordan ser det ud nu? Punditokraterne. http://punditokraterne.dk/2021/06/14/born-et-al-om-epidemien-i-sverige-hvad-er-der-galt-og-hvordan-ser-det-ud-nu/. Accessed 23 August 2021

Bjørnskov, C. (2021a). Did lockdown work? An economist's cross-country comparison. *Cesifo Economic Studies, 00*(00), 14. https://doi.org/10.1093/cesifo/ifab003

Bjørnskov, C., & Voigt, S. (2022a). This time is different?—On the use of emergency measures during the corona pandemic. *European Journal of Law and Economics, 54*(1), 63–81. https://doi.org/10.1007/s10657-021-09706-5

Bjørnskov, C., & Voigt, S. (2022b). Emergencies: On the misuse of government powers. *Public Choice, 190*(1–2), 1–32. https://doi.org/10.1007/s11127-021-00918-6

Blanco, F., Emrullahu, D., & Soto, R. (2020). Do coronavirus containment measures work? *Worldwide Evidence*. https://doi.org/10.1596/1813-9450-9490

Bonardi, J. P., Gallea, Q., Kalanoski, D., & Lalive, R. (2020). Fast and local: How lockdown policies affect the spread and severity of covid-19. *Covid Economics., 28*(23), 325–51.

Book, J. (2020, December 24). Oxford's Stringency Index is Falling Apart—AIER. https://www.aier.org/article/oxfords-stringency-index-is-falling-apart/. Accessed 10 January 2022

Cardella, E., Depew, B., & Williams, R. B. (2024). Behavioral responses of mandatory masking within social interactions. *Public Choice*. https://doi.org/10.1007/s11127-024-01166-0

CDC. (2022, January 7). Burden of influenza. Centers for disease control and prevention. https://www.cdc.gov/flu/about/burden/index.html. Accessed 15 March 2022

Chabé-Ferret, S. (2017). Should we combine difference in differences with conditioning on pre-treatment outcomes? TSE working papers. https://ideas.repec.org//p/tse/wpaper/31798.html. Accessed 18 July 2023

Chernozhukov, V., Kasahara, H., & Schrimpf, P. (2021). Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. *Journal of Econometrics, 220*(1), 23–62. https://doi.org/10.1016/j.jeconom.2020.09.003

Chisadza, C., Clance, M., & Gupta, R. (2021). Government effectiveness and the COVID-19 pandemic. *Sustainability, 13*(6), 3042. https://doi.org/10.3390/su13063042

Clyde, W., Kakolyris, A., & Koimisis, G. (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

Coccia, M. (2023). Sources, diffusion and prediction in COVID-19 pandemic: Lessons learned to face next health emergency. *AIMS Public Health, 10*(1), 145–168. https://doi.org/10.3934/publichealth.2023012

Dave, D., Friedson, A. I., Matsuzawa, K., & Sabia, J. J. (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

Dergiades, T., Milas, C., Panagiotidis, T., & Mossialos, E. (2020). Effectiveness of government policies in response to the COVID-19 outbreak. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3602004

Doucouliagos, H., & Paldam, M. (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

Dur, R. A. J. (2001). Why do policy makers stick to inefficient decisions? *Public Choice, 107*(3/4), 221–234. https://doi.org/10.1023/A:1010305204751

Easterly, W. (2013). *The tyranny of experts: Economists, dictators, and the forgotten rights of the poor*. Basic Books.

Eggers, A. C., & Harding, R. (2021). Rallying in fear? Estimating the effect of the UK COVID-19 lockdown with a natural experiment. *European Journal of Political Research, 1475–6765*, 12467. https://doi.org/10.1111/1475-6765.12467

Engler, S., Brunner, P., Loviat, R., Abou-Chadi, T., Leemann, L., Glaser, A., & Kübler, D. (2021). Democracy in times of the pandemic: Explaining the variation of COVID-19 policies across European democracies. *West European Politics, 44*(5–6), 1077–1102. https://doi.org/10.1080/01402382.2021.1900669

Ertem, Z., Schechter-Perkins, E. M., Oster, E., van den Berg, P., Epshtein, I., Chaiyakunapruk, N., et al. (2021). The impact of school opening model on SARS-CoV-2 community incidence and mortality. *Nature Medicine, 27*(12), 2120–2126. https://doi.org/10.1038/s41591-021-01563-8

Fakir, A. M. S., & Bharati, T. (2021). Pandemic catch-22: The role of mobility restrictions and institutional inequalities in halting the spread of COVID-19. *PLoS ONE, 16*(6), e0253348. https://doi.org/10.1371/journal.pone.0253348

Ferguson, N., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., & Dighe, A. (2020). Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. *Imperial College London., 10*(77482), 491–7.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature, 584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Fowler, J. H., Hill, S. J., Levin, R., & Obradovich, N. (2021). Stay-at-home orders associate with subsequent decreases in COVID-19 cases and fatalities in the United States. *PLoS ONE, 16*(6), e0248849. https://doi.org/10.1371/journal.pone.0248849

Fuller, J. A., Hakim, A., Victory, K. R., Date, K., Lynch, M., Dahl, enjamin, & Henao, O. (2021). Mitigation policies and COVID-19–Associated mortality—37 European Countries, January 23–June 30, 2020. *MMWR. Morbidity and Mortality Weekly Report, 70*(2), 58–62. https://doi.org/10.15585/mmwr.mm7002e4

Gibson, J. (2022). Government mandated lockdowns do not reduce Covid-19 deaths: implications for evaluating the stringent New Zealand response. *New Zealand Economic Papers., 56*(1), 17–28. https://doi.org/10.1080/00779954.2020.1844786

Goldstein, P., Yeyati, E. L., & Sartorio, L. (2021). Lockdown fatigue: The diminishing effects of quarantines on the spread of COVID-19. https://doi.org/10.21203/rs.3.rs-621368/v1

Goodman-Bacon, Andrew, & Marcus, Jan. (2020). Using difference-in-differences to identify causal effects of COVID-19 Policies. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3603970

GRADEpro. (2013, October). GRADE handbook. https://gdt.gradepro.org/app/handbook/handbook.html. Accessed 16 December 2021

Guo, S., An, R., McBride, T. D., Yu, D., Fu, L., & Yang, Y. (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, S., Simon, K., & Wing, C. (2020). Mandated and voluntary social distancing during the COVID-19 Epidemic: A review. NBER Working Paper Series, w28139. https://doi.org/10.3386/w28139

Hale, T., Hale, A. J., Kira, B., Petherick, A., Phillips, T., Sridhar, D., et al. (2020). Global assessment of the relationship between government response measures and COVID-19 Deaths, 2020.07.04.20145334. https://doi.org/10.1101/2020.07.04.20145334

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., et al. (2021a). A global panel database of pandemic policies (Oxford COVID-19 Government response Tracker—OxCGRT). *Nature Human Behaviour, 5*(4), 529–538. https://doi.org/10.1038/s41562-021-01079-8

Hale, T., Angrist, N., Hale, A. J., Kira, B., Majumdar, S., Petherick, A., et al. (2021b). Government responses and COVID-19 deaths: Global evidence across multiple pandemic waves. *PLoS ONE, 16*(7), e0253116. https://doi.org/10.1371/journal.pone.0253116

Hanke, S. H., & Dowd, K. (2022, March 30). Imperial College's Fear Machine. National Review. https://www.nationalreview.com/2022/03/imperial-colleges-fear-machine/. Accessed 16 May 2022

Hatchett, R. J., Mecher, C. E., & Lipsitch, M. (2007). Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proceedings of the National Academy of Sciences, 104*(18), 7582–7587. https://doi.org/10.1073/pnas.0610941104

Havranek, T., Irsova, Z., Janda, K., & Zilberman, D. (2015). Selective reporting and the social cost of carbon. *Energy Economics, 51*, 394–406. https://doi.org/10.1016/j.eneco.2015.08.009

Herby, J. (2021b, September 16). Hvad betyder timingen af nedlukninger? Virker det, hvis man lukker ned tidligt? Punditokraterne. https://punditokraterne.dk/2021/09/16/hvad-betyder-timingen-af-nedlukninger-virker-det-hvis-man-lukker-ned-tidligt/. Accessed 24 February 2022

Herby, J., Jonung, L., & Hanke, S. H. (2022). A Literature Review and meta-analysis of the effects of lockdowns on COVID-19 mortality—II. Studies in applied economics, SAE(210), 115.

Herby, J. (2022). Virker det, hvis man lukker ned tidligt? En kritisk gennemgang af Hatchett et al. (2007). Punditokraterne. https://punditokraterne.dk/2022/07/31/virker-det-hvis-man-lukker-ned-tidligt-en-kritisk-gennemgang-af-hatchett-et-al-2007/

Herby, J. (2020). Economic essay: The civil society – not the regulatory state—Should deal with COVID-19 and the next pandemic. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3689567

Herby, J. (2021a). A first literature review: Lockdowns only had a small effect on COVID-19. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3764553

Herby, J., Jonung, L., & Hanke, S. H. (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

Herby, J., Jonung, L., & Hanke, S. H. (2023). Did Lockdowns work? The verdict on covid restrictions. London: IEA. https://iea.org.uk/wp-content/uploads/2023/06/Perspectives-_1_Did-lockdowns-work__June_web-1.pdf

Higgs, R. (1987). Crisis and leviathan: Critical episodes in the growth of American Government (First Edition.). Oxford University Press.

Iezadi, S., Gholipour, K., Azami-Aghdash, S., Ghiasi, A., Rezapour, A., Pourasghari, H., & Pashazadeh, F. (2021). Effectiveness of non-pharmaceutical public health interventions against COVID-19: A systematic review and meta-analysis. *PLoS ONE, 16*(11), e0260371. https://doi.org/10.1371/journal.pone.0260371

Jefferson, T., Dooley, L., Ferroni, E., Al-Ansary, L. A., van Driel, M. L., Bawazeer, G. A., Jones, M. A., Hoffmann, T. C., Clark, J., Beller, E. M., Glasziou, P. P., & Conly, J. M. (2023). Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD006207.pub6

Johanna, N., Citrawijaya, H., & Wangge, G. (2022). Mass screening vs lockdown vs combination of both to control COVID-19: A systematic review. *Journal of Public Health Research*. https://doi.org/10.4081/jphr.2020.2011

Koppl, R. (2023). Public health and expert failure. *Public Choice, 195*(1–2), 101–124. https://doi.org/10.1007/s11127-021-00928-4

Kuran, T., & Sunstein, C. R. (1998). Availability cascades and risk regulation. *Stan. L. Rev., 51*, 683. https://doi.org/10.2307/1229439

Lee, E. C., Wada, N. I., Grabowski, M. K., Gurley, E. S., & Lessler, J. (2020). The engines of SARS-CoV-2 spread. *Science, 370*(6515), 406–407. https://doi.org/10.1126/science.abd8755

Leeson, P. T., & Rouanet, L. (2021). Externality and COVID-19. *Southern Economic Journal, 87*(4), 1107–1118. https://doi.org/10.1002/soej.12497

Leeson, P. T., & Thompson, H. A. (2023). Public choice and public health. *Public Choice, 195*(1), 5–41. https://doi.org/10.1007/s11127-021-00900-2

Leffler, C. T., Ing, E., Lykins, J. D., Hogan, M. C., McKeown, C. A., & Grzybowski, A. (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

Lendorf, M. (2021). Projekt COVID-19 Mapping: Afdækning af smittespredningsmønsteret i Region Hovedstaden. https://www.nordsjaellandshospital.dk/nyheder/Documents/Projekt%20COVID-19-Mapping%20Data%20rapport_kort_09032021.pdf

Li, Y., Cheng, Z., Cai, X., Mao, Y., & Temkin-Greener, H. (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

Liu, I. T., Prasad, V., & Darrow, J. J. (2021). Evidence for community cloth face masking to limit the spread of SARS-CoV-2: A Critical Review. https://www.cato.org/working-paper/evidence-community-cloth-face-masking-limit-spread-sars-cov-2-critical-review

Markel, H., Lipman, H. B., Navarro, J. A., Sloan, A., Michalsen, J. R., Stern, A. M., & Cetron, M. S. (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

Mavrot, C., & Pattyn, V. (2022). The politics of evaluation. In A. Ladner & F. Sager (Eds.), *Handbook on the politics of public administration* (pp. 244–255). Edward Elgar Publishing. https://doi.org/10.4337/9781839109447.00029

Mccafferty, S., & Ashley, S. (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*, 15–24. https://doi.org/10.2147/POR.S298309

McKenzie, D. (2020, February 3). Revisiting the difference-in-differences parallel trends assumption: Part II What happens if the parallel trends assumption is (might be) violated? World Bank. https://

blogs.worldbank.org/impactevaluations/revisiting-difference-differences-parallel-trends-assumption-part-ii-what-happens. Accessed 17 May 2022
McKenzie, D. (2021, March 10). An adversarial or "long and squiggly" test of the plausibility of parallel trends in difference-in-differences analysis. world bank. https://blogs.worldbank.org/impactevaluations/adversarial-or-long-and-squiggly-test-plausibility-parallel-trends-difference. Accessed 19 May 2022
McKenzie, D. (2022b, January 24). Explaining why we should believe your DiD assumptions. World Bank. https://blogs.worldbank.org/impactevaluations/explaining-why-we-should-believe-your-did-assumptions. Accessed 17 May 2022
McKenzie, D. (2022a, January 10). A new synthesis and key lessons from the recent difference-in-differences literature. World Bank. https://blogs.worldbank.org/impactevaluations/new-synthesis-and-key-lessons-recent-difference-differences-literature. Accessed 17 May 2022
Mistur, E. M., Givens, J. W., & Matisoff, D. C. (2023). Contagious COVID-19 policies: Policy diffusion during times of crisis. *Review of Policy Research, 40*(1), 36–62. https://doi.org/10.1111/ropr.12487
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (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, O. (1963). On the accuracy of economic observations (2d ed., completely rev.). Princeton, N.J: Princeton University Press.
Mulligan, C. B. (2023). Beyond Pigou: Externalities and civil society in the supply–demand framework. *Public Choice, 196*(1), 1–18. https://doi.org/10.1007/s11127-023-01064-x
Munch, P. K., Espenhain, L., Hansen, C. H., Krause, T. G., & Ethelberg, S. (2022a). 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
Munch, P. K., Espenhain, L., Hansen, C. H., Krause, T. G., & Ethelberg, S. (2022b). Case-control study of activities associated with SARS-CoV-2 infection in an adult unvaccinated population and overview of societal COVID-19 epidemic counter measures in Denmark. *PLoS ONE, 17*(11), e0268849. https://doi.org/10.1371/journal.pone.0268849
Murphy, J., Devereaux, A., Goodman, N. P., & Koppl, R. (2021). Expert failure and pandemics: On adapting to life with pandemics. *Cosmos+ Taxis*. https://doi.org/10.2139/ssrn.3773846
Neustadt, R. E., & Fineberg, H. V. (1978). *The swine flu affair: decision-making on a slippery disease*. Washington: National Academies Press. https://doi.org/10.17226/12660
Niskanen, W. A. (2006). The several costs of responding to the threat of terrorism. *Public Choice, 128*(1–2), 351–356. https://doi.org/10.1007/s11127-006-9057-5
Oraby, T., Tyshenko, M. G., Maldonado, J. C., Vatcheva, K., Elsaadany, S., Alali, W. Q., et al. (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
Paldam, M. (2015). Meta-analysis in a nutshell: Techniques and general findings. *Economics, 9*(1), 20150011. https://doi.org/10.5018/economics-ejournal.ja.2015-11
Pan, W. K., Tyrovolas, S., Iago, G. V., Dasgupta, R. R., Daniel, F., Zaitchik, B., Lantos, P. M., & Woods, C. W. (2020). COVID-19: Effectiveness of non-pharmaceutical interventions in the united states before phased removal of social distancing protections varies by region. *MedRxiv., 20*, 2020–08. https://doi.org/10.1101/2020.08.18.20177600
Patt, A., & Zeckhauser, R. (2000). Action bias and environmental decisions. *Journal of Risk and Uncertainty, 21*(1), 45–72. https://doi.org/10.1023/A:1026517309871
Pennington, M. (2021). Hayek on complexity, uncertainty and pandemic response. *The Review of Austrian Economics, 34*(2), 203–220. https://doi.org/10.1007/s11138-020-00522-9
Pennington, M. (2023). Foucault and Hayek on public health and the road to serfdom. *Public Choice, 195*(1), 125–143. https://doi.org/10.1007/s11127-021-00926-6
Pincombe, M., Reese, V., & Dolan, C. B. (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), 1152–1162. https://doi.org/10.1093/heapol/czab054
Rezapour, A., Souresrafil, A., Peighambari, M. M., Heidarali, M., & Tashakori-Miyanroudi, M. (2021). Economic evaluation of programs against COVID-19: A systematic review. *International Journal of Surgery, 85*, 10–18. https://doi.org/10.1016/j.ijsu.2020.11.015
Robert Koch Institut. (2022). RKI—Coronavirus SARS-CoV-2—COVID-19-Fälle, die einem Ausbruch zugeordnet werden, nach Meldewoche und Infektionsumfeld (Tabelle wird jeden Donnerstag aktualisiert). https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Ausbruchsdaten.html. Accessed 21 April 2022

Sears, J., Villas-Boas, J. M., Villas-Boas, S. B., & Villas-Boas, V. (2020). Are we# stayinghome to flatten the curve? *American Journal of Health Economics., 9*(1), 71–95.

Sebhatu, A., Wennberg, K., Arora-Jonsson, S., & Lindberg, S. I. (2020). Explaining the homogeneous diffusion of COVID-19 nonpharmaceutical interventions across heterogeneous countries. *Proceedings of the National Academy of Sciences., 117*(35), 21201–8. https://doi.org/10.1073/pnas.2010625117

Shiva, M., & Molana, H. (2021). The luxury of lockdown. *The European Journal of Development Research*. https://doi.org/10.1057/s41287-021-00389-x

Spiegel, M., & Tookes, H. (2021). Business restrictions and COVID-19 fatalities. *The Review of Financial Studies, 34*(11), 5266–5308. https://doi.org/10.1093/rfs/hhab069

Spiegel, M., & Tookes, H. (2021b). All or nothing? partial business shutdowns and COVID-19 fatality growth. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3835510

Stanley, T. D., & Doucouliagos, H. (2010). Picture this: A simple graph that reveals much ado about research. *Journal of Economic Surveys, 24*(1), 170–191. https://doi.org/10.1111/j.1467-6419.2009.00593.x

Stanley, T. D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. England: Routledge. https://doi.org/10.4324/9780203111710

Stokes, J., Turner, A. J., Anselmi, L., Morciano, M., & Hone, T. (2020). The relative effects of non-pharmaceutical interventions on early Covid-19 mortality: natural experiment in 130 countries, 2020.10.05.20206888. https://doi.org/10.1101/2020.10.05.20206888

Stolper, W. F. (1994). Joseph alois schumpeter: The public life of a private man. *Princeton University Press*. https://doi.org/10.2307/j.ctvbcd24n

Subramanian, S. V., & Kumar, A. (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*. https://doi.org/10.1007/s10654-021-00808-7

Talic, S., Shah, S., Wild, H., Gasevic, D., Maharaj, A., Ademi, Z., Li, X., Wei, X., Mesa-Eguiagaray, I., Rostron, J., Theodoratou, E., Zhang, X., Motee, A., Liew, D., & Ilic, D. (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. *BMJ*. https://doi.org/10.1136/bmj-2021-068302

Toxvaerd, F. (2020). Equilibrium Social Distancing. Cambridge Working Papers in Economics. https://doi.org/10.17863/CAM.52489

UK Department of Health, Social Services, and Public Safety. (2011). UK Influenza Pandemic Preparedness Strategy. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213717/dh_131040.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. Accessed 18 November 2021

WHO. (2022). Burden of influenza. https://www.euro.who.int/en/health-topics/communicable-diseases/influenza/seasonal-influenza/burden-of-influenza. Accessed 15 March 2022

Yang, Q.-C., Chen, X., Chang, C.-P., Chen, D., & Hao, Y. (2021). What is the relationship between government response and COVID-19 pandemics? Global evidence of 118 countries. *Structural Change and Economic Dynamics, 59*, 98–107. https://doi.org/10.1016/j.strueco.2021.08.007

Zhang, M., Wang, S., Hu, T., Fu, X., Wang, X., Hu, Y., Halloran, B., Li, Z., Cui, Y., Liu, H., & Liu, Z. (2022). Human mobility and COVID-19 transmission: a systematic review and future directions. *Annals of GIS., 28*(4), 501–14.

Zhao, P., Zhang, N., & Li, Y. (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