International Journal of Statistics and Probability; Vol. 12, No. 1; January 2023
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Reproducibility of Health Claims in Meta-Analysis Studies of COVID Quarantine (Stay-at-Home) Orders

S. Stanley Young[1] and Warren B. Kindzierski[2]

[1] CGStat, Raleigh, NC, USA

[2] Independent consultant, St Albert, Alberta, Canada

Correspondence: Warren B. Kindzierski, 12 Hart Place, St Albert, Alberta, T8N 5R1, Canada. Email: wbk@shaw.ca or warrenk@ualberta.ca

Received: November 29, 2022 Accepted: December 29, 2022 Online Published: December 30, 2022

doi:10.5539/ijsp.v12n1p54 URL: https://doi.org/10.5539/ijsp.v12n1p54

Abstract

The coronavirus pandemic (COVID) has been an extraordinary test of modern government scientific procedures that inform and shape policy. Many governments implemented COVID quarantine (stay-at-home) orders on the notion that this nonpharmaceutical intervention would delay and flatten the epidemic peak and largely benefit public health outcomes. The overall research capacity response to COVID since late 2019 has been massive. Given a lack of research transparency, only a small fraction of published research has been judged by others to be reproducible before COVID. Independent evaluation of published meta-analysis on a common research question can be used to assess the reproducibility of a claim coming from that field of research. We used a p-value plotting statistical method to independently evaluate reproducibility of specific research claims made in four meta-analysis studies related to benefits/harms of COVID quarantine orders. Research claims that these meta-analyses covered included: mortality, mental health symptoms, incidence of domestic violence, and suicidal ideation (thoughts of killing yourself). Three of the four meta-analysis studies that we evaluated (mortality, mental health symptoms, incidence of domestic violence) raise further questions about benefits/harms of this form of intervention. The fourth meta-analysis study (suicidal ideation) is judged to be unreliable. Given lack of research transparency and irreproducibility of published research, independent evaluation of meta-analysis studies using p-value plotting is offered as a way to strengthen or refute (falsify) claims made in COVID research.

Keywords: COVID, stay-at-home orders, meta-analysis, p-value plot, reproducibility

1. Introduction

1.1 Background

Since late 2019, the coronavirus pandemic (COVID) has been an extraordinary test of modern government scientific procedures that inform and shape policy. Governments worldwide were faced with a disease whose severity was uncertain and was infecting millions. Governments were forced to act quickly given further uncertainties in the capacity of their health care systems to deal with the virus. In many cases, governments relied on public health experts for their policy, and more broadly to the established mechanisms by which scientific and medical expertise inform government policy.

The World Health Organization (WHO) declared COVID a pandemic on March 11, 2020 (Lavezzo et al., 2020; Members, 2020). Many governments subsequently adopted aggressive pandemic policies. Examples of these policies, imposed as large-scale restrictions on people, included (Gostin et al. 2020; Jenson 2020, Magness 2021): quarantine (stay-at-home) orders, masking orders in community settings, nighttime curfews, closures of schools, universities and many businesses, and bans on large gatherings.

Mathematical modelling studies using simulated pandemic scenarios were used to justify durations of restrictions imposed on people, ranging from 2 weeks to months (CDC 2017, Jenson, 2020). These restrictions were intended to "flatten the epidemic curve" (Matrajt & Leung, 2020). The term – flatten the epidemic curve – was originally utilized by the US Centers of Disease Control for pandemic planning (CDC, 2007) to warrant use of targeted antiviral medications and nonpharmaceutical interventions (NPIs) to delay and flatten the epidemic peak.

An intent of flattening the epidemic curve in a pandemic is being able to spread health care demands resulting from a high incidence peak that could potentially overwhelm health care utilization capacity (Jenson, 2020). The restrictions implemented by governments, however, were lengthy as public health official policy targets shifted (Magness 2021). In United States, political influence dominated both the initiation and ultimate duration of these restrictions (Kosnik & Bellas, 2020).

1.2 Research Reproducibility

The overall research capacity response to COVID since late 2019 has been massive (Kinsella et al., 2020; Chu et al., 2021; Ioannidis et al., 2022). To present an estimate of the magnitude of this response, we used the Advanced Search Builder capabilities of freely available PubMed search engine (pubmed.ncbi.nlm.nih.gov/advanced/). We used the terms covid[Title] OR sars-cov-2[Title] for the period 2020-2023 (search performed November 23, 2022). Our search returned 247,597 listings in the National Library of Medicine data base.

As reported in literature, only a small fraction of published research has been judged by others to be reproducible before COVID (Ioannidis, 2005, 2022; Ioannidis et al., 2011; Keown, 2012; Iqbal et al., 2016; Randall & Welser, 2018; Stodden et al., 2018). Landis et al. (2012) suggest that the inability to reproduce findings is due to a lack of research transparency.

Research transparency permits openness of study design, verification of results, synthesis of new findings with previous knowledge, and effective inquiry of research (Munafo et al., 2017). Causes of poor reproducibility of published research are related to aspects of lack of research transparency such as (Ware & Munafo, 2015): biased study designs, flexibility in research practices, low statistical power, and chasing statistical significance.

As indicated above, many research studies have been published in response to COVID. However, there remain concerns about reproducibility of COVID research, particularly where observational data are used to generate results (Bramstedt, 2020; Peng & Hicks, 2021). The current situation of irreproducible research may be that not much has changed during COVID (e.g., Gustot, 2020; Sumner et al., 2020; Paez, 2021).

1.3 Meta-Analysis

Meta-analysis is a statistical method for combining data from individual studies that focus on a common research question (Egger et al., 2001), for example, whether an intervention (or risk factor) is causal of a health outcome. A meta-analysis examines a claim by taking a summary statistic along with a measure of its reliability from multiple individual intervention/risk factor—health outcome studies (called base papers) found in the literature. These statistics are combined to give what is supposed to be a more reliable estimate of an effect (Young & Kindzierski, 2019).

One aspect of replication—the performance of another study statistically confirming the same hypothesis or claim—is a cornerstone of science and replication of research claims is important before causal inference can be made (Moonesinghe et al., 2007). If a replication study result does not conform to a prevailing paradigm, it might not be submitted for publication. Also, if a similar flawed methodology is used in a replication study as in an original study, or if studies with negative findings are not submitted for publication whereas studies with positive findings are, then a false claim can be canonized (Nissen et al., 2016).

Meta-analysis is considered to be at the top of the medical evidence-based pyramid (Murad et al., 2016; Herner, 2019) – ranked higher than randomized trials and case–control and cohort studies. A key assumption of a meta-analysis is that estimates drawn from the base papers for the analysis are unbiased estimates of the effect of interest (Boos & Stefanski, 2013). Given these attributes, independent evaluation of published meta-analysis on a common research question can be used to assess the reproducibility of a claim coming from that field of research (Young & Kindzierski, 2019; Kindzierski et al., 2021; Young & Kindzierski, 2022a).

The objective of this study was to use a p-value plotting statistical method (after Schweder & Spjøtvoll, 1982) to independently evaluate specific research claims of benefits/harms related to COVID quarantine (stay-at-home) orders in published meta-analysis studies. This was done in an effort to illustrate the importance of reproducibility of research claims arising from this nonpharmaceutical intervention in the context of the surge of COVID papers in literature over the past few years.

2. Methods

We first wanted to gauge the number of reports of meta-analysis studies cited in literature related to some aspect of COVID. To do this we again used the Advanced Search Builder capabilities of the PubMed search engine. On November 20, 2022 we used the terms ((covid[Title]) OR (sars-cov-2[Title]) AND (2020:2023[pdat])) AND (meta-analysis[Title] AND (2020:2023[pdat])). Our search returned 3,204 listings in the National Library of Medicine data base. This included 633 listings for 2020, 1,301 listings for 2021, and 1,270 listings thus far for 2022. We find these counts astonishing in that a meta-analysis is a summary of available papers.

Given our understanding of pre-COVID research reproducibility of published literature discussed above, we speculated that there may be numerous meta-analysis studies relating to COVID that are irreproducible. We prepared and posted a research plan – Young & Kindzierski (2022b) – on the *Researchers.One* platform. This plan can be accessed and downloaded without restrictions from the platform. Our plan was to use p-value plotting to independently evaluate four selected published meta-analysis studies specifically relating to possible health outcomes of COVID quarantine (stay-at-home) orders – also referred to as 'lockdowns' or 'shelter-in-place' in literature.

2.1 Data Sets

As stated in our research plan (Young & Kindzierski, 2022b), we considered four meta-analysis studies in our evaluation:

- Herby et al. (2022) – mortality
- Prati & Mancini (2021) – psychological impacts (specifically, mental health symptoms)
- Piquero et al. (2021) – reported incidents of domestic violence
- Zhu et al. (2022) – suicidal ideation (thoughts of killing yourself)

Electronic copies of each meta-analysis study (and any corresponding electronic supplementary information files) were downloaded from the internet and read.

Herby et al. (2022) – The Herby et al. (2022) meta-analysis examined the effect of COVID quarantine (stay-at-home) orders implemented in 2020 on mortality based on available empirical evidence. These orders were defined as the imposition of at least one compulsory, non-pharmaceutical intervention. Herby et al. initially identified 19,646 records that could potentially address their purpose.

After three levels of screening by Herby et al., 32 studies qualified. Of these, estimates from 22 studies could be converted to standardized measures for inclusion in their meta-analysis. For our evaluation, we could only consider results for 20 of the 22 studies (data they provided for two studies could not be converted to p-values).

Herby et al. combined statistics from different models in each base study to understand the effect of lockdowns. The base studies were restricted to European countries and the United States. Most statistics in these base studies were intended to be an intervention relative to 'doing the least'. This, for many Western countries, meant relative to doing as Sweden did during the first pandemic wave. In the first wave Sweden implemented very few restrictions compared to other western countries. While too lengthy to discuss here, Herby et al. explain the statistical methods used for standardization of results of the different base studies in their Table 19.

The Herby et al. meta-analytic statistics involved deriving a combined, standardized estimate of the relative effect of lockdowns on COVID-19 mortality. Each base study statistic was first converted into an estimate of how many percent fewer deaths there were due to lockdowns compared to doing the least; and then all were combined to get a standardized estimate. Their research claim was that… "*lockdowns in the spring of 2020 had little to no effect on COVID-19 mortality*".

Prati & Mancini (2021) – The Prati & Mancini (2021) meta-analysis examined the psychological impact of COVID quarantine (stay-at-home) orders on the general population. This included: mental health symptoms (such as anxiety and depression), positive psychological functioning (such as well-being and life-satisfaction), and feelings of loneliness and social support as ancillary outcomes.

Prati & Mancini initially identified 1,248 separate records that could potentially address their purpose. After screening, they identified and assessed 63 studies for eligibility and ultimately considered 25 studies for their meta-analysis. Their meta-analysis method used a random-effects model with restricted maximum likelihood as a heterogeneity variance estimator after Langan et al. (2015).

The Prati & Mancini meta-analysis computed two summary statistics for each base study – one for positive psychological functioning and another one for mental health symptoms. For our evaluation, we used all 20 results they reported on for mental health symptoms. Their research claim was that… "*lockdowns do not have uniformly detrimental effects on mental health and most people are psychologically resilient to their effects*".

Piquero et al. (2021) – The Piquero et al. (2021) meta-analysis examined the effect of COVID quarantine (stay-at-home) orders on reported incidents of domestic violence. They used the following search terms to identify suitable papers with quantitative data to include in their meta-analysis… "domestic violence", "intimate partner violence", or "violence against women".

Piquero et al. initially identified 22,557 records that could potentially address their purpose. After screening, they assessed 132 studies for eligibility and ultimately considered 18 studies in their meta-analysis. For our evaluation, we used all 17 results (effect sizes) that they presented from the 18 studies. Piquero et al. used a random effects restricted maximum likelihood model after Harbord & Higgins (2008) to estimate their meta-analytic statistics. Their research claim was that… "*incidents of domestic violence increased in response to stay-at-home/lockdown orders*".

Zhu et al. (2021) – The Zhu et al. (2021) meta-analysis examined the effect of COVID quarantine (stay-at-home) orders on suicidal ideation and suicide attempts among psychiatric patients in any setting (e.g., home, institution, etc.). They used the following search terms to identify suitable papers with quantitative data to include in their meta-analysis… "suicide" or "suicide attempt" or "suicidal ideation" or "self-harm", "psychiatric patients" or "psychiatric illness" or "mental disorders" or "psychiatric hospitalization" or "psychiatric department" or "depressive symptoms" or "obsessive-compulsive disorder".

Zhu et al. initially identified 728 records that could potentially address their purpose. After screening, they assessed 83 studies for eligibility and ultimately considered 21 studies in their meta-analysis. For our evaluation, we used all 12 results that they reported on for suicidal ideation among psychiatric patients.

Zhu et al. used STATA16.0 statistical software (StataCorp LLC, College Station, TX) and a random-effects model (DerSimonian & Laird 1986) to estimate their meta-analytic statistics. The random-effects model was used to generalize results by assuming that the base studies they selected for their meta-analysis were random samples from a larger population. Their research claim was that… "*estimated prevalence of suicidal ideation within 12 months* [during COVID] *was… significantly higher than a world Mental Health Survey conducted by the World Health Organization (WHO) in 21 countries* [conducted 2001−2007]".

2.2 P-Value Plots

In epidemiology it is traditional to use risk ratios and confidence intervals instead of p-values from a hypothesis test to demonstrate or interpret statistical significance. Altman & Bland (2011a,b) show that both confidence intervals and p-values are constructed from the same data and they are inter-changeable, and one can be calculated from the other.

Using JMP statistical software (SAS Institute, Cary, NC), we estimated p-values from risk ratios and confidence intervals for all data in each of the meta-analysis studies. In the case of the Herby et al. (2022) meta-analysis, standard error (SE) was presented instead of confidence intervals. Where SE values were not reported, we used the median SE of the other base studies used in the meta-analysis (6.8). The raw data and p-values for each meta-analysis are summarized in an Excel file (.xlsx format) in Appendix A.

We created p-value plots after Schweder & Spjøtvoll (1982). These plots were used to visually examine the p-value distributions for each meta-analysis study. The p-value is a random variable derived from a distribution of the test statistic used to analyze data and to test a null hypothesis (Young & Kindzierski, 2022a).

In a well-designed study, p-values are distributed uniformly over the interval 0 to 1 regardless of sample size under the null hypothesis (Schweder & Spjøtvoll, 1982). A distribution of true null hypothesis points plotted against their ranks in a p-value plot should form a 45-degree line when there are no effects (Schweder & Spjøtvoll, 1982; Hung et al., 1997; Bordewijk et al., 2020). A p-value plot can be used to inspect the heterogeneity of the data (test statistics) being combined in meta-analyses.

The p-value plots we created are interpreted as follows (Young & Kindzierski, 2022a):

- Computed p-values are ranked from smallest to largest and plotted against the integers, 1, 2, 3,…
- If points on the plot follow an approximate 45-degree line, the test statistics result from a random (chance) process and the data support the null hypothesis of no significant association or effect.
- If points on the plot follow approximately a line with a flat/shallow slope, i.e., where most (the majority) of p-values are small (< 0.05), then the test statistics provide evidence for a real, statistically significant, association or effect.
- If points on the plot display a bilinear shape (divide into two lines), the test statistics used for meta-analysis are consistent with a two-component mixture. A general (overall) claim is not supported. In addition, a small p-value reported for the overall claim in the meta-analysis may not be valid (Schweder & Spjøtvoll, 1982).

P-value plot examples derived from real meta-analysis data sets are provided in Appendix B after Young et al. (2022) to assist in interpretation of the p-value plots we created here. Specifically, the plots in Appendix B correspond to 'plausible true null' and 'plausible true alternative' hypothesis outcomes for published meta-analysis studies of observational data sets. As shown in the p-value plots in Appendix B:

- Points representing a plausible true null hypothesis plot as an approximate 45-degree line.
- Points representing a plausible true alternative hypothesis plot as a line with a flat/shallow slope, where most (the majority) of p-values are small (< 0.05).

The distribution of p-values conforming to the alternative hypothesis – where p-values are a measure of evidence against the null hypothesis – is a function of both sample size and the true value or range of true values of the tested parameter (Hung et al., 1997). The p-value plots in Appendix B are examples of distinct (single) sample distributions for each set of conditions – i.e., for true null (random) associations and true effects between two variables. Evidence for p-value plots displaying patterns outside of that shown in Appendix B should initially be treated as ambiguous (uncertain).

3. Results

3.1 Mortality

Our independent evaluation of the effect of COVID quarantine (stay-at-home) orders on mortality – the Herby et al. (2022) meta-analysis – is shown in Figure 1. There are 20 studies that we included in the figure. Six of the 20 studies had p-values below 0.05 while four of the studies had p-values close to 1.00. Ten studies fell roughly on a 45-degree line implying random results.

This data set comprises mostly null associations (14) and with five or six possible non-null associations (1-in-20 of these could be chance, false, positive association). While not perfect, this data set is a closer fit to a sample distribution for a true null association between two variables. Our interpretation of the p-value plot is that COVID quarantine (stay-at-home) orders are not supported for reducing mortality, consistent with Herby et al.'s claim.

3.2 Psychological Impact (Mental Health Symptoms)

Our independent evaluation of the effect of COVID quarantine (stay-at-home) orders on mental health symptoms – the Prati & Mancini (2021) meta-analysis – is shown in Figure 2. Figure 2 data present as a bilinear shape showing a two-component mixture. This data set clearly does not represent a distinct sample distribution for either true null associations or true effects between two variables. Our interpretation of the p-value plot is that COVID quarantine (stay-at-home) orders have an ambiguous (uncertain) effect on mental health symptoms. However as discussed below, there are questions about their research claim.

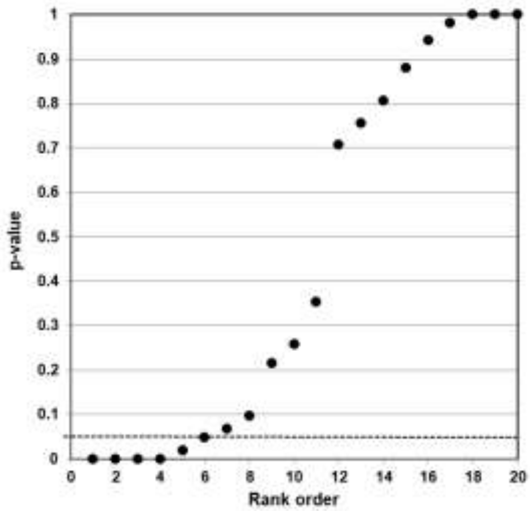

Figure 1. P-value plot (p-value versus rank) for Herby et al. (2022) meta-analysis of the effect of COVID quarantine (stay-at-home) orders implemented in 2020 on mortality. Symbols (circles) are p-values ordered from smallest to largest (n=20)

3.3 Incidents of Domestic Violence

Our independent evaluation of the effect of COVID quarantine (stay-at-home) orders on reported incidents of domestic violence – the Piquero et al. (2021) meta-analysis – is shown in Figure 3. Thirteen of the 17 studies had p-values less than 0.05. While not shown in the figure, eight of the p-values were small (<0.001) (see Appendix A).

This data set comprises mostly non-null associations (13) and with four possible null associations. While not perfect, this data set is a closer fit to a sample distribution for a true alternative association between two variables. Our interpretation of the p-value plot is that COVID quarantine (stay-at-home) have a negative effect (increase) in reported incidents of domestic violence.

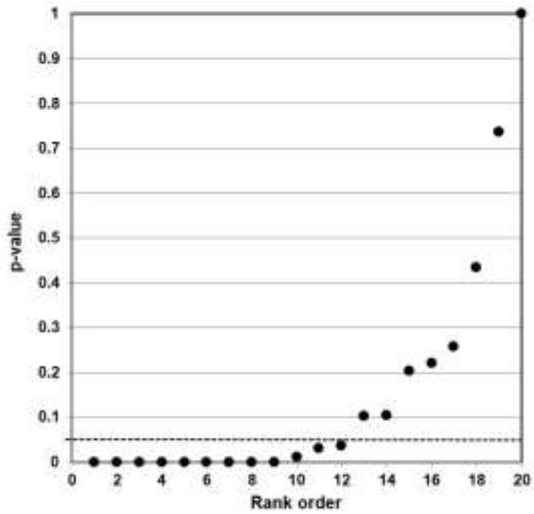

Figure 2. P-value plot (p-value versus rank) for Prati & Mancini (2021) meta-analysis of the effect of COVID quarantine (stay-at-home) orders on mental health symptoms. Symbols (circles) are p-values ordered from smallest to largest (n=20)

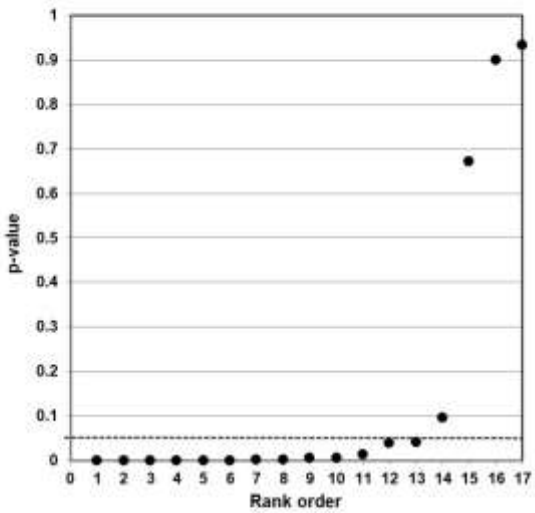

Figure 3. P-value plot (p-value versus rank) for Piquero et al. (2021) meta-analysis of the effect of COVID quarantine (stay-at-home) orders on reported incidents of domestic violence. Symbols (circles) are p-values ordered from smallest to largest (n=17)

3.4 Suicidal Ideation

Our independent evaluation of the effect of COVID quarantine (stay-at-home) orders on suicidal ideation – the Zhu et al. (2021) meta-analysis – is shown in Figure 4. The p-values for all 12 studies were less than 0.05. Ten of the 12 studies had p-values less than 0.05. While not shown in the figure, eight of the p-values were small (<0.001) (see Appendix A).

This data set presents as a distinct sample distribution for true effects between two variables. Our interpretation of the p-value plot is that COVID quarantine (stay-at-home) orders have an effect on suicidal ideation (thoughts of killing yourself). However as discussed below, upon further consideration there are valid questions about how the meta-analysis was formulated.

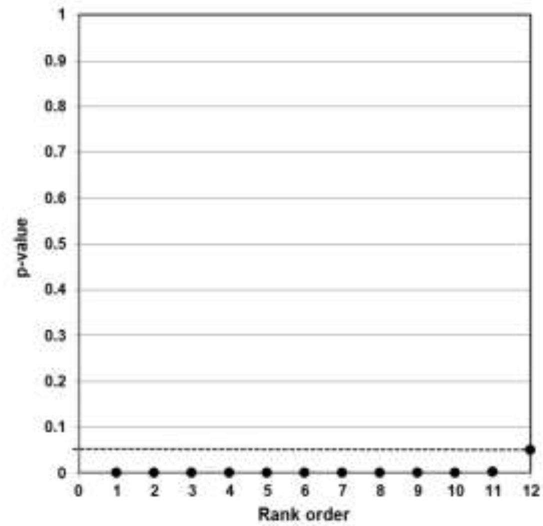

Figure 4. P-value plot (p-value versus rank) for Zhu et al. (2021) meta-analysis of the effect of COVID quarantine (stay-at-home) orders on suicidal ideation (thoughts of killing yourself). Symbols (circles) are p-values ordered from smallest to largest (n=12)

4. Discussion and Implications

As stated previously, independent evaluation of published meta-analysis on a common research question can be used to assess the reproducibility of a claim coming from that field of research. We evaluated four meta-analysis studies of COVID quarantine (stay-at-home) orders implemented in 2020 and corresponding benefits/harms. Our intent was to illustrate the importance of reproducibility of research claims arising from this nonpharmaceutical intervention in the context of the surge of COVID papers in literature over the past few years.

4.1 Mortality

The Herby et al. (2022) meta-analysis examined the effect of COVID quarantine orders on mortality. Their research claim was that… "*lockdowns in the spring of 2020 had little to no effect on COVID-19 mortality*". Here, they imply that the intervention (COVID quarantine orders) had little or no effect on reduction of mortality. The quantitative data Herby et al. present to put their findings into perspective is that they estimated the average lockdown in United States (Europe) in the spring of 2020 avoided 16,000 (23,000) deaths. In contrast, they report that there are about 38,000 (72,000) flu deaths occurring each year in the United States (Europe).

Our evidence agrees with their claim. Our p-value plot (Figure 1) is not consistent with expected behaviour of a distinct sample distribution for a true effect between the intervention (quarantine) and the outcome (reduction in mortality). More importantly, our plot shows considerable randomness (many null associations, p-values > 0.05) supporting no consistent effect. Herby et al. further stated that… "*costs to society must be compared to the benefits of lockdowns, which our meta-analysis has shown are little to none*".

4.2 Psychological Impact (Mental Health Symptoms)

The Prati & Mancini (2021) meta-analysis examined the psychological impact of COVID quarantine orders on the general population. Their research claim was that… "*lockdowns do not have uniformly detrimental effects on mental health and most people are psychologically resilient to their effects*". We evaluated a component of psychological impact – i.e., whether COVID quarantine orders affect mental health symptoms (Figure 2). Figure 2 clearly exhibits a two-component mixture implying an ambiguous (uncertain) effect on mental health symptoms. However, our evidence does not necessarily support their claim.

Digging deep into their study reveals an interesting finding. Their study looked at a variety of psychological symptoms that differed from study to study. Although not shown here, when they examined these symptoms separately – a meta-analysis of each symptom – there was a strong signal for anxiety (p-value less than 0.0001). This is less than the Boos & Stefanski (2011) proposed p-value action level of 0.001 for expected replicability. Here, the term 'action level' means that if a study is replicated, the replication will give a p-value less than 0.05. We note with interest that, at the height of the pandemic, news coverage of COVID was virtually 24/7 saying you could die. It should be no wonder that there was a strong signal for anxiety.

We also note that Prati & Mancini appear to take absence of evidence of a negative mental health effect of COVID quarantine orders in their meta-analysis as implying it does not affect mental health. But "absence of evidence does not imply evidence of absence" (Altman & Bland, 1995, Alderson, 2004; Sedgwick, 2014). Just because meta-analysis failed to find an effect, it does not imply that "…*most people are psychologically resilient to their [lockdown] effects*". A more plausible and valid inference is that this statement of claim is insufficiently researched at this point.

4.3 Incidents of Domestic Violence

The Piquero et al. (2021) meta-analysis examined COVID quarantine orders on reported incidents of domestic violence. Their research claim was that… "*incidents of domestic violence increased in response to stay-at-home/lockdown orders*". Our evidence suggests agreement with this claim. Our p-value plot (Figure 3) is more consistent with expected behaviour of a distinct sample distribution for a true effect between the intervention (quarantine) and the outcome (increase in incidents of domestic violence).

We note that Figure 3 has 13 of 17 p-values less than 0.05 with eight of these less than 0.001, and only a few null association studies (4). Our evidence supports that COVID quarantine orders likely increased incidents of domestic violence.

4.4 Suicidal Ideation

The Zhu et al. (2021) meta-analysis examined COVID quarantine orders on suicidal ideation (thoughts of killing yourself). Their research claim was that… "*estimated prevalence of suicidal ideation within 12 months* [during COVID] *was… significantly higher than a world Mental Health Survey conducted by the World Health Organization (WHO) in 21 countries* [conducted 2001−2007]".

The p-value plot (Figure 4) strongly supports their claim. The plot is very consistent with expected behaviour of a distinct sample distribution for a true effect between the intervention (quarantine) and the outcome (increased prevalence of suicidal ideation). However, digging deep into their study reveals a problem in the formulation of their meta-analysis.

In strong science, a research question being investigated is measured against a control. Zhu et al. effectively ignores controls in their meta-analysis. They compared the incidence of suicidal ideation against a zero standard and not to control groups. The issue is that a pre-COVID (i.e., background) suicidal ideation signal is ignored in their meta-analysis.

Indeed, in their Table 1 they present results from the base papers where data for control groups is available. For example, the Seifert et al. (2021) base paper notes suicidal ideation presented in 123 of 374 patients in the psychiatric emergency department of Hannover Medical School during the pandemic, and 141 of 476 in the same department before the pandemic – 32.9% versus 29.6%. The difference is not significant.

Comparing their Table 1 data set with their Figure 1 forest plot, Zhu et al. only carried 32.9% into their meta-analysis for the Seifert et al. (2021) base paper; in effect they ignored the control data. It is the same situation with all data set entries in their Figure 1. Zhu et al. only considered pandemic incidence in their meta-analysis; they ignored any control data. This approach calls their claims into serious question. We conclude that the Zhu et al. results are unreliable.

4.5 Implications

COVID quarantine orders were implemented on the notion that this nonpharmaceutical intervention would delay and flatten the epidemic peak and benefit public health outcomes overall. Three of the four meta-analyses that we evaluated raise questions about public health benefits/harms of this form of nonpharmaceutical intervention. The fourth meta-analysis study is unreliable.

One meta-analysis that we evaluated – Herby et al. (2022) – questions the benefits of this form of intervention for preventing mortality. Our p-value plot supports their finding that COVID quarantine orders had little or no effect on reduction of mortality.

A second meta-analysis – Prati & Mancini (2021) assessment of mental health symptoms – offers conflicting evidence. Our p-value plot clearly exhibits a two-component mixture implying an ambiguous (uncertain) effect between COVID quarantine orders and mental health symptoms. However, data for a component of mental health symptoms (anxiety) suggests a negative effect from COVID quarantine orders. Further, Prati & Mancini (2021) lack evidence to claim that… "*most people are psychologically resilient to their* [lockdown] *effects*".

Our evaluation of the Piquero et al. (2021) meta-analysis – assessment of domestic violence incidents – supports a true effect between the intervention (quarantine) and the outcome (increase in incidents of domestic violence) with additional confirmatory research needed. Finally, the meta-analysis of Zhu et al. (2021) on suicidal ideation (thoughts of killing yourself) is wrongly formulated. Their results should be disregarded until or unless controls are properly included in their analysis.

Reflecting on the overall findings of our independent evaluation, benefits of COVID quarantine orders remain uncertain for mortality. Also, harms (negative public health consequences) of this intervention cannot be ruled out for mental health symptoms and incidents of domestic violence. Given that the base studies and the meta-analyses themselves were, for the most part, rapidly conducted and published, we acknowledge that confirmatory research is needed for some of these outcomes.

Our interpretations of COVID quarantine benefits/harms are consistent, for example, with research of James (2020) and conventional wisdom on disease mitigation measures used for control of pandemic influenza (Inglesby et al., 2006). James takes a position that is it unclear whether there were benefits from this intervention relative to less restrictive measures aimed at controlling “risky” personal interactions (e.g., mass gatherings and large clusters of individuals in enclosed spaces).

James (2020) also noted numerous economic and public health harms in the United States as May 1, 2020:

- Over 20 million newly unemployed.
- State-wide school closures across the country.
- Increased spouse and child abuse reports.
- Increased divorces.
- Increased backlog of patient needs for mental health services, cancer treatments, dialysis treatments and everyday visits for routine care.
- Increased acute emergency services.

This is consistent with interim quantitative data as of September 2020 presented by the American Institute of Economic Research (2020) on the cost and negative public health implications of pandemic restrictions in United States and around the world.

Acknowledgments

No external funding was provided for this study. The study was conceived based on previous work undertaken by CG Stat for the National Association of Scholars (nas.org), New York, NY.

References

Alderson, P. (2004). Absence of evidence is not evidence of absence. *British Medical Journal*, *328*(7438), 476. https://doi.org/10.1136/bmj.328.7438.476

Altman, D. G., & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *British Medical Journal, 311*(7003), 485. https://doi.org/10.1136/bmj.311.7003.485

Altman, D. G., & Bland, J. M. (2011a). How to obtain a confidence interval from a P value. *British Medical Journal, 343*, d2090. https://doi.org/10.1136/bmj.d2090

Altman, D. G., & Bland, J. M. (2011b). How to obtain the P value from a confidence interval. *British Medical Journal, 343*, d2304. https://doi.org/10.1136/bmj.d2304

American Institute of Economic Research (AIER). (2020). *Cost of Lockdowns: A Preliminary Report*. AIER, Great Barrington, MA. https://www.aier.org/article/cost-of-us-lockdowns-a-preliminary-report/

Boos, D. D., & Stefanski, L. A. (2011). P-value precision and reproducibility. *The American Statistician, 65*(4), 213‒221. https://doi.org/10.1198/tas.2011.10129

Boos, D. D., & Stefanski, L. A. (2013). *Essential Statistical Inference: Theory and Methods*. New York, NY: Springer.

Bordewijk, E. M., Wang, R., & Askie, L. M., et al. (2020). Data integrity of 35 randomised controlled trials in women' health. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *249*, 72–83. https://doi.org/10.1016/j.ejogrb.2020.04.016

Bramstedt, K. A. (2020). The carnage of substandard research during the COVID‐19 pandemic: A call for quality. *Journal of Medical Ethics, 46,* 803‐807. http://doi.org/10.1136/medethics-2020-106494.

Centers for Disease Control and Prevention (CDC). (2007). *Interim pre-pandemic planning guidance: community strategy for pandemic influenza mitigation in the United States: Early, targeted, layered use of nonpharmaceutical interventions*. Atlanta, GA: US CDC. https://stacks.cdc.gov/view/cdc/11425

Centers for Disease Control and Prevention (CDC). (2017). *Community mitigation guidelines to prevent pandemic* *influenza – United States*. Atlanta, GA: US CDC. MMWR Recommendations and reports, 66(No. RR-1)1−36. https://stacks.cdc.gov/view/cdc/45220

Chu, C., Baxamusa, S., & Witherel, C. (2021). Impact of COVID-19 on materials science research innovation and related pandemic response. *MRS Bulletin, 46*, 807–812. https://doi.org/10.1557/s43577-021-00186-1

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*, 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Egger, M., Davey Smith, G., & Altman, D. G. (2001). Problems and limitations in conducting systematic reviews. *In: Egger, M., Davey Smith, G., & Altman, D. G. (eds.) Systematic Reviews in Health Care: Meta−analysis in Context, 2nd ed*. London: BMJ Books.

Gostin, L. O., Friedman, E. A., & Wetter, S. A. 2020. Responding to covid-19: How to navigate a public health emergency legally and ethically. *Hastings Center Report*, *50*, 8−12. http://doi.org/10.1002/hast.1090

Gustot, T. (2020). Quality and reproducibility during the COVID‐19 pandemic. *JHEP Reports*, *2*, 1−3. https://doi.org/10.1016/j.jhepr.2020.100141

Harbord, R. M., & Higgins, J. P. T. (2008) Meta-regression in Stata. *STATA Journal*, *8* 493–519. https://doi.org/10.1177/1536867X0800800403

Herby, J., Jonung, L., & Hanke, S.H. (2022). *A literature review and meta-analysis of the effects of lockdowns on COVID-19 mortality*. SAE./No. 210/May 2022. John Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise https://sites.krieger.jhu.edu/iae/files/2022/06/A-Systematic-Review-and-Meta-Analysis-of-the-Effects-of-Lockdowns-of-COVID-19-Mortality-II.pdf

Herner, M. (2019). Perfect top of the evidence hierarchy pyramid, maybe not so perfect: lessons learned by a novice researcher engaging in a meta-analysis project. *BMJ Evidence-Based Medicine*, *24*(4), 130−132. https://doi.org/10.1136/bmjebm-2018-111141

Hung, H. M. J., O’Neill, R. T., Bauer, P., et al. (1997). The behavior of the p-value when the alternative hypothesis is true. *Biometrics, 53*, 11–22. https://doi.org/10.2307/2533093

Inglesby, T. V., Nuzzo, J. B. O’Tool, T., et al. (2006). Disease mitigation measures in the control of pandemic influenza. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, *4*(4), 366–375. https://doi.org/10.1089/bsp.2006.4.366

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Ioannidis, J. P. A., Tarone, R. E., & McLaughlin, J. K. (2011). The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology, 22*, 450–456. http://doi.org/10.1097/EDE.0b013e31821b506e

Ioannidis, J. P. A. (2022). Correction: Why most published research findings are false. *PLoS Medicine, 19*(8), e1004085. https://doi.org/10.1371/journal.pmed.1004085

Ioannidis, J. P. A., Bendavid, E., Salholz-Hillel, M., et al. 2022. Massive covidization of research citations and the citation elite. *Proceedings of the National Academy of Sciences, 119*, 28 e2204074119. https://doi.org/10.1073/pnas.2204074119

Iqbal, S. A., Wallach, J. D., & Khoury, M. J., et al. (2016). Reproducible research practices and transparency across the biomedical literature. *PLoS Biology, 4*(1), e1002333. http://doi.org/10.1371/journal.pbio.1002333

James, J. J. (2020). Lockdown or lockup. *Disaster Medicine and Public Health Preparedness*, *14*(6), e6−e7. https://doi.org/10.1017/dmp.2020.127

Jenson, H. B. (2020). How did “flatten the curve” become “flatten the economy?” A perspective from the United States of America. *Asian Journal of Psychiatry, 51*, 102165. http://doi.org/10.1016/j.ajp.2020.102165

Keown, S. (2012). *Biases Rife in Research, Ioannidis Says*. NIH Record, VXIV(10). Nih-record.nih.gov/sites/recordNIH/files/pdf/2012/NIH-Record-2012-05-11.pdf (accessed on 10 July 2020).

Kosnik, L. R., & Bellas, A. (2020). Drivers of COVID-19 stay at home orders: Epidemiologic, economic, or political concerns? *Economics of Disasters and Climate Change*, *4*(3), 503–514. https://doi.org/10.1007/s41885-020-00073-0

Kindzierski, W., Young, S., & Meyer, T., et al. (2021). Evaluation of a meta-analysis of ambient air quality as a risk factor for asthma exacerbation. *Journal of Respiration*, *1*(3), 173−196. https://doi.org/10.3390/jor1030017

Kinsella, C. M., Santos, P. D., Postigo-Hidalgo, I., et al. (2020). Preparedness needs research: How fundamental science and international collaboration accelerated the response to COVID-19. *PLoS Pathogens, 16*(10), e1008902. https://doi.org/10.1371/journal.ppat.1008902

Landis, S. C., Amara, S. G., & Asadullah, K., et al. (2012). A call for transparent reporting to optimize the predictive value of preclinical research. *Nature, 490*(7419), 187–191. https://doi.org/10.1038/nature11556

Langan, D., Higgins, J. P. T., & Simmonds, M. (2015). An empirical comparison of heterogeneity variance estimators in 12894 meta-analyses. *Research Synthesis Methods*, *6*(2), 195–205. https://doi.org/10.1002/jrsm.1140

Lavezzo, E., Franchin, E., & Ciavarella, C., et al. (2020). Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo. *Nature, 584*, 425–429. https://doi.org/10.1038/s41586-020-2488-1

Magness, P. (2021). *The Failures of Pandemic Central Planning*. October 1, 2021. http://doi.org/10.2139/ssrn.3934452

Matrajt, L., & Leung, T. (2020). Evaluating the effectiveness of social distancing interventions to delay or flatten the epidemic curve of Coronavirus disease. *Emerging Infectious Diseases, 26*(8), 1740−1748. https://doi.org/10.3201/eid2608.201093

Members, W.-C. J. M. (2020). *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). World Health Organization (WHO)*. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf

Moonesinghe, R., Khoury, M. J., & Janssens, A. C. J. W. (2007). Most published research findings are false—But a little replication goes a long way. *PLoS Medicine, 4*, e28. https://doi.org/10.1146/10.1371/journal.pmed.0040028

Munafo, M. R., Nosek, B. A., Bishop, D. V. M., et al. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, 0021. https://doi.org/10.1038/s41562-016-0021

Murad, M. H., Asi, N., & Alsawas, M., et al. (2016). New evidence pyramid. *BMJ Evidence-Based Medicine, 21*(4), 125−127. http://dx.doi.org/10.1136/ebmed-2016-110401

Nissen, S. B., Magidson, T., & Gross, K., et al. (2016). Publication bias and the canonization of false facts. *eLife, 5*, e21451. https://doi. org/10.7554/elife.21451

Paez, A. (2021). Reproducibility of research during COVID-19: Examining the case of population density and the basic reproductive rate from the perspective of spatial analysis. *Geographical Analysis, 54*, 860–880. https://doi.org/10.1111/gean.12307

Peng, R. D., & Hicks, S. C. (2021). Reproducible research: A retrospective. *Annual Review of Public Health, 42*, 79−93. https://doi.org/10.1146/annurev-publhealth-012420-105110

Piquero, A. R., Jennings, W. G., & Jemison, E., et al. (2021). Domestic violence during the COVID-19 pandemic - Evidence from a systematic review and meta-analysis. *Journal of Criminal Justice, 74*, 101806. https://doi.org/10.1016/j.jcrimjus.2021.101806

Prati, G., & Mancini, A. D. (2021). The psychological impact of COVID-19 pandemic lockdowns: a review and meta-analysis of longitudinal studies and natural experiments. *Psychological Medicine, 51*, 201–211. https://doi.org/10.1017/S0033291721000015

Randall, D., & Welser, C. (2018). *The Irreproducibility Crisis of Modern Science: Causes, Consequences, and the Road to Reform*. New York, NY: National Association of Scholars. Nas.org/reports/the-irreproducibility-crisis-of-modern-science

Schweder, T., & Spjøtvoll, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika, 69*, 493−502. https://doi.org/10.1093/biomet/69.3.493

Sedgwick, P. (2014). Understanding why “absence of evidence is not evidence of absence”. *British Medical Journal, 349*, g4751. https://doi.org/10.1136/bmj.g4751

Seifert, J., Meissner, C., & Birkenstock, A., et al. (2021). Peripandemic psychiatric emergencies: impact of the COVID-19 pandemic on patients according to diagnostic subgroup. *European Archives of Psychiatry and Clinical Neuroscience, 271*(2), 259–270. https://doi.org/10.1007/s00406-020-01228-6

Stodden, V., Seiler, J., & Ma, Z. K. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences, 115*, 2584–2589. https://doi.org/10.1073/pnas.1708290115

Sumner, J., Haynes L., & Nathan S., et al. (2020). Reproducibility and reporting practices in COVID‐19 preprint manuscripts. medRxiv, 2020.03.24.20042796. https://doi.org/10.1101/2020.03.24.20042796

Ware, J. J., & Munafo, M. R. (2015). Significance chasing in research practice: Causes, consequences and possible solutions. *Addiction, 110*(1), 4–8. https://doi.org/10.1111/add.12673

Young, S. S., Cheng, K.-C., Chen, J. H., et al. (2022). Reliability of a meta-analysis of air quality−asthma cohort studies. *International Journal of Statistics and Probability, 11*(2), 61−76. https://doi.org/10.5539/ijspv11n2p61

Young, S. S., & Kindzierski, W. B. (2019). Evaluation of a meta-analysis of air quality and heart attacks, a case study. *Critical Reviews in Toxicology, 49*(1), 85–94. https://doi.org/10.1080/10408444.2019.1576587

Young, S. S., & Kindzierski, W. B. (2022a). Statistical reliability of a diet-disease association meta-analysis. *International Journal of Statistics and Probability, 11*(3), 40–50. https://doi.org/10.5539/ijsp.v11n3p40

Young, S. S., & Kindzierski, W. B. (2022b). Research Plan Lockdowns. Researchers.One. https://researchers.one/articles/22.11.00005v1

Zhu, Y., Li, Y., & Xu, X. (2022). Suicidal ideation and suicide attempts in psychiatric patients during the COVID-19: A systematic review and meta-analysis. *Psychiatry Research, 317*, 114837. https://doi.org/10.1016/j.psychres.2022.114837