International Journal of Statistics and Probability; Vol. 12, No. 4; July 2023
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Statistical Reproducibility of Meta-Analysis for Medical Mask Use in Community Settings to Prevent Airborne Respiratory Virus Infection

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, wbk@shaw.ca

Received: April 12, 2023 Accepted: July 4, 2023 Online Published: July 17, 2023

doi:10.5539/ijsp.v12n4p1 URL: https://doi.org/10.5539/ijsp.v12n4p1

Abstract

Many US states, cities, and counties implemented public masking orders during the coronavirus (COVID) pandemic on the notion that this intervention would delay and flatten the epidemic peak and largely benefit public health outcomes. A p-value plot can provide insights into possible inappropriateness (incorrectness) of assumptions of a statistical model. It can be used to confirm, disprove, or identify ambiguity (uncertainty) in a meta-analytic finding and research claim. P-value plotting was used to evaluate statistical reproducibility of meta-analysis studies for disposable medical (surgical) mask use in community settings to prevent airborne respiratory virus infection. Eight studies (seven meta-analysis, one systematic review) published between 1 January 2020 and 7 December 2022 were evaluated. Base studies were randomized control trials with outcomes of medical diagnosis or laboratory-confirmed diagnosis of viral (Influenza or COVID) illness. Self-reported viral illness outcomes were excluded from the evaluation because of awareness bias. No evidence was observed for a medical mask benefit to prevent respiratory virus infection in six p-value plots (five meta-analysis and one systematic review). Research claims of no benefit in three meta-analysis and the systematic review were reproduced in p-value plots. Research claims of a benefit in two other meta-analysis were not reproduced in p-value plots suggesting irreproducibility of these claims. Insufficient data was available to construct p-value plots for two other meta-analysis because of over-reliance on self-reported outcomes. Independent findings of p-value plotting show that meta-analysis of existing randomized control trials fail to demonstrate a benefit of medical mask use in community settings to prevent airborne respiratory virus infection.

Keywords: respiratory virus, COVID, medical mask, meta-analysis, p-value plot, reproducibility

1. Introduction

1.1 Background

The 2020 coronavirus pandemic (COVID) was an exceptional test of current scientific evidence that informed and shaped government policy. Governments around the world were faced with a disease whose significance was initially uncertain and they acted swiftly given further uncertainties in the capacity of their health care systems to respond to the virus.

On March 11, 2020 the World Health Organization (WHO) declared COVID a pandemic (CDC, 2022). Many governments followed up with pandemic policies. Examples of policies imposed as large-scale restrictions on populations included (Gostin et al., 2020; Jenson, 2020; Magness, 2021): public orders for mask wearing in community settings; stay-at-home orders; curfews at night; school, university, and many business closures; and restrictions on large gatherings.

Early in the pandemic, the U.S. Centers for Disease Control and Prevention (CDC) recommended a cautious approach that patients in health care settings under investigation for symptoms of suspected COVID should wear a medical mask as soon as they are identified (Patel et al., 2020). On April 30, 2020, the CDC recommended that all people wear a mask outside of their home (CDC, 2022).

This recommendation came about after emerging data reported transmission of the COVID virus from persons without symptoms and recognition that there was airborne transmission. Initial recommendations were on using cloth face coverings that could be made more widely available in the community than medical masks (Furukawa et al., 2020). Although, Balazy et al. (2006) and Inglesby et al. (2006) had previously reported that medical masks do little to prevent inhalation of small droplets bearing Influenza virus. Cloth face covering used was intended to preserve personal

protective equipment such as medical masks and N95 respirators to highest-risk exposures in health care settings (Furukawa et al., 2020).

Mathematical modelling studies using simulated pandemic scenarios were used to restrict movement of people for durations ranging from 2 weeks to months (Qualls et al., 2017; Jenson, 2020) in order to "flatten the epidemic curve" (Matrajt & Leung, 2020). The phrase "flatten the epidemic curve" was initially used by the CDC (2007) in pandemic preparation to support using nonpharmaceutical interventions (NPIs) and antiviral medications to delay and flatten the epidemic peak. Inglesby et al. (2006) had articulated early criticisms of the efficacy and appropriateness of some of these NPIs (i.e., quarantine; closing of schools and universities, cancelling or postponing meetings or events in-volving large gatherings) for the control of pandemic Influenza.

A rationale for flattening the epidemic curve in a pandemic is spreading out health care demands resulting from a high incidence peak that could potentially overwhelm health care capacity (Jenson, 2020). Restrictions implemented by governments – including public masking in community settings, however, became lengthy impositions as the policy targets developed by public health official shifted (Magness, 2021). Political influence dominated both the initiation and ultimate duration of these restrictions in the US (Kosnik & Bellas, 2020).

1.2 Research Reproducibility

The research reaction to COVID since the beginning of 2020 has been vast (Kinsella et al., 2020; Chu et al., 2021; Iaonnidis et al., 2022). To show the extent of this reaction, the Advanced Search Builder capabilities of the PubMed search engine was used to approximate the number of COVID publications. The terms covid[Title] OR sars-cov-2[Title] were used for the period 2020-2023 (search performed December 7, 2022). The search returned 250,492 listings in the National Library of Medicine data base.

Prior to the COVID pandemic, researchers have increasingly recognized that only a small portion of published research may be reproducible (Ioannidis, 2005, 2022; Ioannidis et al., 2011; Keown, 2012; Begley & Ioannidis, 2015; Iqbal et al., 2016; Randall & Welser, 2018; Stodden et al., 2018). Lack of research transparency is a reason for research irreproducibility (Landis et al., 2012), due to biased study designs, flexibility in research practices, low statistical power, chasing statistical significance, and selective analysis and reporting (Kavvoura et al., 2007; Ioannidis, 2008; Ioannidis et al., 2011; Ware & Munafo, 2015).

Many research studies have been published in response to COVID. The reproducibility of some of this research is uncertain (Bramstedt, 2020; Peng & Hicks, 2021) and others have suggested that irreproducible research may be a common feature of published COVID literature (Gustot, 2020; Paez, 2021; Sumner et al., 2021).

1.3 Meta-analysis

Meta-analysis is a procedure for statistically combining data (test statistics) from multiple studies that address a common research question or claim (Egger et al., 2008). A general example of a research question or claim (i.e., cause−effect science claim) addressed in meta-analysis is whether an intervention/risk factor prevents/causes a disease.

A meta-analysis assesses a claim by taking a test statistic (e.g., risk ratio, odds ratio, hazard ratio, etc.) along with a measure of its reliability (e.g., confidence interval) from multiple individual intervention—health outcome studies (called base papers) identified in the literature. These statistics are combined to give a supposedly more reliable estimate of an effect (Young & Kindzierski, 2019).

It first involves a systematic review. The systematic review of a clearly formulated research question is intended to use organized and specific methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the identified studies (Moher et al., 2009). A meta-analysis then selects and then combines test statistics of the identified studies from the systematic review.

One component of replication – performance of another study statistically confirming the same hypothesis or research claim – is a foundation of science and replication of research is important before cause−effect claims can be made (Moonesinghe et al., 2007). However, if a replication study result does not concur with a prevailing paradigm, it might not be published. Also, if a similar faulty method is used in a replication study as in an original study, or if studies with negative findings are not published whereas studies with positive findings are, then a false claim can be taken as fact, canonized (Nissen et al., 2016).

Well-designed meta-analyses are ranked high in the medical evidence-based pyramid – similar to well-designed randomized trials, and above observational (case–control and cohort) studies (van Wely, 2014; Murad et al., 2016; Herner, 2019). A major assumption of meta-analysis is that a summary statistic drawn from a base paper for analysis is an unbiased estimate of an effect of interest (Boos & Stefanski, 2013). Given these characteristics, independent evaluation of published meta-analysis on a common research question has been used to assess the statistical

reproducibility of a claim coming from that field of research (Young & Kindzierski, 2019, 2022, 2023; Kindzierski et al., 2021).

Study Objective – Given potentially large data sets available to medical researchers today, intervention−health outcome studies require robust statistical support to establish informative and interpretable intervention−risk/benefit associations and research claims made from these associations. A statistical approach, p-value plotting (Schweder & Spjøtvoll, 1982), was used in this study to evaluate reproducibility of meta-analysis research claims related to benefit of mask use in community settings to prevent airborne respiratory virus infection. The focus was on disposable medical (surgical) masks of the type shown in Figure 1.

Figure 1. Disposable medical (surgical) mask

2. Method

We initially wanted to show the number of listings of meta-analysis studies cited in literature related to COVID. The PubMed search engine was used. The terms ((covid[Title]) OR (sars-cov-2[Title])) AND (meta-analysis[Title]) [timeline 2020-2023] were used on December 7, 2022. The search returned 3,256 listings in the National Library of Medicine data base. This included 633 listings for 2020, 1,300 listings for 2021, and 1,323 listings thus far for 2022. This is considered an astonishing amount in that a meta-analysis is a summary of available papers.

The COVID (sars-cov-2) virus has a reported size range of 60−160 nm (0.06−0.16 μm) (Bar-On et al., 2020; Menter et al., 2020; Zhu et al., 2020). This is similar to the reported size range of Influenza respiratory viruses (80–120 nm, 0.08−0.12 μm) (Stanley, 1944; Mosley & Wyckoff, 1946; NCBI, 2017). Regardless of differing virus sizes, most respiratory viruses are transmitted through secretion fluids during breathing in the form of aerosols (<5 μm) or droplets (>5 μm) rather than isolated viruses (Tellier, 2006, 2009; Clase et al., 2020; Meyerowitz et al., 2020; Prather et al., 2020; Wang et al., 2021). Additional details about respiratory virus airborne transmission characteristics are provided in Appendix 1.

2.1 Study Selection

A well-designed and conducted randomized controlled trial (RCT) is acknowledged as a gold standard for assessing the efficacy of an intervention (O'Conner et al., 2008). For this evaluation, interest was in meta-analysis or systematic review studies of RCTs investigating community medical mask use for prevention of respiratory virus infection. The focus was on Influenza and COVID viruses because of their similar size ranges; keeping in mind it is not the virus itself but airborne transmission of aerosols or droplets containing these viruses that is important for infection.

Another distinction in this evaluation is the type of the outcome for assessing the potential benefit of mask use. Study outcomes that were based solely on self-reported symptoms of viral illness were excluded because of awareness bias. Awareness bias is the tendency of a study participant to self-report a symptom or effect (e.g., a sickness or disease) because of concerns arising from prior knowledge of an environmental hazard that may cause the symptom (Shusterman, 1992; Moffatt & Bhopal, 2000; Smith-Sivertsen et al., 2000; Rabinowitz et al., 2015).

Perception of exposure, causal beliefs, and media coverage have a role in study participants self-reporting symptoms (Borlee et al., 2019). Separating a true biological effect from reporting that is increased because of awareness bias is a problem where study participants are aware of their potential exposure (Moffatt et al., 2000). Using objective health outcomes is recommended in studies to rule out awareness bias (Marcon et al., 2015). Self-reported symptoms, e.g., for viral illness, cannot be considered objective unless it can be corroborated with other more credible outcome measures, e.g., laboratory confirmation (Michaud et al., 2018).

Two online data bases – The Cochrane Central Register of Controlled Trials (CENTRAL) and PubMed – were used to identify eligible studies as both data bases are freely available for public use. Other data bases – such as Scopus, Web of Science, EMBASE – were not considered as these were not accessible to us. CENTRAL was used to identify registered,

controlled trials. PubMed was used because it is acknowledged as an optimal tool in biomedical electronic research (Falagas et al., 2008) and it focuses mainly on life sciences and biomedical disciplines (AlRyalat et al., 2019).

CENTRAL and PubMed data bases were searched for meta-analysis or systematic reviews of randomized controlled trials investigating medical face mask use and Influenza or COVID (sars-cov-2) infections published from January 1, 2020 up to December 7, 2022.

The CENTRAL search strategy was relaxed in that it excluded targeted search terms such as mask, masks, facemasks, nonpharmaceutical, randomized or randomised. Here it was anticipated that there would not be many listings in the CENTRAL data base. The search was performed using the following terms: “influenza A” OR “influenza B” OR “covid” OR “sars-cov-2” OR “respiratory” in Title Abstract Keyword AND “infectious disease” Topic AND “01 January 2020 to 07 December 2022” Custom date range.

Due to the potentially large number of COVID meta-analysis studies in the PubMed data base, the search strategy differed, and it included more targeted terms. These terms included: (((((((influenza[Title]) OR (covid[Title])) OR (sars-cov-2[Title])) OR (respiratory[Title])) OR (viral transmission[Title])) AND ((((nonpharmaceutical[Title]) OR (mask[Title])) OR (masks[Title])) OR (facemasks[Title]))) AND ((randomized[Title/Abstract]) OR (randomised[Title/Abstract]))) AND (("2020/01/01"[Date - Entry] : "2022/12/07"[Date - Entry])).

A potentially eligible systematic review was identified in gray literature during online searches. This review was published by the CATO Institute (Washington, DC) during the 01 January 2020 to 07 December 2022 period (Liu et al., 2021). This review was not captured by searches of the CENTRAL or PubMed data bases. It examined RCTs of medical mask use and viral (including Influenza and COVID) infections.

For each study identified through the searches, titles and full abstracts were read online. Based upon this, electronic copies of eligible meta-analysis or systematic review studies were then downloaded and read. The following criteria was used to determine eligibility of studies for the evaluation:

- Base studies were randomized controlled trials (RCTs) or cluster RCTs.
- Meta-analysis or systematic review.
- Compared the efficacy of medical masks with not wearing masks. Studies were excluded if they did not specify mask type used or present isolated outcomes for individual mask types.
- Included Influenza and/or COVID (sars-cov-2) viruses. Studies were excluded if they did not present isolated outcomes for these viruses.
- Intervention and control groups included community participants. Studies were excluded if they only involved workers in healthcare settings or they did not present isolated outcomes for community participants.
- Included credible outcome measures – i.e., medical diagnosis of respiratory virus illness or lab-confirmed diagnosis of respiratory virus illness.

2.2 P-value Plots

In epidemiology it is traditional to use risk ratios or odds ratios & confidence intervals instead of p-values from a hypothesis test to demonstrate or interpret statistical significance. Both risk ratios or odds ratios & confidence intervals and p-values are constructed from the same data, and they are interchangeable. Formulae exist showing how one can be calculated from the other (Altman & Bland, 2011a,b). Commercial statistical software packages – e.g., SAS and JMP (SAS Institute, Cary, NC) or STATA (StataCorp LLC, College Station, TX) – can also be used to estimate p-values from risk ratios or odds ratios & confidence intervals.

Here, p-values were estimated using JMP statistical software from risk ratios or odds ratios & confidence intervals for all data in each of the studies evaluated. P-value plots (Schweder & Spjøtvoll, 1982) were developed to inspect the distribution of the set of p-values for each 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, 2019).

In a well-designed and conducted study, the p-value is distributed uniformly over the interval 0 to 1 regardless of sample size under the null hypothesis (Schweder & Spjøtvoll, 1982). Suitably scaled, a distribution of p-values 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). Researchers can use a p-value plot to inspect the heterogeneity of the test statistics combined in a meta-analysis.

The p-value plots constructed here were interpreted as follows (Young & Kindzierski, 2019):

- Computed p-values were ordered from smallest to largest and plotted against the integers, 1, 2, 3,…

- If p-value points on the plot followed an approximate 45-degree line, it is concluded that test statistics resulted from a random (chance) process and the data supported the null hypothesis of no significant association or effect.
- If p-value points on the plot followed approximately a line with a flat/shallow slope, where most (the majority) of p-values were small (< 0.05), then test statistic data set provided evidence for a real, statistically significant, association or effect.
- If p-value points on the plot exhibited a bilinear shape (divided into two lines), the data set of test statistics used for meta-analysis is consistent with a two-component mixture and 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).

Examples of p-value plots are provided in Appendix 2 to assist in interpretation of the p-value plots constructed here. Specifically, p-value plots in Appendix 2 represent 'plausible null' and 'plausible true alternative' hypothesis outcomes based on meta-analysis studies of observational data sets. Plausible null hypothesis outcomes plot as an approximate 45-degree line and plausible true alternative hypothesis outcomes plot as a line with a flat/shallow slope, where most (the majority) of p-values are small (< 0.05).

The distribution of the p-value under 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 2 are examples of distinct (single) sample distributions for each condition – i.e., for null (chance or random) associations and true effects between two variables. Evidence for p-value plots exhibiting behaviors outside of that shown in Appendix 2 should be treated as ambiguous (uncertain). A research claim based on ambiguous evidence is unproven.

P-value plots are primarily intended for informal inference (Schweder & Spjøtvoll, 1982). Informal inference involves making generalizations based on data (samples) about a wider universe (population/process), while considering uncertainty without use of a formal statistical procedure or method. Also, it is difficult to make exact probability statements using p-value plots (Schweder & Spjøtvoll, 1982). A p-value plot or other graphical plotting method (e.g., scatter plot, probability plot, residual plot, volcano plot, p-curve plot, etc.) provides insights into possible inappropriateness (incorrectness) of assumptions of a statistical model. Here, p-value plots were used to confirm, disprove, or identify ambiguity (uncertainty) in a meta-analytic finding and research claim. The p-value plots were not used to make exact probability statements.

3. Results

3.1 Search Results

CENTRAL – Sixty-one Cochrane Review articles published for the 01 January 2020 to 07 December 2022 period were identified. These search results are listed in Appendix 3. From examining full abstracts for these articles online, one eligible meta-analysis study that met the search criteria was found – Jefferson et al. (2021).

PubMed (medical research literature) – From the PubMed search, 73 records published for the period were identified (refer to Appendix 3). From examining full abstracts for these studies online, six eligible meta-analysis studies that met the search criteria were found – Aggarwal et al. (2020), Xiao et al. (2020), Nanda et al. (2021), Tran et al. (2021), Kim et al. (2022), and Ollila et al (2022). Coincidentally, one meta-analysis (Xiao et al. 2020) used the exact same RCT data as an earlier World Health Organization (WHO, 2019).

Gray literature – A final study included from gray literature was a systematic review by the public policy research organization CATO Institute (Liu et al., 2021).

3.2 P-value Plots

Descriptive information about characteristics for all eight studies evaluated is presented in Appendix 4. This information includes the following for each study: the data bases searched to identify potential base studies to include, details about viral illness outcomes reported, tables of outcome measures (risk ratio & 95% confidence intervals) and estimated p-values, and other unique evidence and/or limitations worth noting.

P-value plots were constructed and presented here for six meta-analysis: Jefferson et al. (2020) (Figure 2a), Xiao et al. (2020) (Figure 2b), Nanda et al. (2021) (Figure 3a), Tran et al. (2021) (Figure 3b), Kim et al. (2022) (Figure 4a), and Liu et al (2021) (Figure 4b). P-value plots were not constructed for two meta-analysis – Aggarwal et al. (2020) and Ollila et al. (2022) – because of over-reliance on self-reported outcomes in their meta-analysis and/or irregularities or biases worth noting (discussed further in Appendix 4).

3.2.1 Cochrane Review Literature

Jefferson et al. (2020) – Fifteen community (non-healthcare worker) RCTs – base studies – comparing medical masks to no masks were used in this meta-analysis (Appendix 4, Table A1). Their research claim – i.e., cause−effect scientific claim – was (Authors' conclusions, p3 of Jefferson et al. (2020)… "*pooled results of randomised trials did not show a clear reduction in respiratory viral infection with the use of medical/surgical masks during seasonal influenza*". The p-value plot for this study is presented in Figure 2a.

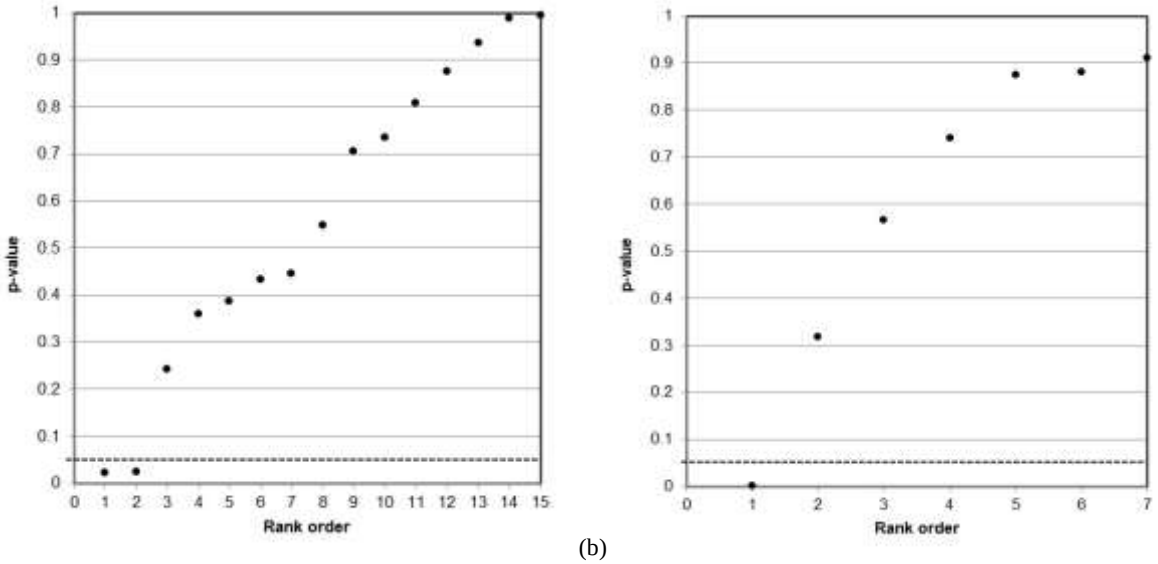
3.2.2 PubMed Literature

Aggarwal et al. (2020) – Five cluster-RCT base studies comparing medical masks to no masks were used in this meta-analysis (Appendix 4, Table A2). The research claim, taken from their Abstract, was… "*data pooled from randomized controlled trials do not reveal a reduction in occurrence of ILI* [Influenza-like illness] *with use of facemask alone in community settings*". A p-value plot for this study was not constructed because two of the five outcome measures failed to meet the eligibility criteria as they were based on self-reported outcomes (with attendant awareness bias) (Table A2).

Xiao et al. (2020) – Seven RCT base studies comparing medical masks to no masks were used in this meta-analysis (Appendix 4, Table A3). The research claim, taken from their Abstract, was… "*Although mechanistic studies support the potential effect of hand hygiene or face masks, evidence from 14 randomized controlled trials of these measures did not support a substantial effect on transmission of laboratory-confirmed influenza*". The p-value plot for this study is presented in Figure 2b.

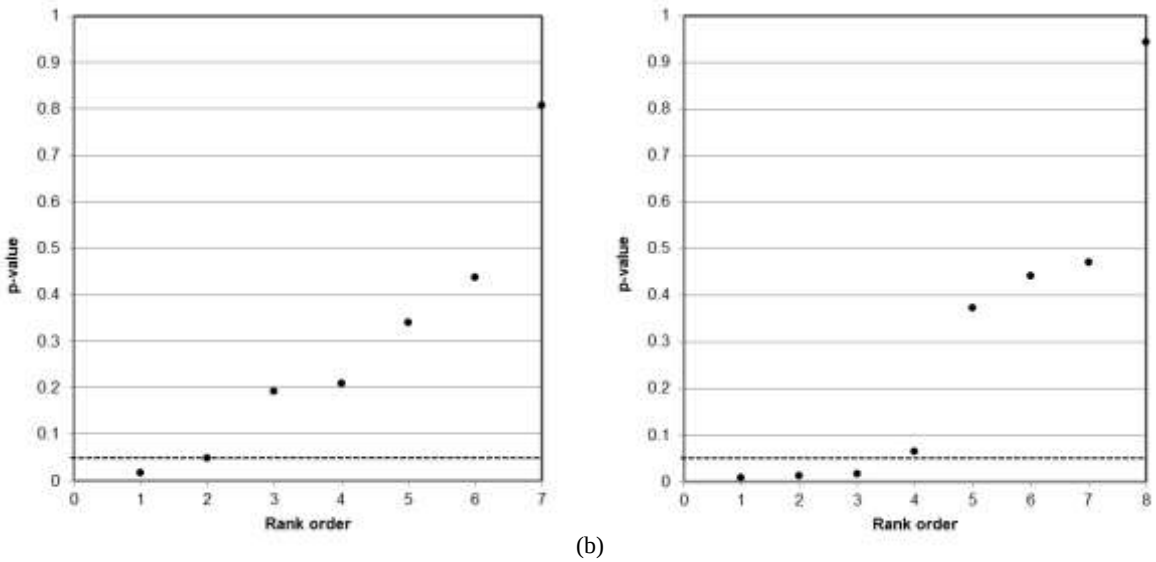
Nanda et al. (2021) – Seven RCT base studies comparing medical masks to no masks were used in this meta-analysis (Appendix 4, Table A4). These are the same seven base studies as Xiao et al. (2020). However, data extracted by Nanda et al. from the base studies and used for calculating risk ratios and confidence interval differed compared to Xiao et al. The research claim, taken from their Abstract, was… "*There is limited available preclinical and clinical evidence for face mask benefit in sars-cov-2. RCT evidence for other respiratory viral illnesses shows no significant benefit of masks in limiting transmission*". The p-value plot for this study is presented in Figure 3a.

Tran et al. (2021) – Eight RCT base studies comparing medical masks to no masks were used in this meta-analysis (Appendix 4, Table A5). Seven of the eight RCT base studies used in their meta-analysis were the exact same as those used by Xiao et al. and Nanda et al. The research claim, taken from their Abstract, was… "*Given the body of evidence through a systematic review and meta-analyses, our findings supported the protective benefits of MFMs* [medical face masks] *in reducing respiratory transmissions, and the universal mask-wearing should be applied—especially during the COVID-19 pandemic*". The p-value plot for this study is presented in Figure 3b.



(a) (b)

Figure 2. Meta-analysis p-value plots: (a) 15 RCT base studies (Jefferson et al. 2020), (b) 7 RCT base studies (Xiao et al. 2020)



(a) (b)

Figure 3. Meta-analysis p-value plots: (a) 7 RCT base studies (Nanda et al. 2021), (b) 8 RCT base studies (Tran et al. 2021)

Kim et al. (2022) – Seven RCT base studies comparing medical masks to no masks were used in this meta-analysis (Appendix 4, Table A6). The viral illness outcome they reported was lab-confirmed infection for Influenza (6 base studies) and COVID (1 base study). The research claim, taken from their Abstract, was… "*Evidence supporting the use of medical or surgical masks against influenza or coronavirus infections (SARS, MERS and COVID‐19) was weak*". The p-value plot for this study is presented in Figure 4a.

Ollila et al. (2022) – Eight RCT base studies comparing medical masks to no masks were used in this meta-analysis (Appendix 4, Table A7). The research claim, taken from their Abstract, was… "*Our findings support the use of face masks particularly in a community setting and for adults*". A p-value plot for this study was not constructed because six of the eight outcome measures failed to meet the eligibility criteria. Specifically, five of these measures were based on self-reported symptoms, and the origin of one measure that Ollila et al. used for another base study could not be confirmed.

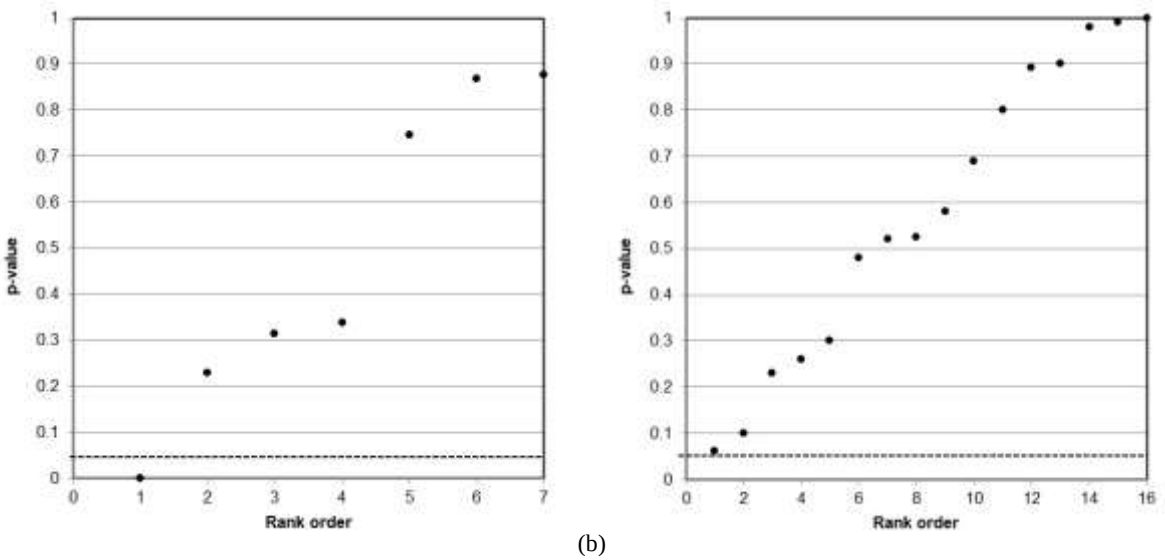
Ollila et al. initially registered a protocol for their study in PROSPERO on 16 November 2020 and changed the protocol on 12 May 2022 and again on 22 September 2022 (Ollila et al., 2020) before it was published on 1 December 2022. Also, test statistics used for three of the base studies for self-reported symptoms showing a benefit of mask use in Table A7 are opposite to other published data of lab-confirmed statistics for the same studies. A more-detailed explanation is presented in Appendix 4.

3.2.3 Gray Literature

Liu et al. (2021) – The Liu et al. systematic review involved examining available clinical evidence of the impact of face mask use in community settings on respiratory infection rates, including by COVID. This review was different than meta-analyses evaluated here in that it did not specify methodologies for identification of RCT base studies. However, they did present and discuss results of RCTs that they identified.

As a result of their different methodology, an attempt was made to obtain original copies of the base studies to confirm their results. They reported outcome measures as p-values for 16 RCT base papers. Only 14 of the 16 base papers were obtained. These results for the 14 base papers are presented in Appendix 4, Table A8.

The research claim, taken from their Abstract, was… "*Of sixteen quantitative meta-analyses, eight were equivocal or critical as to whether evidence supports a public recommendation of masks, and the remaining eight supported a public mask intervention on limited evidence primarily on the basis of the precautionary principle*". The p-value plot for this study, showing results from 14 of the 16 base papers is presented in Figure 4b.



(a) (b)

Figure 4. Meta-analysis p-value plots: (a) 7 RCT base studies (Kim et al. 2022), (b) 14 RCT base studies (Liu et al. 2021)

For all the p-value plots presented here (Figures 2, 3, and 4), no evidence was observed of distinct sample distributions for true effects between two variables (i.e., p-value points forming a line with a flat/shallow slope, where most (the majority) of p-values are small, < 0.05). Again, the reader is referred to Appendix 2 for p-value plots showing true effects between two variables.

The Jefferson et al. (Figure 2a) and Liu et al. (Figure 4b) p-value plots show evidence of distinct sample distributions for null effects – chance or random associations – between two variables (i.e., p-value points plot as an approximate 45-degree line).

The Xiao et al. (Figure 2b) and Kim et al. (Figure 4a) p-value plots are only based on seven points and yet both show evidence of distinct sample distributions for null effects between two variables. The Nanda et al. p-values (Figure 3a) plot closer to a 40-degree line; however, it is still clearly supportive of null effects versus true effects.

The Tran et al. p-value plot (Figure 3b) exhibits a bilinear shape (divides into two lines) – three p-values are small (<0.05) and five p-values >0.05 are oriented on an approximate 45-degree line. This data set of test statistics is consistent with a two-component mixture and thus a general (overall) claim is unproven.

Recall, p-values are interchangeable with traditional epidemiology risk statistics (i.e., risk ratios or odds ratios & confidence intervals). Table 1 presents a summary of p-values that were estimated for risk statistics drawn from base studies used in six meta-analyses. P-values were not estimated for the Ollila et al. (2022) meta-analysis and p-values for the Liu et al. (2021) systematic review are not shown. For eight studies listed in Tables 1, A7 (Ollila et al. 2022) and A8 (Liu et al. 2021), a total of 18 RCT base studies were used across the seven meta-analysis and one systematic review.

Recall that a meta-analysis first involves a systematic review. The meta-analysis then integrates results of identified studies from the systematic review. One would anticipate that well-conducted, independent meta-analyses published within the same period examining the same research question – does medical mask use in community settings prevent airborne respiratory virus infection – should identify and use similar or even the same base study risk statistics for their analysis. Table 1 shows that while most of the base studies identified and used are similar across the meta-analyses, risk statistics that are drawn into meta-analysis are not the same for the exact same research question.

Table 1. Summary of p-values used in six meta-analysis studies

Meta-analysis: Base study, 1st Author Year	**Jefferson et al.**	**Aggarwal et al.**	**Xiao et al.**	**Nanda et al.**	**Tran et al.**	**Kim et al.**
Aiello 2010a		0.0369	0.5663	0.1926	0.007	
Aiello 2012	0.4334, 0.7056	0.5046	0.3187	0.4368	0.373	0.3148
Alfelali 2020						0.7452

Barasheed 2014	0.0222		0.8815	0.2095	0.0155	
Bundgaard 2020						0.2994
Canini 2010	0.9367				0.9432	
Cowling 2008	0.8074, 0.8763	0.2812	0.8746	0.8063	0.4417	0.8763
Jacobs 2009	0.9882					
MacIntyre 2009	0.7342, 0.4456	0.4744	0.9115	0.3404	0.4695	0.8671
MacIntyre 2015	0.2421, 0.5483					
MacIntyre 2016	0.3868, 0.9939		0.7411	0.485	0.0116	
Simmerman 2011						
Suess 2012	0.36, 0.0241	0.4785	0.0009	0.0167	0.0648	0.0002

Take the Aiello 2010a base study, which is used in four meta-analyses in Table 1. Two meta-analyses used risk statistics that are significant (i.e., p-value < 0.05) – Aggarwal et al. and Tran et al. The other two meta-analyses used risk statistics that are non-significant (i.e., p-value > 0.05) – Xiao et al. and Nanda et al. Upon further inspection, Aiello 2010a base study data inputs for estimating risk statistics differed for the meta-analyses.

Specifically, Aiello 2010a base study number of influenza illness events per number of total subjects for the mask group and control group – used for computing risk statistics by the meta-analytic researchers – were as follows:

- For the Aggarwal et al. & Tran et al. meta-analysis – 99/347 (masks) versus 177/487 (controls) for self-reported influenza-like illness (ILI) + laboratory-confirmed influenza.
- For the Xiao et al. & Nanda et al. meta-analysis – 5/347 (masks) versus 3/487 (controls) for laboratory-confirmed influenza only.

Thus, the different (non-significant versus significant) results relate to whether self-reported ILI was used in their calculation.

4. Discussion

The objective of this evaluation was to evaluate the reproducibility of research claims in meta-analysis or systematic review studies of mask use in community settings to prevent airborne respiratory virus infection. Eight eligible studies – seven meta-analyses and one systematic review – were identified and evaluated. These studies were published between the period 1 January 2020 to 7 December 2022. P-value plots were constructed to visually inspect the heterogeneity of test statistics combined in six of these studies. Table 2 compares research claims made in the seven meta-analysis and one systematic review to findings using p-value plots.

For six p-value plots constructed (five meta-analysis and one systematic review), no evidence of distinct sample distributions for true effects between two variables was observed. Five of these plots showed points aligned approximately with 45 degrees – indicating null effects. These p-value plots are consistent with chance or random associations (i.e., no benefit) for medical mask use in community settings to prevent airborne respiratory virus, including COVID, infection.

One other plot (data set of Tran et al. 2021, Figure 3b) had p-value points divided into two lines; consistent with a heterogenic or dissimilar data set (two-component mixture). Here there is insufficient evidence to make a research claim because of ambiguity (uncertainty) in the data set used for meta-analysis. P-value plots for two other meta-analysis – Aggarwal et al. (2020) and Ollila et al. (2022) – were not constructed because of over-reliance on self-reported outcomes (with attendant awareness bias).

Table 2. Comparison of meta-analysis research claims to independent results using p-value plots

Study detail*	Study research claim+	Independent finding of p-value plot	Research claim supported?
Cochrane review literature:			
Jefferson et al. (2020) MA	no significant benefit	null (no) effect	yes
Medical research literature:			
Aggarwal et al. (2020) MA	no significant benefit	insufficient data to examine	unable to determine
Xiao et al. (2020) MA	"	null effect	yes
Nanda et al. (2021) MA	"	null effect	yes
Tran et al. (2021) MA	benefit to mask use	finding is ambiguous (uncertain)	no
Kim et al. (2022) MA	"	null effect	no
Ollila et al. (2022) MA	"	insufficient data to examine	unable to determine
Gray literature:			
Liu et al. (2021) SR [69]	no significant benefit	null effect	yes

Note: *all studies examined randomized control trials (RCTs) of medical mask versus no mask use in community settings for prevention of airborne respiratory (Influenza or COVID) virus infection, + benefit≡reduces respiratory virus infection, MA=meta-analysis, SR=systematic review.

Ample evidence exists of airborne transmission for many respiratory viruses (Wang et al., 2021). These include Influenza virus, respiratory syncytial virus (RSV), human Rhinovirus, severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), SARS-CoV-2 (COVID), measles virus, adenovirus, and enterovirus.

COVID RNA fragments have been identified and infectious COVID virus has been found in airborne aerosols from 0.25 to >4 mm (Wang et al., 2021). This is consistent with the Influenza virus, where RNA has been identified in both $\leqslant 5$ μm and >5 μm aerosols respired from infected hosts, with more Influenza virus RNA found in the $\leqslant 5$ μm aerosols (Fennelly, 2020; Wang et al., 2021). The World Health Organization chief scientist recently acknowledged that COVID was an airborne virus spread by aerosols (Kupferschmidt, 2022).

These observations highlight the importance of airborne aerosol transmission and infection for respiratory viruses, including COVID. Medical mask randomized trials of Influenza infection are directly applicable for understanding the benefit of their use to prevent COVID infection. Again, it is not the virus itself but airborne transmission of aerosols or droplets containing viruses that is important for infection.

Where observational data are used in randomized (or even non-randomized) medical intervention studies, a strong statistical component is required to establish informative and interpretable intervention−risk/benefit associations. This is also the case for research claims made from these associations. For a research claim to be considered valid, it must defeat randomness (i.e., a statistical outcome due to chance).

The p-value plots for five studies – Jefferson et al. (2020), Figure 2a; Xiao et al. (2020), Figure 2b; Nanda et al. (2021), Figure 3a; Kim et al. (2022), Figure 4a; Liu et al. (2021), Figure 4b – show results that look random. The findings of randomness are consistent with research claims made by Jefferson et al. (2020), Xiao et al. (2020), Nanda et al. (2021), and Lui et al. (2021), i.e., no significant benefit to medical mask use (refer to Table 2).

The p-value plot finding of randomness in Figure 4a is opposite to the research claim of Kim et al. (2022) (benefit to medical mask use). This implies irreproducibility of their research claim. The reproducibility of research claims by Aggarwal et al. (2020) (no benefit to medical mask use) and Ollila et al. (2022) (benefit to medical mask use) were not evaluated because of over-reliance on self-reported outcomes (with attendant awareness bias).

Overall, p-value plotting supports that meta-analysis of the randomized trials described above fails to demonstrate a benefit of medical mask use in community settings to prevent airborne respiratory virus infection. This deficiency has been reported in the past (Inglesby et al., 2006; Hardie, 2016) and recently (Drummond, 2022; Miller, 2022; Jefferson et al., 2023). This is consistent with studies showing that medical masks efficiently retain large droplets; however, they

allow airborne particles and aerosols ⩽3 μm or smaller containing viruses to penetrate through the filter material (Davis, 1991; Chen & Willeke, 1992; Weber et al., 1993; Oberg & Brosseau, 2008).

For an intervention to be useful and practical to a population, any benefit of the intervention must be of sufficient magnitude to be able to observe a difference between the intervention group and a control group at the population level. Consider a natural experiment comparing the countries of Germany and Sweden during the pandemic. Germany had a mask mandate for its population whereas Sweden did not.

Survey data on mask compliance during the pandemic was captured in many countries by the University of Maryland (UMD) Social Data Science Center working in collaboration with Facebook (UMD, 2022). One of the daily UMD survey questions asked Facebook users ⩾18 years old if they wore a mask most or all the time in the previous five days. Figure 5 shows Facebook user reported monthly average mask compliance (%) during the second COVID wave – October 2020 through to July 2021 – in Germany and Sweden. Figure 5 shows that monthly average mask compliance in Germany was never less than 80%; whereas in Sweden it was never more than 21%.

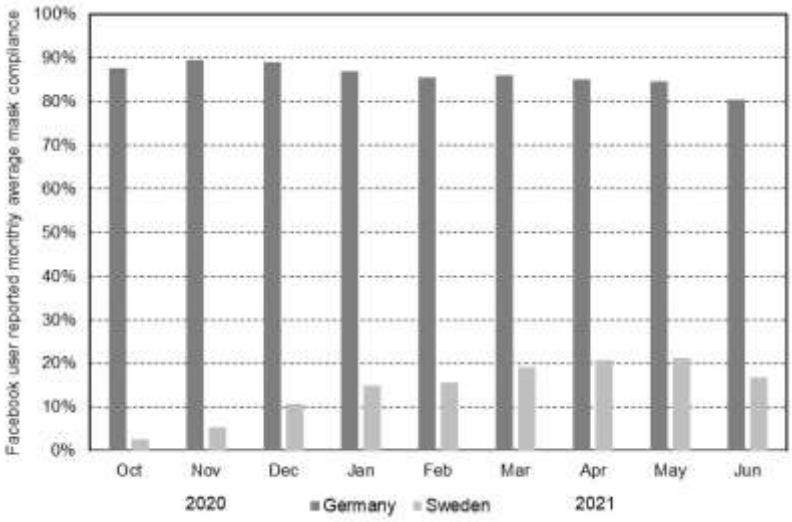


Figure 5. Facebook user reported monthly average mask compliance (%) during the second COVID wave in the countries of Germany and Sweden

Notes: Mask compliance data shown here is averaged from daily data representing percent of Facebook respondents that reported wearing a mask most or all the time in the previous 5 days; data downloaded from UMD COVID Trends and Impacts Survey (https://gisumd.github.io/COVID-19-API-Documentation/) (UMD, 2022); data were adjusted by Facebook for selection biases (i.e., non-response and sampling frame coverage bias). Facebook used survey weights for different countries' regions participating in the survey to adjust data for sampling frame coverage bias (e.g., mask wearing differences in Facebook users ≥18 years old and the target population ≥18 years old) (Meta (2022). Thus, while the UMD COVID Trends and Impacts Survey data would not be expected to be free of sampling frame coverage bias, efforts were made by Facebook to adjust for this bias among different countries' regions.

What role might have masks played in COVID outcomes during the second wave? Consider Figure 6, depicting a severe pandemic outcome measure – daily new COVID deaths per million population in Germany and Sweden. Figure 6 was originally derived by another researcher (Miller, 2022) and it was recreated here using data downloaded from the WHO Coronavirus (COVID-19) Dashboard (https://covid19.who.int/) (WHO, 2023). It shows that WHO-reported daily COVID deaths per million population are not much different during the second wave in the two countries.

Both countries had considered and implemented a variety of pandemic policies, including masking policies, by the second wave of COVID infection (fall 2020). Germany by April 2020 (Schlette et al., 2020) and Sweden by March 2020 (Claeson & Hanson, 2021a,b) had public policies on masks prior to the second wave. Additional information on practices and evolution of Germany and Sweden's COVID pandemic prevention measures are available elsewhere (Schlette et al., 2020; Claeson & Hanson, 2021a,b).

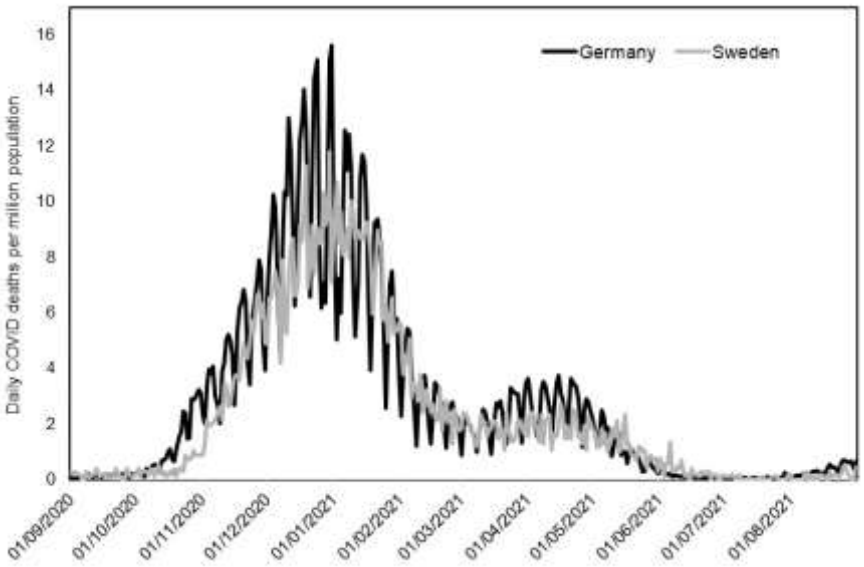


Figure 6. Daily new daily COVID deaths per million population during the second wave in Germany and Sweden

Note: Daily death data downloaded from WHO Coronavirus (COVID-19) Dashboard (https://covid19.who.int/) (WHO, 2023).

Public health risk factors for morbidity and mortality are multi-factorial. Numerous features may be at play in the risk factor (i.e., airborne respiratory virus)−health outcome (i.e., mortality) chain across a population. These can include access to health care, health status, lifestyle, quality of life, standard of living, etc. Germany and Sweden are members of the European Union with similar national health policies, and similar laws and standards for health products and services (EU, 2023). Both should have had similar health care capacities to respond to the COVID pandemic.

Also, in 2020 both Germany and Sweden ranked closely in the top 10 countries of the world with the United Nations Human Development Index (HDI) – Germany 6th, Sweden 7th (World Population Review, 2023). The United Nations HDI tracks measures of life expectancy at birth (health status measure), years of schooling (knowledge measure), and gross national income per capita (standard of living measure).

At the population level, a first impression of Figures 5 and 6 is that mask use had little or no benefit in preventing COVID deaths during the second wave. Despite similar health care capacities, similar United Nations' HDI measures, and obvious differences in mask compliance for these countries (Figure 5), WHO-reported daily COVID deaths per million population are not much different (Figure 6).

Absence of evidence does not imply evidence of absence. Undoubtedly, more meta-analysis on the medical mask−airborne respiratory virus infection research question, including COVID, will be presented in literature using RCTs in the coming years to add to the body of literature. This should allow future researchers to independently re-examine and update the medical mask−airborne respiratory virus infection research question.

5. Conclusions

Research claims of a benefit for medical mask use in community settings to prevent airborne respiratory virus infection were evaluated in eight studies using p-value plots. These studies – seven meta-analysis and one systematic review – used randomized control trials and were published between 1 January 2020 and 7 December 2022. No evidence was observed for a medical mask benefit in six p-value plots (five meta-analysis, one systematic review). Insufficient data were available to construct p-value plots for two other meta-analysis because of over-reliance on self-reported outcomes. Independent findings of p-value plotting show that meta-analysis of existing randomized control trials fail to demonstrate a benefit of medical mask use in community settings to prevent airborne respiratory virus infection.

Acknowledgments

This study was conceived based on previous work undertaken by CG Stat for the National Association of Scholars (nas.org), New York, NY.

References

Abaluck, J., Kwong, L. H., & Styczynski, A., et al. (2022). Impact of community masking on COVID-19: A

cluster-randomized trial in Bangladesh. *Science*, *375*(6577), eabi9069. https://doi.org/10.1126/science.abi9069

Aggarwal, N., Dwarakanathan, V., & Gautam, N., et al. (2020). Facemasks for prevention of viral respiratory infections in community settings: A systematic review and meta-analysis. *Indian Journal of Public Health*, *64*(Supplement), S192–S200. https://doi.org/10.4103/ijph.IJPH_470_20

Aiello, A. E., Murray, G. F., & Perez, V., et al. (2010a). Mask use, hand hygiene, and seasonal influenza-like illness among young adults: A randomized intervention trial. *Journal of Infectious Diseases*, *201*(4), 491−498. https://doi.org/10.1086/650396

Aiello, A. E., Coulborn, R. M., & Perez, V., et al. (2010b). A randomized intervention trial of mask use and hand hygiene to reduce seasonal influenza-like illness and influenza infections among young adults in a university setting. *International Journal of Infectious Diseases*, *14* (Supplement 1), e320. https://doi.org/10.1016/j.ijid.2010.02.2201

AlRyalat, S. A. S., Malkawi, L. W., & Momani, S. M. (2019). Comparing bibliometric analysis using PubMed, Scopus, and Web of Science Databases. *Journal of Visualized Experiments*, 152. https://doi.org/10.3791/58494

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

Balazy, A., Toivola, M., & Adhikari, A., et al. (2006). Do N95 respirators provide 95% protection level against airborne viruses, and how adequate are surgical masks? *American Journal of Infection Control*, *34*(2), 51−57. https://doi.org/10.1016/j.ajic.2005.08.018

Bar-On, Y. M., Flamholz, A., & Phillips, R., et al. (2020). SARS-CoV-2 (COVID-19) by the numbers. *eLife*, *9*, e57309. https://doi.org/10.7554/eLife.57309

Begley, C. G., & Ioannidis, J. P. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research*, *116*(1), 116−126. https://doi.org/10.1161/CIRCRESAHA.114.303819

Belkin, N. (1996). A century after their introduction, are surgical masks necessary? *AORN Journal*, *64*(4), 602−607. http://doi.org/10.1016/s0001-2092(06)63628-4

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

Bordewijk, E. M., Wang, R., Aski, 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. http://dx.doi.org/10.1016/j.ejogrb.2020.04.016

Borlee, F., Yzermans, C. J., & Oostwegel, F. S. M., et al. (2019). Attitude toward livestock farming does not influence the earlier observed association between proximity to goat farms and self-reported pneumonia. *Environmental Epidemiology*, *3*(2), e041. http://doi.org/10.1097/EE9.0000000000000041

Brailovskaia, J., & Margraf, J. (2016). Comparing Facebook users and Facebook non-users: Relationship between personality traits and mental health variables - An exploratory study. *PLoS One*, *11*(12), e0166999. http://doi.org/10.1371/journal.pone.0166999

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

Canini, L., Andreoletti, L., Ferrari, P., et al. (2010). Surgical mask to prevent influenza transmission in households: A cluster randomized trial. *PLoS One*, *5*(11), e13998. http://doi.org/10.1371/journal.pone.0013998

Carp, J. (2012). The secret lives of experiments: Methods reporting in the fMRI literature. *NeuroImage*, *63*(1), 289−300. http://dx.doi.org/10.1016/j.neuroimage.2012.07.004

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*. CDC, Atlanta, GA. https://stacks.cdc.gov/view/cdc/11425

Centers for Disease Control and Prevention (CDC). (2022). *CDC Museum COVID-19 Timeline*, August 16, 2022. CDC, Altanta, GA. https://www.cdc.gov/museum/timeline/covid19.html

Chen, C. C., & Willeke, K. (1992). Aerosol penetration through surgical masks. *American Journal of Infection Control*, *20*(4), 177−184. http://doi.org/10.1016/s0196-6553(05)80143-9

Chikina, M., Pegden, W., & Recht, B. (2022). Re-analysis on the statistical sampling biases of a mask promotion trial in Bangladesh: A statistical replication. *Trials*, *23*(1), 786. https://doi.org/10.1186/s13063-022-06704-z

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

Claeson, M., & Hanson, S. (2021a). COVID-19 and the Swedish enigma. *Lancet*, *397*(10271), 259−261. http://doi.org/10.1016/S0140-6736(20)32750-1

Claeson, M., & Hanson, S. (2021b). The Swedish COVID-19 strategy revisited. *Lancet*, *397*(10285), 1619. http://doi.org/10.1016/S0140-6736(21)00885-0

Clase, C. M., Fu, E. L., & Joseph, M., et al. (2020). Cloth masks may prevent transmission of COVID-19: An evidence-based, risk-based approach. *Annals of Internal Medicine*, *173*(6), 489–491. https://doi.org/10.7326/M20-2567

Contopoulos-Ioannidis, D. G., Karvouni, A., & Kouri, I., et al. (2009). Reporting and interpretation of SF-36 outcomes in randomised trials: Systematic review. *British Medical Journal*, *338*, a3006. https://doi.org/10.1136/bmj.a3006

Davis, W. T. (1991). Filtration efficiency of surgical face masks: The need for more meaningful standards. *American Journal of Infection Control*, *19*(1), 16−18. http://doi.org/10.1016/0196-6553(91)90156-7

de Souto Barreto, P., Rolland, Y., & Vellas, B., et al. (2019) Association of long-term exercise training with risk of falls, fractures, hospitalizations, and mortality in older adults: A systematic review and meta-analysis. *JAMA Internal Medicine*, *179*(3), 394−405. https://doi.org/10.1001/jamainternmed.2018.5406

Drummond, H. (2022). *The Face Mask Cult*. ISBN: 978-1-9999907-9-4. UK: CantusHead Books. https://hectordrummond.com/the-face-mask-cult/

Egger, M., Davey Smith, G., & Altman, D. G. (2008). *Systematic Reviews in Health Care: Meta−analysis in Context, 2nd ed*. Chichester, UK: Wiley-Blackwell. pp. 43−68.

European Union (EU). (2023). *Health: Supporting Public Health in Europe, 2023*. Brussels, Belgium: European Union. https://european-union.europa.eu/priorities-and-actions/actions-topic/health_en

Falagas, M. E., Pitsouni, E. I., & Malietzis, G. A., et al. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *Federation of American Societies for Experimental Biology Journal*, *22*(2), 338−342. https://doi.org/10.1096/fj.07-9492LSF

Fennelly, K. P. (2020). Particle sizes of infectious aerosols: Implications for infection control. *Lancet Respiratory Medicine*, *8*(9), 914−924. https://doi.org/10.1016/S2213-2600(20)30323-4

Furukawa, N. W., Brooks, J. T., & Sobel, J. (2020). Evidence supporting transmission of severe acute respiratory syndrome coronavirus 2 while presymptomatic or asymptomatic. *Emerging Infectious Diseases*, *26*(7), e201595. https://doi.org/10.3201/eid2607.201595

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

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

Han, Z. Y., Weng, W. G., & Huang, Q. Y. (2013). Characterizations of particle size distribution of the droplets exhaled by sneeze. *Journal of the Royal Society Interface*, 10(88), 20130560. http://dx.doi.org/10.1098/rsif.2013.0560

Hardie, J. (2016). Why face masks don't work: A revealing review. *Oral Health*, October 18, 2016. https://web.archive.org/web/20200509053953/https:/www.oralhealthgroup.com/features/face-masks-dont-work-revealing-review/

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. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*(5), 640–648. https://doi.org/10.1097/EDE.0b013e31818131e7

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., Tarone, R. E., McLaughlin, J. K. (2011). The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology*, *22*(4), 450–456. http://doi.org/10.1097/EDE.0b013e31821b506e

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 USA*, *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

Jefferson, T., Del Mar, C. B., & Dooley, L., et al. (2020). Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database of Systematic Reviews 2020*, *11*(11), CD006207. http://doi.org/10.1002/14651858.CD006207.pub5

Jefferson, T., Del Mar, C. B., & Ferroni, E., et al. (2023). Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database of Systematic Reviews 2023*, *1*, CD006207. http://doi.org/10.1002/14651858.CD006207.pub6

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

Kavvoura, F. K., Liberopoulos, G., & Ioannidis, J. P. (2007). Selection in reported epidemiological risks: An empirical assessment. *PLoS Medicine*, *4*(3), e79. http://doi.org/10.1371/journal.pmed.0040079

Keown, S. (2012). Biases rife in research, Ioannidis says. *NIH Record*, LXIV(10). https://nihrecord.nih.gov/sites/recordNIH/files/pdf/2012/NIH-Record-2012-05-11.pdf

Kim, M. S., Seong, D., & Li, H., et al. (2022). Comparative effectiveness of N95, surgical or medical, and non-medical facemasks in protection against respiratory virus infection: A systematic review and network meta-analysis. *Reviews in Medical Virology*, *32*(5), e2336. https://doi.org/10.1002/rmv.2336

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

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

Kupferschmidt, K. (2022). WHO's departing chief scientist regrets errors in debate over whether SARS-CoV-2 spreads through air. *Science*, 22 November 2022. https://doi.org/10.1126/science.adf9731

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

Lee, P. N., Forey, B. A., & Coombs, K.J. (2012). Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. *BMC Cancer*, *12*, 385. https://doi.org/10.1186/1471-2407-12-385

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*. CATO Working Paper No. 64. Washington, DC: The CATO Institute. https://www.cato.org/sites/cato.org/files/2021-11/working-paper-64.pdf

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

Marcon, A., Nguyen, G., & Rava, M., et al. (2015). A score for measuring health risk perception in environmental surveys. *Science of the Total Environment*, *527-528*, 270–278. https://doi.org/10.1016/j.scitotenv.2015.04.110

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

Menter, T., Haslbauer, J. D., & Nienhold, R., et al. (2020). Postmortem examination of COVID-19 patients reveals diffuse alveolar damage with severe capillary congestion and variegated findings in lungs and other organs suggesting vascular dysfunction. *Histopathology*, *77*(2), 198−209. https://doi.org/10.1111/his.14134

Meta (2022). User Guide for the COVID-19 Trends and Impact Survey Weights. https://dataforgood.facebook.com/dfg/resources/user-guide-for-ctis-weights

Meyerowitz., E. A., Richterman, A., & Gandhi, R. T., et al. (2020). Transmission of SARS-CoV-2: A review of viral, host, and environmental factors. *Annals of Internal Medicine*, *174*(1), 69−79. https://doi.org/10.7326/M20-5008

Michaud, D. S., Feder, K., & Voicescu, S. A. (2018). Clarifications on the design and interpretation of conclusions from Health Canada's study on wind turbine noise and health. *Acoustics Australia*, *46*, 99–110. https://doi.org/10.1007/s40857-017-0125-4

Miller, I. (2022). *Unmasked: The Global Failure of COVID Mask Mandates*. Brentwood, TN: Post Hill Press.

Moffatt, S., & Bhopal, R. (2000). Study on environmental hazards is flawed. *British Medical Journal*, *320*(7244), 1274. https://doi.org/10.1136/bmj.320.7244.1274

Moffatt, S., Mulloli T. P., & Bhopal R., et al. (2000). An exploration of awareness bias in two environmental epidemiology studies. *Epidemiology*, *11*(2), 199−208. https://doi.org/10.1097/00001648-200003000-00020

Moher, D., Liberati, A., & Tetzlaff, J., et al. (2009). The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

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

Mosley, V., & Wyckoff, R. (1946). Electron micrography of the virus of influenza. *Nature*, *157*, 263. https://doi.org/10.1038/157263a0

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

Nanda, A., Hung, I., & Kwong, A., et al. (2021). Efficacy of surgical masks or cloth masks in the prevention of viral transmission: Systematic review, meta-analysis, and proposal for future trial. *Journal of Evidence-Based Medicine*, *14*(2), 97–111. https://doi.org/10.1111/jebm.12424

National Center for Biotechnology Information (NCBI). (2017). Influenza Virus Resource help center, Influenza virus biology and epidemiology. Bethesda, MD: National Library of Medicine: https://www.ncbi.nlm.nih.gov/genome/viruses/variation/help/flu-help-center/influenza-virus-biology/

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

Oberg, T., & Brosseau, L. M. (2008). Surgical mask filter and fit performance. *American Journal of Infection Control*, *36*(4), 276−282. http://doi.org/10.1016/j.ajic.2007.07.008

O'Conner, D. S., Green, S., & Higgins, J. P. T. (2008). *Cochrane Handbook of Systematic Reviews of Intervention*. Chichester, UK: Wiley-Blackwell.

Ollila, H. M., Laine, L., Koskela, J., et al. (2020). Systematic review and meta-analysis to examine the use of face mask inter-vention in mitigating the risk of spread of respiratory infections and if the effect of face mask use differs in different exposure settings and age groups. PROSPERO 2020 CRD42020205523. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020205523

Ollila, H. M., Partinen, M., & Koskela, J., et al. (2022). Face masks to prevent transmission of respiratory infections: Systematic review and meta-analysis of randomized controlled trials on face mask use. *PLoS One*, *17*(12), e0271517. https://doi.org/10.1371/journal.pone.0271517

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

Patel, A., Jernigan, D. B., & 2019-nCoV CDC Response Team. (2020). Initial public health response and interim clinical guidance for the 2019 novel coronavirus outbreak – United States, December 31, 2019–February 4, 2020.

Morbidity and Mortality Weekly Report, *69*(5), 140–146. https://doi.org/10.15585/mmwr.mm6905e1

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

Prather, K. A., Wang, C. C., & Schooley, R. T. (2020). Reducing transmission of SARS-CoV-2. *Science*, *368*(6498), 1422–1424. https://doi.org/10.1126/science.abc6197

Qualls, N., Levitt, A., & Kanade, N., et al. (2017). Community mitigation guidelines to prevent pandemic influenza – United States. *MMWR Recommendation Report*, *66*(RR-1), 1–34. http://dx.doi.org/10.15585/mmwr.rr6601a1

Rabinowitz, P. M., Slizovskiy, I. B., & Lamers, V., et al. (2015). Proximity to natural gas wells and reported health status: Results of a household survey in Washington County, Pennsylvania. *Environmental Health Perspectives*, *123*(1), 21–26. https://doi.org/10.1289/ehp.1307732

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

Scheuch, G. (2020). Breathing is enough: For the spread of Influenza virus and SARS-CoV-2 by breathing only. *Journal of Aerosol Medicine and Pulmonary Drug Delivery*, *33*(4), 230−234. https://doi.org/10.1089/jamp.2020.1616

Schlette S., Henke K.-D., & Klenk T., et al. (2020). Germany's response to the coronavirus pandemic. https://www.cambridge.org/core/blog/2020/04/08/germanys-response-to-the-coronavirus-pandemic/

Schnatter, A. R., Chen, M., & DeVilbiss, E. A., et al. (2018). Systematic review and meta-analysis of selected cancers in petroleum refinery workers. *Journal of Occupational and Environmental Medicine*, *60*(7), e329−e342. https://doi.org/10.1097/JOM.0000000000001336

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

Seong, D., Shin, J., & Kim, M. (2020). Comparative efficacy of N95, surgical, medical, and non-medical facemasks in respiratory virus transmission prevention: A living systematic review and network meta-analysis. PROSPERO 2020 CRD42020214729. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020214729

Shusterman, D. (1992). Critical review: The health significance of environmental odor pollution. *Archives of Environmental Health*, *47*, 76–87. https://doi.org/10.1080/00039896.1992.9935948

Smith-Sivertsen, T., Tchachtchine, V., & Lund, E. (2000). Self-reported airway symptoms in a population exposed to heavy industrial pollution: What is the role of public awareness? *Epidemiology*, *11*(6), 739−740. https://doi.org/10.1097/00001648-200011000-00027

Stanley, W. M. (1944). The size of the Influenza virus. *Journal of Experimental Medicine*, *79*(3), 267–283. https://doi.org/10.1084/jem.79.3.267

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 USA*, *115*, 2584–2589. https://doi.org/10.1073/pnas.1708290115

Stott, E. J., & Killington, R. A. (1972). Rhinoviruses. *Annual Review of Microbiology*, *26*(1), 503−524. https://doi.org/10.1146/annurev.mi.26.100172.002443

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

Tellier, R. (2006). Review of aerosol transmission of Influenza A virus. *Emerging Infectious Diseases*, *12*(11), 1657–1662. https://doi.org/10.3201/eid1211.060426

Tellier, R. (2009). Aerosol transmission of Influenza A virus: A review of new studies. *Journal of the Royal Society Interface*, *6*, S783–S790. https://doi.org/10.1098/rsif.2009.0302.focus

Tran, T. Q., Mostafa, E. M., & Ravikulan, R., et al. (2020). Efficacy of facemasks against airborne infectious diseases: A systematic review and network meta-analysis of randomized-controlled trials. PROSPERO 2020 CRD42020178516. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020178516

Tran, T. Q., Mostafa, E. M., & Tawfik, G. M., et al. (2021). Efficacy of face masks against respiratory infectious diseases: A systematic review and network analysis of randomized-controlled trials. *Journal of Breath Research*, *15*, 047102. https://doi.org/10.1088/1752-7163/ac1ea5

University of Maryland (UMD). (2022). Global COVID-19 Trends and Impact Survey. University of Maryland (UMD) Joint Program in Survey Methodology, College Park, MD. https://jpsm.umd.edu/landingtopic/global-covid-19-trends-and-impact-survey

van Wely, M. (2014). The good, the bad and the ugly: Meta-analyses. *Human Reproduction*, *29*(8), 1622–1626. https://doi.org/10.1093/humrep/deu127

Wang, C. C., Prather, K. A., & Sznitman, J. (2021). Airborne transmission of respiratory viruses. *Science*, *373*(6558), eabd9149. https://doi.org/10.1126/science.abd9149

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

Weber, A., Willeke, K., & Marchioni, R., et al. (1993). Aerosol penetration and leakage characteristics of masks used in the health care industry. *American Journal of Infection Control*, *21*(4), 167−173. http://doi.org/10.1016/0196-6553(93)90027-2

Wein, H. (2009). Understanding a Common Cold Virus. NIH Research Matters, National Library of Medicine. https://www.nih.gov/news-events/nih-research-matters/understanding-common-cold-virus

World Health Organization (WHO). (2019). *Non-pharmaceutical Public Health Measures for Mitigating the Risk and Impact of Epidemic and Pandemic Influenza, Annex: Report of Systematic Literature Reviews.* No. WHO/WHE/IHM/GIP/2019.1. Geneva, Switzerland: World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/329439/WHO-WHE-IHM-GIP-2019.1-eng.pdf

World Health Organization (WHO). (2023). WHO Coronavirus (COVID-19) Dashboard. Geneva, Switzerland: World Health Organization. https://www.who.int/

World Population Review. (2023). *Top 10 Countries with Highest Human Development Index, 2020 Report – United Nations; Standard of Living by Country; Quality of Life by Country*. Walnut, CA: World Population Review. https://worldpopulationreview.com/country-rankings/standard-of-living-by-country

Xiao, J., Shiu, E. Y. C., Gao, H., et al. (2020). Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings – Personal protective and environmental measures. *Emerging Infectious Diseases*, *26*(5), 967–975. https://doi.org/10.3201/eid2605.190994

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. (2022). 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. (2023). Reproducibility of health claims in meta-analysis studies of COVID quarantine (stay-at-home) orders. *International Journal of Statistics and Probability*, *12*(1), 54–65. https://doi.org/10.5539/ijsp.v12n1p54

Zhu, N., Zhang, D., & Wang, W., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, *382*, 727–733. https://doi.org/10.1056/NEJMoa2001017