

INPLASY

INPLASY202540045

doi: 10.37766/inplasy2025.4.0045

Received: 15 April 2025

Published: 15 April 2025

Corresponding author:

Trisha Greenhalgh

trish.greenhalgh@phc.ox.ac.uk

Author Affiliation:

University of Oxford.

Mask mandates for the prevention of respiratory infections: a mechanism-informed narrative synthesis

Greenhalgh, T; Williamson, J; Helm, R; Ratnayake, S; Poliseli, L; Trofimov A.

ADMINISTRATIVE INFORMATION

Support - This study is supported by UKRI Cross Research Council Responsive Mode (CRCRM) grant 25130 for Interdisciplinary Research (CI Jon Williamson, University of Manchester). The funder had no role in developing the protocol or the conduct of the study.

Review Stage at time of this submission - Preliminary searches.

Conflicts of interest - TG is a member of Independent Sage. She has previously led a narrative review of mask efficacy. All other authors declare no conflict of interest.

INPLASY registration number: INPLASY202540045

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 15 April 2025 and was last updated on 29 September 2025. The amendment records changes made during the review.

1. Method. The method described as Systematic Review+ (SR+) was developed into Mechanism-Informed Narrative Synthesis (MINS). A draft explanatory account of how a mandate might change mask wearing, built from 28 state-of-the-science reviews and seven systematic reviews, was prepared before primary studies were appraised, and five candidate mechanistic pathways replace the list of mechanisms in the registered protocol.
2. Intervention and eligibility. Only mandates issued by a government body under statutory or regulatory authority qualified. Strong public health advice, non-binding recommendations and institutional mask rules (schools, universities, employers, carriers) did not. Healthcare and clinical settings were excluded. Evidence syntheses were used as sources of primary studies and background evidence rather than included as studies.
3. Appraisal. Instead of GRADE checklists, primary studies were appraised with bespoke design-specific tools developed from generic risk-of-bias tools and GRADE guidance, and assigned to one of three credibility tiers within one of four design categories.
4. Synthesis. No meta-analysis was performed, because the studies were too heterogeneous for pooling. Quantitative bias analysis was used as a stress test of direction and robustness and was not used to adjust effect estimates.
5. Search. The databases searched differ from those originally listed; the updated list and search strings are given below.
6. Use of AI. A large language model (Claude Sonnet 4.6, Anthropic) supplemented searching, stress-tested the explanatory account, and drafted a structured appraisal of every study using a purpose-built appraisal tool (TERN, version 7.4.9). Human reviewers checked every appraisal and over-ruled the tool in seven studies; these will be listed with reasons in the published paper.
7. Qualitative studies. Qualitative evidence on mask mandates will be reported separately.

INTRODUCTION

Review question / Objective Empirical objectives (this protocol):

1. To determine the effectiveness of face mask mandates in reducing the spread of

respiratory infections in different contexts and circumstances, and their effect on mask wearing.

2. To identify the mechanisms through which any effect is achieved, or which explain why it is not achieved.

Methodological objective (described in detail in a separate paper). Using this mask mandate review as a worked example:

1. To produce a potentially generalisable methodology for combining associative and mechanistic evidence in systematic review, originally described as Systematic Review+ (SR+) and now developed as Mechanism-Informed Narrative Synthesis (MINS).
2. To examine the extent to which this methodology mitigates key criticisms of orthodox systematic reviews, both epistemological and ethical.

Rationale The question of whether face masks and face mask mandates are effective in reducing transmission of respiratory diseases generated much controversy during the COVID-19 pandemic. An orthodox systematic review, analysing only randomised controlled trials (RCTs) and published in the Cochrane Library, concluded that the existing evidence on face mask efficacy was not definitive (i.e. the authors decided that they could not firmly conclude either that masks were effective or that they were ineffective).[1] One review which included non-RCT evidence, especially evidence of mechanism ('mechanistic evidence'), concluded that there was evidence of efficacy; it proposed explanations for why some RCTs demonstrated efficacy while others did not. [2] A key area of contestation is whether trials were testing masking itself or 'advice to mask'. If masks work, and people are reluctant to follow mere advice, the question then arises: what happens if we require people to wear masks.

To demonstrate causality, we typically need two kinds of evidence: associative (to show that a change in one phenomenon is associated with a change in another) and mechanistic (to be confident that any observed association is genuinely causal). In biomedicine, the former is, ideally, obtained from RCTs but the latter may come from a wide range of study designs. Orthodox methods of evidence synthesis, such as those used for Cochrane reviews and informed by the GRADE criteria, classify mechanistic evidence as inherently lower quality than RCT evidence. But this position can be challenged on both epistemological grounds (since mechanistic evidence can raise or lower confidence in a causal claim) and on ethical grounds (since dismissing certain forms of evidence is a form of epistemic injustice). In this study, which seeks to extend the emerging scholarly tradition of 'EBM+' with 'SR+'. SR+ will formally and systematically evaluate mechanistic evidence and use that evidence to

complement and challenge findings from RCTs and other associative evidence.

The evidence pertaining to whether and how mask mandates work is heterogeneous and comes from multiple disciplines. The associative evidence consists mostly of observational (before and after) studies from the COVID-19 pandemic, where such mandates were introduced either country-wide or regionally (e.g. in particular US states or German regions), sometimes but not always with strict enforcement sanctions. These studies are of variable methodological quality (for example, some used only self-report data to capture frequency of mask usage). Mechanistic evidence (e.g. on the legal, institutional, economic, communicative, psychological, cultural and other ways in which mask mandates might affect compliance with mandates) comes from a wide range of (mostly) qualitative studies including interview, ethnographic and documentary data.

1. Jefferson T, Dooley L, Ferroni E, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database Syst Rev* 2023;1(1):Cd006207. doi:10.1002/14651858.CD006207.pub6 [published Online First: 20230130]
2. Greenhalgh T, MacIntyre CR, Baker MG, et al. Masks and respirators for prevention of respiratory infections: a state of the science review. *Clinical Microbiology Reviews* 2024:e00124-23.

Condition being studied Face mask mandates in reducing transmission of respiratory infections in the context of respiratory infection outbreaks. (Note: the question of how effectively face masks filter out pathogens from the air, while relevant to the related question of "[how effectively] do masks work?", is outwith the scope of this protocol; it is considered in a separate protocol, INPLASY 202550024.)

Examining the relationship between mask mandates and rates of mask usage (and, where measured, incidence of respiratory infections) requires both associative evidence (whether and to what extent mask mandates affect mask use and disease outcomes) and mechanistic evidence (how mask mandates might change mask use). A draft explanatory account, prepared before primary studies were appraised, identified five candidate mechanistic pathways through which a mandate might change masking behaviour:

- M1, legitimacy-based self-regulation: people comply because they regard the issuing authority as legitimate.
- M2, expressive coordination and norm activation: writing the rule into law states publicly what

conduct is expected, making that expectation common knowledge.

- M3, risk appraisal and cognitive endorsement: people comply because they accept that the risk is real, that masking works against it and that the mandate is a proportionate response.
- M4, identity coding and moralisation: masking becomes coded as identity-expressive or morally loaded, which can amplify compliance in some groups and provoke backlash in others.
- M5, deterrence and the enforcement climate: the threat of sanctions motivates people who would not otherwise comply.

Not all pathways need to be activated, but if none is, behaviour is not expected to change. The account also identified cross-cutting moderators: time and epidemic phase; risk communication and the information environment; setting and context; and equity and structural constraints on adherence.

METHODS

Search strategy Searching was iterative and continued to saturation, using multiple complementary methods:

- sources known to the authors, including studies identified in an earlier narrative review of mask efficacy;
- asking a leading scholar in each relevant discipline to recommend scientific overviews, systematic reviews and primary studies;
- mining existing systematic reviews and meta-analyses as a source of primary studies;
- manual searching of the reference lists of highly informative papers;
- forward citation tracking of highly informative papers in Google Scholar;
- conversational prompting of a large language model (Claude Sonnet 4.6) to identify weaknesses, missing citations and untested assumptions in the emerging explanatory account;
- AI citation matching (Research Rabbit), used towards the end of the review as a check on the other methods;
- two formal database searches, as follows.

November 2025, Web of Science (multiple databases including Medline, Science Citation Index, Social Science Citation Index and Chinese Science Citation Database): ALL=((mask OR facemask) AND (adherence or mask-use or mask use or mask usage) AND (COVID-19 or coronavirus OR SARS-COV-2) AND (mandate OR policy OR mandates) AND results).

June 2026, PubMed, PsycINFO, Embase and Cochrane Library. Four clusters, searched in title or

abstract and combined with AND: (1) mask terms (face mask or face cover or cloth mask or medical mask or surgical mask or respirator or respiratory protective device); (2) infection terms (covid* or influenza or SARS or severe acute respiratory syndrome or SARS-CoV-2 or respiratory infection); (3) policy terms (mandate or legal requirement or mask law or masking law); (4) outcome terms (transmission or infection or case rate or death or mortality or efficacy or effectiveness or impact or hospitalization or reduc or reproduction number). Excluded with NOT: randomised controlled trials and case-control studies of mask efficacy (the scope of the companion protocol); editorials, letters, comments, reviews, guidelines, news, meeting abstracts and book chapters; retracted publications and errata.

Participant or population The general population in any setting where a government mask mandate was in force for all or part of the period studied. Mandates could be national, regional or municipal, and could be limited to particular settings (for example, public transport) or to a defined population. Healthcare and clinical settings are excluded.

Intervention Mask mandates, defined as a mask requirement issued by a government body under statutory or regulatory authority (national, state or municipal). Non-binding recommendations and guidance do not qualify, nor does masking imposed by an institution under its own powers (for example, schools, universities, employers or carriers). Mandate status in each study setting was verified against the Oxford COVID-19 Government Response Tracker (OxCGRT) H6 indicator.

Comparator Settings or periods without a mask mandate in force: for example, jurisdictions without a mandate, the period before a mandate began or after it was withdrawn, or, in modelling studies, a scenario in which the mandate is switched off.

Study designs to be included No restriction on study design. Included quantitative studies were classified into four categories: quasi-experimental; mechanistic modelling; ecological association; and cross-sectional and behavioural. Qualitative studies are reported separately. Randomised trials and case-control studies of mask efficacy (masks versus no masks) fall under the companion protocol, INPLASY 202550024.

Eligibility criteria Primary studies that examined mask policy or mask wearing in a general-population setting under a legal mandate, and met the following conditions.

1. Mask policy or mask use enters the study's own analysis, as exposure, outcome, covariate, moderator or stratifier, counterfactual contrast or scenario parameter, or primary descriptive quantity.
2. A mask requirement issued by a government body under statutory or regulatory authority was in force in the setting studied for all or part of the period covered. It need not be measured by the study or be comprehensive. Where the data fall wholly after a mandate ended, the mandate or its withdrawal must be something the study observed.
3. For modelling studies, a mask policy is a distinct structural element of the model, such as a mandate date coded as an intervention or a scenario switching the policy off. In multi-component studies this applies to the modelling component alone.

Systematic reviews and state-of-the-science reviews were used as sources of primary studies and as background evidence for the explanatory account, not as included studies. Peer-reviewed literature was prioritised, and preprints were considered where evidence was otherwise insufficient [confirm].

Information sources Electronic databases (listed under Search strategy), sources known to the authors, topic experts in the field, existing systematic reviews, reference lists, forward citation tracking, and AI-assisted searching.

Main outcome(s) Whether and to what extent mask mandates affect the incidence of respiratory infections and related disease outcomes (for example, cases, hospitalisations, deaths, reproduction number or epidemic growth rate).

Additional outcome(s) Whether and to what extent mask mandates affect mask wearing, whether self-reported or directly observed. In addition, evidence supporting or refuting the five candidate mechanistic pathways and the cross-cutting moderators set out under Condition being studied.

Data management Eligible papers were stored, organised and coded in an EndNote database. Structured appraisals were produced as Word documents. All 185 appraisals, as generated by the appraisal tool (TERN version 7.4.9), and the tool's knowledge documents are deposited in a public repository (Zenodo, [DOI]); appraisals that the review team over-ruled are annotated to that effect. Data are stored on University of Oxford, University of Manchester and University of Exeter computers in line with university data policies.

Quality assessment / Risk of bias analysis

Instead of GRADE checklists, which assess bodies of evidence rather than individual studies, each primary study was appraised with a bespoke, design-specific tool developed from the draft explanatory account, generic risk-of-bias tools and GRADE guidance. The tools comprised a mechanisms-of-bias table for each design, recording the direction of each bias (towards the null, away from the null, or context-dependent); a topic-specific target design for each category, including a minimally sufficient adjustment set for observational designs; and a structured appraisal template per design.

Each study was assigned to one of four design categories (quasi-experimental; mechanistic modelling; ecological association; cross-sectional and behavioural) and, within it, to one of three credibility tiers: above the line (credible evidence contributing to causal inference), below the line (credible evidence on direction or mechanism of effect only), or uninformative (unable to support a mask-specific claim in either direction).

A large language model applied the tools through a purpose-built appraisal tool (TERN, version 7.4.9), drafting a structured appraisal of each study that included design classification, bias assessment and, where appropriate, quantitative bias analysis. Human reviewers checked every appraisal, resolving discrepancies by discussion and specialist input. Where the review team's judgement departed from the tool's, the reasons are recorded and reported.

Strategy of data synthesis Synthesis followed Mechanism-Informed Narrative Synthesis (MINS). Studies were synthesised first within each design category and then across categories. Within each category, two or three substantive studies were used to build a preliminary narrative covering design, key biases and their likely influence, how the authors addressed them, and the study's contribution to the explanatory account; less informative studies were then added to qualify and stress-test it. Cross-category synthesis identified mechanisms and failure modes recurring across categories, and the five candidate pathways were applied study by study to the above-the-line quasi-experimental studies.

No meta-analysis was performed, because the studies were too heterogeneous for pooling. Quantitative bias analysis was used, for individual studies and across groups of similar studies, as a stress test of whether a finding's direction was robust to plausible biases; it was not used to

adjust effect estimates or upgrade a study's credibility.

A large language model supported this step by organising comparative tables and drafting comparisons and a first-pass narrative. All interpretive judgements remained with the review team. Qualitative evidence will be synthesised and reported separately.

Subgroup analysis No subgroup analyses were pre-specified or performed. Variation in findings across settings and populations was examined through the mechanistic pathways and cross-cutting moderators.

Sensitivity analysis No formal sensitivity analyses were performed across studies. Quantitative bias analysis was used as a stress test of individual findings (see Strategy of data synthesis).

Language restriction No restriction. Studies published in languages not spoken by the review authors will be translated.

Country(ies) involved United Kingdom (University of Oxford).

Keywords See search strategy above.

Dissemination plans The appraisals and the appraisal tool's knowledge documents will be deposited in a public repository (Zenodo) alongside the published paper. Dissemination plans include preparation of academic papers, including methods and publication guidance for MINS to be submitted to the EQUATOR network. In addition, we will produce training materials, build links with stakeholders, and hold workshops for policymakers, academic reviewers, evaluators and lay (e.g. advocacy) groups.

Contributions of each author

Author 1 - Trisha Greenhalgh - Co-investigator on grant. Conceptualised the study with JW and RH. Provides topic expertise in masks. Designed search strategy. With SR, will undertake searches and data extraction and synthesis (including meta-analysis), with input from other authors. Will support postdocs as needed in producing publications.

Email: trish.greenhalgh@phc.ox.ac.uk

Author 2 - Jon Williamson - Chief investigator on grant. Co-developed the philosophical approach to be used and refined in SR+. Conceptualised the study with TG and RH. Leads philosophical aspects and line manages two philosophy postdocs.

Email: jon.williamson@manchester.ac.uk

Author 3 - Rebecca Helm - Co-investigator on grant. Conceptualised the study with JW and TG. Leads on legal and policy aspects of mask mandates.

Email: r.k.helm@exeter.ac.uk

Author 4 - Sahanika Ratnayake - Postdoctoral philosopher. Lead researcher on the mask mandate example. With TG, will undertake searches and data extraction and synthesis (including meta-analysis), with input from other authors. Will lead on some publications.

Email: sahanika.ratnayake@manchester.ac.uk

Author 5 - Luana Poliseli - Postdoctoral philosopher. Will support other authors in all aspects of searching, data extraction, data synthesis and writing up.

Email: luana.poliseli@manchester.ac.uk

Author 6 - Jon Jackson - Joined the author team during the analysis phase. Brought expert knowledge of the psychological mechanisms by which people obey the law (or do not).

Email: j.p.jackson@lse.ac.uk

Author 7 - Amanda Kvalsvig - Epidemiologist. Gave expert advice and undertook human appraisal of studies.

Email: amanda.kvalsvig@otago.ac.nz

Author 8 - Xinchun Gu - Epidemiologist. Gave expert advice and undertook human appraisal of studies.

Email: xinchun.gu@phc.ox.ac.uk

Author 9 - Lei Clifton - Epidemiologist, statistician and AI expert. Gave expert advice and oversaw the work of Dr Gu.

Email: lei.clifton@phc.ox.ac.uk