Meta-Analysis: Quantifying a Systematic Review

October 3, 2023

Ninet Sinaii, PhD, MPH
Biostatistics & Clinical Epidemiology Service
NIH Clinical Center
sinaiin@cc.nih.gov

Please Note

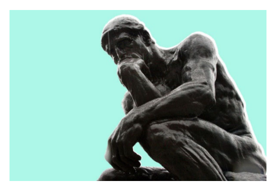
- This lecture covers general concepts about meta-analysis to help you better:
  - understand and prepare your data
  - design and prepare your studies
  - interpret your results and findings
  - make sense of results in published literature

- However, consult a statistician!
  - During the planning/design stage of a study
  - For data analysis and interpretation

Lecture Overview

- Meta-Analysis
  - Review and background
  - Meta-Analysis
  - Examples

Review and Background



Hypothesis Testing

- Process of translating a research question to a hypothesis that can be tested

- Involves conducting a test of statistical significance and quantifying the degree that variability may account for the results observed in a study (so we can make a conclusion about the population based on our study sample)

Hypothesis Testing

- Statistical hypothesis testing automates decision making:
  - the primary goal of analyzing data
  - Examine the evidence provided by the data

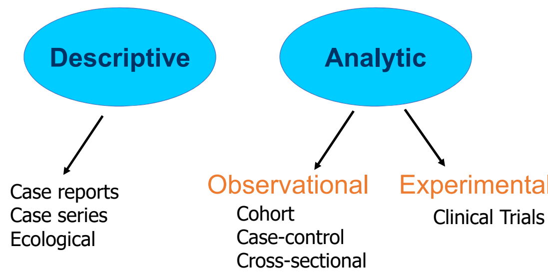
Explicit Statement of the Hypothesis to be Tested

- Null hypothesis (H_0):
  there is no association between IV and DV
  eg, $I_E = I_{\bar{E}}$ or RR = 1 or $\mu_x = \mu_y$
- Alternative (H_a):
  there is an association between IV and DV
  eg, $I_E \neq I_{\bar{E}}$ or RR $\neq$ 1 or $\mu_x \neq \mu_y$

Question: Is there enough evidence to reject H_0?
We expect (hope) to reject H_0 in favor of H_a.

The H_0 is set up with the hope that it will be rejected.

Types of Studies



Descriptive vs. Analytic

- Descriptive:
  - When little is known
  - Provide information on patterns of disease
  - Usually routinely collected data
  - Generate hypotheses

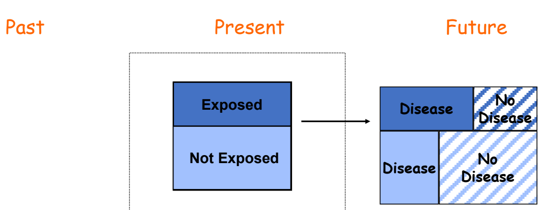
- Analytic:
  - When leads about etiology already known
  - Designed to specifically test hypotheses
  - Often entail new collection of data
  - If well-designed, allow more definitive conclusions

Observational vs. Experimental

Observational studies do not involve any intervention, experimental or otherwise
- Nature is allowed to take its course
- Investigator "passively" observes without intervention other than to record, classify, count, analyze

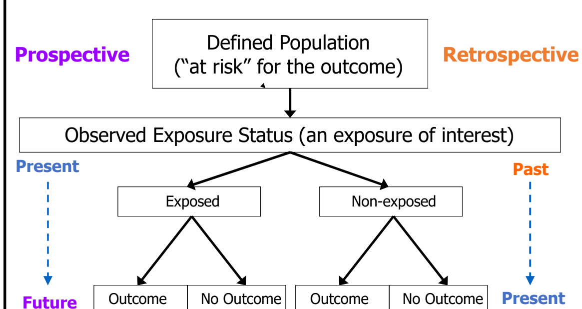
In experimental studies, *investigator has control* over the exposure and all other variables are held constant (equal)
- Typically employ randomization to minimize bias
- Best evidence of causality
- Similar in design to cohort studies

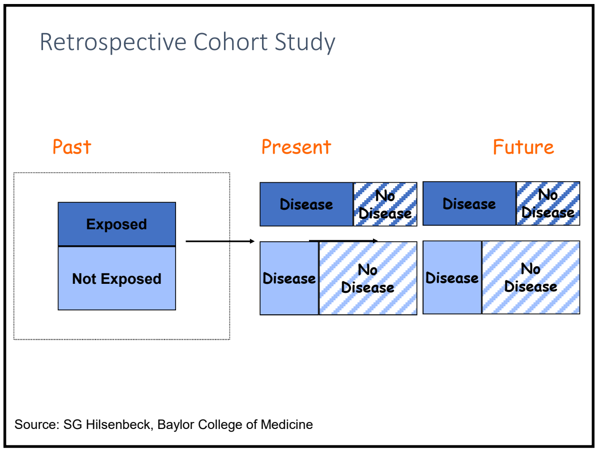
Cohort Study

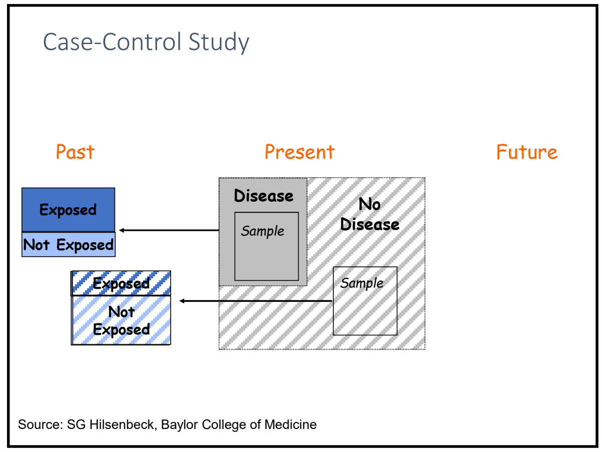


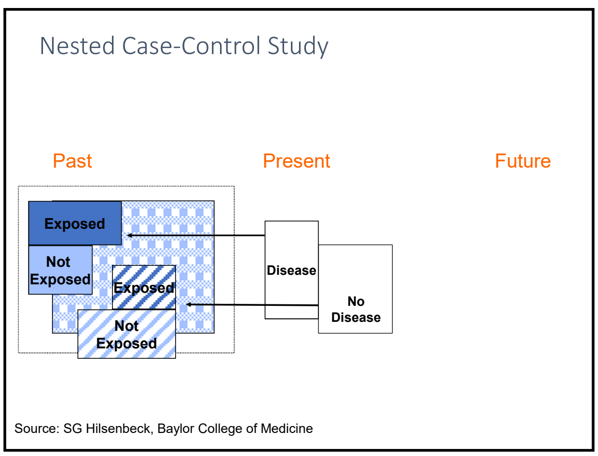
Source: SG Hilsenbeck, Baylor College of Medicine

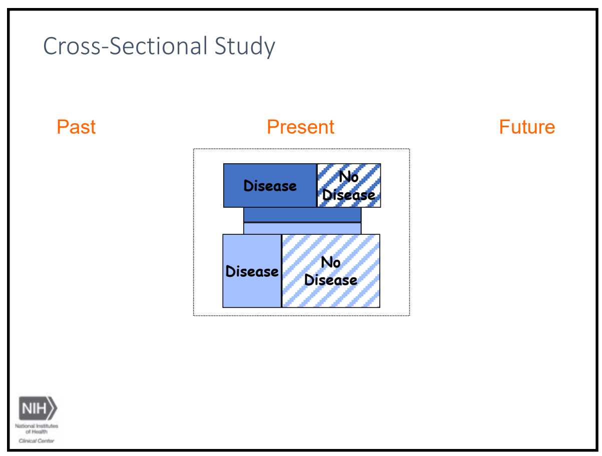
Prospective vs. Retrospective

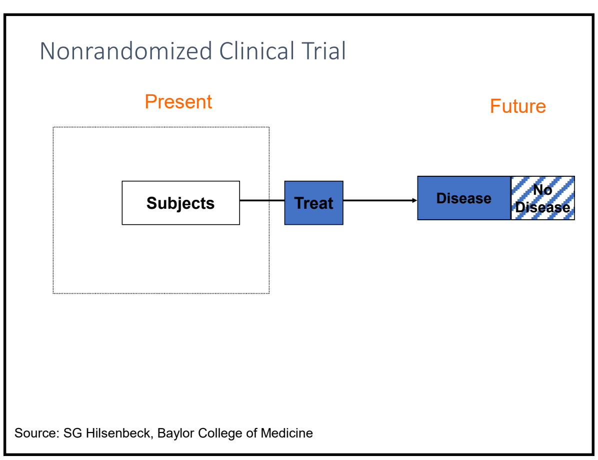


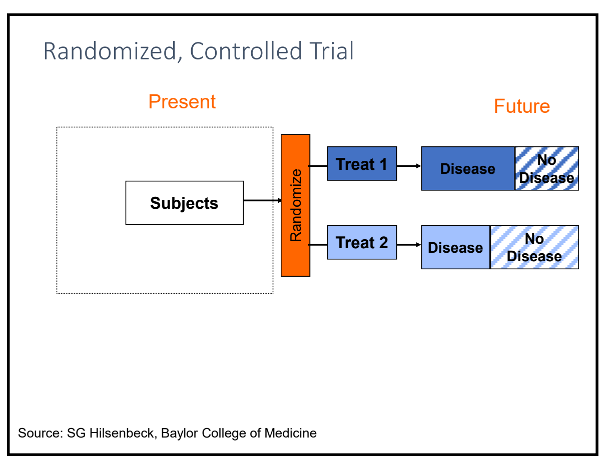

Retrospective Cohort Study
Past
Present
Future
Exposed
Not Exposed
Disease
No Disease
Disease
No Disease
Disease
No Disease
Disease
No Disease
Source: SG Hilsenbeck, Baylor College of Medicine


Case-Control Study
Past
Present
Future
Exposed
Not Exposed
Exposed
Not Exposed
Disease
Sample
No Disease
Sample
Source: SG Hilsenbeck, Baylor College of Medicine


Nested Case-Control Study
Past
Present
Future
Exposed
Not Exposed
Exposed
Not Exposed
Disease
No Disease
Source: SG Hilsenbeck, Baylor College of Medicine


Cross-Sectional Study
Past
Present
Future
Disease
No Disease
Disease
No Disease
NIH


Nonrandomized Clinical Trial
Present
Future
Subjects
Treat
Disease
No Disease
Source: SG Hilsenbeck, Baylor College of Medicine


Randomized, Controlled Trial
Present
Future
Subjects
Randomize
Treat 1
Treat 2
Disease
No Disease
Disease
No Disease
Source: SG Hilsenbeck, Baylor College of Medicine

Spectrum of Study Designs

Most scientifically rigorous

Experimental
- Randomized Controlled Trials
- Community Intervention Trials

Observational
- Cohort
- Case-Control
- Cross-Sectional

Descriptive
- Ecological
- Case reports/series

Least scientifically rigorous

What Follows These Studies?

- Meta-Analysis:
  - Systematic combination of data from several studies to develop a single conclusion that has greater statistical power
- Systematic Review:
  - Document that provides a comprehensive review of all relevant studies on a particular topic
  - Umbrella Review:
    - Compiles evidence from multiple reviews
- Practice Guideline:
  - Statement produced by panel of experts outlining current best practices to inform health care professionals and patients in making clinical decisions

Source: GWU Himmelfarb

Meta-Analysis vs. Systematic Review

- Meta-analysis is "the structuring process through which a thorough review of previous research is carried out" – it is the *statistical procedure*
- Systematic review is the *study design*

Meta-analysis should be as carefully planned as any other research project, with a detailed written protocol being prepared in advance

The a priori definition of eligibility criteria for studies to be included and a comprehensive search for such studies are central to high quality meta-analysis

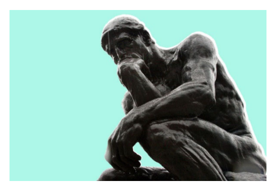
The graphical display of results from individual studies on a common scale is an important intermediate step, which allows a visual examination of the degree of heterogeneity between studies

Different statistical methods exist for combining the data, but there is no single "correct" method

A thorough sensitivity analysis is essential to assess the robustness of combined estimates to different assumptions and inclusion criteria

Figure Ref. Egger et al, BMJ, 1997

Meta-Analysis



What is Meta-Analysis?

- Statistical procedure to combine the results of similar, independent studies in order to obtain a single estimate of the effect that integrates all of the available information
- Pooling results of individual studies "as if they were one much larger study"

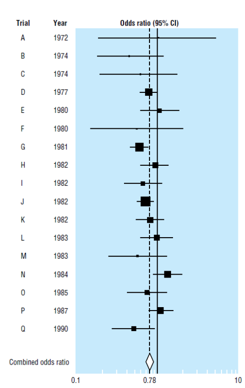


Figure ref: Egger et al, BMJ, 1997

Why Meta-Analysis?

- "Scientists have known for centuries that a single study will not resolve a major issue. Indeed, a small sample study will not even resolve a minor issue. Thus, the foundation of science is the cumulation of knowledge from the results of many studies."

Ref. Hunter JE, Schmidt FL, Jackson GB (1982) Meta-analysis: cumulating research findings across studies. Sage Publications, Beverly Hills (p. 10)

Meta-Analysis Example

Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis

Andrea Cipriani, Toshi A Furukawa*, Georgia Salanti*, Anna Chaimani, Lauren Z Atkinson, Yusuke Ogawa, Stefan Leucht, Henricus G Ruhe, Erick H Turner, Julian P T Higgins, Matthias Egger, Nozomi Takeshima, Yu Hayasaka, Hissei Imai, Kiyomi Shinohara, Aran Tajika, John P A Ioannidis, John R Geddes

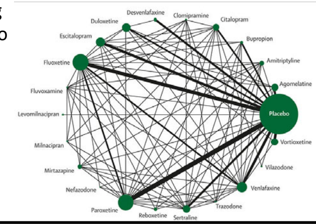
Summary

Background Major depressive disorder is one of the most common, burdensome, and costly psychiatric disorders worldwide in adults. Pharmacological and non-pharmacological treatments are available; however, because of inadequate resources, antidepressants are used more frequently than psychological interventions. Prescription of these agents should be informed by the best available evidence. Therefore, we aimed to update and expand our previous work to compare and rank antidepressants for the acute treatment of adults with unipolar major depressive disorder.

Lancet 2018; 391: 1357–66
Published Online February 21, 2018
http://dx.doi.org/10.1016/S0140-6736(17)32802-7
See Comment page 1333

Meta-Analysis Example

- Considerable debate re: the effectiveness of antidepressants
- Largest meta-analysis of antidepressants
- Study compared 21 commonly used antidepressants from 522 randomized controlled trials (RCTs); n=100,000
- Results: All antidepressants more effective than placebo for short-term treatment of acute depression in adults
- Allowed comparison among medications and with placebo to guide treatment decisions



Ref: Cipriani et al, The Lancet, 2018

Why Meta-Analysis?

- Meta-analysis goes beyond a narrative summary of key findings
  - Adds value by providing an objective and quantitative assessment of relationship between variables of interest or effectiveness of an intervention
  - Can also be used to test competing assumptions
  - Identify moderators, and explain reasons, where results are contradictory from primary studies

When is it Appropriate?

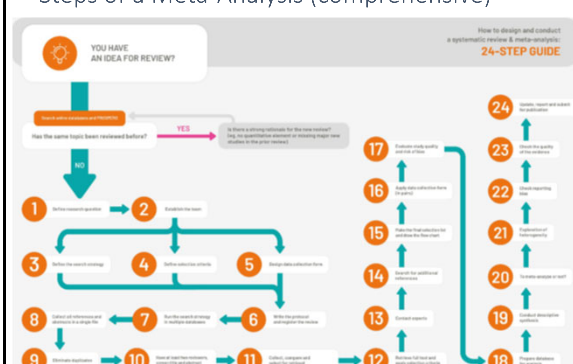
- Whenever there are multiple studies that test the same or similar hypotheses, and the joint findings of the studies do not clearly indicate the results
- When studies are contradictory
- When trying to review complex literature

→ The studies must contain sufficient information for the meta-analysis to be meaningful and appropriate for analysis!!

Only as Good as What Goes In

- Systematic reviews and meta-analyses are only as good as the rigor involved
  - Similar studies
  - Similar interventions
  - Similar outcomes
- Otherwise, not informative

Steps of a Meta-Analysis (comprehensive)



Ref: Muka et al, Eur J of Epidemiol, 2020

Steps of a Meta-Analysis (comprehensive also)

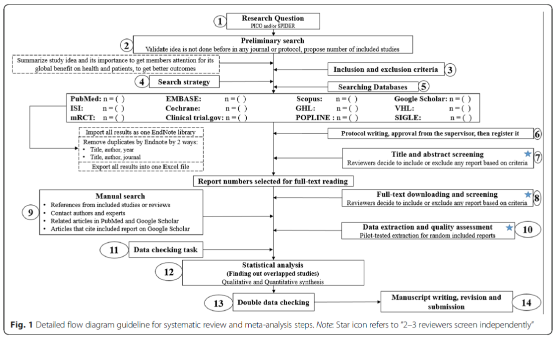


Fig. 1 Detailed flow diagram guideline for systematic review and meta-analysis steps. *Note*: Star icon refers to "2–3 reviewers screen independently"

Ref. Tawfik et al, Tropical Medicine and Health, 2019

Steps of a Meta-Analysis (condensed)

- 1. Define the research question
- 2. Literature search
- 3. Measure of association
- 4. Analytical method
- 5. Preparing data for analysis
- 6. Software
- 7. Analysis
- 8. Reporting results

Image ref. iStock

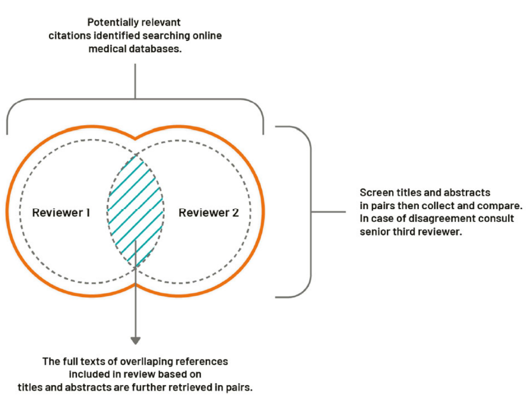
Step 1: Define the Research Question

- Narrow topic and hypothesis being tested
- Specified clearly
- Structured questions before review work
  - Modifications allowed if alternate ways become necessary
- Detailed:
  - Population
  - Interventions, exposures, pertinent variables
  - Outcomes
  - Study design
  - Inclusions/exclusion criteria

Step 2: Literature Search

- Systematic, reproducible, transparent
- Capture as many relevant citations as possible
  - Can use previous related meta analyses as starting point
- Available databases to aid search, use multiple databases if possible
- Don't forget references in articles, abstracts, presentations, unpublished data
- Eliminate duplicates
- Screening process and multiple reviewers
- Evaluate for quality, risk of bias, and whether it contains sufficient information to be included
  - Design threshold for study selection
  - Inclusion/exclusion criteria
  - Should be detailed and reported
- (Start database preparation)

Cochrane.org



Ref: Muka et al, Eur J of Epidemiol, 2020

Assessing the Quality of Studies

- Quality hierarchy can be developed
  - degree of compliance with criteria
  - types of study designs
  - completeness of data
  - appropriateness of population
  - inclusion/exclusion criteria, etc
- Cochrane Collaboration's tool RoB 2 for RCTs; Newcastle-Ottawa Scale for prospective observational studies (see Muka et al ref)

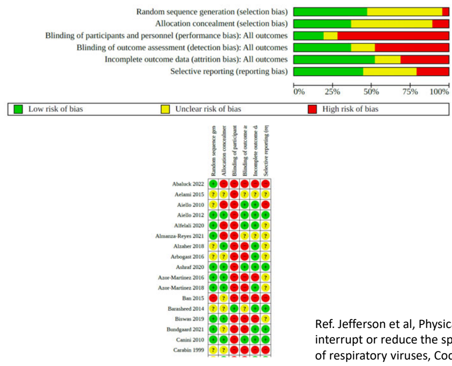
Table 1 **Description of quality assessment of studies on safety of public water fluoridation**

Quality categories	*High*	*Moderate*	*Low*
Prospective design	Prospective	Prospective	Prospective or retrospective
Ascertainment of exposure	Study began within 1 year of fluoridation	Study began within 3 years of fluoridation	Study began > 3 years after fluoridation
Ascertainment of outcome	Follow-up for at least 5 years and blind assessment	Long follow-up and blind assessment	Short follow-up and unblinded assessment
Control for confounding	Adjustment for at least three confounding factors (or use of randomization)	Adjustment for at least one confounding factor	No adjustment for confounding factors

Ref: Khan, et al, J R Soc Med, 2003

Risk of Bias Example

Figure 2. 'Risk of bias' graph: review authors' judgements about each risk of bias item presented as percentages across all included trials.

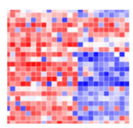


Ref. Jefferson et al, Physical interventions to interrupt or reduce the spread of respiratory viruses, Cochrane Review

Step 3: Measure of Association

- Choice of effect size measure
- Types of effect sizes
  - Correlation coefficient
  - Standardized means (or differences)
  - Odds Ratios
  - Relative Risks
  - Effect Size, etc
- Conversion of effect sizes to common measure

Step 4: Analytical Methods



- Directly related to the research question
  - Relation, effect of intervention, etc
- Main broad categories:
  - Univariable meta-analysis → weighted mean effect size (of study characteristics, sample characteristics, measurement) and provides magnitude of variance among included primary studies
  - Meta-regression analysis → investigates heterogeneity, tests multiple factors simultaneously
  - Meta-analytic structural equation modeling → complex
    - Pooled correlation matrix
    - Matrix used to fit path model
  - Qualitative meta-analysis → provides in-depth information about specific factors, explain heterogeneity or contradictory results among primary studies

Ref: Hansen et al, Manag Rev Q, 2022

Step 5: Preparing Data for Analysis

- Database
- Data details and coding
  - Depends on methods, respective software (next step), research design, measures of association
- Collect details needed for computing effect sizes
  - Sample size, actual numbers with/without a response, numbers in groups, standard deviations (standard errors, confidence intervals), etc
  - Include potential confounders or other pertinent variables
- Can contact investigators of original studies for details!

	A	B	C	D	E	F	G	H	I	J	K	...
1	*ID1*	*ID2*	*STUDY_NAME*	*YEAR*	*ES*	*N*	*COUNTRY*	*SAMPLE_PERIOD*	*FIRM_TYPE*	*TMT_DIVERSITY*	*PROFITABILITY*	
2	1	1	Doe et al. (1995)	1995	0.15	150	USA	1990-1992	Private	Blau index	ROA	...
3	2	1	Average & Ordinary (2010)	2010	-0.05	900	China	2005-2008	Listed	Female ratio	ROA	...
4	3	1	Citizen et al. (2015)	2015	0.07	550	USA	2000-2010	Listed	Female ratio	ROE	...
5	3	2	Citizen et al. (2015)	2015	0.00	550	USA	2000-2010	Listed	Blau index	ROE	...
...	...	...	...	...	...	...	...	...	...	...	...	...
100	75	1	Bloggs & Public (2019)	2019	0.20	275	Italy	2006-2015	Private	Female ratio	ROE	...

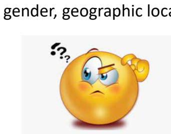
Image Ref: Hansen et al, Manag Rev Q, 2022

Step 6: Software

- Traditional statistical packages range from built-in functions or additional packages focused on meta-analyses
- Examples:
  - SAS (macros for meta-analysis)
  - Stata (built-in functions)
  - SPSS (macros for meta-analysis)
  - R (60+ meta-analysis packages and functionalities)
  - Python (PythonMeta package)
  - Others (metaphor, Comprehensive Meta-Analysis, RevMan, Meta-Essentials)

Step 7: Analysis: Meta-Analyze or Describe?

- Are the data gathered suitable for pooling using quantitative methods?
- When data are combined:
  - Sample size increases and so does statistical power
  - Improves estimates of the effect size
  - Has potential for resolving uncertainty when primary studies disagree
- Decision to pool depends on degree of heterogeneity
  - Design
  - Population characteristics (age, gender, geographic location)
  - Different methods used
  - Adjustments
  - Measures of association



Analysis: Descriptive Synthesis

- Summarize and explain findings
  - Number of references and search strategy
  - Populations studies
  - Types of exposures or variables
  - Intervention details
  - Outcomes
  - Description of results: direct and size of effect, consistency across studies, strength of evidence

Table 1. Characterization of Studies Describing Coronavirus Disease 2019 Prevalence in Health-Care Workers Using Reverse Transcription–Polymerase Chain Reaction, 2020

First Author, Year (Reference No.)	No. of Participants	Setting	Selection Criteria	City, Country
Kluytmans, 2020 (31)	86	Hospital, all services	HCWs with fever or (mild) respiratory symptoms	Breda, The Netherlands
García-Basteiro, 2020 (25)	583	Hospital, all services	All HCWs who deliver care and services to patients, either directly as physicians or nurses, or indirectly as assistants, technicians, stretcher bearers, or other support staff (e.g., administrative officers, cleaning, kitchen, laundry, maintenance)	Barcelona, Spain
Breazzano, 2020 (19)	2,088	Hospitals at the city level. Data reported by each postgraduate director	Medical residents in universities located in New York City. Screened according to symptoms	New York City, NY, USA

Flow Diagram

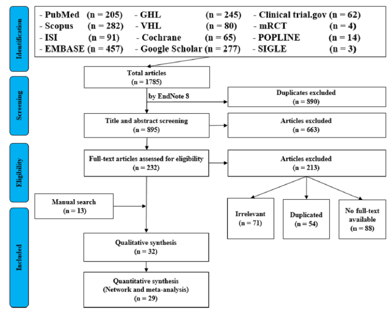


Fig. 2 PRISMA flow diagram of studies' screening and selection

*Please check PRISMA guidelines for appropriate reporting of type of meta-analysis

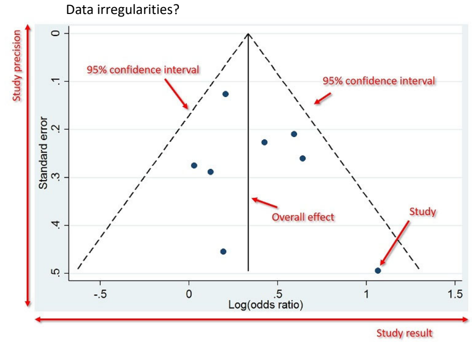
Ref. Tawfik et al, Tropical Medicine and Health, 2019

Step 7: Analysis

- Preliminary sensitivity analyses:
  - Influential outliers
- Tests for publication bias
  - Funnel plot – simple graphical tool that provides an overview of the effect size distribution; can help to identify publication bias, outliers
    - Uses deviation (eg, studentized residuals) and leverage (Cook's distance)
  - Subgroup comparisons between published and unpublished
  - Other specific tests – but should not rely on a single test → use multiple tests that may be conceptually different ("triangulation approach")
- Model choice (fixed effects vs random effects)

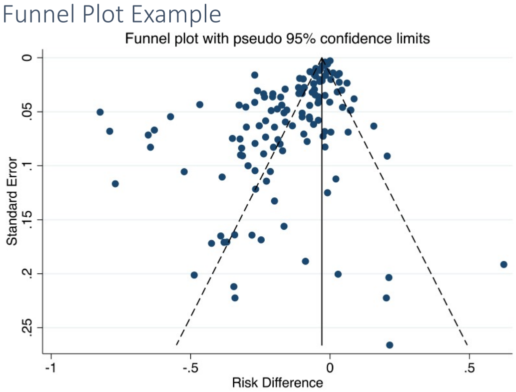
Funnel Plot Example

- Scatterplot of study results (eg, effect size, treatment effect) vs measure of precision
- Asymmetric = Selection/publication bias? Inappropriate effect measure? True heterogeneity? Data irregularities?



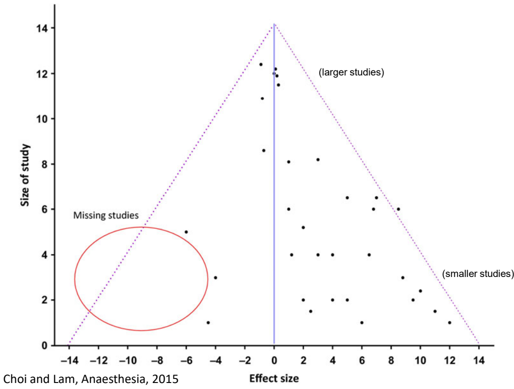
Ref: Egger et al, Bias in meta-analysis detected by a simple, graphical test, BMJ, 1997 Figure Ref. Top Tip Bio

Funnel Plot Example



Ref: WHO Guidelines on Use of Medically Important Antimicrobials in Food-Producing Animals.

Funnel Plot Example



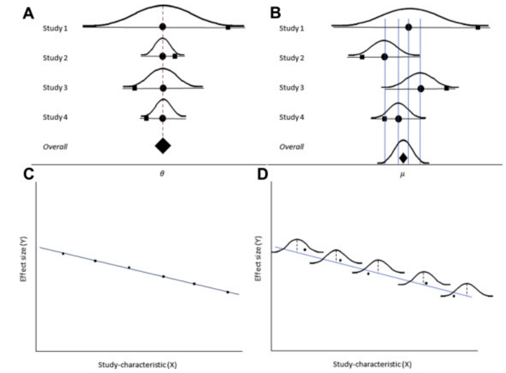
Ref: Choi and Lam, Anaesthesia, 2015

Model Choice

- Fixed effects
  - Assumes all studies share a common true effect size (ie, summary effect size)
  - Assumes differences are due to sampling error
- Random effects
  - Assume normal distribution of true effect size and estimate the mean (ie, summary effect size) and variance of the distribution (ie, heterogeneity)
  - Allow for a variation of the true effect size across studies (eg, different locations, types of participants)

Ref. Spineli and Pandis, Stat and Res Design, 2020

Fixed-Effects vs. Random-Effects



Ref. Spineli and Pandis, AJO-DO Stat and Res Design, 2020

	Fixed effect model (FE)	Random effect model (RE)
Assumption	All of the studies in the meta-analysis have one true effect size, and the observed variation among studies is caused by sampling errors or chance.	Different studies exhibit substantial diversity, and the true effect size may vary from study to study.
Consequence	The weights assigned to each study depend on the study's precision; each study's weight is equal to the inverse of its variance. The larger the sample size in a trial, the smaller the variance of the effect size and the larger the corresponding weight assigned in the meta-analysis. Bigger studies contribute more in the estimation of the summary effect and are assigned larger weights, whereas smaller studies convey less information and are assigned smaller weights. Assesses only intra-study sampling errors (intra-study variation).	The summary effect is estimated as a weighted average and the weights assigned to each observed effect size equal the inverse of their variance plus an additional variance component that reflects heterogeneity. The summary estimate under the obtains more nformation from the larger and more precise studies but the distribution of the weights is not as much contrasted as under an FE model. Assesses both intra-study sampling errors and inter-study variance (between-study variation)
Application	FE model may be used when there is strong evidence that all trials/studies are functionally identical and inference is limited to the population included in the analysis.	1. When there are methodological or clinically relevant differences in the included studies 2. When in between study heterogeneity is high NOTE 1: Test for homogeneity (χ2 test) has low power when studies have small sample sizes or are few in number, and a lack of statistical significance does not guarantee the absence of heterogeneity.
Summary effect interpretation	The summary effect is the best estimate of the common treatment effect and together with its uncertainty they are the only information of relevance	Effect estimate is an estimation of the average of a collection of possible treatment effects in various settings. NOTE 2: RE is not a solution for extreme heterogeneity. If present, differences in effects should be explored via subgroup analyses or meta-regression.

Ref: Muka et al, Eur J of Epidemiol, 2020

Statistical Assessment of Heterogeneity

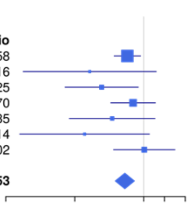
- Cochrane's Chi squared test (Cochran's Q)
  - H_0: all studies are evaluating the same effect
- Higgins's I^2 is the percentage of variation between the sample estimates that is due to heterogeneity rather than sampling error
  - Ranges 0% to 100% (maximum heterogeneity)
  - <25% low; 25-75% moderate; >75% high
- Subgroup or Stratification
- Synthase different study designs?
  - Inclusion of more than one design can improve the quality of systematic review
  - Include RCTs? Randomized vs non-randomized?

Step 8: Reporting Results

- CCC: Clear, complete, comprehensible
- Checklists and guidelines (PRISMA, QUOROM, MOOSE)
- GRADE approach to score quality of evidence in the review
- Tables and figures
  - Forest Plot
  - Include all important information and test statistics
    - Sample size
    - Effect sizes and variability (standard deviation, confidence intervals)
- Open science practices

Forest Plot Example

Study	Deaths (corticosteroid)	Deaths (placebo)	Odds ratio
Auckland	36	60	0.58
Block	1	5	0.16
Doran	4	11	0.25
Gamsu	14	20	0.70
Morrison	3	7	0.35
Papageorgiou	1	7	0.14
Tauesch	8	10	1.02
Summary			0.53



Odds ratio with 95% confidence interval
(1=no effect, <1=treatment has fewer deaths)

Reading a forest plot:
Study identities, effect size, confidence intervals, weight, heterogeneity

Ref. wikipedia

Forest Plot Example

Study	Vaccine A Events	Vaccine A Total	Placebo Events	Placebo Total	OR	95%-CI	Weight
A	30	120	20	100	1.33	[0.70; 2.53]	21.4%
B	15	90	12	70	0.97	[0.42; 2.22]	12.7%
C	25	110	24	90	0.81	[0.42; 1.54]	21.1%
D	17	85	14	80	1.18	[0.54; 2.58]	14.3%
E	14	72	12	60	0.97	[0.41; 2.28]	11.9%
F	23	115	18	100	1.14	[0.57; 2.26]	18.7%
Random effects model		592		500	**1.06**	**[0.79; 1.42]**	**100.0%**

Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0$, $\chi^2_5 = 1.37$ ($p = 0.93$)

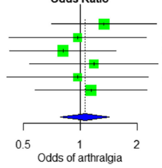
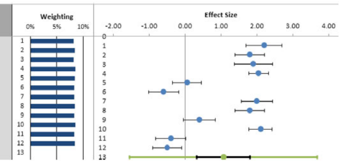


Fig. 3 Random effect model forest plot for comparison of vaccine A versus placebo

Ref. Tawfik et al, Tropical Medicine and Health, 2019

Forest Plot Example, and Prediction Interval

#	Study name	Effect size	CI Lower limit	CI Upper limit	Weight
1	aaaa	2.20	1.70	2.70	8.22%
2	bbbb	1.80	1.39	2.21	8.35%
3	cccc	1.90	1.37	2.43	8.17%
4	dddd	2.05	1.77	2.33	8.49%
5	eeee	0.05	-0.35	0.45	8.35%
6	ffff	-0.60	-1.02	-0.18	8.33%
7	gggg	2.00	1.56	2.44	8.30%
8	hhhh	1.80	1.39	2.21	8.35%
9	iiii	0.40	-0.05	0.85	8.30%
10	jjjj	2.10	1.78	2.42	8.45%
11	kkkk	-0.40	-0.82	0.02	8.33%
12	llll	-0.50	-0.90	-0.10	8.36%
13					



"Prediction interval" around the combined effect size gives a range for the point estimate where 95% of future studies will fall.
But, it can be affected by:
- -assumes true effect sizes are approx. normally distributed
- -based on effect sizes observed in the included studies
- -can "predict" if there's no selection bias, representative studies, other bias

So, recommendation is to interpret the prediction interval as a *description of the range of observed effect sizes* rather than a prediction of the range of the effect sizes to expect in future studies

Refs. Hak et al, The Netherlands: Erasmus Rotterdam Institute of Management, 2016;
Michael Borenstein et al. (2009), *Introduction to Meta-Analysis*, Chichester (UK): Wiley.

Meta-Analysis of Individual Participant Data

- Generate a summary estimate using Individual Participant Data (IPD), may be *published* or *unpublished*
- Take more time to complete, more expensive, more challenges in obtaining data
- Offer more control, may reduce study heterogeneity, can pose new questions
- Procedure same as meta-analysis (of aggregate data), with specialized focus on:
  - Harmonizing data – use of same original measure across studies/individuals
  - Individual estimates calculated and plotted to observe similarities and differences
  - Fitted together in a single model – must account for clustering
- PRISMA-IPD statement specifically addresses IPD

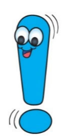
Refs. Columbia Public Health, Population Health Methods, 2022; or Chochrane.org

Meta-Analysis Assumptions

- Researcher is supposed to verify these are true for the meta-analysis:
  - An effect is precisely defined (independent and dependent variables, empirical studies)
  - Specifies type of unit in which the effect occurs (eg, persons, groups, countries) and the domain for which the effect will be meta-analyzed (eg, all persons? Adults only? Women only? Which countries?)
  - All empirical studies of the effect in the applicable domain have been identified (issue: are all possible studies published?)
  - All studies are methodologically sound, well carried-out, valid, reliable, adequate statistics – need methodological evaluation of included studies; otherwise, garbage in, garbage out
  - Effect size measures in the studies are comparable

Ref. Hak et al, The Netherlands: Erasmus Rotterdam Institute of Management, 2016

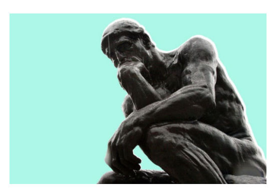
Issues in Meta-Analysis



- When all studies use the same endpoint (rare) or very similar designs
- Studies are heterogeneous (dissimilar)
- Publication bias: "positive" studies are more likely to be published
- Studies may differ widely in quality (but restricting to only studies that are "high quality" can introduce bias)
- There may be important confounders that studies may not have accounted or provide enough information/detail
- Treating studies as fixed vs random effects
- Multi-center trials don't usually report center-specific results
- Assumes studies are independent of each other

NIH National Institutes of Health Clinical Center

Examples



NIH National Institutes of Health Clinical Center

Meta-Analysis Example

American Journal of Epidemiology
Vol. 190, No. 1
DOI: 10.1093/aje/kwaa191
Advance Access publication: September 1, 2020

Systematic Reviews and Meta- and Pooled Analyses

COVID-19 in Health-Care Workers: A Living Systematic Review and Meta-Analysis of Prevalence, Risk Factors, Clinical Characteristics, and Outcomes

Sergio Alejandro Gómez-Ochoa, Oscar H. Franco, Lyda Z. Rojas, Peter Francis Raguindin, Zayne Milena Roa-Díaz, Beatrice Minder Wyssmann, Sandra Lucrecia Romero Guevara, Luis Eduardo Echeverría, Marija Glisic, and Taulant Muka*

* Correspondence to Dr. Taulant Muka, Institute of Social and Preventive Medicine, University of Bern, Mittelstrasse 43, 3012, Bern, Switzerland (e-mail: taulant.muka@ispm.unibe.ch).

Meta-Analysis Example

American Journal of Infection Control 49 (2021) 21–29

Contents lists available at ScienceDirect

American Journal of Infection Control

journal homepage: www.ajicjournal.org

Major Article

Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19

Beatriz Böger Msc [a], Mariana M. Fachi Msc [a], Raquel O. Vilhena PhD [b], Alexandre F. Cobre MSc [a], Fernanda S. Tonin PhD [a], Roberto Pontarolo PhD [b,*]

[a] Pharmaceutical Sciences Postgraduate Program, Health Sciences Sector, Federal University of Paraná, Curitiba, Brazil
[b] Department of Pharmacy, Federal University of Paraná, Curitiba, Brazil

Cochrane Library

Cochrane Database of Systematic Reviews

Physical interventions to interrupt or reduce the spread of respiratory viruses (Review)

Jefferson T, Dooley L, Ferroni E, Al-Ansary LA, van Driel ML, Bawazeer GA, Jones MA, Hoffmann TC, Clark J, Beller EM, Glasziou PP, Conly JM

Cochrane Reviews are considered the gold standard of evidence-based medicine.

Data collection and analysis

We used standard Cochrane methodological procedures.

Main results

We included 11 new RCTs and cluster-RCTs (610,872 participants) in this update, bringing the total number of RCTs to 78. Six of the new trials were conducted during the COVID-19 pandemic; two from Mexico, and one each from Denmark, Bangladesh, England, and Norway. We identified four ongoing studies, of which one is completed, but unreported, evaluating masks concurrent with the COVID-19 pandemic.

Many studies were conducted during non-epidemic influenza periods. Several were conducted during the 2009 H1N1 influenza pandemic, and others in epidemic influenza seasons up to 2016. Therefore, many studies were conducted in the context of lower respiratory viral circulation and transmission compared to COVID-19. The included studies were conducted in heterogeneous settings, ranging from suburban schools to hospital wards in high-income countries; crowded inner city settings in low-income countries; and an immigrant neighbourhood in a high-income country. Adherence with interventions was low in many studies.

The risk of bias for the RCTs and cluster-RCTs was mostly high or unclear.

Medical/surgical masks compared to no masks

We included 12 trials (10 cluster-RCTs) comparing medical/surgical masks versus no masks to prevent the spread of viral respiratory illness (two trials with healthcare workers and 10 in the community). Wearing masks in the community probably makes little or no difference to the outcome of influenza-like illness (ILI)/COVID-19 like illness compared to not wearing masks (risk ratio (RR) 0.95, 95% confidence interval (CI) 0.84 to 1.09; 9 trials, 276,917 participants; moderate-certainty evidence. Wearing masks in the community probably makes little or no difference to the outcome of laboratory-confirmed influenza/SARS-CoV-2 compared to not wearing masks (RR 1.01, 95% CI 0.72 to 1.42; 6 trials, 13,919 participants; moderate-certainty evidence). Harms were rarely measured and poorly reported (very low-certainty evidence).

N95/P2 respirators compared to medical/surgical masks

We pooled trials comparing N95/P2 respirators with medical/surgical masks (four in healthcare settings and one in a household setting). We are very uncertain on the effects of N95/P2 respirators compared with medical/surgical masks on the outcome of clinical respiratory illness (RR 0.70, 95% CI 0.45 to 1.10; 3 trials, 7779 participants; very low-certainty evidence). N95/P2 respirators compared with medical/surgical masks may be effective for ILI (RR 0.82, 95% CI 0.66 to 1.03; 5 trials, 8407 participants; low-certainty evidence). Evidence is limited by imprecision and heterogeneity for these subjective outcomes. The use of a N95/P2 respirators compared to medical/surgical masks probably makes little or no difference for the objective and more precise outcome of laboratory-confirmed influenza infection (RR 1.10, 95% CI 0.90 to 1.34; 5 trials, 8407 participants; moderate-certainty evidence). Restricting pooling to healthcare workers made no difference to the overall findings. Harms were poorly measured and reported, but discomfort wearing medical/surgical masks or N95/P2 respirators was mentioned in several studies (very low-certainty evidence).

One previously reported ongoing RCT has now been published and observed that medical/surgical masks were non-inferior to N95 respirators in a large study of 1009 healthcare workers in four countries providing direct care to COVID-19 patients.

Hand hygiene compared to control

Nineteen trials compared hand hygiene interventions with controls with sufficient data to include in meta-analyses. Settings included schools, childcare centres and homes. Comparing hand hygiene interventions with controls (i.e. no intervention), there was a 14% relative reduction in the number of people with ARIs in the hand hygiene group (RR 0.86, 95% CI 0.81 to 0.90; 9 trials, 52,105 participants; moderate-certainty evidence), suggesting a probable benefit. In absolute terms this benefit would result in a reduction from 380 events per 1000 people to 327 per 1000 people (95% CI 308 to 342). When considering the more strictly defined outcomes of ILI and laboratory-confirmed influenza, the estimates of effect for ILI (RR 0.94, 95% CI 0.81 to 1.09; 11 trials, 34,503 participants; low-certainty evidence), and laboratory-confirmed influenza (RR 0.91, 95% CI 0.63 to 1.30; 8 trials, 8332 participants; low-certainty evidence), suggest the intervention made little or no difference. We pooled 19 trials (71, 210 participants) for the composite outcome of ARI or ILI or influenza, with each study only contributing once and the most comprehensive outcome reported. Pooled data showed that hand hygiene may be beneficial with an 11% relative reduction of respiratory illness (RR 0.89, 95% CI 0.83 to 0.94; low-certainty evidence), but with high heterogeneity. In absolute terms this benefit would result in a reduction from 200 events per 1000 people to 178 per 1000 people (95% CI 166 to 188). Few trials measured and reported harms (very low-certainty evidence).

We found no RCTs on gowns and gloves, face shields, or screening at entry ports.

SUMMARY OF FINDINGS

Summary of findings 1. Medical/surgical masks compared to no masks for preventing the spread of viral respiratory illness

Randomised studies: medical/surgical masks compared to no masks for preventing the spread of viral respiratory illness

Patient or population: general population
Setting: community and hospitals
Intervention: medical/surgical masks
Comparison: no masks

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with no masks	Risk with randomised studies: masks				
Viral respiratory illness - influenza/COVID-like illness	Study population		RR 0.95 (0.84 to 1.09)	276,917 (9 RCTs)	⊕⊕⊕⊝ Moderate[a]	
	160 per 1000	152 per 1000 (134 to 174)				
Viral respiratory illness - laboratory-confirmed influenza/SARS-CoV-2	Study population		RR 1.01 (0.72 to 1.42)	13,919 (6 RCTs)	⊕⊕⊕⊝ Moderate[b]	
	40 per 1000	40 per 1000 (29 to 57)				
Adverse events	-		-	(3 RCTs)	⊕⊝⊝⊝ Very low[a,c]	Adverse events were not reported consistently and could not be meta-analysed. Adverse events reported for masks included warmth, discomfort, respiratory difficulties, humidity, pain, and shortness of breath, in up to 45% of participants.

***The risk in the intervention group** (and its 95% confidence interval) is based on the median observed risk in the comparison group of included studies and the **relative effect** of the intervention (and its 95% CI).
CI: confidence interval; **RCT:** randomised controlled trial; **RR:** risk ratio

Summary of findings 2. N95 respirators compared to medical/surgical masks for preventing the spread of viral respiratory illness

Randomised studies: N95 respirators compared to medical/surgical masks for preventing the spread of viral respiratory illness

Patient or population: general population and healthcare workers
Setting: hospitals and households
Intervention: N95 masks
Comparison: medical/surgical masks

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with medical masks	Risk with randomised studies: N95				
Viral respiratory illness - clinical respiratory illness	Study population		RR 0.70 (0.45 to 1.10)	7799 (3 RCTs)	⊕⊝⊝⊝ Very Low[a,b,c]	All studies were conducted in hospital settings with healthcare workers.
	120 per 1000	84 per 1000 (54 to 132)				
Viral respiratory illness - influenza-like illness	Study population		RR 0.82 (0.66 to 1.03)	8407 (5 RCTs)	⊕⊕⊝⊝ Low[a,b]	1 study was conducted in households (MacIntyre 2009).
	50 per 1000	41 per 1000 (33 to 52)				
Viral respiratory illness - laboratory-confirmed influenza	Study population		RR 1.10 (0.90 to 1.34)	8407 (5 RCTs)	⊕⊕⊕⊝ Moderate[b]	1 study was conducted in households (MacIntyre 2009).
	70 per 1000	77 per 1000 (63 to 94)				
Adverse events	-		-	(5 RCTs)	⊕⊝⊝⊝ Very Low[a,b,c]	There was insufficient consistent reporting of adverse events to enable meta-analysis. Only 1 study reported detailed adverse events: discomfort was reported in 41.9% of N95 wearers versus 9.8% of medical mask wearers ($P < 0.001$); headaches

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Downgraded one level for study limitations (lack of blinding).
Downgraded one level for imprecision (wide confidence interval or no meta-analysis conducted).
Downgraded one level for inconsistency of results (heterogeneity).

AUTHORS' CONCLUSIONS

Implications for practice

The evidence summarised in this review on the use of masks is largely based on studies conducted during traditional peak respiratory virus infection seasons up until 2016. Two relevant randomised trials conducted during the COVID-19 pandemic have been published, but their addition had minimal impact on the overall pooled estimate of effect. The observed lack of effect of mask wearing in interrupting the spread of influenza-like illness (ILI) or influenza/COVID-19 in our review has many potential reasons, including: poor study design; insufficiently powered studies arising from low viral circulation in some studies; lower adherence with mask wearing, especially amongst children; quality of the masks used; self-contamination of the mask by hands; lack of protection from eye exposure from respiratory droplets (allowing a route of entry of respiratory viruses into the nose via the lacrimal duct); saturation of masks with saliva from extended use (promoting virus survival in proteinaceous material); and possible risk compensation behaviour leading to an exaggerated sense of security (Ammann 2022; Brosseau 2020; Byambasuren 2021; Canini 2010; Cassell 2006; Coroiu 2021; MacIntyre 2015; Rengasamy 2010; Zamora 2006).

Robin Monotti @robinmonotti · Follow

"Wearing masks in the community probably makes little or no difference to the outcome of laboratory-confirmed influenza/SARS-CoV-2 compared to not wearing masks"

@CochraneLibrary Review.
Published: 30 January 2023
cochranelibrary.com/cdsr/doi/10.10...

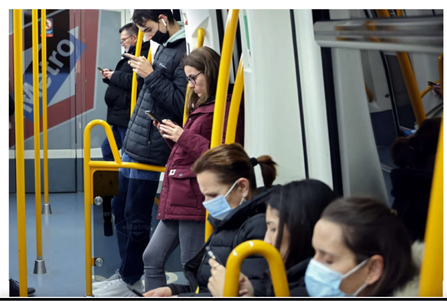
SCIENCE

To Mask or Not to Mask: That Is (Somehow) Still a Question

The latest, highest quality evidence does not show that masks effectively protect against COVID-19.

BY LIZ HIGHLEYMAN

FEB 13, 2023 • 5:45 AM

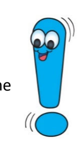


But…

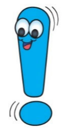
conclusions were misleading

Absence of Evidence is Not Evidence of Absence

- Underlying assumptions about transmission were incorrect
  - Flu/COVID spread through air; only tight-fitting masks prevent air leakage; all masks are not equal
- Different settings have different risks
  - Community compliance was generally poor
  - People don't always wear high-quality masks the right way all the time
- Circulating viruses vary by season/year
- Studies addressed different questions
  - RCTs can be used in meta-analysis if they are addressing the same question in the same way
  - "Partial" use was combined with "continuous" use
  - No proper controls
  - Studies tested protection of wearer from infection (not prevention of infected from spreading)



Absence of Evidence is Not Evidence of Absence



- The review doesn't show that masks definitely do not reduce the spread of influenza/COVID
- The review showed us:
  - There have been very few high-quality studies examining the effectiveness of masks during the COVID pandemic
  - Limited to RCT (observational studies may be more appropriate here)
  - Based on the few high-quality data we have, we don't see large impacts of masking in preventing viral infections on the population level
    - Doesn't mean masks don't protect individuals
    - Danish DANMASK study: overall use in the population was low
    - No study compared N95s to no masks

NIH National Institutes of Health Clinical Center

Statement on 'Physical interventions to interrupt or reduce the spread of respiratory viruses' review

The Cochrane Review 'Physical interventions to interrupt or reduce the spread of respiratory viruses' was published in January 2023 and has been widely misinterpreted.

Karla Soares-Weiser, Editor-in-Chief of the Cochrane Library, has responded on behalf of Cochrane:

> *Many commentators have claimed that a recently-updated Cochrane Review shows that 'masks don't work', which is an inaccurate and misleading interpretation.*
>
> *It would be accurate to say that the review examined whether interventions to promote mask wearing help to slow the spread of respiratory viruses, and that the results were inconclusive. Given the limitations in the primary evidence, the review is not able to address the question of whether mask-wearing itself reduces people's risk of contracting or spreading respiratory viruses.*
>
> *The review authors are clear on the limitations in the abstract: 'The high risk of bias in the trials, variation in outcome measurement, and relatively low adherence with the interventions during the studies hampers drawing firm conclusions.' Adherence in this context refers to the number of people who actually wore the provided masks when encouraged to do so as part of the intervention. For example, in the most heavily-weighted trial of interventions to promote community mask wearing, 42.3% of people in the intervention arm wore masks compared to 13.3% of those in the control arm.*
>
> *The original Plain Language Summary for this review stated that 'We are uncertain whether wearing masks or N95/P2 respirators helps to slow the spread of respiratory viruses based on the studies we assessed.' This wording was open to misinterpretation, for which we apologize. While scientific evidence is never immune to misinterpretation, we take responsibility for not making the wording clearer from the outset. We are engaging with the review authors with the aim of updating the Plain Language Summary and abstract to make clear that the review looked at whether interventions to promote mask wearing help to slow the spread of respiratory viruses.*

THE CONVERSATION

Academic rigor, journalistic flair

Pexels/Uriel Mont

Yes, masks reduce the risk of spreading COVID, despite a review saying they don't

Published: February 6, 2023 2.04pm EST

C Raina MacIntyre
Professor of Global Biosecurity, NHMRC Principal Research Fellow, Head, Biosecurity Program, Kirby Institute, UNSW Sydney

Abrar Ahmad Chughtai
Senior lecturer, UNSW Sydney

David Fisman
Professor in the Division of Epidemiology, University of Toronto

Trish Greenhalgh
Professor of Primary Care Health Sciences, University of Oxford

---

Translational Psychiatry www.nature.com/tp

ARTICLE OPEN

Efficacy and practice of facemask use in general population: a systematic review and meta-analysis

Hui Li[1,9], Kai Yuan[1,9], Yan-Kun Sun[1,9], Yong-Bo Zheng[1], Ying-Ying Xu[2,3], Si-Zhen Su[1], Yu-Xin Zhang[1,3], Yi Zhong[1,4], Yi-Jie Wang[1,4], Shan-Shan Tian[1], Yi-Miao Gong[1,4], Teng-Teng Fan[1], Xiao Lin[1,4], Nina Gobat[5], Samuel Yeung Shan Wong[6], Emily Ying Yang Chan[7], Wei Yan[1], Si-Wei Sun[1], Mao-Sheng Ran[8], Yan-Ping Bao[2,3], Jie Shi[2], and Lin Lu[1,4]

frontiers | Frontiers in Public Health

SYSTEMATIC REVIEW
published: 27 April 2022
doi: 10.3389/fpubh.2022.874693

Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis

Derek K Chu, Elie A Akl, Stephanie Duda, Karla Solo, Sally Yaacoub, Holger J Schünemann, on behalf of the COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors*

Associations Between Wearing Masks and Respiratory Viral Infections: A Meta-Analysis and Systematic Review

Yiming Chen[1], Yuelin Wang[1], Ningbin Quan[1], Jun Yang[2,3]* and Yinyin Wu[1]*

[1] Department of Epidemiology and Health Statistics, School of Public Health, Faculty of Medicine, Hangzhou Normal University, Hangzhou, China, [2] Department of Nutrition and Toxicology, School of Public Health, Faculty of Medicine, Hangzhou Normal University, Hangzhou, China, [3] National Key Laboratory for the Diagnosis and Treatment for Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

International Journal of Nursing Studies 108 (2020) 103629

Contents lists available at ScienceDirect

International Journal of Nursing Studies

journal homepage: www.elsevier.com/ijns

A rapid systematic review of the efficacy of face masks and respirators against coronaviruses and other respiratory transmissible viruses for the community, healthcare workers and sick patients

C. Raina MacIntyre[a], Abrar Ahmad Chughtai[b,*]

Epidemiol. Infect. (2010), 138, 449–456. © Cambridge University Press 2010
doi:10.1017/S0950268809991658

REVIEW ARTICLE

Face masks to prevent transmission of influenza virus: a systematic review

B. J. COWLING[1*], Y. ZHOU[1], D. K. M. IP[1], G. M. LEUNG[1] AND A. E. AIELLO[2]

[1] School of Public Health, The University of Hong Kong, Hong Kong Special Administrative Region, China
[2] Department of Epidemiology, Center for Social Epidemiology & Population Health, School of Public Health, University of Michigan, Ann Arbor, MI, USA

A Randomized Clinical Trial of Three Options for N95 Respirators and Medical Masks in Health Workers

C. Raina MacIntyre[1], Quanyi Wang[2], Holly Seale[1], Peng Yang[2], Weixian Shi[2], Zhanhai Gao[1], Bayzid Rahman[1], Yi Zhang[2], Xiaoli Wang[2], Anthony T. Newall[1], Anita Heywood[1], and Dominic E. Dwyer[3]

[1] School of Public Health and Community Medicine, UNSW Medicine, University of New South Wales, Sydney, Australia; [2] The Beijing Center for Disease Prevention and Control, Beijing, China; and [3] Institute for Clinical Pathology and Medical Research, Westmead Hospital, Sydney, Australia

Research

JAMA | Original Investigation

N95 Respirators vs Medical Masks for Preventing Influenza Among Health Care Personnel
A Randomized Clinical Trial

Lewis J. Radonovich Jr, MD; Michael S. Simberkoff, MD; Mary T. Bessesen, MD; Alexandria C. Brown, PhD; Derek A. T. Cummings, PhD; Charlotte A. Gaydos, MD; Jenna G. Los, MLA; Amanda E. Krosche, BS; Cynthia L. Gibert, MD; Geoffrey J. Gorse, MD; Ann-Christine Nyquist, MD; Nicholas G. Reich, PhD; Maria C. Rodriguez-Barradas, MD; Connie Savor Price, MD; Trish M. Perl, MD; for the ResPECT investigators

---

?

?  ?

?