

SHORT META-ANALYSIS · Peer-reviewed · Published

Study overlap across 501 Cochrane reviews: quantifying non-independence for meta-research

A cross-sectional overlap audit of 501 Cochrane reviews (Pairwise70) for large-scale meta-research.

Mahmood Ahmad¹¹ Royal Free Hospital, London, United Kingdom · ORCID 0009-0003-7781-4478

Correspondence: mahmood.ahmad2@nhs.net

Published 16 June 2026 · Diamond open access · CC BY 4.0

Short Meta-Analysis · Synthesis

ABSTRACT

Objective. To quantify the degree of primary-study overlap across Cochrane systematic reviews and assess whether non-independence threatens the validity of methodological meta-research using these reviews.

Design. Cross-sectional bibliometric analysis.

Data source. 501 Cochrane reviews from the Pairwise70 dataset, containing 10,006 unique primary studies.

Main outcome measures. Corrected Covered Area (CCA) index, pairwise Jaccard similarity coefficients, and study-frequency distributions across all 125,250 review pairs.

Results. Overlap was minimal: CCA was 0.0001 (95% CI 0.00005–0.00018). Only 444 of 10,006 unique studies (4.4%) appeared in more than one review. Of 604 overlapping review pairs, the most overlapping pair shared 43 studies (Jaccard 0.37), while 72% of pairs shared only one study. The median number of reviews per overlapping study was 2 (IQR 2–2).

Conclusions. The Pairwise70 Cochrane review collection exhibits negligible study overlap, supporting its use as a source of largely independent meta-analyses for large-scale methodological benchmarking.

KEY WORDS study overlap; meta-research; Cochrane reviews; corrected covered area; non-independence

Negligible overlap: CCA 0.0001; only 4.4% of studies appear in more than one review.

501 Cochrane reviews · 10,006 unique studies · Pairwise70

Introduction

Large-scale meta-research — studying the methods, quality and reporting of systematic reviews themselves — increasingly relies on collections of Cochrane reviews as a benchmark corpus. Studies examining publication-bias prevalence, heterogeneity patterns and methodological quality across hundreds of reviews implicitly assume that individual reviews represent independent units of analysis. If many reviews share the same primary studies, this non-independence could inflate precision estimates and produce misleading meta-research conclusions.

The extent of primary-study overlap across large Cochrane review collections has not been systematically quantified. We addressed this gap by mapping all primary-study appearances across 501 Cochrane reviews from the Pairwise70 dataset and computing standard overlap metrics.

Methods

Study identification. Primary studies within each Cochrane review were identified using normalised first-author-year keys extracted from the Pairwise70 structured data. Keys were normalised by lowercasing, removing diacritics and standardising name formats. Studies with identical keys across different reviews were flagged as shared.

Overlap metrics. The Corrected Covered Area (CCA) is the proportion of the study-by-review matrix that contains shared entries, corrected for matrix size and ranging from 0 (no overlap) to 1 (complete overlap). Pairwise Jaccard similarity was computed for each review pair (A, B) as $Jaccard = |A \cap B| / |A \cup B|$ across all 125,250 unique review pairs. The study-frequency distribution is the number of reviews in which each primary study appears.

Results

The Pairwise70 dataset contained 10,006 unique primary studies across 501 reviews. Only 444 studies (4.4%) appeared in more than one review. The CCA was 0.0001 (95% CI 0.00005–0.00018), indicating negligible overlap. Of 125,250 possible review pairs, 604 (0.5%) had any study overlap; among these, 72% shared only one study, 18% shared two to five and 10% shared six or more, with a maximum of 43 shared studies (Jaccard 0.37). Among the 444 overlapping studies, the median number of reviews per study was 2 (IQR 2–2), the maximum was 7, and 89% appeared in exactly two reviews.

Table 1. Primary-study overlap metrics across 501 Cochrane reviews (Pairwise70).

Overlap metric	Value	Interpretation
Unique primary studies	10,006	Across 501 reviews
Studies in >1 review	444 (4.4%)	Remainder appear once
Corrected Covered Area (CCA)	0.0001	95% CI 0.00005–0.00018; negligible
Review pairs with any overlap	604 / 125,250	0.5% of all pairs
Pairs sharing only 1 study	72%	18% share 2–5; 10% share ≥6
Median reviews per shared study	2 (IQR 2–2)	Max 7; 89% in exactly 2
Most-overlapping pair	43 studies	Jaccard = 0.37

Study Overlap Across 501 Cochrane Reviews

Quantifying non-independence in the Pairwise70 corpus

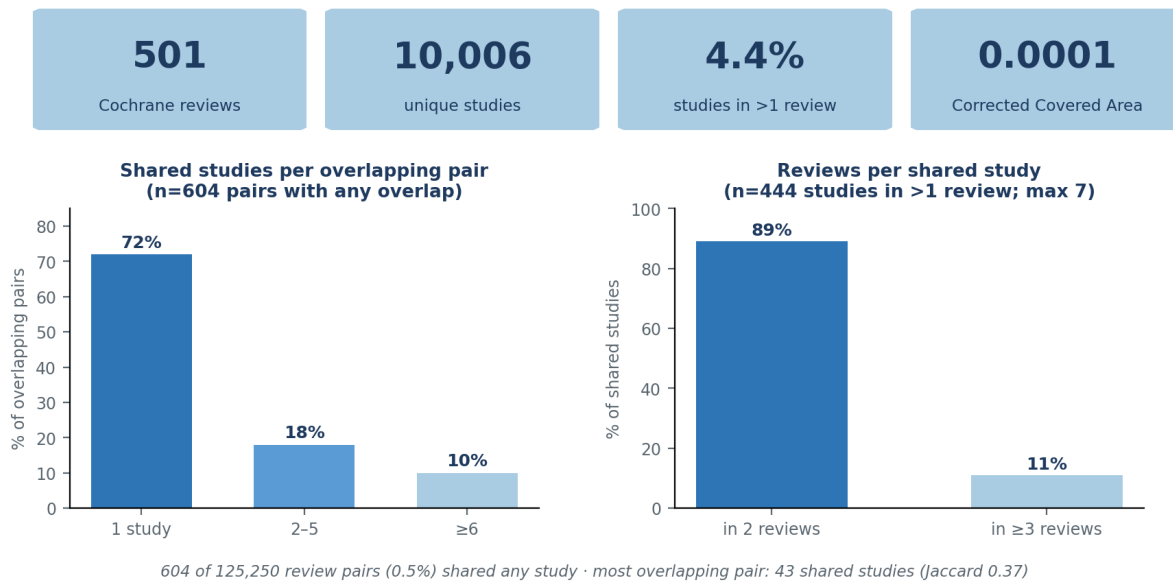


Figure 1. Study overlap, Pairwise70 corpus. Summary of study overlap across the Pairwise70 Cochrane review corpus: headline counts, shared studies per overlapping pair, and reviews per shared study.

Discussion

The negligible overlap (CCA = 0.0001) supports the Pairwise70 collection as a source of largely independent meta-analyses. The 4.4% shared-study rate is concentrated in closely related topic areas (for example, multiple reviews on similar interventions in the same clinical domain) rather than reflecting systematic network-like non-independence; study sharing was predominantly pairwise, with 89% of shared studies appearing in exactly two reviews.

The main limitation is that first-author-year matching may produce false positives when different studies share identical author surnames and publication years. This would overestimate overlap, making the finding of minimal overlap conservative.

Author contributions (CRediT)

Mahmood Ahmad: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization.

Competing interests and editorial handling

Mahmood Ahmad serves on the editorial board of *Synthesis* (Editor-in-Chief). To preserve independence, this manuscript was handled entirely by an independent editor with no involvement of the author in any editorial decision, reviewer selection or acceptance determination. The author declares no other competing interests.

Funding

No specific funding was received for this work.

Data and code availability

Analysis code is available at <https://github.com/mahmood726-cyber/overlap-detector> (MIT licence) and is archived on Zenodo (DOI: 10.5281/zenodo.20763980, <https://doi.org/10.5281/zenodo.20763980>).

References

1. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med.* 2016;13(5):e1002028. doi:10.1371/journal.pmed.1002028. PMID: 27218655.
2. Ioannidis JPA. Why most published research findings are false. *PLoS Med.* 2005;2(8):e124. doi:10.1371/journal.pmed.0020124.
3. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21(11):1539–1558. doi:10.1002/sim.1186.
4. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews. *BMJ.* 2017;358:j4008. doi:10.1136/bmj.j4008. PMID: 28935701.
5. Pieper D, Antoine SL, Mathes T, et al. Systematic review finds overlapping reviews were not mentioned in every other overview. *J Clin Epidemiol.* 2014;67(4):368–375. doi:10.1016/j.jclinepi.2013.11.007.

HOW TO CITE

Ahmad M. Study overlap across 501 Cochrane reviews: quantifying non-independence for meta-research. *Synthesis* 2026;2(4). Available at <https://synthesis-medicine.org/index.php/journal/article/view/overlap-cochrane-nonindependence>. Diamond open access under CC BY 4.0. DOI: not assigned.