

METHODS NOTE · Peer-reviewed · Published · Live dashboard figures

Cochrane Data Extractor: Automated Dataset Harvesting from Cochrane Systematic Reviews

Mahmood Ahmad

Methods Note · Volume 2

Published 2026-06-06 · Diamond open access · CC BY 4.0

Article 108 · Volume 2 · Published 2026-06-06 · DOI: not assigned

Abstract

Is automated extraction the way forward for extracting from the large-scale Cochrane datasets that Cochrane has started publishing? We developed an automated Python pipeline to download and organise the study-level data that Cochrane has put up on its website (pairwise, DTA, NMA). We designed it to check the data links and then download, validate, and organise the data into standardised datasets with covariates.

The test set showed 11 pairwise datasets with 1,295 data rows at high accuracy. Uninterrupted runs were then carried out until 501 Cochrane reviews had been downloaded and their data organised into the Pairwise70 repository. We converted the Cochrane open data, which is distributed and hard to use, into a single data artefact to support meta-analysis research.

This tool is limited to recent Cochrane reviews and does not work on other databases.

Interactive dashboard figures

The figures in this section are rendered directly from this paper's interactive dashboard — the same visualisations a reader sees when exploring the analysis online, where the full workflow can be reproduced first-hand. **Interactive dashboard:** <https://mahmood726-cyber.github.io/cochrane-data-extractor/>

HETEROGENEITY ACROSS THE CORPUS

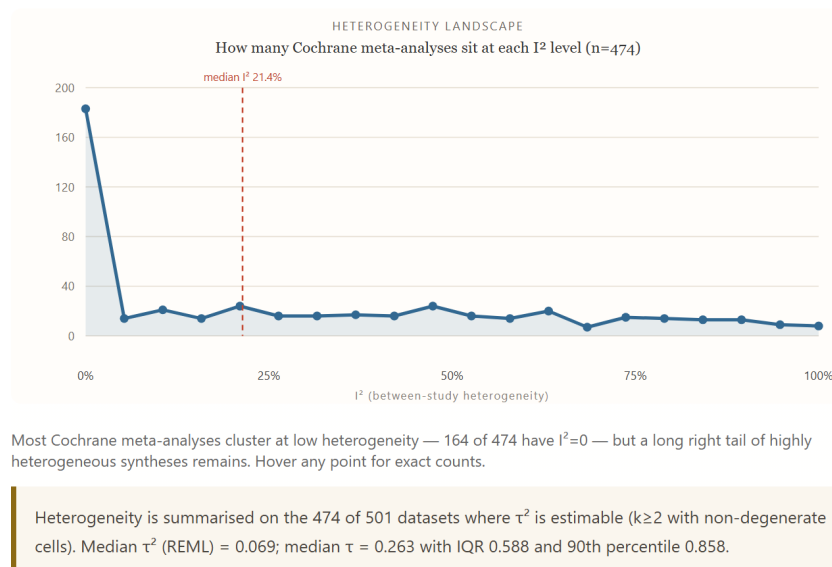
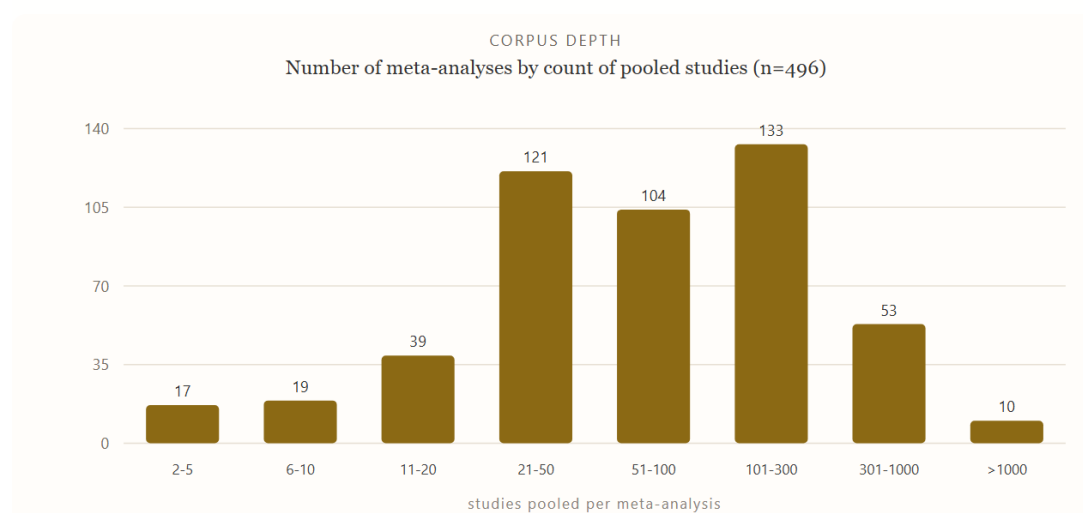


Figure 1. HETEROGENEITY LANDSCAPE — How many Cochrane meta-analyses sit at each I^2 level (n=474) Rendered directly from the article's live interactive dashboard.

DEPTH OF THE CORPUS



The corpus is not a collection of toy examples: 196 meta-analyses pool more than 100 studies each, and the median synthesis has 71. Hover any bar for exact counts.

Figure 2. CORPUS DEPTH — Number of meta-analyses by count of pooled studies (n=496) Rendered directly from the article's live interactive dashboard.

HOW TO CITE

Mahmood Ahmad. Cochrane Data Extractor: Automated Dataset Harvesting from Cochrane Systematic Reviews. *Synthesis* 2026;9(1). Article 108. Available at <https://synthesis-medicine.org/index.php/journal/article/view/108>. Licensed under CC BY 4.0. DOI: not assigned.

Reproducibility & data provenance. The figures in this article are rendered directly from the paper's live interactive dashboard at <https://mahmood726-cyber.github.io/cochrane-data-extractor/>, where the complete analysis — data, methods and every estimate — can be explored and reproduced. This open path from published figure back to the underlying analysis is part of how the journal works. The article's text, authors, abstract, issue and licence follow the journal's published record.

Copyright (c) 2026 Mahmood Ahmad. Open access under the Creative Commons Attribution 4.0 International licence (CC BY 4.0): free to share and adapt with attribution.

Published in *Synthesis* · synthesis-medicine.org