

Title

Meta-Analytical Approaches to Ordinal Outcome Data in Clinical Interventional Studies: A Scoping Review with Reproducible Research

Ali Mulhem

Author information

Ali Mulhem^{1,2} – Corresponding author

Email: ali.mulhem@kellogg.ox.ac.uk

ORCID: 0000-0002-2614-610X

Tel: +49 1782936939 , +49 15142261401

Affiliations:

¹Westfälische Wilhelms-Universität Münster, Münster, Germany

²Department of Continuing Education, DPhil program in Evidence-Based Health Care, & Kellogg College, University of Oxford

Abstract

Background: Conducting meta-analyses on ordinal outcome data is more complex than on binary or continuous data. This study aims to summarise the current biomedical literature on meta-analytical methods for ordinal outcomes and attempts to reproduce the results of previous studies.

Methods: A systematic search was conducted in three databases, MEDLINE, EMBASE, and PsycINFO, from inception to 05/05/2024. Forward and backward citation searches were also performed. The screening was conducted in two phases using Covidence software. Studies were included if they reported or compared methods for meta-analysis of ordinal outcomes in clinical interventional studies. Relevant studies were summarised and discussed. If sufficient data for methods comparison were available, either in the retrieved reports or after contacting the authors, an attempt at reproducible research was made.

Results: 333 records were screened, yielding four methodological studies that met the inclusion criteria. These studies addressed meta-analytical methods for ordinal scales ranging from 5 to 20 ordered categories. The three primary approaches identified were (1) ordinal models (proportional odds and generalised odds), (2) binary models (dichotomisation of ordinal scales), and (3) continuous models (treating ordinal scales as continuous variables). None of the included studies provided a comprehensive comparison of all three approaches. Two studies compared different proportional odds models; one compared binary with proportional odds, but the full results were not published, and one compared continuous and generalised odds models using simulated data with one scenario. The latter study allowed for reproducibility, but our analysis produced different results, and attempts to clarify the discrepancy with the authors were unsuccessful.

Conclusions: A significant knowledge gap exists regarding the optimal meta-analytical method for ordinal outcomes in clinical interventional studies. Further methodological research is required to establish a robust evidence base for choosing the most appropriate approach.

Background

Meta-analysis is a vital research method used to estimate the effect sizes of clinical interventions.^{1,2} The statistical methods employed in meta-analyses vary depending on the type of outcome data.³ For binary and continuous data, robust and well-documented techniques are available.^{1,3} However, in many biological and medical research fields, response variables are frequently measured as ordinal data,⁴ typically scales or scores that categorise variables into ordered groups.⁴ Although ordinal scales are widely used in clinical research, meta-analytical methods for these data types are less straightforward than binary or continuous data.²

This review summarises the current biomedical literature on meta-analytical methods for ordinal outcomes and explores the reproducibility of the reported results.

Methods

Scoping Review

Protocol and Registration

This scoping review was conducted as part of a DPhil thesis, registered with the Open Science Framework under <https://doi.org/10.17605/OSF.IO/WJY5D>.

Eligibility Criteria

Eligible studies for this review compared or reported meta-analytical methods for ordinal outcome data in clinical interventional studies. We restricted the inclusion to only methodological studies; we used the term methodological study to refer to any study that reports on the analysis of research-related reports (i.e. research-on-research).⁵ Thus clinical studies that used different meta-analytical methods to conduct the statistical analysis, not with the aim of methods comparison, were excluded. Studies focused on meta-analysis of non-interventional studies, such as diagnostic accuracy studies, were also excluded. There were no restrictions on publication language or date.

Information Sources

We restricted the search to three databases—MEDLINE, EMBASE, and PsycINFO—which were searched through the Ovid interface from inception to May 5, 2024. Additionally, forward and backward citation searches were performed. The search strategy comprised keywords and MeSH terms as follows: ((meta-analysis OR "exp Meta-Analysis") OR ("systematic review" OR "exp Systematic Review")) AND ("ordinal scale" OR "ordinal score" OR "ordinal data" OR "ordinal outcome").

Selection of Studies

After removing duplicates, one reviewer (AM) conducted a two-stage screening process through Covidence software,⁶ involving title/abstract screening followed by a full-text review.

Data Extraction and Summary

The following data items were extracted by AM: study ID (first author and year of publication), sample size (number of studies included in the methodological analysis), the ordinal outcome scale, the type of data reported for the ordinal scale (raw data vs. summary statistics), methods of meta-analysis, metrics used to calculate the effect size, and, where available, the results of comparisons between different meta-analytical methods concerning the direction of the effect size.

Reproducible Research

Where sufficient data were available in the included studies or after contacting the authors, a reproducible analysis was conducted to replicate the original results. Statistical procedures were performed using RStudio (version 2022.12.0+353)⁷ and STATA 18 BE Basic Edition.⁸

Results

Scoping Review

Flow Diagram

Three hundred thirty-three records were identified: 245 from the database search and 88 from citation searches. Four studies were included in the review following a two-stage screening process (title/abstract and full-text). Out of the 108 studies assessed for eligibility, 104 were excluded. The reasons for exclusion were studies focusing on statistical methods for ordinal data in primary research (i.e., not within a meta-analysis context; $n = 100$), studies that reported meta-analytical methods for ordinal data in diagnostic accuracy studies (i.e., not within an interventional context; $n = 3$), and studies focusing on statistical methods to compute correlations between variables (i.e., not effect size for interventions; $n = 1$).

Figure 1 illustrates this review's PRISMA 2020 flow diagram and the reasons for exclusion.⁹

Among the four included studies, one was available only as a conference paper, obtained after contacting the first and last authors. Despite its limited information, it was included because it contained relevant data.¹⁰

Characteristics of included methodological studies

All ordinal scales meta-analysed in the included studies had between 5 and 10 categories, with increments of either one or half points. So, the scales used in these studies had categories of 5 and 20. The number of meta-analysed studies (as samples) within each methodological study was relatively small, with counts of 5, 8, or 13. The exception was the conference paper, which analysed 216 studies across 24 meta-analyses but did not specify the number of studies in each analysis.¹⁰

Table 1 provides a summary of the characteristics of the included studies.

Methods to meta-analyse ordinal data

Three primary approaches for meta-analysis of ordinal data were identified. The first is the **continuous approach**, where the ordinal scale is treated as a continuous variable to calculate the effect size, using the standardised mean difference (SMD) as a metric. The second is the **binary approach**, where the ordinal scale is dichotomised at a specified cut-point to generate a binary variable, with the effect size

calculated using the logarithmic odds ratio as a metric. The third is the **ordinal approach**, where the ordinal data is analysed in its original form. Within the ordinal approach, either a proportional odds model (frequentist or Bayesian) or a generalised odds model is used to calculate the effect size, with the logarithmic proportional odds ratio (logPOR) or the logarithmic generalised odds ratio (logGeOR) as the respective metrics.^{11,12,13}

For the binary and continuous approaches, the included studies in the meta-analysis reported the ordinal outcomes either as summary statistics (such as means and standard deviations or counts/proportions) or as individual patient data (IPD), which provides the distribution of subject responses across ordered categories for each treatment group. Both reporting styles were suitable for these approaches.

In contrast, the ordinal approach requires individual patient data to perform the analysis. However, the generalised odds model can also be conducted using median and interquartile range (IQR) as summary statistics.

Table 1 summarises the methods with their metrics and the reported data type.

Comparing the meta-analysis methods of ordinal data

The comparison of meta-analysis methods for ordinal data revealed that the direction of the effect size remains consistent regardless of the approach used.

The earliest methodological study in this area was conducted by Whitehead and Jones,¹¹ who utilised a proportional odds model to compare two techniques for calculating effect sizes: maximum likelihood estimates and Fisher's information. This study is summarised in Table 1. A later study by Whitehead et al.¹² examined various models within the proportional odds framework, including comparing Bayesian and frequentist approaches.

The third study by Ashma Krishan and colleagues compared the binary approach with the proportional odds approach.¹⁰ However, this study was published as a conference paper with limited details. Despite contacting the first and last authors, no full-text paper was available for further analysis.

The final study by Cumming, Churilov, and Sena reported a comparison between the continuous method and the ordinal generalised approach (generalised odds model).¹³ This comparison was based on mock data generated from an ordinal outcome scale, with only one iteration of a meta-analysis (as described in Table 1).

Reproducible research analysis

One of the included studies,¹³ provided raw data, allowing us to conduct reproducible research. In this study, the authors used a neurological outcome scale, the Expanded Disability Status Scale (EDSS), to illustrate an ordinal approach (the rank-based generalised odds model) for meta-analysis of ordinal data and compared it with the continuous approach. The EDSS is used for multiple sclerosis and ranges from 0 (normal neurological state) to 10 (death), with half-point increments, except for the transition from 0 to 1, creating a total of 20 ordered categories.¹⁴

To generate the data for their analysis, the authors used the EDSS distribution from a large observational study¹⁵ to create a control group (sample 0) and five intervention groups (samples 1 to 5), each with 100 observations. They conducted a meta-analysis of five simulated studies, comparing each intervention group with the control group (e.g., sample 0 vs. sample 1, sample 0 vs. sample 2, etc.). The analysis was conducted using two approaches: calculating the standardised mean difference (SMD), representing the continuous approach, and calculating the log generalised odds ratio (logGeOR), representing the ordinal approach.

Their results showed statistically significant effect sizes in the same direction for both approaches (SMD = -0.17, 95% CI: -0.29 to -0.04; logGeOR = -0.21, 95% CI: -0.35 to -0.07). The heterogeneity was similar between the approaches ($I^2 = 0\%$ for both), as depicted in Figures 5 and 6 of the published report.¹³

We attempted to reproduce these results using the raw data provided in the supplementary material, conducting meta-analyses with both SMD and logGeOR in two software platforms: STATA and R. Our results differed from the original study, showing statistically insignificant effect sizes in both STATA

and R (SMD in STATA = -0.03, 95% CI: -0.16 to 0.09; logGeOR in STATA = -0.04, 95% CI: -0.19 to 0.10; SMD in R = -0.03, 95% CI: -0.15 to 0.10; logGeOR in R = -0.04, 95% CI: -0.19 to 0.10). The heterogeneity remained consistent across both platforms and approaches ($I^2 = 0\%$). Figures 2 (for forest plots in STATA) and 3 (for forest plots in R) illustrate these results.

To understand the discrepancies, we examined the raw data provided by the authors in the supplementary material, comparing the summary statistics with those reported in the original study. We found inconsistencies in three of the samples (see Figure 4).

Despite these discrepancies, both analyses agreed on the direction of the effect size, and heterogeneity was similar between the two approaches. Further clarification from the original study's authors would be helpful, but our attempts to contact them were unsuccessful.

Discussion

Summarising the Literature

Despite more than thirty years since the foundation of the Cochrane Collaboration, meta-analysing ordinal data methods remain less well-established than binary and continuous data.^{1,2} Our scoping review identified four methodological studies that addressed the meta-analysis of ordinal outcome scales with 5 to 20 ordered categories.

In 1994, Anne Whitehead and Nicola M. B. Jones proposed the stratified proportional odds model, combining multiple studies with differing classifications of ordinal outcomes. They used data from 13 trials examining the prevention of gastrointestinal damage with misoprostol, where the outcome scale ranged from 1 to 5, with one-point increments. This scale varied across studies, with some trials using only 2, 3, or 4 categories. They compared two ways of estimating logPOR and found no change in the direction of the effect size.¹¹

In a subsequent methodological study, Whitehead and colleagues introduced different proportional models to meta-analyse ordinal outcomes, including frequentist and Bayesian techniques.¹² They used individual patient data (IPD) for their analyses, allowing them to count the number of patients in each ordered category for each treatment group. The study included data from a subset of the same 13 trials from 1994 and another dataset from five trials evaluating the effect of tacrine in Alzheimer's disease patients. The outcome scale in the second dataset was the Clinical Global Impression of Change (CGIC), a seven-point scale with one-point increments.¹⁷ This comparison also showed no change in the effect size between the two techniques of the proportional approach.

In 2013, Ashma Krishan and colleagues compared ordinal methods (using a proportional odds model) with binary methods (dichotomisation of the ordinal scale) using data from stroke trials in Cochrane systematic reviews.¹⁰ The outcome scale was the modified Rankin Scale (mRS), which has seven ordered categories from 0 to 6 with one-point increments.¹⁷ This study was available only as a conference paper, limiting its results' depth, which indicates no change in the effect-size direction.

Finally, Cumming, Churilov, and Sena used the Expanded Disability Status Scale (EDSS) for multiple sclerosis, where EDSS ranges from 0 to 10 with half-point increments, to present the rank-based generalised odds model as an ordinal approach and compared it with the continuous approach. They also found no change in the direction of effect size between the two approaches.^{13,14}

Critical shortcomings in the existing literature

The existing methodological studies on the meta-analysis of ordinal outcomes have several shortcomings, limiting their capacity to provide robust empirical evidence for preferring one approach over another. The two studies by Anne Whitehead focused exclusively on different techniques and models within the proportional method, offering limited scope for broader comparisons. While the proportional odds method might be statistically appealing, in spite of its underlying assumptions,¹⁸ it requires effective implementation of raw data.^{12,18,19} This reliance on individual patient data (IPD) makes it challenging to apply, as primary studies often do not provide sufficient data and there is

currently no trend of complete sharing of raw data from clinical trials,²⁰ which leads to the exclusion of studies from meta-analysis.

On the other hand, the binary and continuous approaches can be performed using either IPD or summary statistics, offering more flexibility.² However, the binary approach entails dichotomization, which in turn leads to loss of information,²¹ and the continuous approach entails the assumption of normality of distribution, which does not always hold.^{21,22} The study by Ashma Krishan, which compared the proportional odds approach with the binary method, was published as a conference paper with incomplete reporting, raising concerns about the validity of its findings. And the study by Cumming and others attempted a comparison between the continuous approach and the generalised ordinal approach (generalised odds model). However, this analysis was based on only one simulated meta-analysis with no iterations. Moreover, our attempts to reproduce their results yielded different outcomes, casting doubt on the reliability of their conclusions.

Strengths and limitations

In this review, we aimed to be comprehensive by conducting both electronic and citation searches, including screening the references of relevant studies. However, one limitation is that the screening and data extraction were performed solely by the main author due to the unavailability of other reviewers. Despite this, we believe this limitation did not significantly influence our findings, as the total number of citations screened was manageable for one reviewer.

A key strength of our review is our attempt to reproduce the results of previous studies, which is crucial for assessing the credibility of the evidence. However, we faced a limitation in this area as we were unable to contact the authors of the reproduced study to clarify the discrepancies in our results.

Conclusions

Given the current literature, there's a significant knowledge gap regarding the optimal method for meta-analysing ordinal outcome data, particularly for scales with 5 to 20 categories. The existing studies do

not provide robust empirical evidence to support one approach over another, with many suffering from methodological limitations and a lack of comprehensive comparisons.

A new methodological study is needed to address the knowledge gap. This study should overcome the shortcomings of previous research and systematically compare the results of the meta-analysis for ordinal scales in clinical interventional trials using three different approaches: continuous, binary, and ordinal (proportional and generalised). The study should be based on clinical data and validated through rigorous simulations with sufficient iterations to ensure reliability and reproducibility.

References

1. Deeks JJ, Higgins JPT, Altman DG, Higgins J, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6. 2022.
2. Sutton AJ, Higgins JPT. Recent developments in meta-analysis. *Stat Med* [Internet]. 2008;27(5):625–50. Available from: <http://dx.doi.org/10.1002/sim.2934>
3. Borenstein M, Hedges LV, Higgins J, Rothstein HR. Introduction to meta-analysis. 2nd ed. Nashville, TN: John Wiley & Sons; 2021.
4. Agresti A. Analysis of ordinal categorical data. Wiley; 2014.
5. Mbuagbaw L, Lawson DO, Puljak L, Allison DB, Thabane L. A tutorial on methodological studies: the what, when, how and why. *BMC Med Res Methodol* [Internet]. 2020;20(1):226. Available from: <http://dx.doi.org/10.1186/s12874-020-01107-7>
6. Babineau J. Product Review: Covidence (Systematic Review Software). *J Can Health Libr Assoc* [Internet]. 2014;35(2):68. Available from: <http://dx.doi.org/10.5596/c14-016>
7. The R project for statistical computing [Internet]. R-project.org. [cited 2024 Jun 3]. Available from: <https://www.R-project.org>
8. Stata Statistical Software: Release 18.

9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev* [Internet]. 2021;10(1). Available from: <http://dx.doi.org/10.1186/s13643-021-01626-4>
10. Krishan A, Weir J, Murray G, Thomas B, Sandercock P, Lewis S. Practical methods for ordinal data meta-analysis in stroke. *Trials* [Internet]. 2013;14(S1). Available from: <http://dx.doi.org/10.1186/1745-6215-14-s1-p120>
11. Whitehead A, Jones NMB. A meta-analysis of clinical trials involving different classifications of response into ordered categories. *Stat Med* [Internet]. 1994;13(23–24):2503–15. Available from: <http://dx.doi.org/10.1002/sim.4780132313>
12. Whitehead A, Omar RZ, Higgins JPT, Savaluny E, Turner RM, Thompson SG. Meta-analysis of ordinal outcomes using individual patient data. *Stat Med* [Internet]. 2001;20(15):2243–60. Available from: <http://dx.doi.org/10.1002/sim.919>
13. Cumming TB, Churilov L, Sena ES. The missing medians: Exclusion of ordinal data from meta-analyses. *PLoS One* [Internet]. 2015;10(12):e0145580. Available from: <http://dx.doi.org/10.1371/journal.pone.0145580>
14. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology* [Internet]. 1983;33(11):1444–1444. Available from: <http://dx.doi.org/10.1212/wnl.33.11.1444>
15. Hohol MJ, Orav EJ, Weiner HL. Disease steps in multiple sclerosis: a simple approach to evaluate disease progression. *Neurology* [Internet]. 1995;45(2):251–5. Available from: <http://dx.doi.org/10.1212/wnl.45.2.251>
16. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* [Internet]. 1988;19(5):604–7. Available from: <http://dx.doi.org/10.1161/01.str.19.5.604>
17. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)*. 2007;4(7):28–37.
18. McNulty K. Proportional odds logistic regression for ordered category outcomes. In: *Handbook of Regression Modeling in People Analytics*. Boca Raton: Chapman and Hall/CRC; 2021. p. 143–62.

19. Morris P, Wang C, O'Connor A. Network meta-analysis for an ordinal outcome when outcome categorization varies across trials. *Syst Rev* [Internet]. 2024;13(1):128. Available from: <http://dx.doi.org/10.1186/s13643-024-02537-w>
20. Vickers AJ. Sharing raw data from clinical trials: what progress since we first asked “Whose data set is it anyway?” *Trials* [Internet]. 2016;17(1):227. Available from: <http://dx.doi.org/10.1186/s13063-016-1369-2>
21. Forrest M, Andersen B. Ordinal scale and statistics in medical research. *BMJ* [Internet]. 1986;292(6519):537–8. Available from: <http://dx.doi.org/10.1136/bmj.292.6519.537>
22. Verhulst B, Neale MC. Best practices for binary and ordinal data analyses. *Behav Genet* [Internet]. 2021;51(3):204–14. Available from: <http://dx.doi.org/10.1007/s10519-020-10031-x>

Figure 1. PRISMA flow diagram of the Scoping review

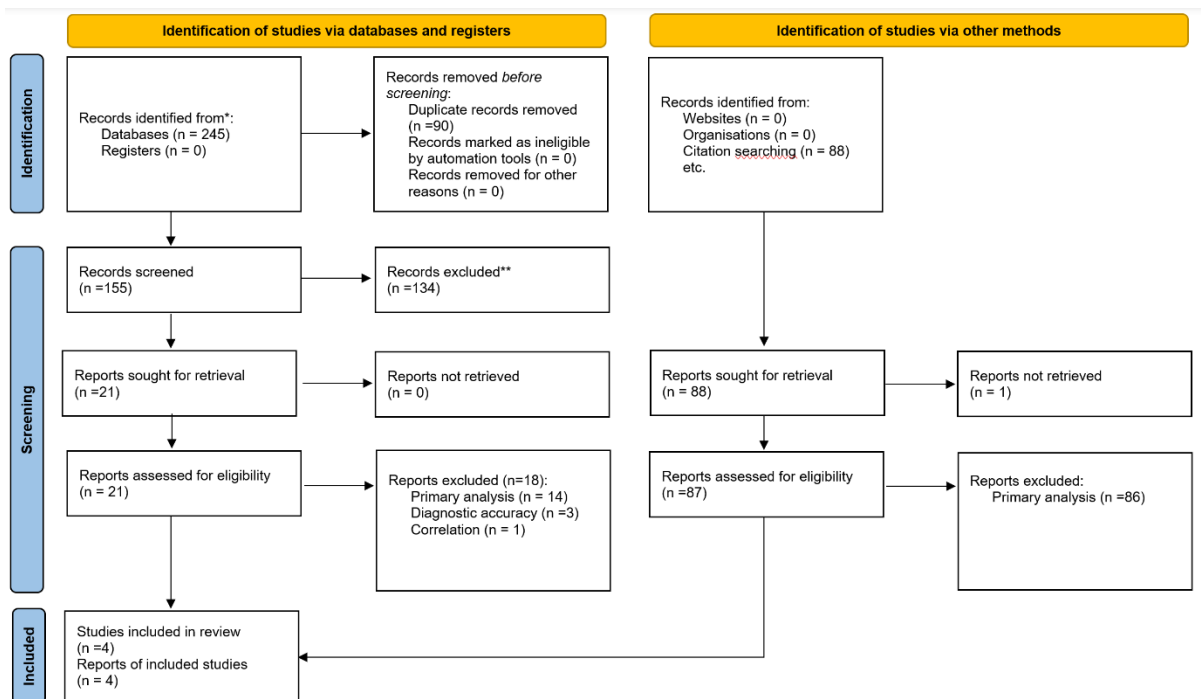


Figure 2. Forest plots for our re-run of mocked-data in Cumming et al. 2015 using STATA, left using ordinal generalised odds model (with log GeOR as effect-size estimation) and right using continuous approach (with standardised mean difference as effect-size estimation)

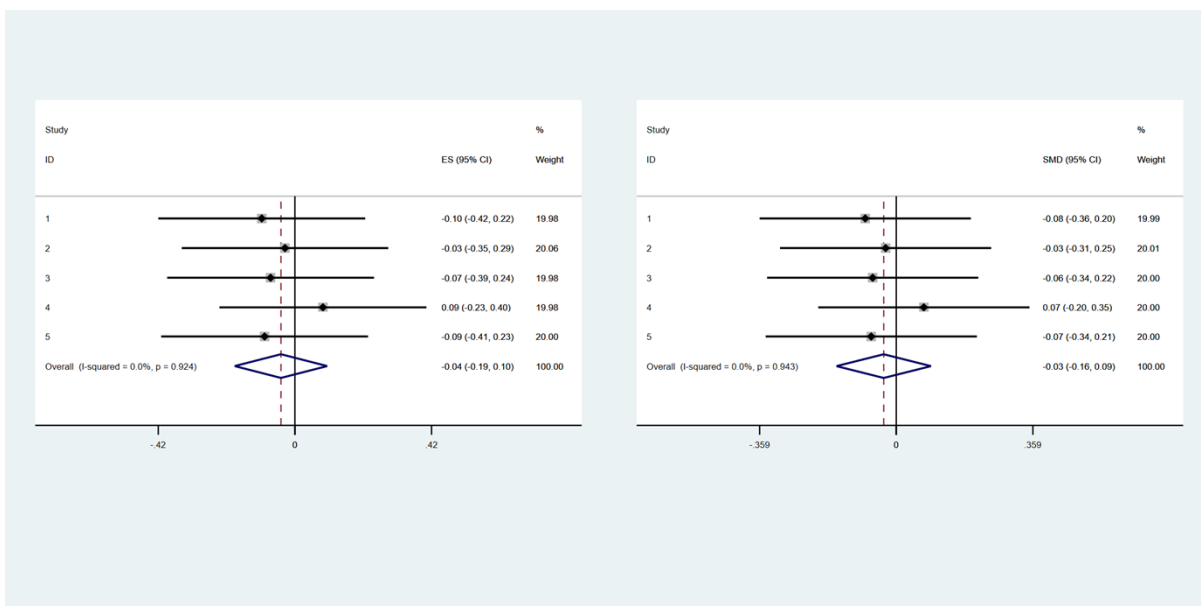


Figure 3. Forest plots for our re-run of mocked-data in Cumming et al. 2015 using R, left using ordinal generalised odds model (with log GeOR as effect-size estimation) and right using continuous approach (with standardised mean difference as effect-size estimation)

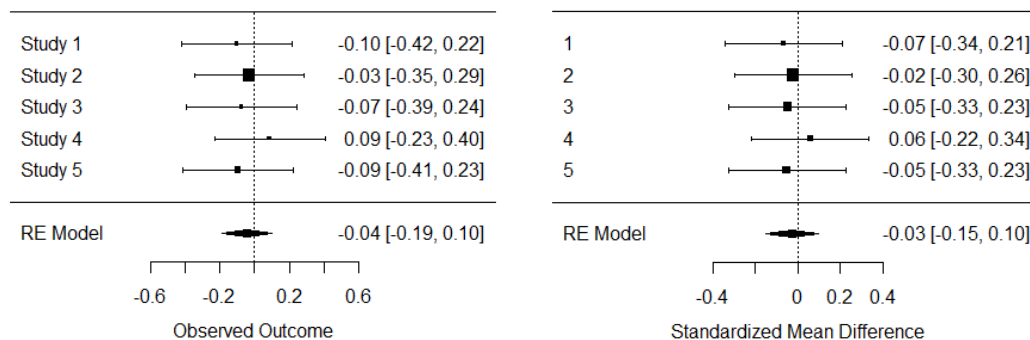


Figure 4. Comparison of summary statistics of mocked data provided by Cumming et al. 2015, the upper table is from their report, and the lower table is through our calculations (the discrepancies are in red).

Sample	N	Median (IQR)	Mann-Whitney U (vs sample 0)
0	100	6.0 (2.5, 6.5)	
1	100	4.25 (2.0, 6.0)	4155, z = -2.08 (p = .038)
2	100	6.0 (2.5, 6.5)	4782, z = -0.54 (p = .592)
3	100	4.5 (1.625, 6.375)	4186, z = -2.00 (p = .045)
4	100	5.25 (2.125, 6.5)	4490, z = -1.26 (p = .209)
5	100	5.75 (2.5, 6.5)	4764, z = -0.58 (p = .562)

doi:10.1371/journal.pone.0145580.t001

Sample	N	Median	25th	75th	Their	Mean
0	100	6	2,5	6,5	Correct	4,84
1	100	6	2,5	6,5	False	4,65
2	100	6	2,5	6,5	Correct	4,775
3	100	6	2,5	6,5	False	4,695
4	100	6	3	6,5	False	5,01
5	100	5,75	2,5	6,5	Correct	4,685

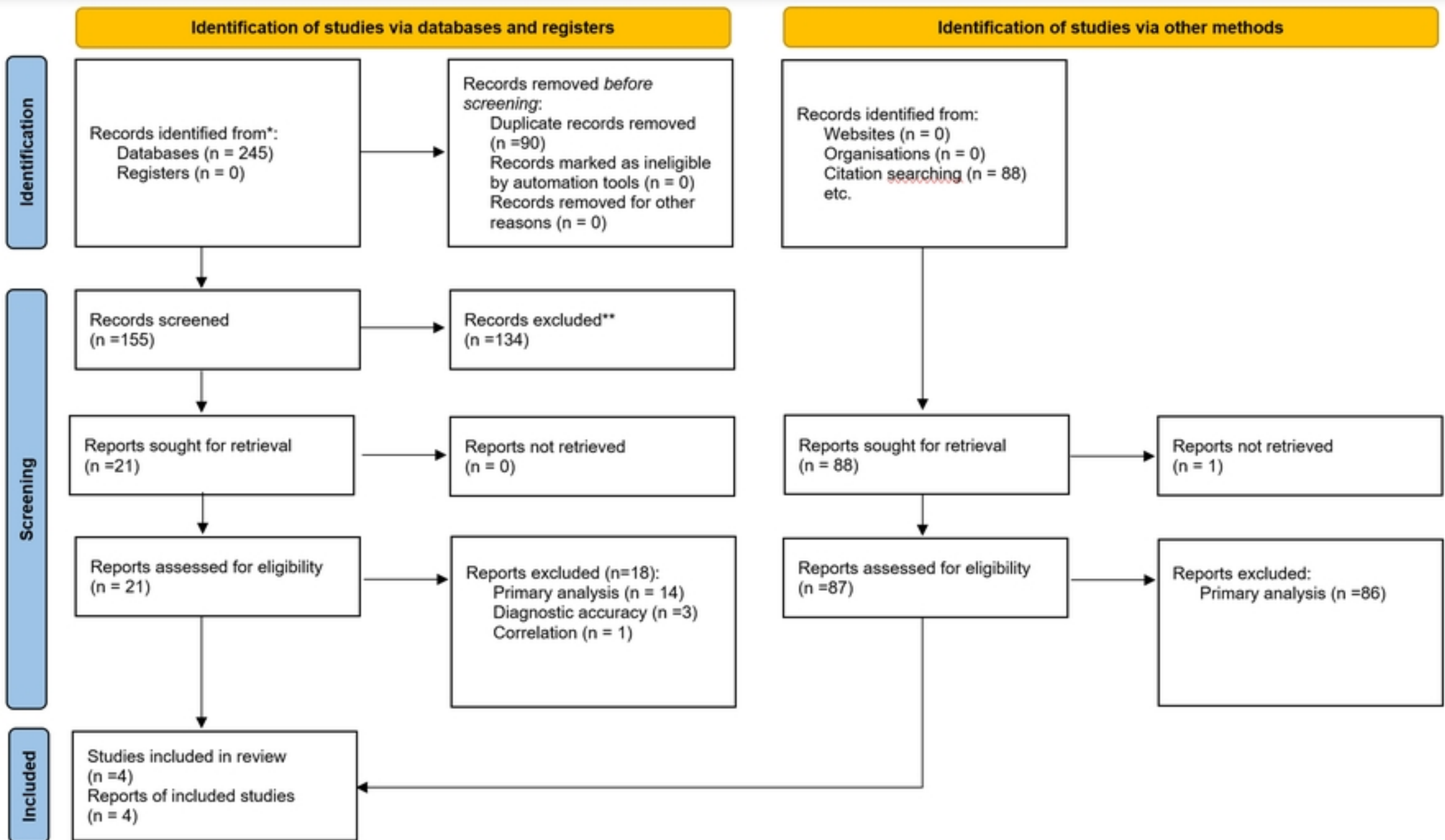


Figure 1