

RESEARCH ARTICLE

The Missing Medians: Exclusion of Ordinal Data from Meta-Analyses

Toby B. Cumming¹, Leonid Churilov¹, Emily S. Sena^{2*}

1 Florey Institute of Neuroscience and Mental Health, University of Melbourne, Melbourne, Australia, **2** Centre for Clinical Brain Sciences, School of Clinical Sciences, University of Edinburgh, Edinburgh, United Kingdom

* emily.sena@ed.ac.uk



 OPEN ACCESS

Citation: Cumming TB, Churilov L, Sena ES (2015) The Missing Medians: Exclusion of Ordinal Data from Meta-Analyses. PLoS ONE 10(12): e0145580. doi:10.1371/journal.pone.0145580

Editor: Jakob Pietschnig, Universitat Wien, AUSTRIA

Received: August 11, 2015

Accepted: December 4, 2015

Published: December 23, 2015

Copyright: © 2015 Cumming et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: TC and LC received no specific funding for this work. ES was supported by the UK National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) infrastructure award: ivSyRMAF—the CAMARADES—NC3Rs in vivo systematic review and meta-analysis facility, grant number NC/L000970/1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Background

Meta-analyses are considered the gold standard of evidence-based health care, and are used to guide clinical decisions and health policy. A major limitation of current meta-analysis techniques is their inability to pool ordinal data. Our objectives were to determine the extent of this problem in the context of neurological rating scales and to provide a solution.

Methods

Using an existing database of clinical trials of oral neuroprotective therapies, we identified the 6 most commonly used clinical rating scales and recorded how data from these scales were reported and analysed. We then identified systematic reviews of studies that used these scales (via the Cochrane database) and recorded the meta-analytic techniques used. Finally, we identified a statistical technique for calculating a common language effect size measure for ordinal data.

Results

We identified 103 studies, with 128 instances of the 6 clinical scales being reported. The majority—80%—reported means alone for central tendency, with only 13% reporting medians. In analysis, 40% of studies used parametric statistics alone, 34% of studies employed non-parametric analysis, and 26% did not include or specify analysis. Of the 60 systematic reviews identified that included meta-analysis, 88% used mean difference and 22% employed difference in proportions; none included rank-based analysis. We propose the use of a rank-based generalised odds ratio (WMW GenOR) as an assumption-free effect size measure that is easy to compute and can be readily combined in meta-analysis.

Conclusion

There is wide scope for improvement in the reporting and analysis of ordinal data in the literature. We hope that adoption of the WMW GenOR will have the dual effect of improving the

reporting of data in individual studies while also increasing the inclusivity (and therefore validity) of meta-analyses.

Introduction

Combining results from multiple studies to more powerfully estimate effect size can be very informative. Meta-analysis, however, has several well documented weaknesses: bias in the original studies will flow through to the pooled data, and non-publication of negative studies (the ‘file drawer’ problem [1]) can lead to over-estimation of effect. Less well documented is the omission of ordinal data from meta-analyses (the ‘missing medians’ problem). In reporting study-specific summary statistics, many authors present means and standard deviations on the assumption that their data are continuous and normally distributed. Deriving effect sizes from such studies and pooling them in meta-analysis is straightforward [2]. In many cases, however, outcome measures are ordinal rather than continuous; a scale’s categories have a natural order, but it cannot be assumed that differences between the categories are equivalent. This is particularly common in clinical research, where scales are designed to evaluate impairment and be clinically meaningful. One example is in the stroke literature, where the modified Rankin Scale is the primary outcome of choice in the vast majority of trials. It is a measure of functional disability, and has a 7-point ordinal scale ranging from 0 (no symptoms) to 6 (dead). By generating disability weights from WHO Global Burden of Disease data, Hong & Saver [3] demonstrated empirically that the points on the scale are not equally spaced. Yet prominent recent stroke trials have summarised ordinal modified Rankin Scale data using means [4]. Additionally, many clinical scales suffer from ceiling or floor effects, yielding data that are not normally distributed. Interpreting means and standard deviations in these conditions is problematic; medians and inter-quartile ranges are statistically more valid.

These reporting considerations have important implications for meta-analysis. Where ordinal data are reported appropriately in individual studies, they are often excluded from meta-analysis due to the difficulty in pooling them. Alternatively, where study authors report means and standard deviations, often inappropriately, these data can be included in meta-analysis but the validity of the pooled results is questionable. Meta-analytical results are heavily influenced by treatment of outliers and by parametric versus non-parametric estimation [5]. The Cochrane collaboration acknowledge the problem with meta-analysis of ordinal or non-parametric data in their handbook (“difficulties will be encountered if studies have summarised their results using medians”, section 9.2.4[2]), but do not propose a solution. In practice, investigators often dichotomise data from shorter ordinal scales, and treat data from longer ordinal scales as continuous. Both of these approaches are sub-optimal. Dichotomising scales necessitates a loss of detail, and participants close to but on opposite sides of the split are characterised as very different rather than very similar. Statistical power is lost: a median split has been equated to discarding one-third of the data [6]. Treating data as continuous implies a consistent relationship between each level of the scale, which is not true of ordinal scales, and assumptions of normality are often violated. In the context of meta-analysis, it may be argued that, due to central limit theorem, mean values across a group of studies (and hence mean differences) will be approximately normally distributed, rendering any concerns about violation of normality invalid. Although this may be true, it fails to acknowledge that it is inappropriate to use means as a measure of central tendency for scales where we know only the order of levels on the scale, and not the distance between them.

Attempts to address these limitations were made more than 30 years ago. Kraemer and Andrews published a paper entitled “A non-parametric technique for effect size calculation” in 1982 [7]. This was followed in 1983 by a technical report from Hedges and Olkin entitled “Nonparametric estimators of effect size in meta-analysis” [8]. Yet there remains little agreement on how to pool ordinal data appropriately in meta-analysis. Some have suggested using a proportional odds model for combining data from ordinal scales [9,10], although this assumes a consistent relationship between levels on the scale. Others have developed techniques to estimate mean and standard deviation from the median and range [11,12], but this is an inexact solution, with estimates affected by sample size and normality of the data. Data may be pooled based on the Fisher’s exact test rather than the t-statistic [13], though this requires dichotomisation. A Bayesian approach that is not reliant on an assumption of normality has been proposed [14], but this necessitates complex modelling of the data.

Methods

First, to identify the most widely used clinical rating scales, we analysed all the studies included in a systematic review of clinical trials of oral neuroprotective therapies in multiple sclerosis, Alzheimer’s disease, amyotrophic-lateral sclerosis, Parkinson’s disease, and Huntington’s disease [15]. These studies provide an insight into the use of neurological rating scales; they may not be representative of all conditions. We selected the 6 most common rating scales, and extracted data from each of the individual studies that included these scales on (a) the measure of central tendency reported and (b) the type of analysis used. Second, to assess how these data are pooled in meta-analysis, we searched the Cochrane Database (in September 2013) for systematic reviews using each of the 6 rating scales as a keyword. For each of the relevant systematic reviews, we extracted data on the meta-analytic approach used. Third, we identified an effect size measure for use in meta-analysis of ordinal data that is assumption-free and easy to compute. To demonstrate feasibility, we generated mock data and used this effect size measure in meta-analysis.

Results and Discussion

Reporting and analysis of individual study data

The 6 most commonly used rating scales were: the Mini-Mental State Examination (MMSE), the Expanded Disability Status Scale (EDSS), the Unified Parkinson’s Disease Rating Scale (UPDRS), the Alzheimer’s Disease Assessment Scale (ADAS), the Unified Huntington’s Disease Rating Scale (UHDRS) and the Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS). Properties of these 6 scales are outlined in the boxed text (Box 1).

Across 103 studies, we identified 128 instances of scale data being reported (MMSE 34, EDSS 27, UPDRS 26, ADAS-cog 20, UHDRS 11, ALSFRS 10). Of these, 7% did not include a measure of central tendency, 13% reported medians (either alone or alongside means) and 80% reported means alone (Fig 1).

Of the 128 instances of scale data being reported, 26% did not include analysis or did not specify analysis type, 34% featured non-parametric (rank-based) analysis—either alone or alongside other analysis, and 40% included parametric (reliant on assumption of normality) analysis alone (Fig 2).

Meta-analysis of rating scale data

We identified 70 relevant systematic reviews (MMSE 24, EDSS 11, UPDRS 10, ADAS-cog 20, UHDRS 1, ALSFRS 4), 60 of which incorporated meta-analysis. Of these 60, 88% included

Box 1. Clinical Rating Scales.

MMSE: Cognitive screening tool scored out of 30 [16], often treated as a continuous scale. Data are typically skewed towards ceiling, as illustrated in a post-stroke population (median 26, IQR 22–27, skewness -1.09) [17].

EDSS: Ordinal rating scale widely used to evaluate function in multiple sclerosis [18]. Ranges from 0–10 in half-point increments. Distribution of EDSS scores is rarely normal, with the predominant pattern being bimodal [19].

UPDRS: Ordinal rating scale widely used to assess function in Parkinson’s disease [20]. Includes 42 items that are mostly scored on a 5 point scale (0 normal, 4 most severe).

ADAS: Designed to screen for early Alzheimer’s disease. Includes a cognitive subscale (ADAS-cog) [21]. ADAS-cog is scored from 0–70 and is generally treated as a continuous scale.

UHDRS: Ordinal rating scale used to test function in Huntington’s disease [22]. Each of its 6 components has a different scoring format; they are typically reported separately.

ALSFRS: Ordinal rating scale used to measure function in amyotrophic lateral sclerosis. Revised to ALSFRS-R [23]. Contains 12 items, scored from 0 (most severe) to 4 (normal). Data are often skewed, with >80% of patients classed as ‘mild’ or ‘moderate’ (>24) [24].

mean difference and 22% included difference in proportions (Fig 3). None included medians or other rank-based effect size measures in their meta-analyses. In fact, across all the systematic reviews, only a single study was identified that reported median difference as an outcome (on the UPDRS) [25].

Why the status quo is a problem

In 4 out of every 5 studies, means were the single measure of central tendency reported, while only approximately 1 in every 8 studies included medians. Approach to data analysis revealed a mis-match; in studies where analyses were specified, almost half (46%) included a non-parametric technique. Acknowledgement that the data may not conform to parametric

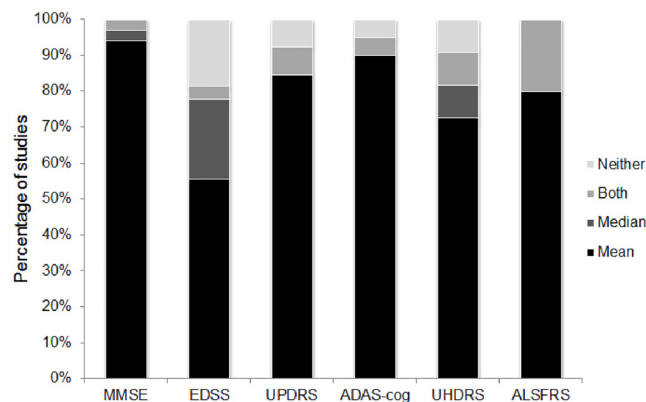


Fig 1. Reporting of central tendency for the 6 clinical rating scales.

doi:10.1371/journal.pone.0145580.g001

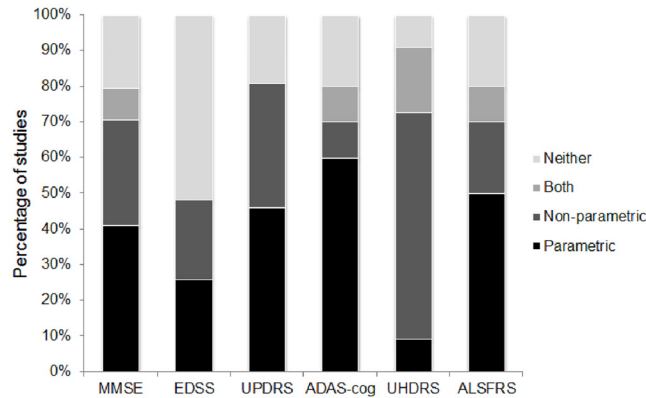


Fig 2. Approach to statistical analysis for the 6 clinical rating scales.

doi:10.1371/journal.pone.0145580.g002

assumptions was stronger at the analysis stage than at the descriptive stage. Reliance on means, standard deviations and standardised mean differences in the context of ordinal assessment scales that often yield non-normally distributed data is likely to violate statistical assumptions and bias interpretation of results. Yet correct reporting of medians and inter-quartile ranges will disqualify the data from subsequent meta-analysis, not only reducing the available information but potentially selecting out higher quality studies. What is required is a method for reporting ordinal data that does not assume a certain underlying distribution, does not necessitate dichotomisation, that facilitates pooling of results in meta-analysis, and does not inflict too harsh a reporting burden on individual study authors.

A simple solution: The generalised odds ratio

In pursuit of our third aim, we propose the use of a rank-based generalised odds ratio as a non-parametric equivalent to the standardised mean difference for use in meta-analysis. Agresti [26] outlined a generalisation of the odds ratio (GenOR) for ordinal outcomes. The method involves considering all possible pairs of observations, one from the treatment group and one from the control group. Agresti’s GenOR measures the ratio of the probabilities that a randomly chosen pair favours the treatment or control group. For example, if GenOR = 2, the clinical interpretation is that “compared to a person who does not receive the treatment, you are twice as likely to have a better outcome as you are to have a worse outcome.” Agresti’s GenOR

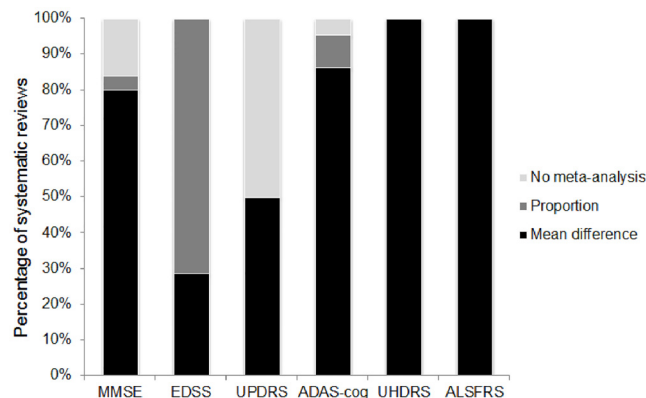


Fig 3. Approach to pooling data from the 6 clinical rating scales in systematic reviews.

doi:10.1371/journal.pone.0145580.g003

measure, however, discards data where observations are tied. This can involve discarding a considerable amount of data, thus reducing precision in the estimate of treatment effect, and potentially biasing clinical interpretation. If GenOR = 2 and 50% of the observations are tied, the above interpretation still applies but is made more accurate with the addition of “. . . but there's also a 50% chance you will have the same outcome.”

The proposed Wilcoxon-Mann-Whitney generalised odds ratio (WMW GenOR) follows the same logic as Agresti's GenOR but does not ignore the ties; tied observations are split evenly between better and worse outcomes [27]. We have published full details on how to calculate the WMW GenOR elsewhere [28] (S2 File). For a 2-group comparison, the WMW GenOR can be used as a natural effect size measure to accompany the WMW test. Importantly, the WMW GenOR can also be calculated for other outcomes, including continuous ones [27]. Interpretation is straightforward, as the odds-based measure can be directly translated to the 'common language effect size' measure [29]. Assuming normally distributed data with similar variances, this common language effect size can then be translated to the standardized mean difference. For example, for a certain fixed sample size, if the probability that a randomly selected person from the treatment group will score higher than a randomly selected person from the control group is 0.58, this equates to a standardized mean difference of 0.3; a probability of 0.71 equates to a standardized mean difference of 0.8 [30]. The WMW GenOR also has a natural relationship to the Number-Needed-to-Treat calculated on an ordinal scale [31]. Confidence intervals and p-values for WMW GenOR are implementable in any statistical software, including standard spreadsheets, given the existence of conveniently computable closed-form asymptotic expressions. Importantly, for the analysis of research trials where adjustment for prognostic covariates is often recommended, these formulas easily extend to stratified analyses.

Where individual studies have computed the WMW GenOR, the $\ln(\text{GenOR})$ and the standard error can be inserted directly into meta-analysis (e.g., using *metan* command in Stata). While the sampling distribution of the odds ratio may be skewed, the sampling distribution of the log odds ratio follows an approximately normal distribution [26]. Where individual studies have not computed the WMW GenOR, it remains possible to calculate it retrospectively and include the data in meta-analysis. This requires the following information: sample size of the groups (N_1 , N_2), Mann-Whitney U statistic and 2-tailed p-value. $\ln(\text{GenOR})$ is calculated as $\ln[(U/(N_1*N_2))/(1-(U/(N_1*N_2)))]$ and standard error as $ABS[\ln(\text{GenOR})/Invnormal(1-p/2)]$. Although the WMW GenOR is different to an odds ratio generated from a study with a binary outcome, it can be meta-analysed and visually represented in forest plots in the same way.

Implementation: An example using mock EDSS data

Using the basic distribution of EDSS scores identified in a large multiple sclerosis study [19] as a starting point, we generated mock data for a control group (sample 0) and 5 treatment groups (samples 1–5). Each group consisted of 100 patients. Shapiro-Wilk tests indicated that distribution of data in all samples was significantly non-normal. The histograms of EDSS scores are presented in Fig 4.

These data were used to simulate 5 studies, with EDSS scores compared between control and treatment groups (Study 1: sample 0 versus sample 1, Study 2: sample 0 versus sample 2, etc.). Using these raw data, we pooled the 5 studies in 2 different ways: (1) a typical random effects meta-analysis based on the standardised mean difference (Fig 5), and (2) an ordinal meta-analysis based on the WMW GenOR (Fig 6).

In practice, the WMW GenOR is rarely calculated at the study level and so this information is unavailable to authors of systematic reviews. To provide a more realistic implementation, we used the same mock data but assumed that only summary statistics were available from the 5

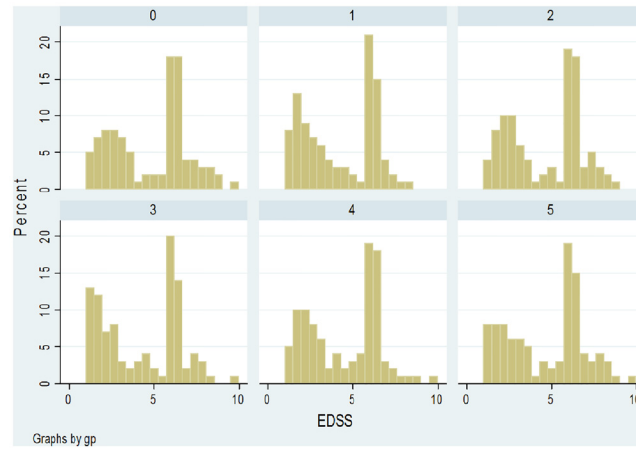


Fig 4. Histograms of mock EDSS data in a control group (sample 0) and 5 treatment groups (samples 1–5).

doi:10.1371/journal.pone.0145580.g004

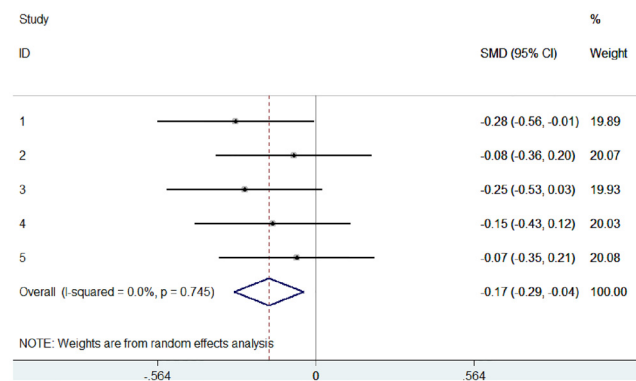


Fig 5. Random effects meta-analysis, based on standardised mean difference, of 5 simulated studies using mock EDSS data.

doi:10.1371/journal.pone.0145580.g005

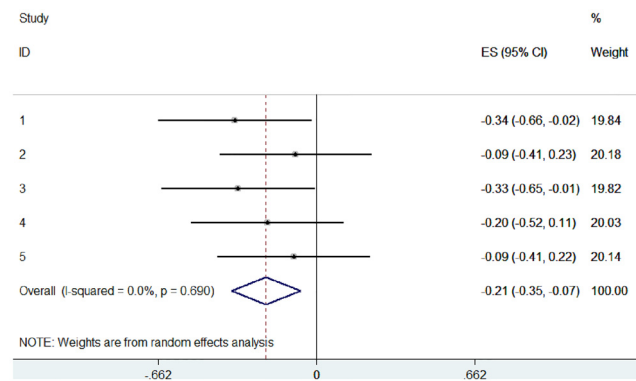


Fig 6. Ordinal meta-analysis, based on WMW GenOR, of 5 simulated studies using mock EDSS data.

doi:10.1371/journal.pone.0145580.g006

Table 1. Mock EDSS data: Summary statistics from the 5 ‘studies’.

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

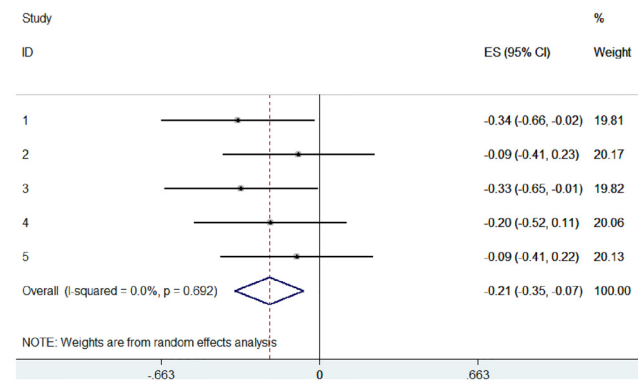


Fig 7. Ordinal meta-analysis, based on WMW GenOR calculated from summary statistics, of 5 simulated studies using mock EDSS data.

doi:10.1371/journal.pone.0145580.g007

‘studies’ (Table 1). These Mann-Whitney U and p-values were used to calculate WMW GenORs, which were then combined in ordinal meta-analysis (Fig 7). Calculating WMW GenORs retrospectively from summary data made no difference to effect sizes (compare Figs 6 and 7), indicating that meta-analysis is feasible even when WMW GenORs are not reported at the study level.

These forest plots show that ordinal data can be visually represented and interpreted using meta-analysis in the same way as continuous or dichotomous data. WMW GenOR-based meta-analysis is not only the more appropriate approach for ordinal data, in this case it was more sensitive to group differences than the standardised mean difference approach (compare Figs 5 and 6). In Study 3 (sample 0 versus sample 3), group difference was significant according to Mann-Whitney U (p = 0.045) but not according to t-test (p = 0.073). It is often assumed that parametric analyses are more likely to detect significant treatment effects, but this is not necessarily true in the context of non-normal distributions.

Conclusions

Meta-analysis is a vital tool for research and clinical decision-making. At the individual study level, it is important that ordinal data can be reported appropriately and not be excluded from meta-analysis. Using means and standard deviations in the context of ordinal scales gives an impression of exactness, but this is false precision; we do not know the distance between points on the scale, only their order. Our results demonstrate that there is wide scope for improvement in the reporting and analysis of ordinal and non-parametric data in the literature. A solution to this problem is the WMW GenOR. This odds-based effect size measure is assumption-free, easy to compute and can be readily combined in meta-analysis. Such a probability-based

effect size is not limited to ordinal data; it is easily applied to continuous data too. We hope that adoption of the WMW GenOR will have the dual effect of improving the reporting of ordinal data in individual studies while also increasing the inclusivity (and therefore validity) of meta-analyses.

Supporting Information

S1 File. Excel file containing study data.

(XLSX)

S2 File. PDF file of related article (Churilov et al., 2014).

(PDF)

Acknowledgments

Emily Sena is supported by funding from Research Councils UK. We thank Mahsa Keshtkaran for assistance with data extraction and Hayden Johns for development of statistical programming.

Author Contributions

Conceived and designed the experiments: TBC ESS LC. Performed the experiments: TBC ESS LC. Analyzed the data: TBC LC. Contributed reagents/materials/analysis tools: LC. Wrote the paper: TBC ESS LC.

References

1. Rosenthal R. The file drawer problem and tolerance for null results. *Psychological Bulletin*. 1979; 86(3):638–41.
2. *Cochrane Handbook for Systematic Reviews of Interventions*: The Cochrane Collaboration; 2011. Available from: www.cochrane-handbook.org.
3. Hong KS, Saver JL. Quantifying the value of stroke disability outcomes: WHO global burden of disease project disability weights for each level of the modified Rankin Scale. *Stroke*. 2009; 40(12):3828–33. doi: [10.1161/STROKEAHA.109.561365](https://doi.org/10.1161/STROKEAHA.109.561365) PMID: [19797698](https://pubmed.ncbi.nlm.nih.gov/19797698/)
4. Kidwell CS, Jahan R, Gornbein J, Alger JR, Nenov V, Ajani Z, et al. A trial of imaging selection and endovascular treatment for ischemic stroke. *New England Journal of Medicine*. 2013; 368(10):914–23. doi: [10.1056/NEJMoa1212793](https://doi.org/10.1056/NEJMoa1212793) PMID: [23394476](https://pubmed.ncbi.nlm.nih.gov/23394476/)
5. Kaul S, Boyle KJ, Pope JC, Parmeter CF, Kuminoff NV. Robust meta-analysis using median-quantile and nonparametric regression procedures: Investigating the validity of benefit transfers. US: EPA, 2011.
6. Cohen J. The Cost of Dichotomization. *Applied Psychological Measurement*. 1983; 7(3):249–53.
7. Kraemer HC, Andrews G. A nonparametric technique for meta-analysis effect size calculation. *Psychological Bulletin*. 1982; 91(2):404–12.
8. Hedges LV, Olkin I. Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin*. 1984; 96(3):573–80.
9. Agresti A, Coull BA. Order-restricted tests for stratified comparisons of binomial proportions. *Biometrics*. 1996; 52(3):1103–11. PMID: [8805770](https://pubmed.ncbi.nlm.nih.gov/8805770/)
10. Whitehead A, Jones NM. A meta-analysis of clinical trials involving different classifications of response into ordered categories. *Stat Med*. 1994; 13(23–24):2503–15. PMID: [7701150](https://pubmed.ncbi.nlm.nih.gov/7701150/)
11. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology*. 2005; 5(13).
12. Bland M. Estimating mean and standard deviation from the sample size, three quartiles, minimum, and maximum. *International Journal of Statistics in Medical Research*. 2015; 4:57–64.
13. Kong X, Mas V, Archer KJ. A non-parametric meta-analysis approach for combining independent microarray datasets: application using two microarray datasets pertaining to chronic allograft nephropathy. *BMC Genomics*. 2008; 9(98).

14. Branscum AJ, Hanson TE. Bayesian nonparametric meta-analysis using Polya tree mixture models. *Biometrics*. 2008; 64(3):825–33. PMID: [18162115](#)
15. Vesterinen HM, Connick P, Irvine CMJ, Sena ES, Egan KJ, Carmichael GG, et al. Drug repurposing: A systematic approach to evaluate candidate oral neuroprotective interventions for secondary progressive multiple sclerosis. *PLoS One*. 2015; 10(4), article e0117705, doi: [10.1371/journal.pone.0117705](#) PMID: [25856304](#)
16. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12(3):189–98. PMID: [1202204](#)
17. Cumming TB, Churilov L, Linden T, Bernhardt J. Montreal Cognitive Assessment and Mini-Mental State Examination are both valid cognitive tools in stroke. *Acta Neurologica Scandinavica*. 2013; 128:122–9. doi: [10.1111/ane.12084](#) PMID: [23425001](#)
18. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*. 1983; 33(11):1444–52. PMID: [6685237](#)
19. Hohol MJ, Orav EJ, Weiner HL. Disease steps in multiple sclerosis: A simple approach to evaluate disease progression. *Neurology*. 1995; 45(2):251–5. PMID: [7854521](#)
20. Fahn S, Elton RL, Committee UD. Unified Parkinson's Disease Rating Scale. In: Fahn S MC, Calne DB, Goldstein M, eds., editor. *Recent Developments in Parkinson's Disease*. Florham Park, NJ: Macmillan Health Care Information; 1987. p. 153–63.
21. Mohs RC, Knopman D, Petersen RC, Ferris SH, Ernesto C, Grundman M, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: Additions to the Alzheimer's Disease Assessment Scale that broaden its scope. *The Alzheimer's Disease Cooperative Study. Alzheimer Disease & Associated Disorders*. 1997; 11(Suppl 2):s13–21.
22. Anonymous. Unified Huntington's Disease Rating Scale: Reliability and consistency. *Huntington Study Group. Movement Disorders*. 1996; 11(2):136–42. PMID: [8684382](#)
23. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, et al. The ALSFRS-R: A revised ALS functional rating scale that incorporates assessments of respiratory function. *BDNF ALS Study Group (Phase III). Journal of the Neurological Sciences*. 1999; 169(1–2):13–21. PMID: [10540002](#)
24. Mioshi E, Lillo P, Kiernan M, Hodges J. Activities of daily living in motor neuron disease: Role of behavioural and motor changes. *Journal of Clinical Neuroscience*. 2012; 19(4):552–6. doi: [10.1016/j.jocn.2011.07.042](#) PMID: [22249015](#)
25. Guttman M. Double-blind comparison of pramipexole and bromocriptine treatment with placebo in advanced Parkinson's disease. *International Pramipexole-Bromocriptine Study Group. Neurology*. 1997; 49(4):1060–5. PMID: [9339690](#)
26. Agresti A. Generalized odds ratios for ordinal data. *Biometrics*. 1980; 36:59–67.
27. O'Brien RG, Castelloe J, editors. *Exploiting the link between the Wilcoxon-Mann-Whitney Test and a simple odds statistic*. Thirty-first Annual SAS Users Group International Conference; 2006: Cary, NC: SAS Institute Inc.
28. Churilov L, Arnup S, Johns H, Leung T, Roberts S, Campbell BCV, et al. An improved method for simple, assumption-free ordinal analysis of the modified Rankin Scale using generalized odds ratios. *International Journal of Stroke*. 2014; 9(8):999–1005. doi: [10.1111/ijis.12364](#) PMID: [25196780](#)
29. McGraw KO, Wong SP. A common language effect size statistic. *Psychological Bulletin*. 1992; 111(2):361–5.
30. Coe R. It's the effect size, stupid: What effect size is and why it is important. *British Educational Research Association annual conference*; Exeter 2002.
31. Optimising the Analysis of Stroke Trials Collaboration, Bath P, Hogg C, Tracy M, Pocock S. Calculation of numbers-needed-to-treat in parallel group trials assessing ordinal outcomes: Case examples from acute stroke and stroke prevention. *International Journal of Stroke*. 2011; 6(6):472–9. doi: [10.1111/j.1747-4949.2011.00614.x](#) PMID: [21645271](#)