The data extraction challenge

Observer variation when extracting data for the calculation of a standardized mean difference

Britta Tendal[1], Julian P. T. Higgins[4], Peter Jüni[2,3], Asbjørn Hróbjartsson[1], Sven Trelle[2,3], Eveline Nüesch[2,3], Simon Wandel[2,3], Anders W. Jørgensen[1], Katarina Gesser[5], Søren Ilsøe-Kristensen[5] and Peter C. Gøtzsche[1]

[1] The Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
[2] Division of Clinical Epidemiology and Biostatistics, University of Bern, Switzerland
[3] CTU Bern, Bern University Hospital, Switzerland
[4] MRC Biostatistics Unit, Institute of Public Health, University of Cambridge, United Kingdom
[5] The Faculty of Pharmaceutical Sciences, University of Copenhagen, Denmark

What is known?

- Incorrect data extraction can lead to false results.
- Multiplicity in trial reports invites variation in data extraction.
- The impact of these factors on meta-analyses is not clear.

What this study adds

- There is considerable observer variation in data extraction.
- The reasons for disagreement are data extraction errors and multiplicity of data.
- The impact on meta-analyses is potentially large.

Background

Data extraction for meta-analysis can be complicated. Many decisions need to be made and there is often a multiplicity of data related to different time points, outcomes, scales and groups; also calculations may be necessary. These choices can lead to bias, and researchers can make errors[1] or disagree. Previous studies on observer variation are limited by few observers[2,3] and few trials[2]. We aimed to better quantify the effect of observer variation in data extraction for meta-analyses.

Objectives

We investigated the observer variation when extracting data for the calculation of a standardized mean difference (SMD), its impact on meta-analyses, and the reasons for disagreement.

Methods

We selected a random sample of 10 recent Cochrane reviews, which presented a result as an SMD and retrieved the protocols for the reviews and the trial reports (n=45) that corresponded to the first SMD result in each review. Ten observers (five experienced methodologists and five PhD students) independently extracted the necessary data from the trial reports for calculation of the SMD. The observers did not have access to the original review and were only given the review protocols and the trial reports as PDF files. In the protocols, an additional researcher had highlighted the relevant outcome and other important factors. Based on the extracted data, this researcher calculated the SMDs. Agreement was defined as SMDs that differed less than 0.1 in their point estimates or confidence intervals.

Results

The 10 meta-analysis results from each of the 10 observers are shown in figure 1 along with the results from the originally published meta-analyses. Out of the total of 100 meta-analyses, 7 originally statistically significant results became non-significant, 3 non-significant results became significant, and 4 meta-analyses, which were based on the same protocol, showed a significantly beneficial effect for the control group whereas the corresponding originally published review[4] had found a significantly beneficial effect for the experimental group.

To assess the inter-observer variation, the results from the 10 observers were paired in all possible ways (45 pairs). Thus, the 10 meta-analyses yielded a total of 450 pairs and the 45 trials yielded 2025 pairs for agreement analysis. The agreement was 31% for the meta-analyses and 53% for the trials. The experts agreed somewhat more than the PhD students at trial level, but not at meta-analysis level (Table 1). The magnitudes of the disagreements on meta-analyses level are shown in figure 2. Fifty-six per cent had a disagreement that ranged from 0.1 to 0.99 SMD and 10% had disagreements of 1 SMD or more (4% of the meta-analyses were excluded since one of the observers excluded all the underlying trials). Important reasons for disagreement were calculation errors, oversights, and differences in selection of time points, scales, control groups and type of calculations (Table 2).

Limitations of the study

The researchers were in an experimental setting where they were presented with protocols they had not themselves developed based on research questions that they had not themselves asked. Furthermore, in several cases the observers were not content area experts. Another limitation is that in real life some of the errors we made would probably have been detected, as two observers are generally recommended. Finally, even though it was an exclusion criterion for our study that the meta-analyses had to be based on published data only, it became apparent during data extraction that some of the trial reports did not contain the data needed for the calculation of the SMD. Usually, one would contact the authors to clear up details.

Conclusions

There was considerable disagreement between observers. Despite the limitations of this study due to its experimental design, it is clear that the multiplicity of data in trial reports calls for more detailed review protocols that take into account the possibility of different time points, scales and groups. The potential for error highlights the need for more than one observer and statistical knowledge or help from a statistician.

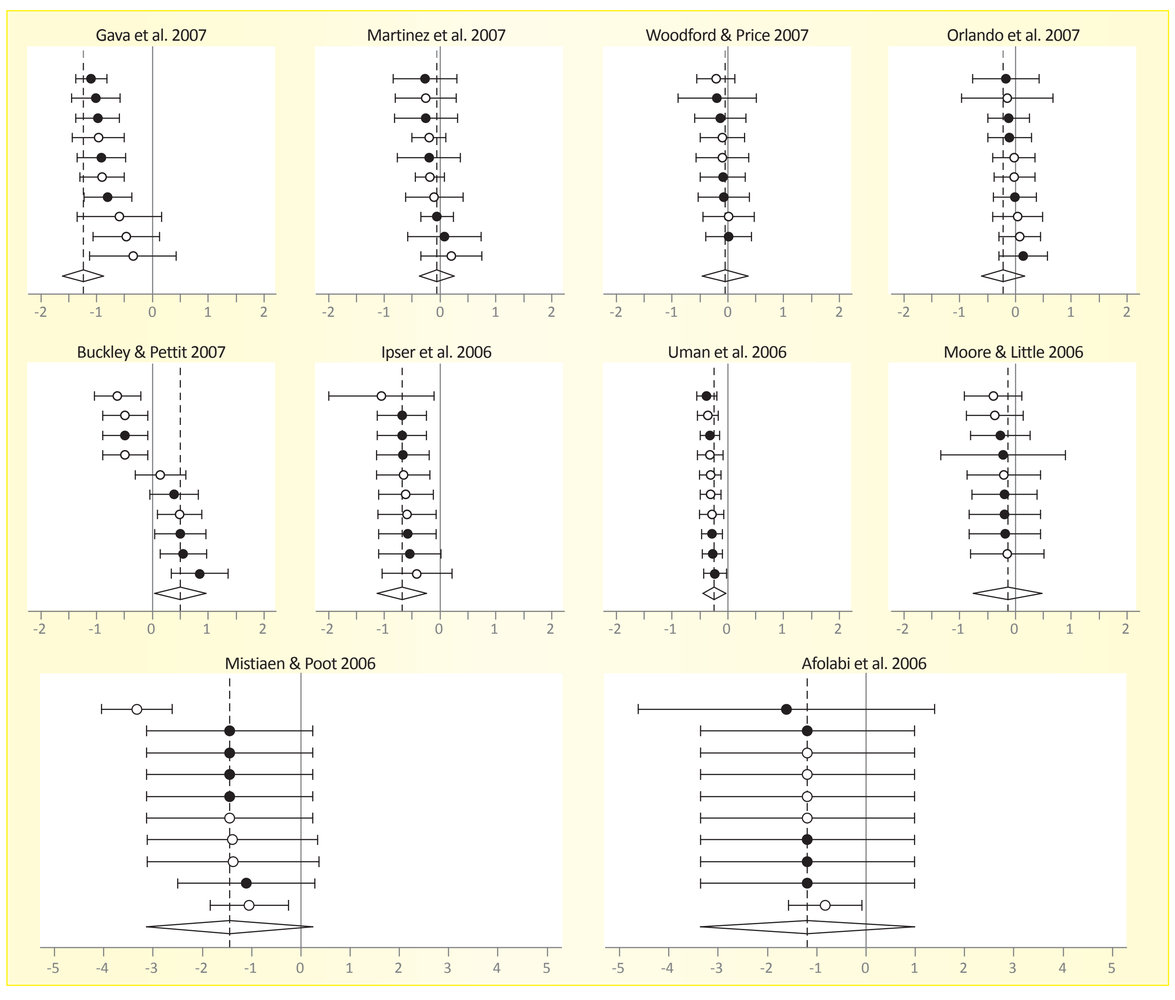
Table 1. Levels of overall agreement, number (%) of observer pairs that agreed on the SMD result and its 95% CI within the 0.1 cut off.

Trial level	
All pairs (n=2025)	1070 (53%)
Methodologist (n=450)	273 (61%)
Students (n=450)	209 (46%)
Mixed pairs (n=1125	586 (52%)
Meta-analysis level	
All pairs (n=450)	137 (31%)
Methodologist (n=100)	33 (33%)
Students (n=100)	27 (27%)
Mixed pairs (n=250)	78 (31%)

Table 2. Reasons for disagreement among the 41 trials on which the observers disagreed. The table shows the numbers of trials for which the listed reasons for disagreement were relevant. There may be more than one reason for disagreement per trial.

Reasons	No. of trials
Different choices regarding:	
Groups, pooling, splitting	15
Timing	9
Scales	6
Different calculations/imputations	6
Dropouts	4
Change or after treatment	4
Individual patient data	1
Exclusion of trials due to:	
Did not meet protocol inclusion criteria	14
Reporting unclear	14
Missing data	7
Could not/would not calculate	2
Only change after treatment	2
Errors due to:	
Misreading/typing error	4
Direction of effect	4
Standard error taken as standard deviation	2
Calculation error	1
Rounding	1

Figure 1. Forest plots of the meta-analyses (SMDs and 95% confidence intervals). The black circles represent experienced methodologists, the white circles the PhD students. The diamonds represents the results from the published meta-analyses.



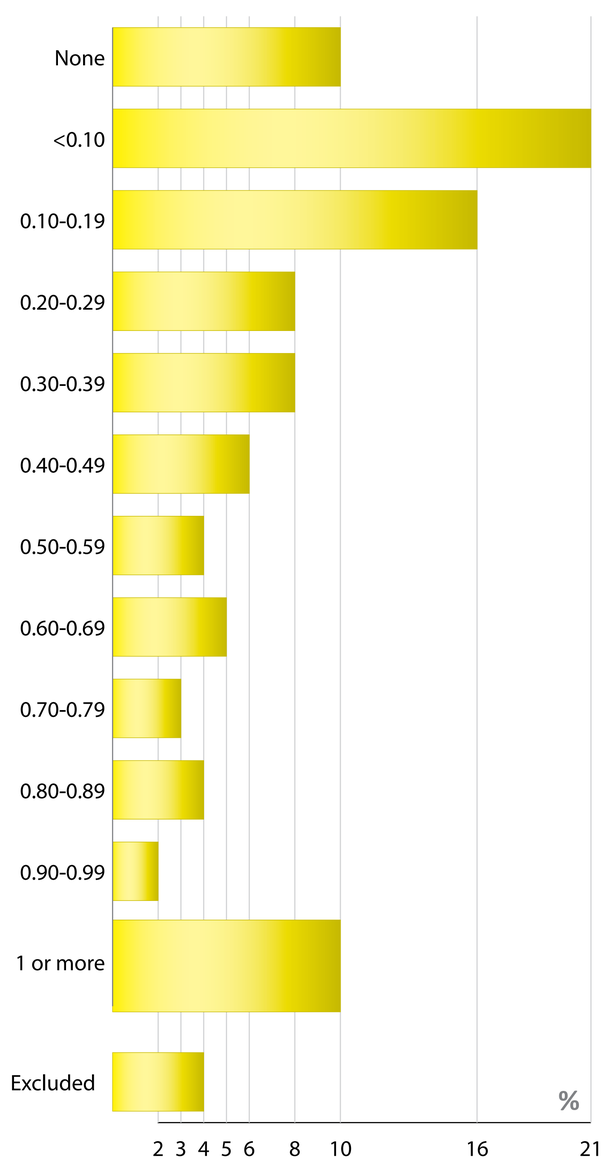
What is a standardised mean difference?

The standardized mean difference is used as a summary statistic in meta-analysis when the studies all assess outcomes of a similar nature, but measured on different scales, e.g. pain on a 10-point ranking scale or on a 100 mm visual analogue scale.

In such cases, it is necessary to standardise the measurements on a uniform scale before they can be pooled in a meta-analysis. This is done by calculating the standardized mean difference (SMD) for each trial, which is the difference in means between the two groups, divided by the pooled standard deviation of the measurements[5].

By this transformation, the outcome becomes dimension-less and the scales become uniform, e.g. for the same degree of pain, values measured on a 100 mm analogue scale would be expected to be 10 times larger than values measured on a 10-point ranking scale, but the standard deviation would also be expected to be 10 times larger.

Figure 2. Magnitudes of the disagreements on meta-analysis level. The figure displays percent of pairs for each level of disagreement.



References

1 Gøtzsche PC, Hróbjartsson A, Maric K, Tendal B. Data extraction errors in meta-analyses using standardised mean differences. JAMA 2007 Jul; 98(4): 430-7.

2 Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. J Clin Epidemiol 2006 Jul;59(7):697-703.

3 Jones AP, Remmington T, Williamson PR, Ashby D, Smyth RL. High prevalence but low impact of data extraction and reporting errors were found in Cochrane systematic reviews. J Clin Epidemiol 2005 Jul;58(7):741-2.

4 Buckley LA, Pettit T, Adams CE. Supportive therapy for schizophrenia. Cochrane Database Syst Rev 2007;(3):CD004716

5 Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.0 [updated February 2008]. The Cochrane Collaboration, 2008. Available from www.cochrane-handbook.org.

Design: Jacob Riis, Nordic Cochrane Centre