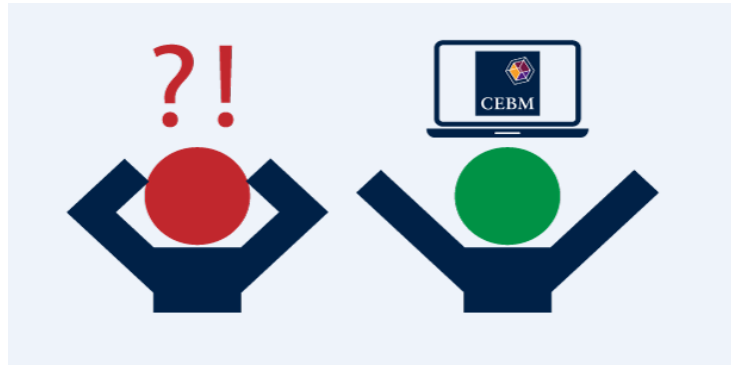


## Tip for data extraction for meta-analysis – G1



### How can you make data extraction more efficient?

Kathy Taylor

Data extraction that's not well organised will add delays to the analysis of data. There are a number of ways in which you can improve the efficiency of the data extraction process and also help in the process of analysing the extracted data. The following list of recommendations is derived from a number of different sources and my experiences of working on systematic reviews and meta-analysis.

1. **Highlight the extracted data on the pdfs of your included studies.** This will help if the other data extractor disagrees with you, as you'll be able to quickly find the data that you extracted. You obviously need to do this on your own copies of the pdfs.
2. **Keep a record of your data sources.** If a study involves multiple publications, record what you found in each publication, so if the other data extractor disagrees with you, you can quickly find the source of your extracted data.
3. **Create folders to group sources together.** This is useful when studies involve multiple publications.
4. **Provide informative names of sources.** When studies involve multiple publications, it is useful to identify, through names of files, which is the protocol publication (source of data for quality assessment) and which includes the main results (sources of data for meta-analysis).
5. **Document all your calculations and estimates.** Using calculators does not leave a record of what you did. It's better to do your calculations in EXCEL, or better still, in a computer program. Calculations coded (written) in a computer program are easier to read and manage than formulae in an EXCEL cell, particularly as some calculations for data extraction involve long equations with multiple brackets.



6. **Extract data into an EXCEL spreadsheet.** The data can be imported into computer programs or copied into [Review Manager](#), the software used in Cochrane Reviews.
7. **Extract only one number per cell.** The two sets of extracted data will need to be compared and the final data set agreed for the meta-analysis. In order to do both these analyses, the data has to be presented in a particular form, with only one number per cell in the EXCEL spreadsheet. Let me show you, using data from a [review](#) that I worked on.

The EXCEL entries in Figure 1 looks very neat and tidy, but they're not ready for analysis. The study name and year need to be separated to enable subgroup analyses on year of publication, and whilst both studies have years recorded with 4 characters, the study names have different numbers of characters. The means and standard deviations (SDs) also need to be separated and this is complicated by the fact that the means are recorded to two decimal places in one study and one decimal place in the other study. Therefore, time will need to be spent extracting the extracted data in Figure 1 before the data are ready for analysis.

	A	B	C	D	E
1		<b>Intervention group</b>		<b>Comparator group</b>	
2	<b>Study</b>	<b>n</b>	<b>mean (SD)</b>	<b>n</b>	<b>mean(SD)</b>
3	Atlantis 2000	28	53.49 (26.59)	37	110.98 (147.11)
4	Bakris 1994	8	33.1 (37.4)	7	421.9 (46.1)
5					

Figure 1. Neat and tidy, but not ready for analysis

The data in Figure 2 look ugly, compared to the data in Figure 1, but they're ready for analysis.

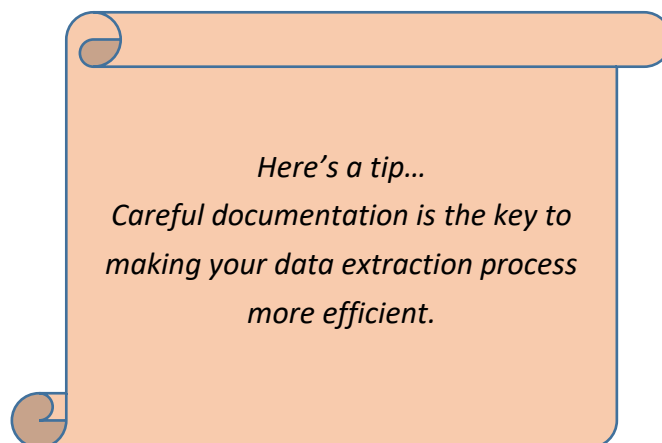
	A	B	C	D	E	F	G	H
1	<b>Study</b>	<b>year</b>	<b>ni</b>	<b>meani</b>	<b>Sdi</b>	<b>nc</b>	<b>meanc</b>	<b>Sdc</b>
2	Atlantis	2000	28	53.49	26.59	37	110.98	17.11
3	Bakris	1994	8	33.1	37.4	7	421.9	46.1
4								

Figure 2. Ready for analysis

8. **If a variety of data is reported, tabulate.** When a variety of data is reported across trials, in terms of different measures of the same outcome, or different stages of the trial (e.g. baseline, or endpoint), the number of columns in your data extraction file in EXCEL will

multiply. This can be overwhelming, and to help decide how to analyse the data (in terms of what effect estimate to use) it might be useful to produce a simple table to provide an overview of what data are available. Let me give you an example. In a [systematic review](#) I worked on, we were interested in the effect of treatment on albumin excretion, but this was reported in some studies as albumin excretion rate and in other studies as albumin creatinine ratio. Between the included studies, data was reported in terms of endpoint data, change from baseline, percentage change from baseline and percentage difference between the two treatment arms. In our review, we found it helpful just considering these six variables, and we decided that using a ratio of means effect measure (which will be covered in a future blog post) we were able to pool more data.

9. **Flag studies for which you have made estimates.** These studies will need to be excluded as a sensitivity analysis to ensure that your conclusions are not sensitive to the estimates that you've made. So it's important that you can quickly identify them when you're doing the meta-analysis.
  
10. **Flag studies which have low assessments of quality.** You may also wish to exclude studies with low quality as a sensitivity analysis, so also flagging these will ensure that all the information that's required for the sensitivity analysis is together.



In my next post I will give some tips on how to reduce the risk of bias and errors in data extraction.

Dr Kathy Taylor teaches data extraction in [Meta-analysis](#). This is a short course that is also available as part of our [MSc in Evidence-Based Health Care](#), [MSc in EBHC Medical Statistics](#), and [MSc in EBHC Systematic Reviews](#).

Follow updates on this blog, related news, and to find out about other examples of statistics being made more broadly accessible on Twitter [@dataextips](#)