



The  
University  
Of  
Sheffield.

# Data Extraction

Myfanwy Lloyd Jones, Senior Research  
Fellow, ScHARR, University of Sheffield

# The purpose of data extraction

- “is to describe the study in general, to extract the findings from each study in a consistent manner to enable later synthesis, and to extract information to enable quality appraisal so that the findings can be interpreted. Ideally this should be undertaken in such a way as to require minimal reference to the original papers at data synthesis stage.” (Social Care Institute for Excellence, 2006)

# The challenge is.....

- To data extract honestly and consistently as a precursor to the interpretative phase.
- Not to prejudge the value or meaning of data as it is extracted (or as an alternative to extraction)
- **Query:** Use of multiple extractors?
  - Ideally, to minimise bias, data should be extracted independently by two reviewers, who should then agree on a final version (SCIE, 2006)

# Why do we do data extraction?

- Articles are in different formats and use different styles of reporting
- Need to highlight main data elements of interest
- Need to provide standardisation
- Need to aid pattern recognition and analysis

# Data extraction is:

- An attempt to reduce a complex, messy, context-laden and quantification-resistant reality to a matrix of categories and numbers
- Time consuming
- Often difficult

# Two levels of data extraction

## Rapid:

- Extracting data direct to tables:
  - Study characteristics
  - Methodological quality
  - Findings
- Practical where clear idea of review output (eg mapping existing research)
- Synthesis impossible or limited

## Comprehensive:

- Extracting data to data extraction forms
- One form per study
- Facilitates more detailed analysis and synthesis
- Valuable for multi-use of data
- Resource intensive

# Stages of Data Extraction

- Stage 1 - Assessment of eligibility
- Stage 2 - Assessment of quality
- Stage 3 - Assessment of study characteristics
- Stage 4 - Extraction of study findings

# Assessment of eligibility

- Most articles identified by the literature searches will have been eliminated in the sifting process at the title or abstract stage
- Data extraction and sifting overlap if data extraction is used as a means of documenting why studies which were excluded at a full reading were excluded



# Assessment of eligibility -

C:\Documents and Settings\Mr Booth\My Documents\Data\_extraction.zip-CC-#21 Oldham Data extraction form.doc - Microsoft Word

File Edit View Insert Format Tools Table Window Help Type a question for help

Data extracted by: CC	Date: 11 March 2009	Ref ID: 2786 (Oldham 2004)
-----------------------	---------------------	----------------------------

**Full paper screen:**

Question		If Yes	If No
1	Is the study 2003 onwards	Continue	Exclude
2	Does the intervention being evaluated include one or more of the following: Aspirin, NSAIDs, Vitamins, Minerals, Calcium, Folic acid, dietary supplements generally	Continue	Exclude
3	Does the study report, as an outcome, people's attitudes, perceptions or beliefs concerning the taking of one or more of the agents listed above	Continue	Exclude
		Include	
	<i>Note: any population group may be included</i>		

Page 1 Sec 1 1/7 At 15.4cm Ln 26 Col 1 REC TRK EXT OVR English (U.K)

Start Unused Desktop Shortcuts blueyonder broadband I... ESPRX420 Reference Guide 15:31

# Issues re: eligibility

- Qualitative research versus qualitative data (surveys, quantitative studies, audit/evaluations)
- Reports which include original data versus those which include author interpretations
- “Thickness of detail”?

# Examples of definitions of qualitative evidence

- Any study that utilised both qualitative data collection and analysis methods (Munro 2007, Noyes 2007)
- Studies in which qualitative methods were used to describe people's experiences (Briggs 2007)
- Any study reporting empirical, non-numerical data (Marston 2006)
- "Papers had to report results of qualitative (i.e., textbased and interpretive) analysis based on qualitative methods of data collection." (Smith 2006 Lancet p826)
- "Qualitative methods were used to describe people's experience of living with a leg ulcer e.g. phenomenological studies; grounded theory; descriptive; focus groups or interview studies." (Briggs 2007, p320)

# Issues re assessment of quality

- Is quality assessment a separate process, or is it part of data extraction?
- Should it be used to determine an inclusion/exclusion quality threshold?
- Can a poor quality study yield a valid finding/contribution?

# Assessment of study characteristics

- Research question
- Study location (country, setting)
- Time frame
- Population (number, age, gender, ethnicity etc; how the sample was recruited)
- Study type
  - Data collection – methods
  - Analysis – methods
- Researcher (demographic data, disciplinary background, source of funding, etc)

# Issues re study characteristics

- Study location, time frame, and population may explain differences in findings
- Authors' definition of study type?
  - Not always accurate
- Reviewer's assessment of study type?
- Importance of theoretical/conceptual underpinning?

# Study findings (or data?)

- Sandelowski & Barroso (2003) differentiate between:
  - Findings: “the data-driven and integrated discoveries, judgments, and/or pronouncements researchers offer about the phenomena, events, or cases under investigation”
  - Data: “case descriptions or histories, quotes, incidents, and stories obtained from participants” – ie the empirical material on which findings should be based
- Some articles only report data

# Which to include?

- Many researchers feel that syntheses of qualitative studies should only draw on study findings – ie the categories, themes, metaphors, interpretations or explanations presented by the authors of the primary research
- Should study findings be included only if they are supported by a quote from a participant?
- What if the data (eg quotations) suggest to you a finding which isn't identified by the study authors?



# Issues re: study findings

- The data which are extracted should be determined by the review question/aim
- Should one approach data extraction with a pre-existing framework (Framework Analysis) or with a blank sheet (grounded theory type approach)?
- “Translation” between studies?
- Identification of new themes – iterative readings versus single pass?

# Data extraction forms

- Must balance detail with usefulness (look back at original protocol to identify main variables and hypotheses)
- Overly inclusive data extraction forms result in too much “white space” and consume valuable time
- Overly minimalist data extraction forms result in too many marginal notes and may require revisiting the articles

# Software for data extraction

- No set requirement for software, principally determined by desired means of presentation
- MS Word
- MS Excel
- MS Access (possible links with Reference Manager)
- SUMARI (Joanna Briggs Institute)

# Summary

- Data extraction approach must be appropriate to the review question, the type of review and the available evidence
- Regardless of approach, data extraction needs to be systematic and transparent and described in detail in final review document ('audit trail' between primary studies, data extraction and synthesis findings)
- Formal, technical process is a necessary, but not sufficient, element of synthesis. Provides substrate for subsequent interpretive and creative element of giving meaning to data
- Close attention to data extraction will facilitate initial understanding and description of characteristics of body of evidence. Paves way for more analytic and interpretive process of synthesis to follow



# Exercise

- To extract data from a published paper which has been identified as relevant to a systematic review of people's attitudes, perceptions, or beliefs regarding dietary supplements

# References

- Briggs M, Flemming K. Living with leg ulceration: a synthesis of qualitative research. *Journal of Advanced Nursing*. 2007;59(4):319-28.
- Marston C, King E. Factors that shape young people's sexual behaviour: a systematic review. *Lancet*. 2006;368(9547):1581-6.
- Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A. et al. Patient adherence to tuberculosis treatment: A systematic review of qualitative research. *PLoS Med*. 2007;4:e238

# References

- Noyes J. & Popay J. Directly observed therapy and tuberculosis: How can a systematic review of qualitative research contribute to improving services? A qualitative meta-synthesis. *Journal of Advanced Nursing* 2007;57(3):227-43.
- Sandelowski M. & Barroso J. Classifying the findings in qualitative studies. *Qualitative Health Research* 2003;13(7):905-923

# References

- Smith L., Pope C., Botha J. Patients' help-seeking experiences and delay in cancer presentation: a qualitative synthesis. *The Lancet* 366(9488):825-831
- Social Care Institute for Excellence (2006) *The conduct of systematic research reviews for SCIE knowledge reviews*  
<http://www.scie.org.uk/publications/researchresources/rr01.pdf>