Data Extraction

Systematic review and Meta-analysis

Shahab Rezaeian

PhD in epidemiology

School of Health, Kermanshah University of Medical Sciences

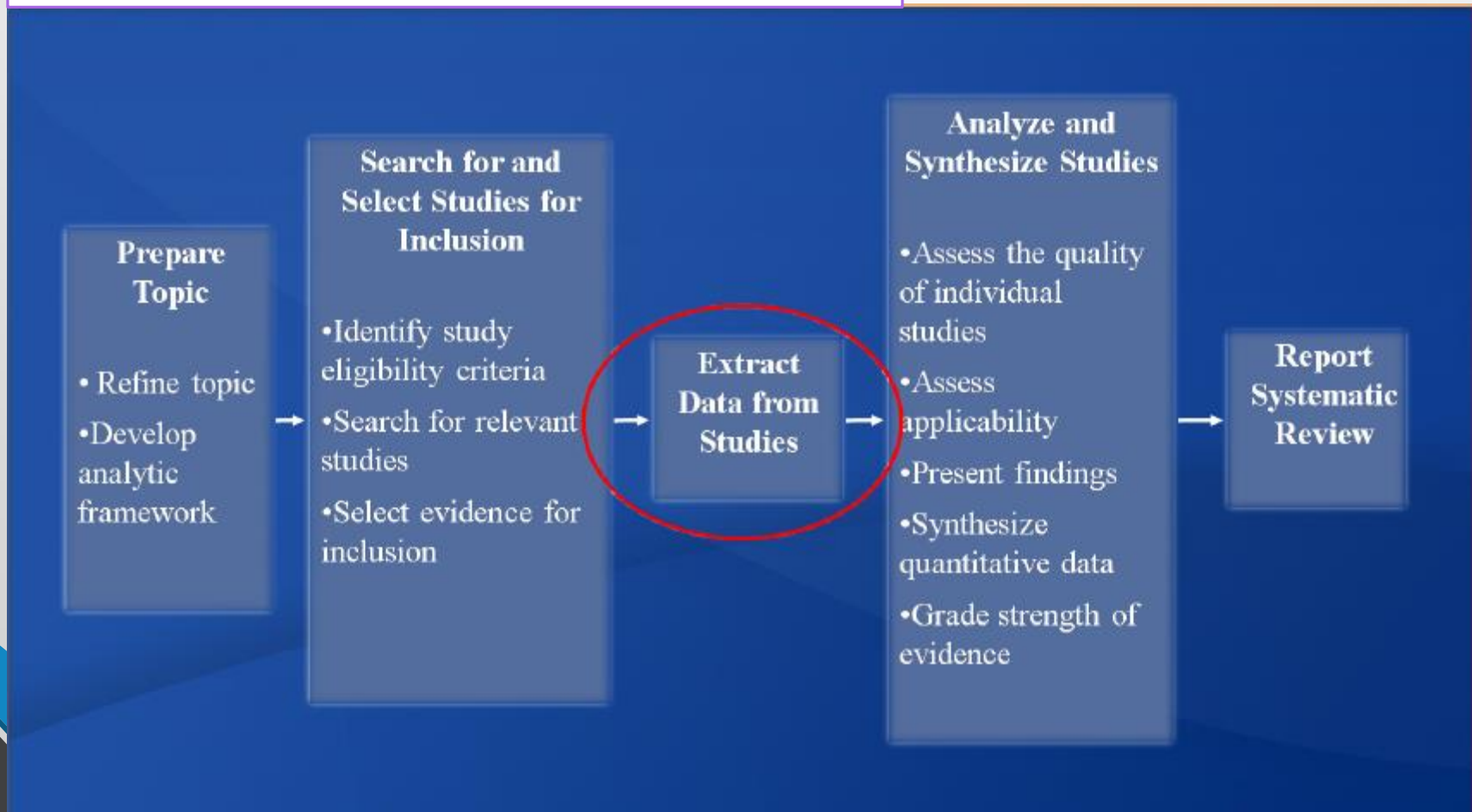
Overview

- The Purpose
- Data Extraction Forms
- Example Data Extraction Forms

Data Extraction

Purpose

Data collection and data extraction



Data Extraction

Purpose

- To summarize studies in a common format to facilitate synthesis and coherent presentation of data
- To identify numerical data for meta-analyses
- To obtain information to assess more objectively the risk of bias in and applicability of studies
- To identify systematically missing or incorrectly assessed data, outcomes that are never studied, and underrepresented populations

The Rationale for Data Extraction Processes

- “The findings of a systematic review depend critically:
  - on decisions relating to which studies are included,
  - on decisions relating to which data from these studies are presented and analyzed
  - methods used for these decisions must be transparent, and they should be chosen to minimize biases and human error” (Cochrane Handbook)

Data Extraction

Single vs. Double Extraction

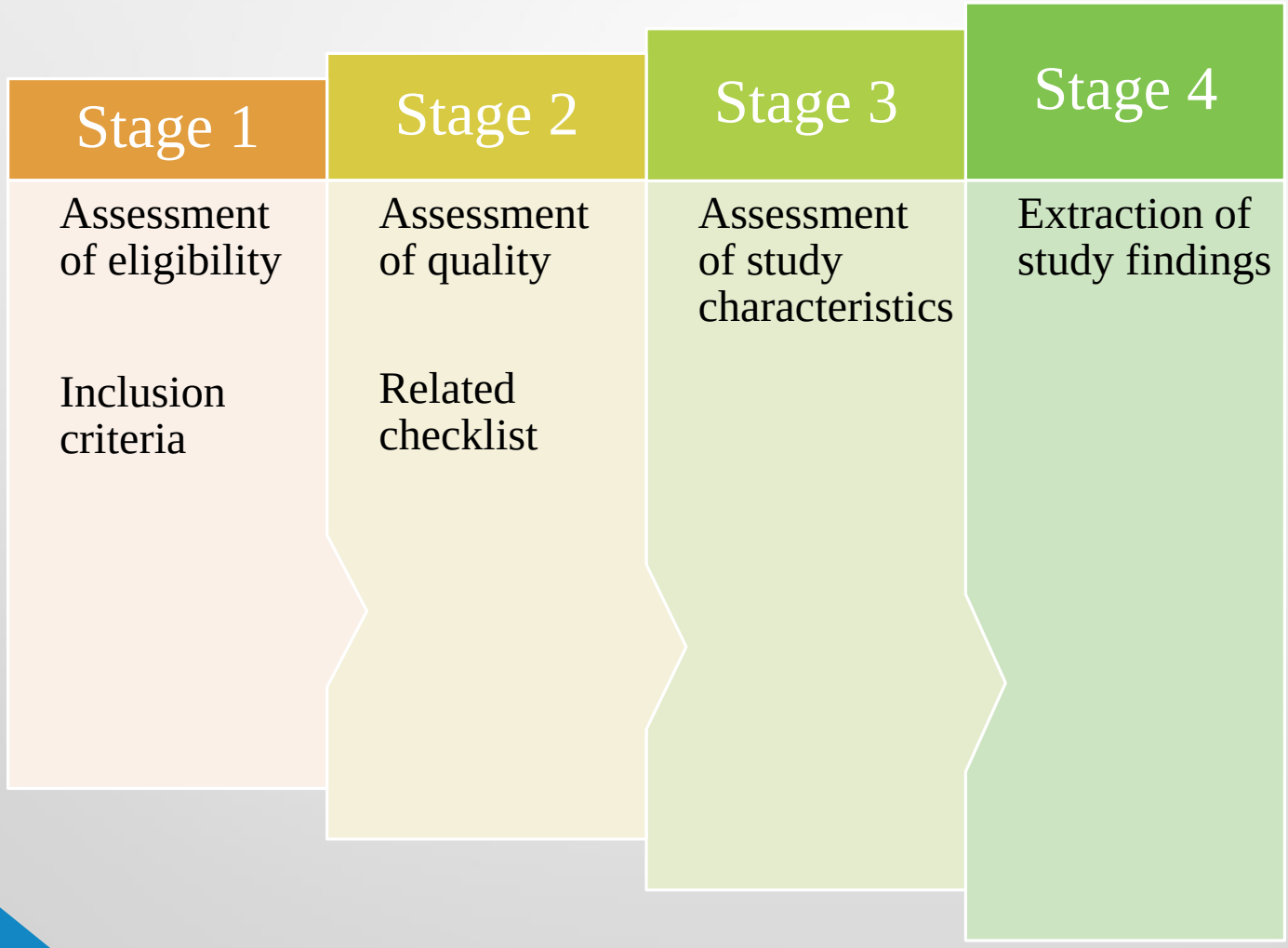
- Independent extraction of data by at least two experienced reviewers is ideal but is also resource intensive.
- There is a tradeoff between cost and the quality of data extraction.
  - Data extraction often takes longer than 2 hours per paper.
  - A reduction in the scope of the work may be necessary if independent data extraction is desired.
- Careful single extraction by experienced reviewer, with or without crosschecking of selected items by a second reviewer, is a good compromise.

The Process of Data Extraction

- Plan data collection in advance
  - Data – “any information about (or deriving from) a study, including details of methods, participants, setting, context, interventions, outcomes, results, publications and investigators”
- Develop an extraction strategy
  - Replicable
  - Transparent
- Use resources
  - Checklists and Review Group guidelines
  - Topical systematic reviews and primary studies
  - Data extraction forms
- Pilot test method

Data Extraction

Stages of Data Extraction



Two types of tables in the Evidence-based Practice

- Evidence Tables
  - Essentially are data extraction forms
  - Typically are study specific, with data from each study extracted into a set of such tables
  - Are detailed and typically not included in main reports
- Summary Tables
  - Are used in main reports facilitate the presentation of the synthesis of the studies
  - Typically contain context-relevant pieces of the information included in study-specific evidence tables
  - Address particular research questions

Two types of tables

Summary Tables

Table 1

Characteristics of included studies

Author	Methods	Participants	Intervention		Primary outcome	Evaluation of re-infection
			Gentamicin	Comparator		
Hira et al. (1984) [36]	Quasi-random (treatment assigned to alternate consecutive patients)	Men with uncomplicated gonorrhoea infection (gram-negative diplococci on urethral smear), Lusaka, Zambia	Single-dose gentamicin 280 mg intramuscular injection (n = 302)	Single-dose kanamycin 2 g intramuscular injection (n = 113)	Cure Patients in whom *N. gonorrhoea* persisted or re-appeared (as determined by a positive result of a smear or culture) in the absence of sexual activity during the follow-up period were considered to be treatment failure	All patients advised to abstain from sexual activity for 2 weeks after therapy. Patients excluded if reported sexual activity during 2 weeks follow-up period with or without persistent or re-appearing gonorrhoea on culture
Iskandar et al. (1978) [33]	RCT (randomly allocated to 3 groups of 30 patients)	Men with acute gonorrhoea infection (gonorrhoea on Gram stain of urethral smears), Egypt	Single-dose gentamicin 240 mg intramuscular injection (n = 30)	Co-trimoxazole (Bactrim, Roche) 8 tablets daily divided into 2 doses for 2 days (n = 30). Trimethoprim-sulphametrol (Lidaprim, Ciba) 8 tablets divided into 2 doses for 2 days (n = 30)	Cure Cases with negative smears plus resolution of discharge on day 7 were considered cured	One case of re-infection reported in which there was a history of re-exposure. Safe sex advice and assessment of re-infection not described
Pareek and Chowdhury	Non-randomised,	Men with urethral gonorrhoea infection	Single-dose gentamicin	Single-dose spectinomycin 2 g intramuscular injection (n = 20)	Cure	Safe sex advice, definition and assessment of re-infection not described

Data Extraction

What Data To Collect?

- Use key questions and eligibility criteria as a guide
- Anticipate what data summary tables should include:
  - To describe studies
  - To assess outcomes, risk of bias, and applicability
  - To conduct meta-analyses
- Use the PICOTS framework to choose data elements:
  - Population
  - Intervention (or exposure)
  - Comparator (when applicable)
  - Outcome (remember numerical data)
  - Timing
  - Study design (study setting)

What Data To Collect?

General Information

1. Date
2. Name/ID of person extracting data
3. Report title (title of paper/ abstract/ report that data are extracted from)
4. Report contact details of person extracting data
5. Publication type (e.g. full report, abstract, letter)
6. Study ID (e.g. plus surname of first author and year of study was published e.g. Smith 2001)
7. Country in which the study conducted

What Data To Collect?

Population

- Participant characteristics?
- Total number?
- Sub-group analyses? (Age, Sex, Ethnicity, Country, Co-morbidity, etc.)
- Disease stage
- More specific items may be needed, depending upon the topic.

What Data To Collect?
Intervention

- Intervention or exposure and comparator items depend upon the extracted study
- Study types include randomized trial, observational study, diagnostic test study, prognostic factor study, …
  - Number of int. or expo. groups?
  - Routes of delivery? (Oral or intravenous, Surgical technique, etc.)
  - Dosage? (Amount, Intensity, Frequency, etc.)
  - Timing? (Between diagnosis and treatment, etc.)
  - Length of treatment?

What Data To Collect?

Outcome

- Outcomes should be determined a priori with the Technical Expert Panel
- Definition? (Diagnostic method, scale, threshold, or behaviour)
- Unit of measurement?
- Scale interpretation? (Upper/Lower limits; Favorable outcomes, etc.)
- Method of collection?
- Number of outcomes reported?

What Data To Collect?

Type of outcomes

- Dichotomous variables (e.g., deaths, patients with at least one stroke)
- Count data (e.g., number of strokes, counting multiple ones)
- Continuous variables (e.g., mm Hg, pain score)
- Survival data
- Sensitivity, specificity, receiver operating characteristic
- Correlations
- Slopes

What Data To Collect?

Comparing two groups

Prospective studies

- Events and sample size in each group
- Non-events and sample size in each group
- Events and non-events in each group
- Event rate and sample size in each group
- Chi-squared and total sample size

Retrospective studies

- Exposed and unexposed for cases and controls
- Exposed and total for cases and controls
- Proportion exposed and total for cases and controls

What Data To Collect?

Comparing two groups

Effect size

- Odds ratio and confidence limits
- Log odds ratio and standard error
- Log odds ratio and variance
- Peto's (O-E) and V
- Risk ratio and confidence limits
- Log risk ratio and standard error
- Log risk ratio and variance
- Risk difference and confidence limits
- Risk difference and standard error
- Risk difference and variance

What Data To Collect?

Comparing two groups

Continuous outcome for post test

- Mean, SD and sample size in each group
- Difference in means, common SD, and sample size
- Cohen's d (standardized by pooled within-groups SD) and sample size
- Means, sample size, and t-value
- Difference in means, sample size, and t-value
- Sample size and t-value
- Means, sample size, and p-value
- Difference in means, sample size, and p-value
- Sample size and p-value

Unmatched groups, pre and post data

What Data To Collect?

Comparing two groups

Continuous outcome for pre- and post-test

- Means, SD pre and post, N, in each group, Pre/Post Corr
- Means, SD difference, N, in each group, Pre/Post Corr
- Means pre and post in each group, t within groups, N
- Means pre and post in each group, p within groups, N
- Means pre and post in each group, F for difference between changes, N
- Mean change, SD pre and post, N, in each group, Pre/Post Corr
- Mean change, SD difference, N, in each group, Pre/Post Corr
- Mean change in each group, t within groups, N
- Mean change in each group, p within groups, N
- Mean change in each group, F for difference between changes, N
- F for difference between changes, N

What Data To Collect?

Estimating proportion or mean

- One Group
  - Dichotomous (number of events)
    - Raw data
      - Events and sample size
      - Non-events and sample size
      - Events and non-events
      - Event rate and sample size
  - Continuous (means)
    - Raw data
      - Mean, SD and sample size
    - Computed effect sizes
      - Mean and confidence limits
      - Mean and standard error
      - Mean and variance
  - Rates (events by person years)

Types of Data Extraction Forms

Paper Advantages

- convenience or preference;
- data extraction can be undertaken almost anywhere;
- easier to create and implement (no need for computer programming or specialist software);
- provides a permanent record of all manipulations and modifications (providing these manipulations and modifications are not erased); and
- simple comparison of forms completed by different review authors

Electronic Advantages

- combines data extraction and data entry into one step;
- forms may be programmed (e.g. using Microsoft Access) to 'lead' the author through the data collection process, for example, by posing questions that depend on answers to previous questions;
- data from reviews involving large numbers of studies are more easily stored and retrieved;
- allows simple conversions at the time of data extraction (recode)
- rapid comparison of forms completed by different review authors

Data Collection Checklist

Source
•Study ID (created by review author).
•Report ID (created by review author).
•Review author ID (created by review author).
•Citation and contact details.
Eligibility
•Confirm eligibility for review.
•Reason for exclusion.
Methods
•Study design.
•Total study duration.
•Sequence generation*.
•Allocation sequence concealment*.
•Blinding*.
•Other concerns about bias*.
Participants
•Total number.
•Setting.
•Diagnostic criteria.
•Age.
•Sex.
•Country.
•[Co-morbidity].
•[Socio-demographics].
•[Ethnicity].
•[Date of study].
Interventions
•Total number of intervention groups.
For each intervention and comparison group of interest:
•Specific intervention.
• Intervention details (sufficient for replication, if feasible).
• [Integrity of intervention].

Outcomes
•Outcomes and time points (i) collected; (ii) reported*.
For each outcome of interest:
•Outcome definition (with diagnostic criteria if relevant).
•Unit of measurement (if relevant).
•For scales: upper and lower limits, and whether high or low score is good.
Results
•Number of participants allocated to each intervention group.
For each outcome of interest:
•Sample size.
•Missing participants*.
•Summary data for each intervention group (e.g. 2×2 table for dichotomous data; means and SDs for continuous data).
•[Estimate of effect with confidence interval; P value].
•[Subgroup analyses].
Miscellaneous
•Funding source.
•Key conclusions of the study authors.
•Miscellaneous comments from the study authors.
•References to other relevant studies.
•Correspondence required.
•Miscellaneous comments by the review authors.

Data Extraction Form Examples

- Sample Data Extraction Forms:
  - Petticrew and Roberts Systematic Reviews in the Social Sciences: A Practical Guide: http://onlinelibrary.wiley.com/doi/10.1002/9780470754887.app4/pdf
  - Cochrane Schizophrenia Group: http://www.cochrane-net.org/openlearning/Other/Example_from_a_review_in_schizophrenia.pdf
  - https://training.cochrane.org/sites/training.cochrane.org/files/public/uploads/resources/downloadable_resources/English/Collecting%20data%20-%20form%20for%20RCTs%20only.doc.

Study	Setting	Sample	Subgroups	Measurement Tools	Analyses	Results
Al-Krenawi & Slonim-Nevo, 2008	The Negev Desert Region, Israel	Quasi-Random Sample; Bedouin Arab Women Recruited by Phone in Conjunction with a Randomized Trial Involving Their Children (N=315)	159 Monogamous and 156 Senior Wives of Two Wife Families	A Socio-demographic Questionnaire, the Self-Esteem scale (SE), the Brief Symptom Inventory (BSI), the McMaster Family Assessment Device (FAD), the ENRICH questionnaire, and the Index of Parental Attitudes	Independent Samples t-test; Multiple Linear Regressions: Full Model and Best Predictive Model	"The findings show that the wives in polygamous marriages suffered from more psychological difficulties than those in monogamous marriages" with higher levels of somatization, depression, anxiety, hostility, and paranoid ideation ($p<0.05$); greater general severity (GSI) and total number of symptoms ($p<0.05$); lower self-esteem, less marital satisfaction, and more problematic family functioning ($p<0.001$). Family functioning was the strongest and most consistent predictor of mental distress; when family functioning was held as a dependent variable, polygamy was only significantly predictive of self-esteem ($p<0.001$)

References

- Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org
- Social Care Institute for Excellence (2006). *The conduct of systematic research reviews for SCIE knowledge reviews. Available from http://www.scie.org.uk/publications/researchresources/rr01.pdf*


Thank You