Unit Nine: Synthesising the Evidence

Learning Objectives

- To understand the different methods available for synthesising evidence
- To understand the terms: meta-analysis, confidence interval, heterogeneity, odds ratio, relative risk, narrative synthesis

Generally, there are two approaches to synthesising the findings from a range of studies:

**Narrative synthesis** – findings are summarised and explained in words

**Quantitative/statistical synthesis** – data from individual studies are combined statistically and then summarised

The Cochrane Reviewers’ Handbook\(^1\) suggests the following framework for synthesis of primary studies (regardless of the method (narrative/meta-analysis) used to synthesise data):

- What is the direction of the effect?
- What is the size of the effect?
- Is the effect consistent across studies?
- What is the strength of evidence for the effect?

Before deciding which synthesis approach to use it is important to tabulate the findings from the studies. This aids the reviewer in assessing whether studies are likely to be homogenous or heterogenous, and tables greatly assist the reader in eyeballing the types of studies that were included in the review. Reviewers should determine which information should be tabulated; some examples are provided below:

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Intervention details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison details</td>
<td>Theoretical basis</td>
<td>Study design</td>
</tr>
<tr>
<td>Quality assessment</td>
<td>Outcomes</td>
<td>Setting/context (incl. country)</td>
</tr>
<tr>
<td>Population characteristics</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example: An example of tabulating studies can be found in the following systematic review:


The choice of analysis usually depends on the diversity of studies included in the review. Diversity of studies is often referred to as ‘heterogeneity’. Because some reviews may include studies that differ in such characteristics as design, methods, or outcome measures, a quantitative synthesis of studies is not always appropriate or meaningful.

Is there heterogeneity?
- No → Meta-analysis
- Yes → Narrative synthesis → Deal with heterogeneity (eg. subgroup analyses)
Where studies are more homogenous, i.e., we can compare like with like, it may be appropriate to combine the individual results using a meta-analysis. If the results are similar from study to study we can feel more comfortable that a meta-analysis is warranted. Heterogeneity can be determined by presenting the results graphically and examining the overlap of confidence intervals (CI) (if CI overlap studies are more likely to be homogenous) and by calculating a statistical measure of heterogeneity. Both of these methods are further outlined in Chapter Eight of the Cochrane Reviewers’ Handbook (Analysing and presenting results).

Meta-analysis produces a weighted summary result (more weight given to larger studies). By combining results from more than one study it has the advantage of increasing statistical power (which is often inadequate in studies with a small sample size). The final estimate is usually in the form of an odds ratio: the ratio of the probability of an event happening to that of it not happening. The odds ratio is often expressed together with a confidence interval (CI). A confidence interval is a statement of the range within which the true odds ratio lies - within a given degree of assurance (eg. usually estimates of effect like odds ratios are presented with a 95% confidence interval).

Guidelines for narrative synthesis are not yet available, although research is currently underway to develop guidelines for systematic reviews. Ideally, the reviewer should:
- Describe studies
- Assess whether quality is adequate in primary studies to trust the results
- Demonstrate absence of data for planned comparisons
- Demonstrate degree of heterogeneity
- Stratify results by populations, interventions, settings, context, outcomes, validity (if appropriate)

Example: A number of Cochrane systematic reviews of health promotion and public health topics synthesise the results narratively. Visit The Cochrane Library to read examples. Another example can be found in the following article: Riemsma RB, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, Walker A. Systematic review of the effectiveness of stage based interventions to promote smoking cessation. BMJ 2003;326:1175-77.

**Integrating qualitative and quantitative data**

The Evidence for Policy and Practice Information and Co-ordinating Centre has developed methods for synthesising the findings from diverse types of studies within one review3. These methods involve conducting three types of syntheses in the same review: 1) a statistical meta-analysis to pool trials of interventions tackling particular problems (or a narrative synthesis when meta-analysis is not appropriate or possible); 2) a synthesis of studies examining people’s perspectives or experiences of that problem using qualitative analysis (‘views’ studies); and 3) a ‘mixed methods’ synthesis bringing the products of 1) and 2) together. These developments have been driven by particular review questions rather than methodology; ‘users’ of the reviews want to know about the effects of interventions, but also want to know which interventions will be most appropriate and relevant to people. However, they do illustrate how qualitative studies can be integrated into a systematic review as ‘views’ studies are often, but not always, qualitative in nature. The methods for each of the three syntheses are described in brief below:

Synthesis 1) Effectiveness synthesis for trials
Effect sizes from good quality trials are extracted and, if appropriate, pooled using statistical meta-analysis. Heterogeneity is explored statistically by carrying out sub-group analyses on a range of categories specified in advance (eg. study quality, study design, setting and type of intervention).
Synthesis 2) Qualitative synthesis for ‘views’ studies
The textual data describing the findings from ‘views’ studies are copied verbatim and entered into a software package to aid qualitative analysis. Two or more reviewers undertake a thematic analysis on this data. Themes are descriptive and stay close to the data, building up a picture of the range and depth of people’s perspectives and experiences in relation to the health issue under study. The content of the descriptive themes are then considered in the light of the relevant review question (eg. what helps and what stops children eating fruit and vegetables?) in order to generate implications for intervention development. The products of this kind of synthesis can be conceptualised as ‘theories’ about which interventions might work. These theories are grounded in people’s own understandings about their lives and health. These synthesis methods have much in common with the work of others who have emphasised the theory building potential of synthesis.4

Synthesis 3) A ‘mixed methods’ synthesis
Implications for interventions are juxtaposed against the interventions which have been evaluated by trials included in Synthesis 1. Using the descriptions of the interventions provided in the reports of the trials, matches, miss-matches and gaps are identified. Gaps are used for recommending what kinds of interventions need to be newly developed and tested. The effect sizes from interventions which matched implications for interventions derived from people’s views can be compared to those which do not, using sub-group analysis. This provides a way to highlight which types of interventions are both effective and appropriate. Unlike Bayesian methods, another approach to combining ‘qualitative’ and ‘quantitative’ studies within systematic reviews which translates textual data into numerical data, these methods integrate ‘quantitative’ estimates of benefit and harm with ‘qualitative’ understanding from people’s lives, whilst preserving the unique contribution of each.3

REFERENCES


EXERCISE

1. Read the methods and results section of the article: “Riemsma RB, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, Walker A. Systematic review of the effectiveness of stage based interventions to promote smoking cessation. BMJ 2003;326:1175-77”. Look at the type of narrative synthesis of results and contrast with the meta-analysis approach.
Synthesising the evidence

**Steps**

1. **Table of study data**
2. **Check for heterogeneity**
   - No – meta-analysis
   - Yes – identify factors, subgroup analysis or narrative synthesis
3. **Sensitivity analyses**
4. **Explore publication bias**

**Step One:**

1. Table of study data
   - Year
   - Setting
   - Population details (including any baseline differences)
   - Study design
   - Intervention details (including theory)
   - Control group details
   - Results
   - Study quality

**Table:**

<table>
<thead>
<tr>
<th>Study</th>
<th>Setting</th>
<th>Setting details</th>
<th>Sample size and characteristics</th>
<th>Unit of randomisation and analysis</th>
<th>Sample size and characteristics</th>
<th>Length of follow-up</th>
<th>Meta-analysis</th>
<th>Study quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aarons et al. 2000</td>
<td>Junior High Schools, Washington DC</td>
<td>3869 grade 9 students, mean age 15 years, 53% female, 31% white, 27% Hispanic, 16% African American</td>
<td>Grade 9, 10 lessons for grade 9 and 10; lessons for grade 10 on knowledge and skills and led by trained peers and teachers</td>
<td>Group randomisation</td>
<td>School Analysis, Individual</td>
<td>3869 grade 9 students, mean age 15 years, 53% female, 31% white, 27% Hispanic, 16% African American</td>
<td>Individual analysis at 3 months after intervention</td>
<td>Social Cognitive Theory</td>
</tr>
<tr>
<td>Coyle et al. 2004</td>
<td>Junior High Schools, Texas and California</td>
<td>684 grade 7 students, mean age 12.8 years, 52% female, 84% African-American, 13% low socioeconomic status</td>
<td>Junior High School, 582 grade 7 students, mean age 12.8 years, 52% female, 84% African-American, 13% low socioeconomic status</td>
<td>Group randomisation</td>
<td>School Analysis, Individual</td>
<td>684 grade 7 students, mean age 12.8 years, 52% female, 84% African-American, 13% low socioeconomic status</td>
<td>Group analysis at 3 months after intervention</td>
<td>Social Learning Theory</td>
</tr>
</tbody>
</table>

**Step Two:**

**Check for heterogeneity**

- Are the results consistent?

**Yes**

- **Meta-analysis**

**No**

- **Narrative synthesis or subgroup analysis**

**Explain causes of heterogeneity**

---

**Terminology**

**Homogeneity** = similar

Homogenous studies - if their results vary no more than might be expected by the play of chance (opposite = heterogeneity)

---

**Investigating heterogeneity**

Graphically:
- If homogenous studies -
  - Point estimates on same side of line of unity
  - CI should overlap to a large extent
  - Lack of outliers

---

**Heterogeneity - the eyeball test**

Investigating heterogeneity

Statistically:
- $p<0.1$ would indicate heterogeneity
- But test has low power when there are a few studies
- Lack of statistical significance does not imply homogeneity

---

**Sources of heterogeneity**

Examples:
- Populations
- Interventions
- Outcomes
- Study designs
- Study quality

Need to identify which factors contribute to heterogeneity

---

**Identifying and dealing with heterogeneity**

Subgroup analyses

By gender, age group, quality, type of intervention.....but keep analyses to a minimum!
Dealing with heterogeneity

Not all systematic reviews are meta-analyses

“...it is always appropriate and desirable to systematically review a body of data, but it may sometimes be inappropriate, or even misleading, to statistically pool results from separate studies. Indeed, it is our impression that reviewers often find it hard to resist the temptation of combining studies even when such meta-analysis is questionable or clearly inappropriate.”


Yes – heterogeneity present

Narrative synthesis

- Describes studies
- Assesses whether quality is adequate in primary studies to trust their results
- Demonstrates absence of data for planned comparisons
- Demonstrates degree of heterogeneity
- Stratify by – populations, interventions, settings, context, outcomes, validity, etc

No – heterogeneity not present

Meta-analysis

- What comparisons should be made?
- What study results should be used in each comparison?
- Are the results of studies similar within each comparison?
- What is the best summary of effect for each comparison?

Cochrane Reviewers’ Handbook

Meta-analysis

- Weighted average of effect sizes
- Weighted by study size, events

Study results

- For dichotomous/binary outcomes (Y/N) use: Relative Risk or Odds Ratio

Risk = \frac{\text{number of events}}{\text{total number of observations}}

Odds = \frac{\text{number of events}}{\text{number without the event}}
Relative risk and Odds ratio

RR – Risk of the event in one group divided by the risk in the other group.

OR – Odds of the event occurring in one group divided by the odds occurring in the other group.

Let’s try!

<table>
<thead>
<tr>
<th></th>
<th>ABC</th>
<th>No ABC</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention Group</td>
<td>2</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Control Group</td>
<td>4</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>121</td>
<td>127</td>
</tr>
</tbody>
</table>

Calculation

Relative risk

\[
\frac{a/(a+b)}{c/(c+d)} = \frac{2/(2+62)}{4/(4+59)} = 0.49
\]

The risk of developing ABC was 49% of the risk in the control group.

The intervention reduced the risk by 51% of what it was in the control group.

Calculation

Odds ratio

\[
\frac{a/b}{c/d} = \frac{2/62}{4/59} = 0.48
\]

The intervention reduced the odds of having ABC by about 50% of what they were.

Odds Ratio Graph
**Odds Ratio - with pooled effect size**

- Best/point estimate
- Confidence Interval

Is the range within which the true size of effect (never exactly known) lies, with a given degree of assurance (usually 95%).

**Confidence Interval (CI)**

- How ‘confident’ are we that the results are a true reflection of the actual effect/phenomena?
  - the shorter the CI the more certain we can be about the results
  - if it crosses the line of unity (no treatment effect) the intervention might not be doing any good and could be doing harm

**The p-value in a nutshell**

- Could the result have occurred by chance?
  - The result is unlikely to be due to chance
  - The result is likely to be due to chance

  - p < 0.05  
    - a statistically significant result
    - for 1 in 20 result fairly unlikely to be due to chance
  - p > 0.05  
    - not a statistically significant result
    - p = 0.05
    - for 1 in 2 result quite likely to be due to chance

**Continuous data**

- Data which is normally presented with means and SDs (i.e. height, BMI)
- For each study you need means and SDs to calculate difference
- Difficult if continuous data arise from different scales

**Statisticians are your best friend!**
Statistical software for Meta-analysis

The meta-analysis

Steps

Step Three:
Sensitivity analysis
- How sensitive are the results of the analysis to changes in the way it was done?
- Changing inclusion criteria for types of studies
- Including or excluding studies where there is ambiguity
- Reanalysing the data imputing using a reasonable range of values for missing data
- Reanalysing the data using different statistical approaches

Steps

Step Four:
Explore publication bias
- Is there a possibility I have missed some studies?

Sensitivity analysis

Publication bias
- Funnel plot
- Studies with significant results are more likely to be
  - Published
  - Published in English
  - Cited by others
Funnel plots

No publication bias = symmetrical inverted funnel
Effect size vs. sample size
i.e. Smaller studies without statistically significant effects remain unpublished, gap in bottom corner of graph

Synthesis of qualitative research

- In its infancy
- Widely varying theoretical perspectives
- Unit of analysis is concept/theme
- Secondary summary of research
- Most developed method is meta-ethnography
- Help is around the corner - many research projects in progress!!