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 *(meta-analysis)* 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):
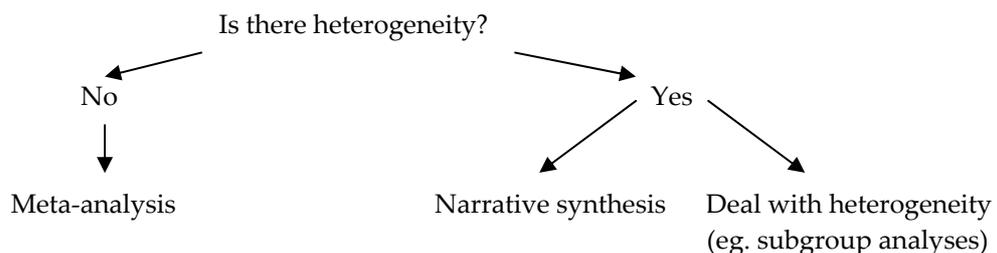
- What is the direction of the effect?
- What is the size of the effect?
- Is the effect consistent across studies?
- What is the strength of evidence for the effect?

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:

Authors	Year	Intervention details
Comparison details	Theoretical basis	Study design
Quality assessment	Outcomes	Setting/context (incl. country)
Population characteristics		

Example: An example of tabulating studies can be found in the following systematic review:
DiCenso A, Guyatt G, Willan A, Griffith L. Interventions to reduce unintended pregnancies among adolescents: systematic review of randomised controlled trials. BMJ. 2002 Jun 15;324(7351):1426.

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

Yes → 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[2]:
- 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 review[3]. 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

1. Clarke M, Oxman AD, editors. Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]. http://www.cochrane.dk/cochrane/handbook/handbook.htm

2. Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for those Carrying Out or Commissioning Reviews. CRD Report Number 4 (2nd Edition). NHS Centre for Reviews and Dissemination, University of York. March 2001. http://www.york.ac.uk/inst/crd/report4.htm

3. Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, Brunton G, Kavanagh J. Integrating qualitative research with trials in systematic reviews. BMJ 2004;328:1010-2.

4. Harden A, Garcia J, Oliver S, Rees R, Shepherd J, Brunton G, Oakley A. Applying systematic review methods to studies of people's views: an example from public health research. J Epidemiol Community Health. 2004 Sep;58(9):794-800.

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
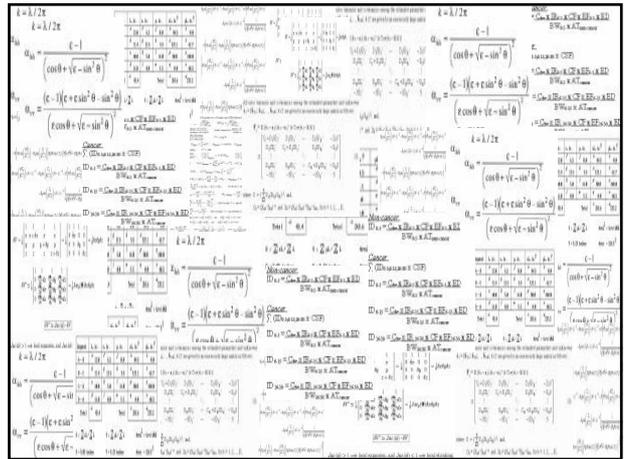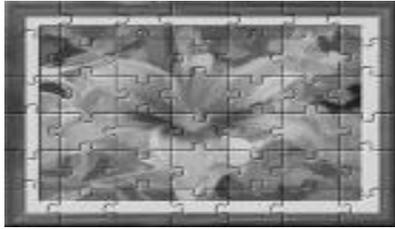
EBPH Steps

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

EBPH Steps

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

Study	Setting	Sample size and characteristics	Unit of randomisation and analysis	Theory	Intervention	Length of follow-up	Outcome	Baseline characteristics
Aarons et al 2000	6 Junior High Schools Washington DC	582 grade 7 students Mean age 12.8 years 52% female, 84% African-American, 13% low socioeconomic status	Randomisation: School Analysis: Individual	Social Cognitive Theory	3 reproductive health lessons taught by health professionals, 5 sessions of postponing sexual involvement curriculum. Control group: conventional programme	3 months, 96.4% followed	Intercourse. Use of birth control at last intercourse	Favour intervention
Coyle et al 2001	20 urban high schools Texas and California	3869 grade 9 students, mean age 15 years, 53% female, 31% white, 27% Hispanic, 16% African American	Randomisation: School Analysis: Adjusted	Social Learning Theory	Safe choices: 10 lessons for grade 9 and 10: lessons for grade 10 on knowledge and skills and led by trained peers and teachers	31 months, 79% followed	Intercourse. Use of birth control at last intercourse	Favour control

DiCenso A, Guyatt G, Willan A, Griffith L. Interventions to reduce unintended pregnancies among adolescents: systematic review of randomised controlled trials. BMJ 2002;324:1426-34

EBPH Steps

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)
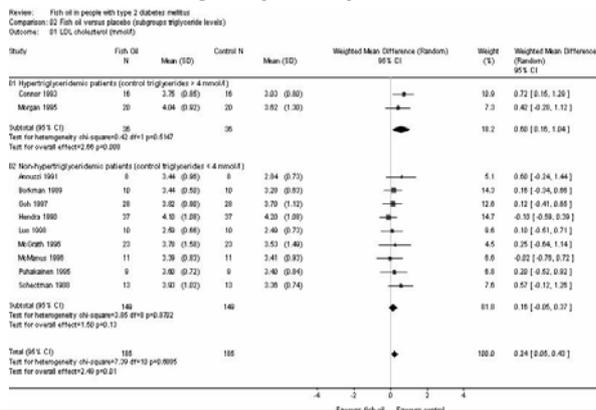
Investigating heterogeneity

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
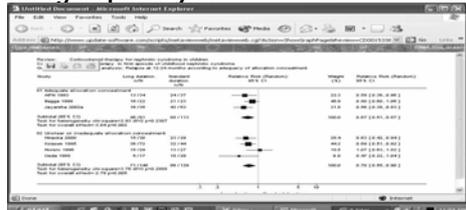
Sources of heterogeneity

Examples:

- Populations
- Interventions
- Outcomes
- Study designs
- Study quality
  Need to identify which factors contribute to heterogeneity
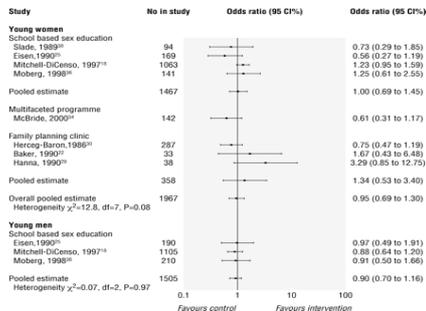
Identifying and dealing with heterogeneity

Subgroup analyses

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

Dealing with heterogeneity

Study	No in study	Odds ratio (95 CI%)	Odds ratio (95 CI%)
Young women			
School based sex education			
Slade, 1989[35]	94		0.73 (0.29 to 1.85)
Eisen,1990[15]	169		0.56 (0.27 to 1.19)
Mitchell-DiCenso, 1997[18]	1063		1.23 (0.95 to 1.59)
Moberg, 1998[36]	141		1.25 (0.61 to 2.55)
Pooled estimate	1467		1.00 (0.69 to 1.45)
Multifaceted programme			
McBride, 2000[34]	142		0.61 (0.31 to 1.17)
Family planning clinic			
Herceg-Baron,1986[30]	287		0.75 (0.47 to 1.19)
Baker, 1990[22]	33		1.67 (0.43 to 6.48)
Hanna, 1990[35]	38		3.29 (0.85 to 12.75)
Pooled estimate	358		1.34 (0.53 to 3.40)
Overall pooled estimate	1967		0.95 (0.69 to 1.30)
Heterogeneity χ^2=12.8, df=7, P=0.08			
Young men			
School based sex education			
Eisen,1990[15]	190		0.97 (0.49 to 1.91)
Mitchell-DiCenso, 1997[18]	1105		0.88 (0.64 to 1.20)
Moberg, 1998[36]	210		0.91 (0.50 to 1.66)
Pooled estimate	1505		0.90 (0.70 to 1.16)
Heterogeneity χ^2=0.07, df=2, P=0.97		0.1 1 10 100	
		Favours control Favours intervention	

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."

Egger et al. Systematic reviews in health care. London: BMJ Books, 2001:5

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

Adapted from NHS CRD Report No. 4, 2nd Ed. www.york.ac.uk/inst/crd/report4.htm

---

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 = $\dfrac{\text{number of events}}{\text{total number of observations}}$
- Odds = $\dfrac{\text{number of events}}{\text{number without the event}}$

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!

	ABC	No ABC	Total
Intervention Group	2 **a**	62 **b**	64
Control Group	4 **c**	59 **d**	63
Total	9	121	127

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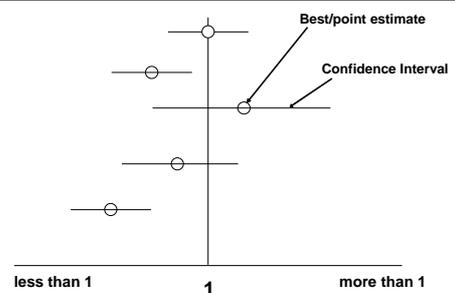- The intervention reduced the risk by 51% of what it was in the control group

Calculation

- Odds ratio
  - $\frac{a/b}{c/d}$ $\frac{2/62}{4/59}$ = 0.48
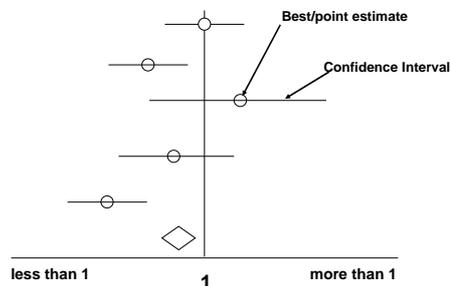- The intervention reduced the odds of having ABC by about 50% of what they were

Odds Ratio Graph
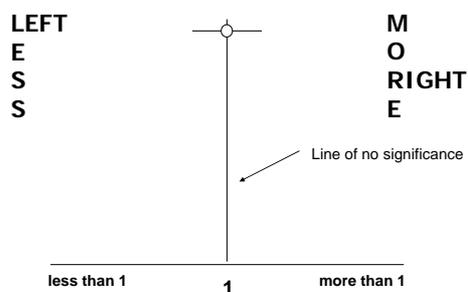
Odds Ratio
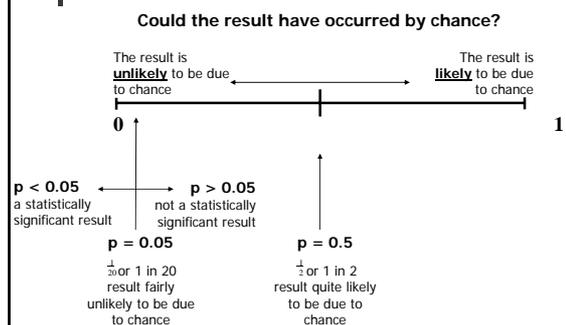
Odds Ratio – with pooled effect size

Best/point estimate

Confidence Interval

less than 1   **1**   more than 1

Confidence Interval (CI)

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

0   **1**

$p < 0.05$
a statistically significant result

$p > 0.05$
not a statistically significant result

$p = 0.05$

$\frac{1}{20}$ or 1 in 20 result fairly unlikely to be due to chance

$p = 0.5$

$\frac{1}{2}$ or 1 in 2 result quite likely to be due to chance
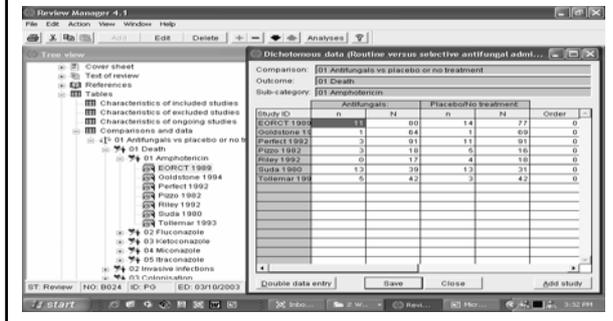

Continuous data

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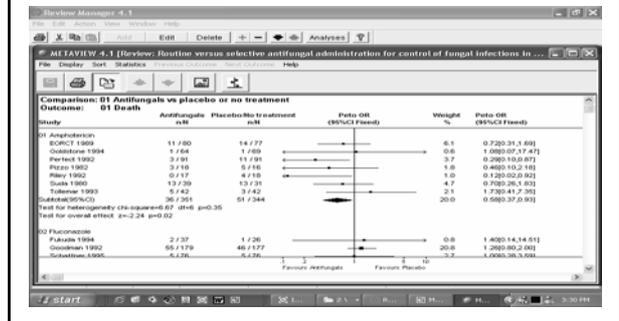- Difficult if continuous data arise from different scales

---

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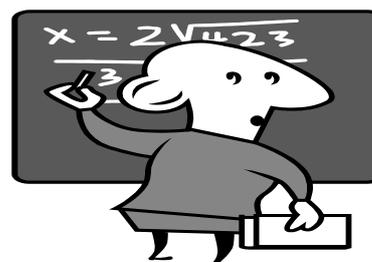
Statistical software for Meta-analysis

The meta-analysis

Steps

Step Three:

Sensitivity analysis

- How sensitive are the results of the analysis to changes in the way it was done?

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
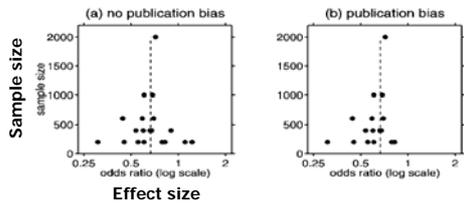 - Reanalysing the data using different statistical approaches

Steps

Step Four:

Explore publication bias

- Is there a possibility I have missed some studies?

Publication bias

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

Funnel plots

(a) no publication bias (b) publication bias

Sample size — sample size 2000, 1500, 1000, 500, 0; odds ratio (log scale) 0.25, 0.5, 1, 2

Effect size

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!!