Cochrane Training

Synthesising and presenting results using other methods

Trusted evidence.
Informed decisions.
Better health.

Cochrane Training

Steps in a Cochrane Review

1. define the question
2. plan eligibility criteria
3. plan methods
4. search for studies
5. apply eligibility criteria
6. collect data
7. assess studies for risk of bias
8. **analyse and present results**
9. interpret results and draw conclusions
10. improve and update review

Cochrane Training

Session outline

- **Introduction**
- Reasons for not undertaking a meta-analysis of effect estimates
- Structured summaries
- Acceptable synthesis methods and accompanying visual displays
- What to include in the protocol
- Take home messages

Cochrane Training

An overview of available methods for summary and synthesis*

Table 9.5.a Overview of available methods for summary and synthesis

	Summary	Acceptable statistical synthesis methods			Preferred statistical synthesis methods		
Methods	Text/Tabular	Vote counting	Combining P values	Summary of effect estimates	Pairwise meta-analysis	Network meta-analysis	Subgroup analysis/ meta-regression
Questions addressed	Structured summary of evidence presented in either text of tabular form	Is there any evidence of an effect?	Is there evidence that there is an effect in at least one study?	What is the range and distribution of observed effects?	What is the common intervention effect? (fixed-effect model) What is the average intervention effect? (random effects model)	Which intervention of multiple is most effective?	What factors modify the magnitude of the intervention effects?
Example plots	Forest plot (plotting individual study effects without a combined effect estimate)	Harvest plot Effect direction plot	Albatross plot	Box and whisker plot Bubble plot	Forest plot	Forest plot Network diagram Rankogram plots	Forest plot Box and whisker plot Bubble plot

* Table 9.5.a (modified) from Chapter 9, 2019 Cochrane Handbook

Cochrane Training

Session outline

- Introduction
- **Reasons for not undertaking a meta-analysis of effect estimates**
- Structured summaries
- Acceptable synthesis methods and accompanying visual displays
- What to include in the protocol
- Take home message

Cochrane Training

Reasons why meta-analysis of effect estimates may not be possible

Legitimate reasons	Commonly cited reasons where meta-analysis should be considered
1. Limited evidence for a pre-specified comparison (one or no studies)	**5. Clinical or methodological diversity** (consider modifying comparisons with rationale)
2. Incompletely reported outcome or effect estimate	**6. Statistical heterogeneity** (attempt to reduce or explain; present prediction intervals)
3. Different effect measures across studies (e.g. some dichotomize time-to-event, others do not, leading to hazard ratios and risk ratios)	
4. Bias in the evidence (missing studies, missing outcomes, or bias in studies)	

Arguments against using meta-analysis because of too much diversity equally apply to other synthesis approaches.

Scenarios reviewers may encounter

Consider a single study and outcome

Estimate of intervention effect	Variance of the estimate	Direction of effect (benefit or harm)	Statistical significance or P-value	Example of reporting	Can be included in a meta-analysis of effect estimates?
✓	✓	(✓)	(✓)	"The intervention had a small effect on anxiety levels compared with the control (**−0.52 (95%CI −1.34 to 0.30**); scale 0-21, **higher scores indicate greater anxiety**)."	Yes

Cochrane Training

Scenarios reviewers may encounter

Consider a single study and outcome

Estimate of intervention effect	Variance of the estimate	Direction of effect (benefit or harm)	Statistical significance or P-value	Example of reporting	Can be included in a meta-analysis of effect estimates?
✓	✓	(✓)	(✓)	"The intervention had a small effect on anxiety levels compared with the control (**−0.52 (95%CI −1.34 to 0.30**); scale 0-21, **higher scores indicate greater anxiety**)."	Yes
✓		(✓)		"The intervention **reduced anxiety levels by -0.52** compared with the control (scale 0-21, **higher scores indicate greater anxiety**)."	No

Cochrane Training

Scenarios reviewers may encounter

Consider a single study and outcome

Estimate of intervention effect	Variance of the estimate	Direction of effect (benefit or harm)	Statistical significance or P-value	Example of reporting	Can be included in a meta-analysis of effect estimates?
✓	✓	(✓)	(✓)	"The intervention had a small effect on anxiety levels compared with the control (**−0.52 (95%CI −1.34 to 0.30**); scale 0-21, **higher scores indicate greater anxiety**)."	Yes
✓		(✓)		"The intervention **reduced anxiety levels by -0.52** compared with the control (scale 0-21, **higher scores indicate greater anxiety**)."	No
		✓	✓	"The **intervention reduced anxiety levels**, but was **not statistically significant**."	No

Cochrane Training

Scenarios reviewers may encounter

Consider a single study and outcome

Estimate of intervention effect	Variance of the estimate	Direction of effect (benefit or harm)	Statistical significance or P-value	Example of reporting	Can be included in a meta-analysis of effect estimates?
✓	✓	(✓)	(✓)	"The intervention had a small effect on anxiety levels compared with the control (**−0.52 (95%CI −1.34 to 0.30**); scale 0-21, **higher scores indicate greater anxiety**)."	Yes
✓		(✓)		"The intervention **reduced anxiety levels by -0.52** compared with the control (scale 0-21, **higher scores indicate greater anxiety**)."	No
		✓	✓	"The **intervention reduced anxiety levels**, but was **not statistically significant**."	No
		✓		**"The intervention reduced anxiety levels."**	No

Cochrane Training

Scenarios reviewers may encounter

Consider a multiple studies and a single outcome

Study	Estimate of intervention effect	Variance of the estimate	Direction of effect (benefit or harm)	Statistical significance or P-value
Study 1			✓	
Study 2			✓	✓
Study 3	✓		(✓)	
Study 4	✓	✓	(✓)	(✓)
Study 5	✓	✓	(✓)	(✓)

Across studies the:

- **completeness of the reported statistics may vary**
- **effect measures could vary (e.g. mix of SMDs, MDs)**

In these circumstances, consideration might be given to using other synthesis methods beyond meta-analysis

Cochrane Training

Session outline

- Introduction
- Reasons for not undertaking a meta-analysis of effect estimates
- **Structured summaries**
- Acceptable synthesis methods and accompanying visual displays
- What to include in the protocol
- Take home messages

Cochrane Training

An overview of available methods for summary and synthesis*

Table 9.5.a Overview of available methods for summary and synthesis

	Summary	Acceptable statistical synthesis methods			Preferred statistical synthesis methods		
Methods	Text/Tabular	Vote counting	Combining P values	Summary of effect estimates	Pairwise meta-analysis	Network meta-analysis	Subgroup analysis/ meta-regression
Questions addressed	Structured summary of evidence presented in either text of tabular form	Is there any evidence of an effect?	Is there evidence that there is an effect in at least one study?	What is the range and distribution of observed effects?	What is the common intervention effect? (fixed-effect model) What is the average intervention effect? (random effects model)	Which intervention of multiple is most effective?	What factors modify the magnitude of the intervention effects?
Example plots	Forest plot (plotting individual study effects without a combined effect estimate)	Harvest plot Effect direction plot	Albatross plot	Box and whisker plot Bubble plot	Forest plot	Forest plot Network diagram Rankogram plots	Forest plot Box and whisker plot Bubble plot

* Table 9.5.a (modified) from Chapter 9, 2019 Cochrane Handbook

Structured summary of effects

Consider ordering studies in text, tables and figures to increase the prominence of the most relevant and trustworthy evidence.

Options include ordering by:

- certainty of the evidence (individual studies if no synthesis)
- risk of bias, study size or study design characteristics
- characteristics that determine how directly a study addresses the review question
  (e.g. relevance and validity of the outcome measures, time of publication)

Cochrane Training

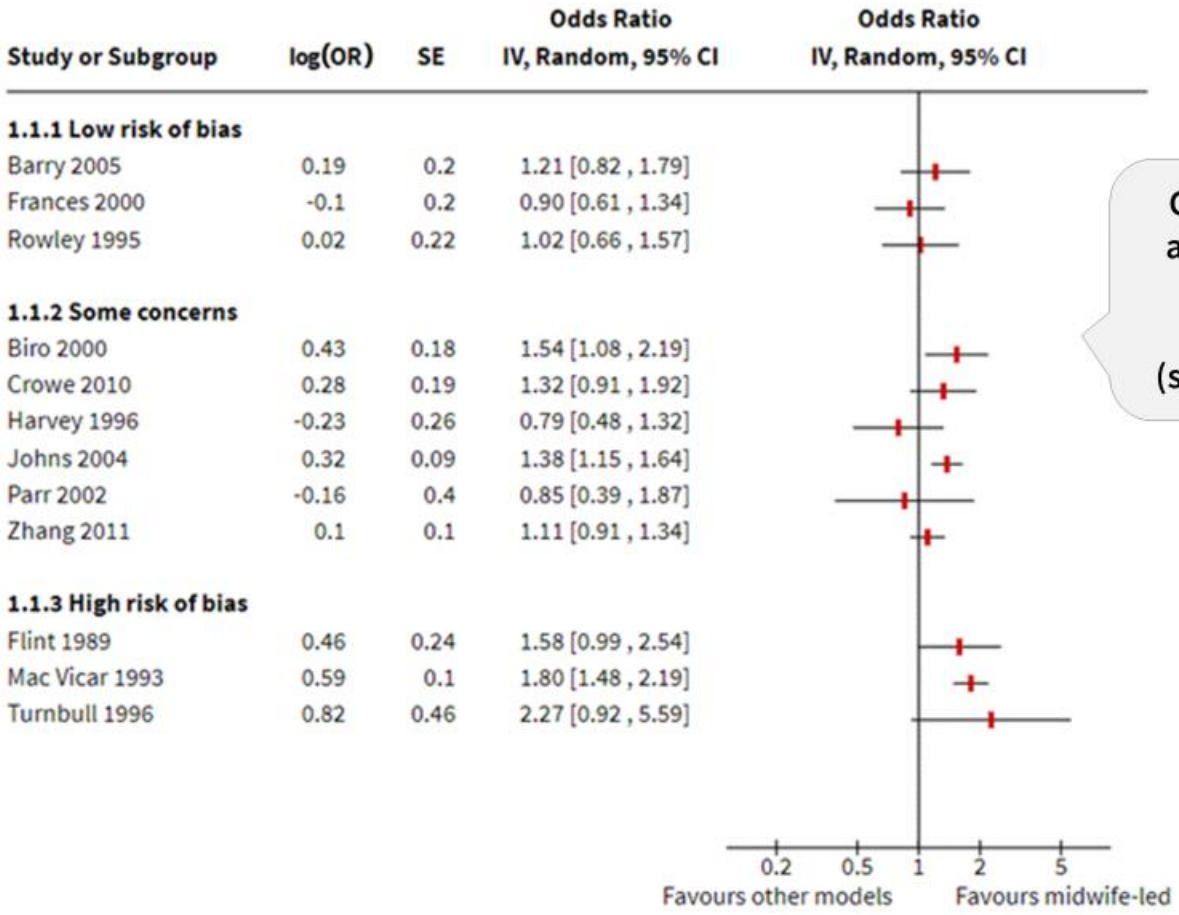
Outcome* (scale details)	Intervention	Control	Mean difference (95% CI)**	Odds ratio (95% CI)†
Low risk of bias				
Barry 2005	n/N	n/N		
Experience of labour	90/246	72/223		1.21 (0.82 to 1.79)
Frances 2000	n/N	n/N		
Communication: labour / birth				0.90 (0.61 to 1.34)
Rowley 1995	n/N	n/N		
Encouraged to ask questions [during labour/birth]				1.02 (0.66 to 1.58)
Some concerns				
Biro 2000	n/N	n/N		
Perception of care: labour / birth	260/344	192/287		1.54 (1.08 to 2.19)
Crowe 2010	Mean (SD) N	Mean (SD) N		
Experience of labour/ birth (0 to 18 points)	9.8 (3.1) 182	9.3 (3.3) 186	0.5 (-0.15 to 1.15)	1.32 (0.91 to 1.92)
Harvey 1996	Mean (SD) N	Mean (SD) N		
Labour & Delivery Satisfaction Index (37 to 222 points)	182 (14.2) 101	185 (30) 93	-3 (-10 to 4)	0.79 (0.48 to 1.32)
Johns 2004	n/N	n/N		
Satisfaction with intrapartum care	605/1163	363/826		1.38 (1.15 to 1.64)
Parr 2002	n/N	n/N		
Experience of childbirth				0.85 (0.39 to 1.87)
Zhang 2011	n/N	n/N		
Perception of care: labour and birth	N = 355	N = 320		POR 1.11 (0.91 to 1.34)
High risk of bias				
Flint 1989	n/N	n/N		
Care from staff during labour	240/275	208/256		1.58 (0.99 to 2.54)
Mac Vicar 1993	n/N	n/N		
Birth satisfaction	849/1163	496/826		1.80 (1.48 to 2.19)
Turnbull 1996	Mean (SD) N	Mean (SD) N		
Intrapartum care rating (-2 to 2 points)	1.2 (0.57) 35	0.93 (0.62) 30	0.27 (-0.03 to 0.57)	2.27 (0.92 to 5.59)

Results ordered by risk of bias to increase prominence of the most 'trustworthy' evidence

Calculate a standardized effect metric (measure)

Aids interpretation

Cochrane Training



Calculating a standardized metric allows results to be presented in a forest plot

(studies ordered by RoB as in table)

No pooled estimate

Figure 12.4.a: Forest plot depicting standardized effect estimates (odds ratios) for satisfaction

Cochrane Training

Structured summary: how to write it up?

Scenario 1. Structured reporting of effects (no synthesis)

> Table 12.4.b and Figure 12.4.a present results for the 12 included studies that reported a measure of maternal satisfaction with care during labour and birth (hereafter 'satisfaction'). Results from these studies were not synthesized for the reasons reported in the data synthesis methods. Here, we summarize results from studies providing high or moderate certainty evidence (based on GRADE) for which results from a valid measure of global satisfaction were available. Barry 2015 found a small increase in satisfaction with midwife-led care compared to obstetrician-led care (4 more women per 100 were satisfied with care; 95% CI 4 fewer to 15 more per 100 women; 469 participants, 1 study; moderate certainty evidence). Harvey 1996 found a small possibly unimportant decrease in satisfaction with midwife-led care compared to obstetrician-led care (3-point reduction on a 185-point LADSI scale, higher scores are more satisfied; 95% CI 10 points lower to 4 higher; 367 participants, 1 study; moderate certainty evidence). The remaining 10 studies reported specific aspects of satisfaction (Frances 2000, Rowley 1995,...), used tools with little or no evidence of validity and reliability (Parr 2002, ...) or provided low or very low certainty evidence (Turnbull 1996, ...).

Cochrane Training

Structured summary: how to write it up?

Scenario 1. Structured reporting of effects (no synthesis)

Table 12.4.b and Figure 12.4.a present results for the 12 included studies that reported a measure of maternal satisfaction with care during labour and birth (hereafter 'satisfaction'). Results from these studies were not synthesized for the reasons reported in the data synthesis methods. Here, we summarize results from studies providing high or moderate certainty evidence (based on GRADE) for which results from a valid measure of global satisfaction were available. Barry 2015 found a small increase in satisfaction with midwife-led care compared to obstetrician-led care (4 more women per 100 were satisfied with care; 95% CI 4 fewer to 15 more per 100 women; 469 participants, 1 study; moderate certainty evidence). Harvey 1996 found a small possibly unimportant decrease in satisfaction with midwife-led care compared to obstetrician-led care (3-point reduction on a 185-point LADSI scale, higher scores are more satisfied; 95% CI 10 points lower to 4 higher; 367 participants, 1 study; moderate certainty evidence). The remaining 10 studies reported specific aspects of satisfaction (Frances 2000, Rowley 1995,...), used tools with little or no evidence of validity and reliability (Parr 2002, ...) or provided low or very low certainty evidence (Turnbull 1996, ...).

Cochrane Training

Structured summary: how to write it up?

Scenario 1. Structured reporting of effects (no synthesis)

Table 12.4.b and Figure 12.4.a present results for the 12 included studies that reported a measure of maternal satisfaction with care during labour and birth (hereafter 'satisfaction'). Results from these studies were not synthesized for the reasons reported in the data synthesis methods. Here, we summarize results from studies providing high or moderate certainty evidence (based on GRADE) for which results from a valid measure of global satisfaction were available. Barry 2015 found a small increase in satisfaction with midwife-led care compared to obstetrician-led care (4 more women per 100 were satisfied with care; 95% CI 4 fewer to 15 more per 100 women; 469 participants, 1 study; moderate certainty evidence). Harvey 1996 found a small possibly unimportant decrease in satisfaction with midwife-led care compared to obstetrician-led care (3-point reduction on a 185-point LADSI scale, higher scores are more satisfied; 95% CI 10 points lower to 4 higher; 367 participants, 1 study; moderate certainty evidence). The remaining 10 studies reported specific aspects of satisfaction (Frances 2000, Rowley 1995,...), used tools with little or no evidence of validity and reliability (Parr 2002, ...) or provided low or very low certainty evidence (Turnbull 1996, ...).

Cochrane Training

Structured summary: how to write it up?

Scenario 1. Structured reporting of effects (no synthesis)

> Table 12.4.b and Figure 12.4.a present results for the 12 included studies that reported a measure of maternal satisfaction with care during labour and birth (hereafter 'satisfaction'). Results from these studies were not synthesized for the reasons reported in the data synthesis methods. Here, we summarize results from studies providing high or moderate certainty evidence (based on GRADE) for which results from a valid measure of global satisfaction were available. Barry 2015 found a small increase in satisfaction with midwife-led care compared to obstetrician-led care (4 more women per 100 were satisfied with care; 95% CI 4 fewer to 15 more per 100 women; 469 participants, 1 study; moderate certainty evidence). Harvey 1996 found a small possibly unimportant decrease in satisfaction with midwife-led care compared to obstetrician-led care (3-point reduction on a 185-point LADSI scale, higher scores are more satisfied; 95% CI 10 points lower to 4 higher; 367 participants, 1 study; moderate certainty evidence). The remaining 10 studies reported specific aspects of satisfaction (Frances 2000, Rowley 1995,...), used tools with little or no evidence of validity and reliability (Parr 2002, ...) or provided low or very low certainty evidence (Turnbull 1996, ...).

Cochrane Training

In any systematic review we are aiming to:

- make best use of available research
- fairly represent the results
- be transparent about the methods used, interpreting results appropriately and acknowledging any limitations of the methods.

Structured summaries may help with this, but it is more challenging to achieve these aims without synthesis (especially without meta-analysis)

Cochrane Training

Session outline

- Introduction
- Reasons for not undertaking a meta-analysis of effect estimates
- Structured summaries
- **Acceptable synthesis methods and accompanying visual displays**
- What to include in the protocol
- Take home messages

Cochrane Training

An overview of available methods for summary and synthesis*

Table 9.5.a Overview of available methods for summary and synthesis

	Summary	Acceptable statistical synthesis methods			Preferred statistical synthesis methods		
Methods	Text/Tabular	Vote counting	Combining P values	Summary of effect estimates	Pairwise meta-analysis	Network meta-analysis	Subgroup analysis/ meta-regression
Questions addressed	Structured summary of evidence presented in either text of tabular form	Is there any evidence of an effect?	Is there evidence that there is an effect in at least one study?	What is the range and distribution of observed effects?	What is the common intervention effect? (fixed-effect model) What is the average intervention effect? (random effects model)	Which intervention of multiple is most effective?	What factors modify the magnitude of the intervention effects?
Example plots	Forest plot (plotting individual study effects without a combined effect estimate)	Harvest plot Effect direction plot	Albatross plot	Box and whisker plot Bubble plot	Forest plot	Forest plot Network diagram Rankogram plots	Forest plot Box and whisker plot Bubble plot

* Table 9.5.a (modified) from Chapter 9, 2019 Cochrane Handbook

Cochrane Training

Introduction to the acceptable synthesis methods

Similar to meta-analysis of effect estimates, the acceptable synthesis methods involve two primary steps:

1. A standardised summary statistic is calculated for each study that quantifies the impact of the intervention
2. The summary statistics are combined across the studies using a mathematical formula

Unlike meta-analysis of effect estimates, the acceptable synthesis methods differ in the data they require, the questions they address and the conclusions and recommendations that can be drawn

Cochrane Training

Synthesis methods: Question addressed and minimum data

Method	Question addressed	Estimate of effect	Variance	Direction of effect	Precise P value
Meta-analysis of effect estimates and extensions	What is the common (or average) effect? What intervention is most effective? What factors modify the magnitude of intervention effects?	✓	✓		
Summarizing effect estimates	What is the range and distribution of observed effects?	✓			
Combining P values	Is there evidence that there is an effect in at least one study?			✓	✓
Vote counting based on direction of effect	Is there any evidence of an effect?			✓	

Summarising effect estimates

Question addressed: What is the range and distribution of observed effects?

Purpose

Can be used to synthesize when difficult to undertake a meta-analysis (e.g. missing variances of effects, unit of analysis errors)

Provides information on the magnitude and range of observed effects

Effects are summarised using descriptive statistics such as the median, interquartile range, and the range

Cochrane Training

Summarising effect estimates

			Scenario 2. Summary statistics	
Study ID	Outcome (scale details*)	Overall RoB judgement	Available data**	Stand. Metric OR (SMD)
Continuous			Mean (SD)	
Crowe 2010	Expectation of labour/birth (0 to 18 points)	Some concerns	Intervention 9.8 (3.1); Control 9.3 (3.3)	1.3 (0.16)
Finn 1997	Experience of labour/ birth (0 to 24 points)	Some concerns	Intervention 21 (5.6); Control 19.7 (7.3)	1.4 (0.20)
Binary				
Frances 2000	Communication: labour / birth	Low	OR 0.90	0.90
Mac Vicar 1993	Birth satisfaction	High	OR 1.80, P < 0.001	1.80
Parr 2002	Experience of childbirth	Some concerns	OR 0.85	0.85
Rowley 1995	Encouraged to ask questions	Low	OR 1.02, NS	1.02
Ordinal				
Zhang 2011	Perception of care: labour / birth	Low	POR 1.10, P > 0.05	1.10

Required data: effect estimates without measures of precision

A measure of variance was not reported and couldn't be calculated for 10 of 15 studies

Standardised estimates (OR)

Summary estimate: median OR and interquartile range

Cochrane Training

Summarising effect estimates: how to write it up?

Scenario 2. Synthesis of summary statistics

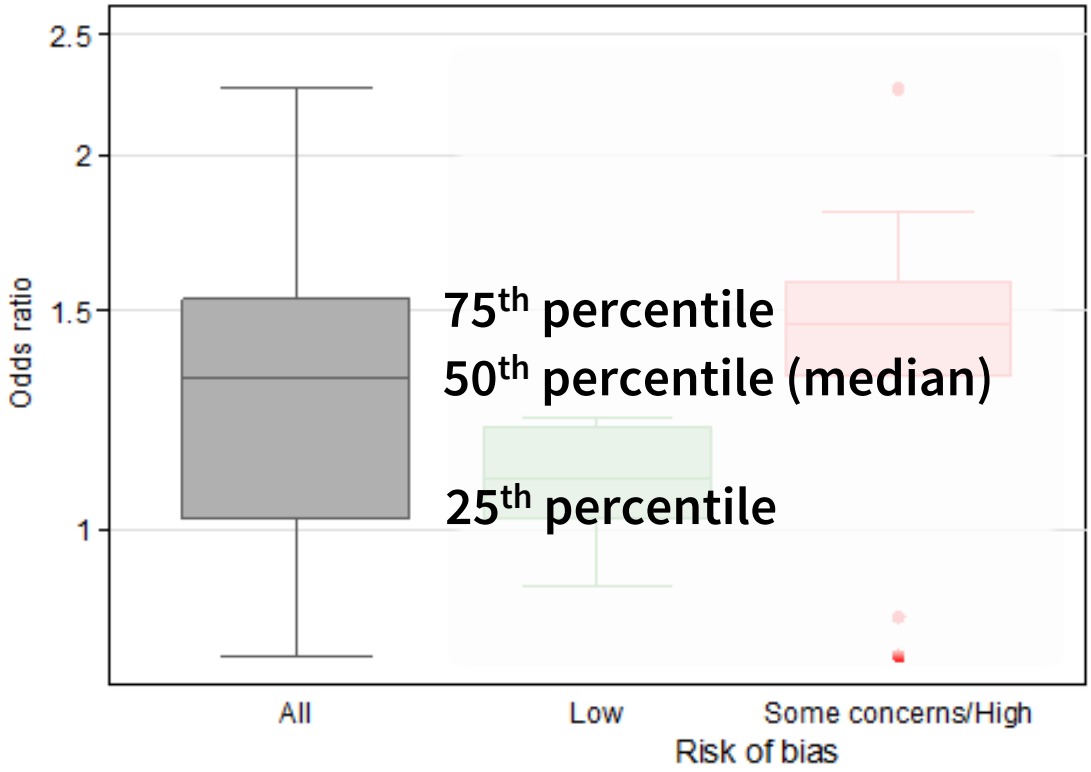
'The median odds ratio of satisfaction was 1.32 for midwife-led models of care compared with other models (interquartile range 1.02 to 1.53; 15 studies). Only five of the 15 effects were judged to be at a low risk of bias, and informal visual examination suggested the size of the odds ratios may be smaller in this group.'

Cochrane Training

Summarising effect estimates: visual display of results

Box-and-whisker plots

(odds ratios for all outcomes & separately by overall RoB)

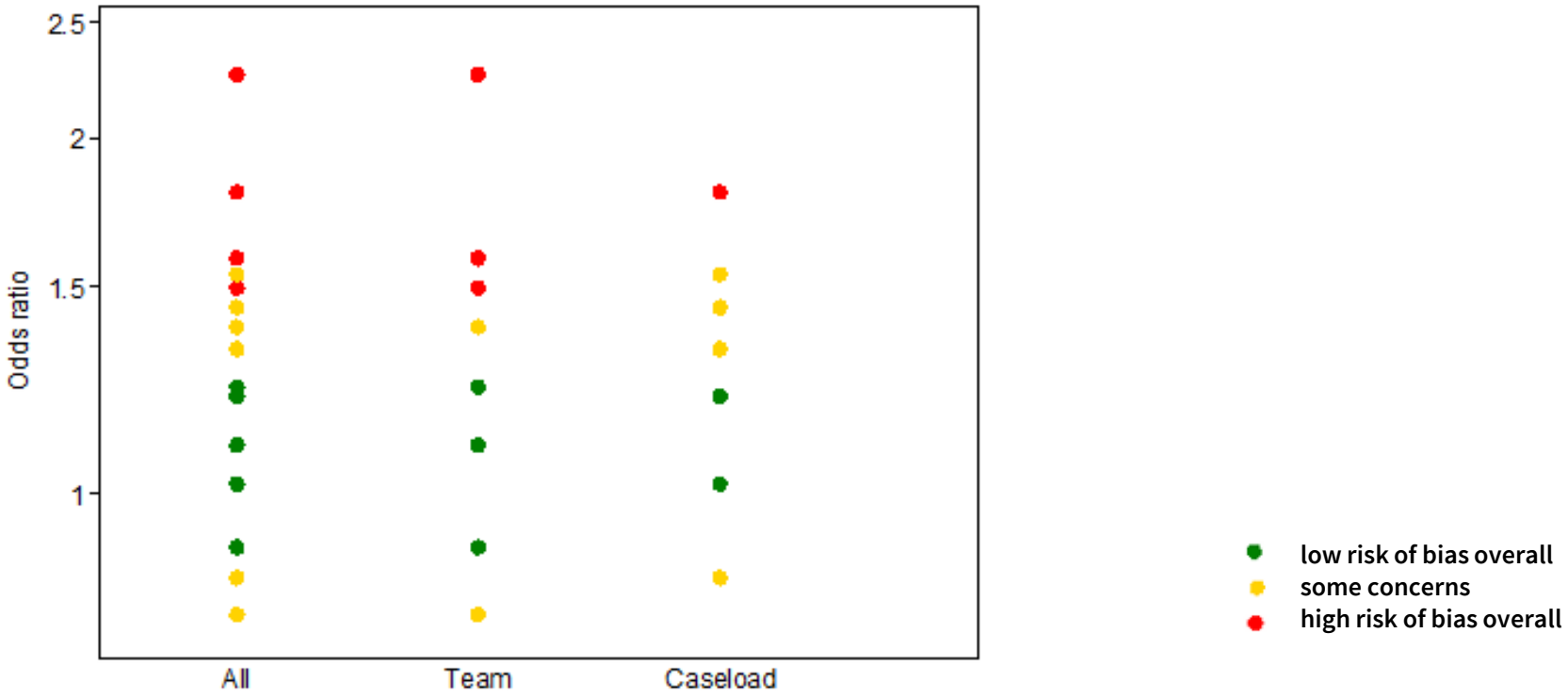


Cochrane Training

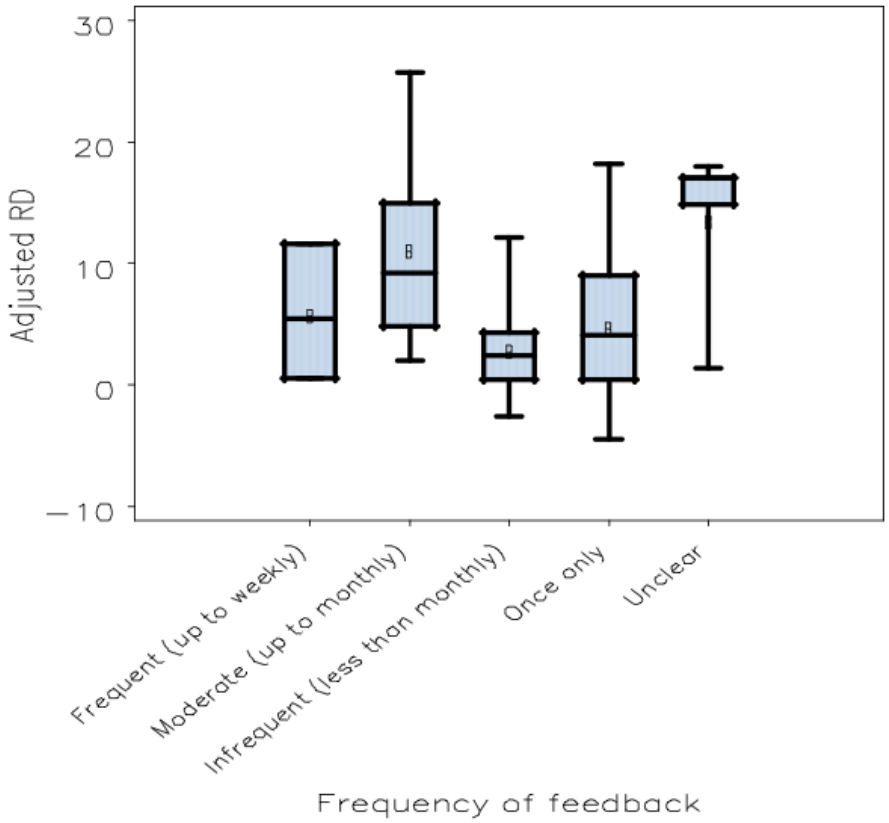
Summarising effect estimates: visual display of results

Bubble plot of odds ratios

(odds ratios for all outcomes and separately by the model of care)



Summarising effect estimates: visual display of results



Audit and feedback: effects on professional practice and healthcare outcomes [Ivers 2012]

Figure 7. risk difference in adherence to recommended practice by frequency of feedback

Summarising effect estimates

Question addressed: What is the range and distribution of observed effects?

Purpose

Can be used to synthesize when difficult to undertake a meta-analysis (e.g. missing variances of effects, unit of analysis errors)

Provides information on the magnitude and range of effects

Effects are summarised using descriptive statistics such as the median, interquartile range, and the range

Limitations

Does not account for differences in the relative size of studies

Cochrane Training

Synthesis methods: Question addressed and minimum data

Method	Question addressed	Estimate of effect	Variance	Direction of effect	Precise P value
Meta-analysis of effect estimates and extension	What is the common (or average) effect? What intervention is most effective? What factors modify the magnitude of intervention effects?	✓	✓		
Summarizing effect estimates	What is the range and distribution of observed effects?	✓			
Combining P values	Is there evidence that there is an effect in at least one study?			✓	✓
Vote counting based on direction of effect	Is there any evidence of an effect?			✓	

Cochrane Training

Combining P values

Question addressed: Is there evidence that there is an effect in at least one study?

Purpose

Can be used to synthesize results when studies report:

- no or minimal information beyond P values and direction of effect
- results of non-parametric analyses
- results of different types of outcomes and statistical tests
- outcomes are different across studies (e.g. serious side effects)

Combining P values

			Scenario 3. Combining P values	
Study ID	Outcome (scale details*)	Overall RoB judgement	Available data** (2-sided P value)	Stand. Metric (1-sided P value)
Continuous				
Crowe 2010	Expectation of labour/birth (0 to 18 points)	Some concerns	Favours intervention, P = 0.135, N = 368	0.068
Finn 1997	Experience of labour/ birth (0 to 24 points)	Some concerns	Favours intervention, P = 0.061, N = 331	0.030
Harvey 1996	Labour & Delivery Satisfaction Index (37 to 222 points)	Some concerns	MD -3, P = 0.368, N = 194	0.816
Kidman 2007	Control during labour/ birth (0 to 18 points)	High	MD 0.8, P = 0.035, N = 368	0.017
Turnbull 1996	Intrapartum care rating (-2 to 2 points)	High	MD 0.27, P = 0.072, N = 65	0.036
Binary				
Barry 2005	Experience of labour	Low	NS	—
Biro 2000	Perception of care: labour / birth	Some concerns	RR 1.13, P = 0.018	0.009
Johns 2004	Satisfaction with intrapartum care	Some concerns	Favours intervention, P < 0.001	0.0005

Required data:
P value and direction of effect

Minimal reporting of data and statistical tests vary

11 of 15 studies provide a precise P value and direction of effect.
2 report a P value less than a threshold (< 0.001)

Convert to one-sided P values which incorporate direction of effect

Combine P values

Combining P values

			Scenario 3. Combining P values	
Study ID	Outcome (scale details*)	Overall RoB judgement	Available data** (2-sided P value)	Stand. Metric (1-sided P value)
Continuous				
Crowe 2010	Expectation of labour/birth (0 to 18 points)	Some concerns	Favours intervention, P = 0.135, N = 368	0.068
Finn 1997	Experience of labour/ birth (0 to 24 points)	Some concerns	Favours intervention, P = 0.061, N = 351	0.030
Harvey 1996	Labour & Delivery Satisfaction Index (37 to 222 points)	Some concerns	MD -3, P = 0.368, N = 194	0.816
Kidman 2007	Control during labour/ birth (0 to 18 points)	High	MD 0.8, P = 0.035, N = 368	0.017
Turnbull 1996	Intrapartum care rating (-2 to 2 points)	High	MD 0.27, P = 0.072, N = 65	0.036
Binary				
Barry 2005	Experience of labour	Low	NS	—
Biro 2000	Perception of care: labour / birth	Some concerns	RR 1.13, P = 0.018	0.009
Johns 2004	Satisfaction with intrapartum care	Some concerns	Favours intervention, P < 0.001	0.0005

- Fisher's method for meta-analyzing P values from statistical tests for *k* studies:

$$observed\ test\ statistic = -2\sum_{i=1}^{k}\ln(P_i)$$

- The *observed test statistic* (above) is compared with the chi-squared distribution with *2k* degrees of freedom
- From this comparison, a P value is obtained
- When this P value is small, it provides evidence that there is an effect in at least one study

Cochrane Training

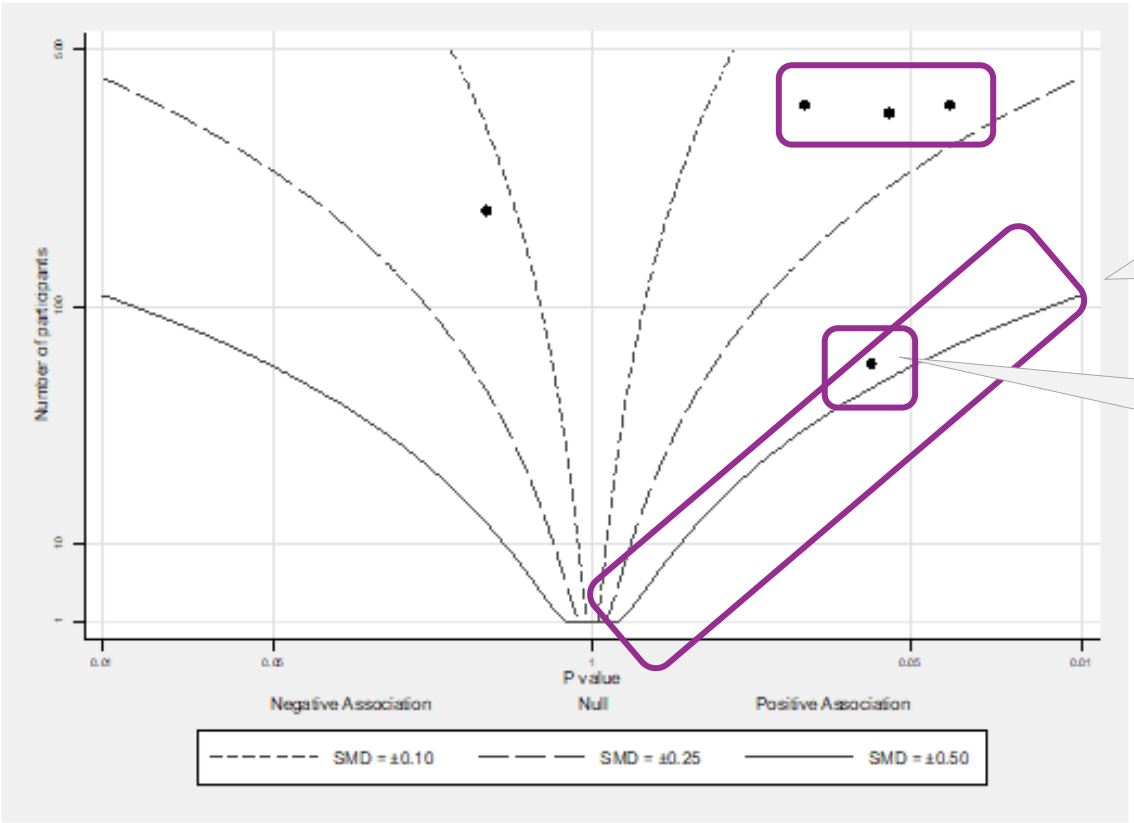
Combining P values: how to write it up?

Scenario 3. Synthesis of P values

'There was strong evidence of benefit of midwife-led models of care in at least one study ($P < 0.001$, 13 studies). However, a sensitivity analysis restricted to studies with an overall low risk of bias suggested there was no effect of midwife-led models of care in any of the trials ($P = 0.314$, 3 studies). Estimated standardized mean differences for five of the outcomes were small (ranging from -0.13 to 0.45) (Figure 12.4.b, Panel C).'

Combining P values: visual display of results

Albatross plot of the study sample size against P values
(for the five continuous outcomes)



Allows approximate examination of the underlying intervention effects

The contours on the plot reflect standardized mean differences (SMD)

Each point reflects a study's result

The location of the points relative to the SMD contours indicates their likely size

Cochrane Training

Combining P values: how to write it up?

Scenario 3. Synthesis of P values

'There was strong evidence of benefit of midwife-led models of care in at least one study ($P < 0.001$, 13 studies). However, a sensitivity analysis restricted to studies with an overall low risk of bias suggested there was no effect of midwife-led models of care in any of the trials ($P = 0.314$, 3 studies). Estimated standardized mean differences for five of the outcomes were small (ranging from -0.13 to 0.45) (Figure 12.4.b, Panel C).'

Eur J Clin Pharmacol (2000) 56: 27–33

SPECIAL ARTICLE

M. Cucherat · M. C. Haugh · M. Gooch
J.-P. Boissel, for the HMRAG group

Evidence of clinical efficacy of homeopathy

A meta-analysis of clinical trials

Cochrane Training

Table 1 Description of the study designs and patients' characteristics for trials included in the meta-analysis. *NA* not available

Trials	Disease setting	Homeopathic treatment	Primary outcome	No. of patients evaluated/ randomised	Result (*P* value)	Blinding	Lost to follow-up (%)[a]	Placebo
Mössinger 1980 [39]	Boils and pyoderma	Hepar sulfuris calcareum D4	Healing time	46/NA	0.318	Double	NA	NA
Coudert 1981 [40]	Dystocia	Caulophyllum 5 °C	Success within 2 h	34/34	0.00055	Double	0	Identical pellet
Reilly 1986 [41]	Active hay fever	Fixed, mixed grass pollens 30 °C	Visual analogue scale of overall symptom intensity	102/158	0.018	Double	35	NA
Grecho 1988 [29, 31, 42, 43]	Post-surgery ileus	Opium 15 °C	Delay to the first stool	300/300	0.699	Double	0	Identically prepared globules but without active constituent
Grecho 1988 [29]	Post-surgery ileus	Raphanus 15 °C + Opium 15 °C	Delay to the first stool	300/300b	0.358	Double	0	Identically prepared globules but without active constituent
Zell 1988 [44, 45]	Acute ankle sprains	Traumel ointment	Composite criteria of treatment success	69/NA	0.028	Double	NA	Ointment base without active constituent
Ferley 1989 [46]	Influenza-like syndrome	Fixed, Oscillococcinum	Recovery rate with in 48 h of treatment	462/478	0.032	Double	3	Identical pellet
Alfred 1990 [47]	Post-operative pain agitation	Aconit 4 °C	Sedation within 15 min	47/50	0.002	Double	6	NA
Thiel 1991 [48]	Knee joint haematoma	Intraarticular Traumel R	Joint mobility	73/80	0.026	Double	9	Intraarticular injection of NaCl
Lièvre 1992 [49]	2nd and 3rd degree burns	Calendula	Composite criteria of treatment success	103/103	0.147	Open	0	Vaseline
Gaus 1993 [50–52]	Rheumatoid arthritis	Rheumaselect	Composite criteria of treatment success	176/176	0.018	Double	0	NA
Whitmarsh 1993 [53]	Headache	Individualised	Change in mean attack frequency over the course of the trial	NA/64	0.83	Double	NA	Identical pellet
Jacobs 1994 [54]	Acute childhood diarrhoea	Individualised	Duration of diarrhoea	81/92	0.048	Double	12	Identical pellet
Reilly 1994 [55]	Allergic asthma	Individualised homeopathic immunotherapy	Visual analogue scale of overall symptom intensity	24/28	0.003	Double	14	Identically prepared globules but without active constituent
Weiser 1995 [56, 57]	Chronic sinusitis	Euphorbium compositum S nasal spray	Cumulative score	155/172	0.016	Double	10	NA
Diefenbach 1997 [30]	Bronchitis	Bronchiselect	Length of productive cough	209/258	0.86	Double	19	NA
Papp 1998 [58]	Influenza-like syndromes	Oscillococcinum	Multiple endpoint: rate of patients affected and duration of disease	334/372	0.0257	Double	10	Identical pellet

[a] Randomised patients for whom the outcome measure was missing were considered as 'lost to follow-up'
[b] Same control group as in Grecho[a]
[c] Holm's procedure was used to perform simultaneous inference. The *P* value given is an adjusted *P* value to take into account the two statistical tests performed

Statistical analysis

"The statistical approach used, therefore, was the combination of the significance levels (P values) [6-8]. The rationale for this choice is that all the trials explored the same broad question, i.e. ``is homeopathic treatment efficacious?", even if, for individual trials, the question asked expressed more specific terms and focused on a given treatment of a particular disease. Thus, unlike in the conventional meta-analytical methods, the method used does not involve pooling the numerical estimates of treatment effect sizes obtained, in our case, in very different situations. Using this approach, the null hypothesis tested is that the effect of interest (in this case, the efficacy of homeopathic treatment) is not present in any of the trials considered. If the null hypothesis is rejected, we can conclude that in at least one trial there is a non-null effect."

Cochrane Training

Table 2 Pooled *P* values obtained from all eight methods investigated for the 17 comparisons

Method	*P* value (two tailed)
Weighted sum Z	0.000036
Mean *P*	1.7×10^{-6}
Mean Z	7.8×10^{-8}
Logit	8.7×10^{-12}
Sum log	4.7×10^{-12}
Sum Z	5.9×10^{-12}
Sum t	3.2×10^{-13}
Count	2.8×10^{-29}

All results suggested that there was strong evidence that homeopathy was effective in at least one study

Table 3 Sensitivity analysis by stepwise removal of lower methodological quality comparisons (weighted sum of Z)

Class	Comparisons	No. of trials	Combined 2-tailed *P* value
Randomised, blind or open	See Table 1	17	0.000036
Randomised, double blind	As above except Lièvre 1992 [49][a]	16	0.000068
Randomised, double blind, with less than 10% of lost to follow-up	Coudert 1981 [40], Thiel 1991 [48], Whitmarsh 1993 [53], Ferley 1989 [46], Alibeu 1990 [47], Gaus 1993 [50–52], Grecho 1988 [29], Weiser 1995 [56, 57]	9	0.0084
Randomised, double blind, with less than 5% of lost to follow-up	Coudert 1981 [40], Ferley 1989 [46], Gaus 1993 [50–52], Grecho 1988 [29]	5	0.082

[a] The trial by Liévre used as control a treatment without active component but distinguishable from the active treatment

When the analysis was restricted to studies with fewer methodological limitations, there was no evidence to suggest in that homeopathy was effective in any trial

Cochrane Training

Combining P values

Question addressed: Is there evidence that there is an effect in at least one study?

Purpose

Can be used to synthesize results when studies report

- no or minimal information beyond P values and direction of effect
- results of non-parametric analyses
- results of different types of outcomes and statistical tests
- outcomes are different across studies (e.g. serious side effects)

Limitations

Provides no information on magnitude of effects

Does not distinguish between evidence from large studies with small effects and small studies with large effects

Difficult to interpret when statistically significant (null hypothesis can be rejected on the basis of an effect in only one study)

When combining few, small studies a non statistically significant test should not be interpreted as evidence of no effect in all studies

Cochrane Training

Synthesis methods: Question addressed and minimum data

Method	Question addressed	Estimate of effect	Variance	Direction of effect	Precise P value
Meta-analysis of effect estimates and extension	What is the common (or average) effect? What intervention is most effective? What factors modify the magnitude of intervention effects?	✓	✓		
Summarizing effect estimates	What is the range and distribution of observed effects?	✓			
Combining P values	Is there evidence that there is an effect in at least one study?			✓	✓
Vote counting based on direction of effect	Is there any evidence of an effect?			✓	

Vote counting based on direction of effect

Question addressed: Is there any evidence of an effect?

Purpose

Can be used to synthesize results when only direction of effect is reported, or there is inconsistency in the effect measures or data reported across studies

For each study, the effect is categorised as beneficial or harmful based on the direction of effect.

Statistical significance is NOT considered.

Can apply a sign test to test if there is any evidence of an effect (equivalent to testing if the true proportion of effects favouring the intervention (or comparator) is equal to 0.5).

Vote counting based on direction of effect

			Scenario 4. Vote counting	
Study ID	Outcome (scale details*)	Overall RoB judgement	Available data**	Stand. Metric
Continuous				
Crowe 2010	Expectation of labour/birth (0 to 18 points)	Some concerns	NS	—
Finn 1997	Experience of labour/ birth (0 to 24 points)	Some concerns	MD 1.3, NS	1
Harvey 1996	Labour & Delivery Satisfaction Index (37 to 222 points)	Some concerns	MD -3, NS	0
Kidman 2007	Control during labour/ birth (0 to 18 points)	High	MD 0.8 (95% CI 0.1 to 1.5)	1
Turnbull 1996	Intrapartum care rating (-2 to 2 points)	High	MD 0.27 (95% CI -0.03 to 0.57)	1
Binary				
Barry 2005	Experience of labour	Low	RR 1.13, NS	1
Biro 2000	Perception of care: labour / birth	Some concerns	RR 1.13, P < 0.05	1
Flint 1989	Care from staff during labour	High	RR 1.07, (95% CI 1.00 to 1.16)	1

Required data: direction of effect

Minimal reporting of data and type of effect measure varies across studies.

12 of 15 reported direction of effect (3 did not).

For each study, the effect is categorised as beneficial or harmful based on direction of effect

Statistical significance is NOT considered

Cochrane Training

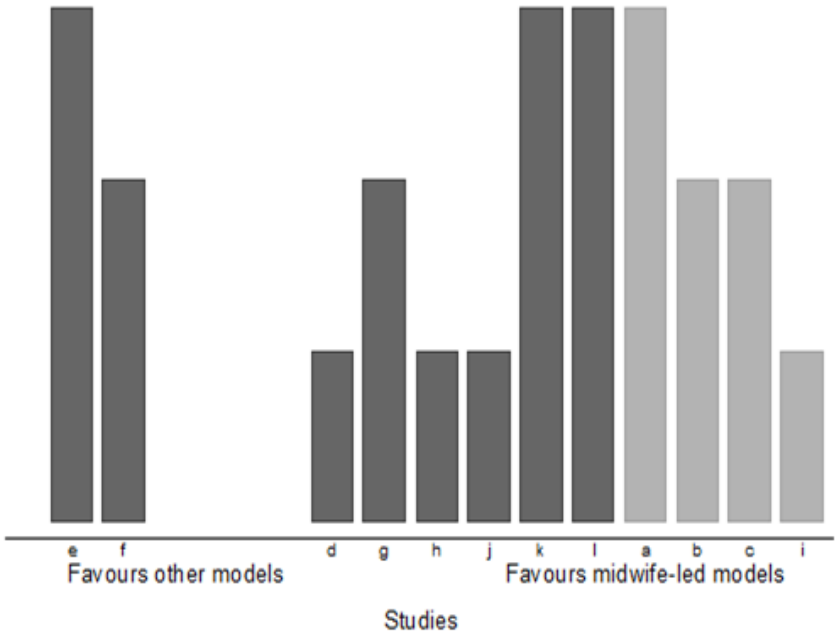
Vote counting based on direction of effect: how to write it up?

Scenario 4. Synthesis using vote counting based on direction of effects

'There was evidence that midwife-led models of care had an effect on satisfaction, with 10 of 12 studies favouring the intervention (83% (95% CI 55% to 95%), P = 0.039) (Figure 12.4.b, Panel D). Four of the 12 studies were judged to be at a low risk of bias, and three of these favoured the intervention. The available effect estimates are presented in [review] Table X.'

Vote counting based on direction of effect: Visual display

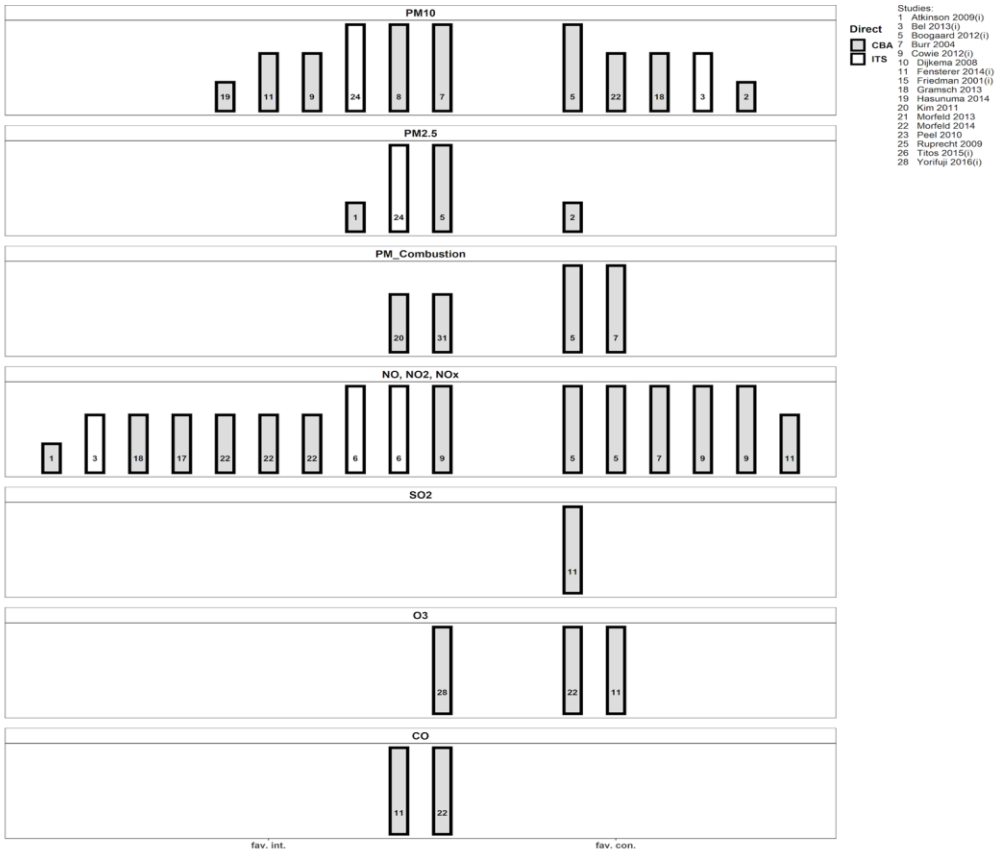
Harvest plot



- Each bar represents a result from one study, denoted by alphabet characters
- The bars are categorized based on their effects, i.e. those where the intervention is favourable, and those where the alternative intervention is favourable
- Height and shading can be used to depict characteristics of the studies. In this example:
  - height depicts overall risk of bias judgement
    (tall = low risk of bias, medium = some concerns, short = high risk of bias)
  - shading depicts model of care
    (light grey = caseload, dark grey = team)

Vote counting based on direction of effect: Visual display

Harvest plot



- A series of harvest plots may be combined in a matrix that displays vote counting results from, for example, different comparisons or outcomes
- In this example, each harvest plot depicts the vote count for one comparison, but different air pollution outcomes

Vote counting based on direction of effect

Question addressed: Is there any evidence of an effect?

Purpose

Can be used to synthesize results when only direction of effect is reported, or there is inconsistency in the effect measures or data reported across studies

For each study, the effect is categorised as beneficial or harmful based on the direction of effect.

Statistical significance is NOT considered.

Can apply a sign test to test if there is any evidence of an effect (equivalent to testing if the true proportion of effects favouring the intervention (or comparator) is equal to 0.5).

Limitations

Provides no information on the magnitude of effects (e.g. equal importance given to risk difference of 5% and 50%).

Does not account for differences in the relative sizes of the studies.

Less powerful than methods used to combine P values.

May be confused with **UNACCEPTABLE forms of vote counting**:

- when statistical significance used to define # positive and # negative studies
- when subjective decisions are used to define 'positive' and 'negative' studies.

Unacceptable methods:
Vote counting based on statistical significance

Create two stacks of papers

Stack 1: studies that show **statistically significant benefit** from the intervention

Stack 2: all other studies

If **Stack 1** is highest, conclude the intervention is beneficial. Otherwise conclude that there is no effect.

Friedman 2001

Cochrane Training

What are the problems with vote counting based on statistical significance?

- Small studies with insufficient power to detect an underlying intervention effect are counted as not showing benefit.
- As the number of studies (and participants) increases, the power of conventional vote counting tends to zero except with large studies and at least moderate intervention effects
- Regardless of the specific formulation, when based on statistical significance, all vote counting methods have serious limitations and can lead to the wrong conclusion.

Cochrane Training

Session outline

- Introduction
- Reasons for not undertaking a meta-analysis of effect estimates
- Structured summaries
- Acceptable synthesis methods and accompanying visual displays
- **What to include in the protocol**
- Take home messages

What to include in the protocol

- Provide details of the planned tabular structure to display results of individual studies
- Provide details of other synthesis and visual display methods that may be used

Cochrane Training

Session outline

- Introduction
- Reasons for not undertaking a meta-analysis of effect estimates
- Structured summaries
- Acceptable synthesis methods and accompanying visual displays
- What to include in the protocol
- **Take home messages**

Cochrane Training

Take home messages

- Other synthesis methods may be needed where there is incompletely reported data in the primary studies
- These methods differ in the completeness of the data they require, the hypotheses they address, and the conclusions that can be drawn from their findings
- Tabulation and visual display of the results should always be presented
  alongside any synthesis
- Synthesis methods that involve vote counting based on statistical significance have serious limitations and are unacceptable

Cochrane Training

Slides prepared by:
Joanne E. McKenzie and Sue E. Brennan
School of Public Health and Preventive Medicine, Monash University, Australia

Based on:

McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition. Chichester (UK): John Wiley & Sons, 2019: 321–348.

Other references

1. Cucherat, M., et al., Evidence of clinical efficacy of homeopathy. A meta-analysis of clinical trials. HMRAG. Homeopathic Medicines Research Advisory Group. *Eur J Clin Pharmacol*, 2000. 56(1): p. 27-33.
2. Friedman: Why vote-count reviews don't count. *Biological Psychiatry* 2001, 49(2):161-162
3. Harrison S, Jones HE, Martin RM, Lewis SJ, Higgins JPT. The albatross plot: A novel graphical tool for presenting results of diversely reported studies in a systematic review. *Res Synth Methods*. 2017;8:281-9.
4. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews*. 2012;6:CD000259.
5. McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnson RV, Thomas J. Chapter 9: Summarizing study characteristics and preparing for synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd ed. Chichester UK: John Wiley & Sons; 2019. p. 12.
6. McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd ed. Chichester UK: John Wiley & Sons; 2019. p. 28.