Deborah Caldwell and Georgia Salanti
d.m.caldwell@bristol.ac.uk

Addressing multiple treatments II: multiple-treatments meta-analysis basic methods

Keystone, 2010

18th Cochrane Colloquium, 18th to 22nd October

University of BRISTOL

Workshop outline

- **The Basics: indirect comparisons**
  - What are indirect comparisons & why are they necessary
  - Exercise: how to do an indirect comparison
- **Slightly more advanced:**
  - Checking assumptions for IC (and MTM) with exercise
  - Checking consistency: Bucher's method
  - Advantages of MTM: applied example
- **Statistical methods for MTM**
  - How to do MTM within a frequentist framework
  - Inconsistency
  - Assumptions of MTM

Multiple treatment decision-making

- For many clinical indications there will often be several possible interventions.
- The Cochrane Database of Systematic Reviews
  - 22 interventions for adult smoking cessation
  - >12 interventions for chronic asthma in adults
- Health care decisions should be based on 'best available' evidence from systematic reviews & meta-analysis of RCTs

Problem…

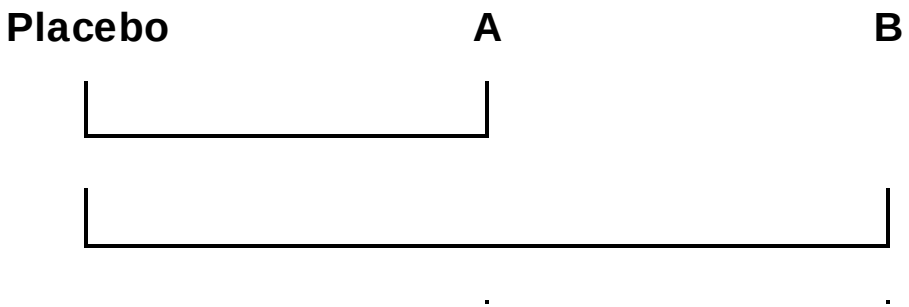
- Systematic reviews focus on direct, head-to-head comparisons of interventions.
  - e.g. NRT vs placebo; Olanzapine vs placebo
  - A vs B; A vs C.
- The evidence base consists of a set of pair-wise comparisons of interventions
  - Placebo comparisons of limited use to the practitioner or policy-maker who wants to know the ‘best’ treatment to recommend/ prescribe.

Problem... (2)

- ‘Best available’ evidence is not always available or sufficient
  - Placebo controlled trials sufficient for regulatory approval of new drugs
  - Even when active comparisons have been made such direct evidence is often limited.
- Therefore, evidence base may not contain treatment comparisons of relevance for clinician or policy maker.

Example evidence structure #1

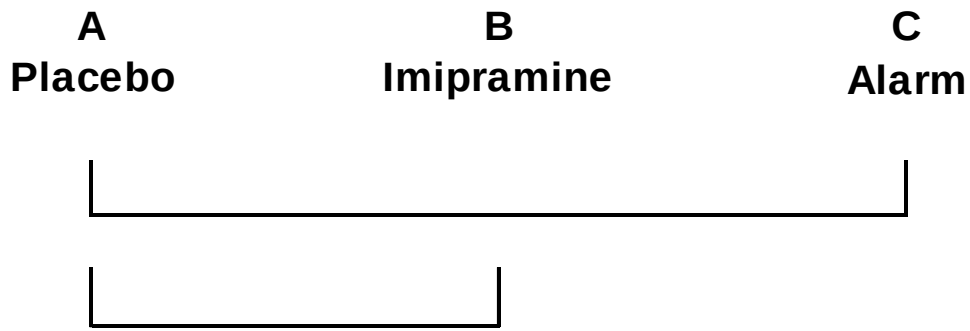
- Common situation is to have multiple competing treatments (often within class) each studied in placebo-controlled RCTs but none compared directly to each other.



- How do we know which treatment to use?

Case study: childhood nocturnal enuresis *

Evidence base: 3 treatment options; 2 comparisons



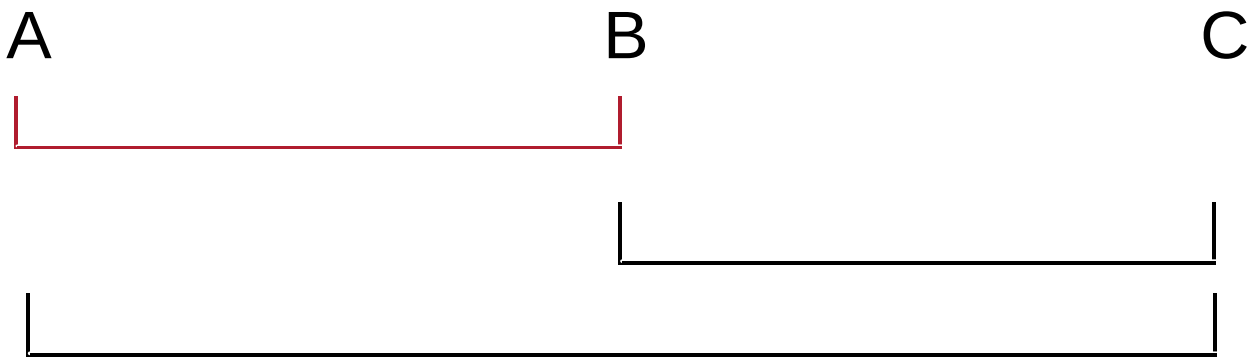
Summary of results from 2 separate enuresis meta-analyses

Comparison	n/ N active	n/ N no.treat	Relative Risk	CIs
Alarm vs no treatment	107/ 316	250/ 260	0.39	(0.33 to 0.46)
Imipramine vs no treatment	314/ 400	391/ 403	0.95	(0.87 to 0.99)

*Source: Russell and Kiddoo (2006)

‘Adjusted’ Indirect comparisons

- In absence of direct evidence for treatments A vs B, an *indirect* estimate of log risk ratio lrr_{AB} can be obtained from RCTs comparing A vs C and B vs C:

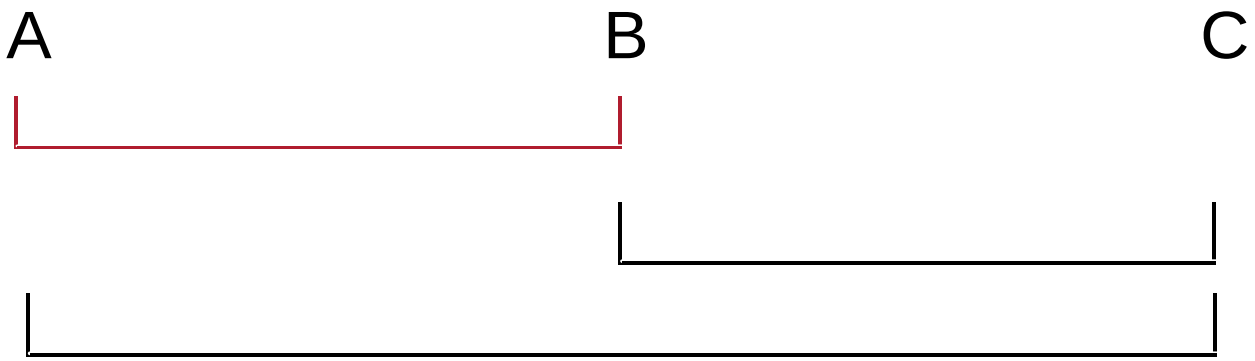


$$LRR_{AB} = LRR_{AC} - LRR_{BC}$$

*Bucher HC, et al.(1997); Glenny et al (2005)

Indirect comparisons

- In absence of direct evidence for treatments A vs B, an *indirect* estimate of log risk ratio lrr_{AB} can be obtained from RCTs comparing A vs C and B vs C:

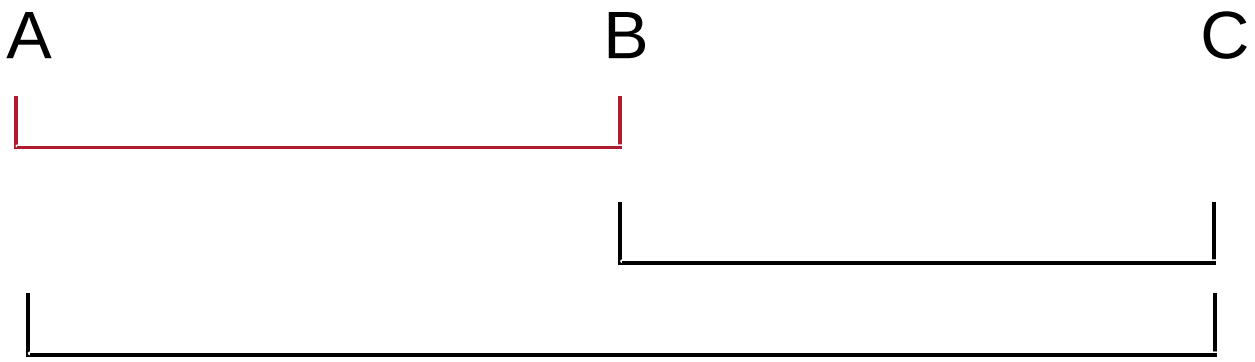


$$LRR_{AB} = LRR_{AC} - LRR_{BC}$$

*Bucher HC, et al.(1997); Glenny et al (2005)

Indirect comparisons

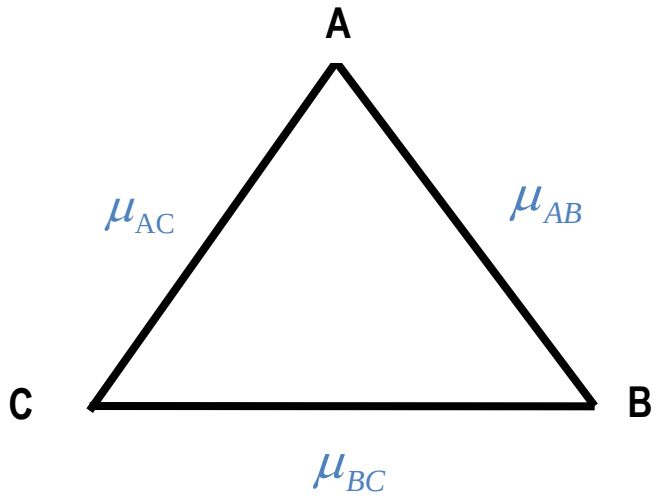
- In absence of direct evidence for treatments A vs B, an *indirect* estimate of log risk ratio lrr_{AB} can be obtained from RCTs comparing A vs C and B vs C:



Consistency equation*

*Lu et al (2007) Journal of the American Statistical Association

3 treatment network



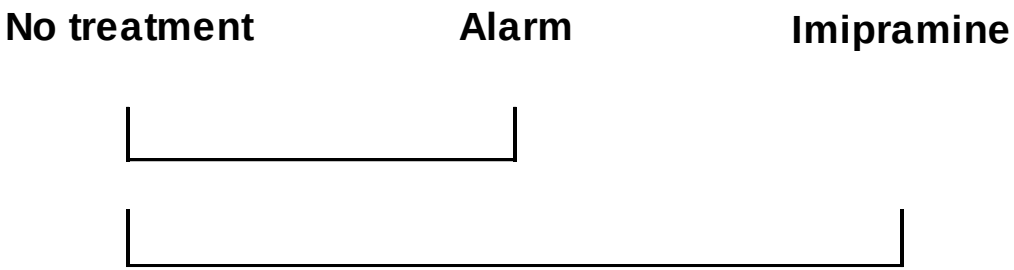
Three possible indirect comparisons, all equivalent:

$$\mu_{AB}^{Indirect} = \mu_{AC} - \mu_{BC};$$

$$\mu_{AC}^{Indirect} = \mu_{BC} - \mu_{AB};$$

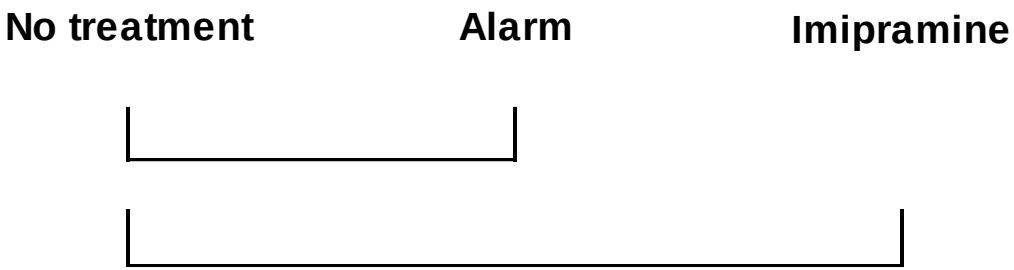
$$\mu_{BC}^{Indirect} = \mu_{AC} - \mu_{AB}$$

Simple exercise



Comparison	RR	CIs
No treatment vs Imipramine	0.95	(0.87 to 0.99)
No treatment vs Alarm	0.39	(0.33 to 0.46)

Simple exercise



Comparison	RR	CIs
No treatment vs Imipramine **AB**	0.95	(0.87 to 0.99)
No treatment vs Alarm **AC**	0.39	(0.33 to 0.46)

Pen and paper exercise.

$$LRR_{BC} = LRR_{AC} - LRR_{AB}$$

lrr_{AB} = -0.06

lrr_{AC} = -0.93

$lrr_{BC} = lrr_{AC} - lrr_{AB} =$

Indirect $RR_{BC} = \exp(lrr_{BC}) =$

Pen and paper exercise.

$$LRR_{BC} = LRR_{AC} - LRR_{AB}$$

lrr_{AB} = -0.06

lrr_{AC} = -0.93

$lrr_{BC} = lrr_{AC} - lrr_{AB}$ = -0.93 – (-0.06) = -0.87

Indirect $RR_{BC} = \exp(lrr_{BC})$ = **0.42**

Confidence intervals and p-value

$$Var(\hat{LRR}_{BC}^{Indirect}) = Var(\hat{LRR}_{AC}^{Direct}) + Var(\hat{LRR}_{AB}^{Direct}) \;=\; 0.007 + 0.001 = 0.008$$

$$SE(\hat{LRR}_{BC}^{Indirect}) = \sqrt{\text{var}(\hat{LRR}_{BC}^{Indirect})} = \quad \sqrt{0.008} = \quad \underline{\mathbf{0.09}}$$

95% CI= LRR $\pm$1.96*SE = 0.35 to 0.50 p= <0.0001 (z = -9.66)

Note: $Var(\hat{LRR}_{BC}^{Indirect}) = Var(\hat{LRR}_{AB}^{Direct}) + Var(\hat{LRR}_{AC}^{Direct})$

Therefore, all things being equal (trials all of same size, equal variance and assuming a common treatment effect) 1 directly randomised trial is as precise as an indirect comparison based on 4 randomised trials (see Glenny, 2005 for more detail)

Assumptions for indirect comparisons

- Validity relies on the AB & AC RCTs being **similar** across factors which may affect the outcome (modify treatment effect).
- A clinical/ epidemiological judgement:
  - No treatment by comparison interaction
  - Assuming inclusion/ exclusion criteria same across comparisons
  - Patients, trial protocols, doses, administration etc are similar in ways which might modify treatment effect.

"Between-trial comparisons [Indirect Comparisons] are unreliable. Patient populations may differ in their responsiveness to treatment. Therefore an apparently more effective treatment may have been tested in a more responsive population"

Cranney, Guyatt et al. *End Rev* 2002, 23; 570-8

"Placebo controlled trials lacking an active control give little useful information about comparative effectiveness. Such information cannot reliably be obtained from cross-study comparisons, as the conditions of the studies may have been quite different"

International Council of Harmonisation E10 2.7.1.4

"Indirect comparisons are observational studies across trials, and may suffer the biases of observational studies, for example confounding"

Cochrane Handbook for systematic reviews of interventions
4.2.5. Cochrane Library Issue 3

Checking assumptions

Exercise:

- Using the forest plots and study characteristics tables provided, work with a neighbour/ in small groups to discuss whether the AB and AC trials are similar enough across factors which may modify treatment effect.
- Suggested time: 10 minutes

Handout: trial characteristics

Alarm vs placebo characteristics of studies

	Age	Boys(%)	Exclusion	Previous treatment	Dropouts	Baseline wetting (SD)	Recruitment/setting
Bennet	8.5 (5-12)	63%	Gross psychopathology	Exc. If previous behavioural	32/40	2.7 in 14 nights	GP referral
Bollard(a)	9.6	71%	No details	No details	3/45	4.97 per week	No details
Bollard(b)	8.9	82%	No details	No details	12/100	5.56 mean wet nights	No details
Houts	5-13	63%	No details	No details	7/56	5.41(1.63) mean wet nights/week	Media/ consultant referral
Jehu	9.3 (4.8-14.6)	64%	No details	Exc. If previous alarm	1/39	4 mean wet nights/week	childrens home
Lynch	5-12	Not clear	Daytime wetting	No details	6/60	11.33 in 14 nights	School/ consultant referral
Moffatt	8-14	Not clear	No details	No details	5/121	64% wet nights	Hospital clinic
Nawaz	7-12	50%	Psychiatric pathology	No details	0/36	5.67 per week	GPs
Ronen	10 (SD 2.28)	48%	Developmental problems <5years	No details	23/77	19.1 days in 3 weeks	Mental health clinic
Sacks	5.5-14	Not clear	Severe psychosis	No details	Not clear	No details	No details
Sloop	12.5(7-18)	52%	Severe behavioural probs. tranquilisers	No previous treatment	Not clear	3.99 Not clear	Residential setting for learning disabled
Wagner	7.9('5-14)	51%	IQ<70	No conditioning treatment	0/39	84% wet nights per week	No details
Wagner(b)	6-16	82%	Daytime wetting	Drugs/alarm in prev. yr	13/49	72% 3x week	Media/consultant referral/school/GP
Werry	9.99 (SD 2.25)	66%	Dry >3months	No details	10/70	Min 1x per week	Hospital clinic

Imipramine vs placebo characteristics of studies

	Age	Boys(%)	Exclusion	Previous treatment	Dropouts	Baseline wetting (SD)	Recruitment/setting
Argawala	6-12	52%	Mental disability	Some patients had imipramine	29	No details	No details
Forsythe	4-15	64%	No UTI	No details	51/298	>6x per week/ for 1yr	Children's hospital
Hodes	5-15	Not clear	No details	No details	No details	No details	GP
Khorana	8.2 (5-15)	74%	Severe mental disability	No details	24/100	No details	Psychiatric inpatients (India)
Manhas	5-15	43%	No details	No details	No details	No details	No details
Poussaint	5-16	77%	No details	3 had psychotherapy	7/47	5.6 per week	No details
Schroder	3.5-10	No details	Organic causes	Resistant to previous therapy	34/62	No details	No details
Smellie	5-13	81%	Organic causes	No details	4/80	1.4 Dry nights	No details
Tahmaz	6-14	100%	Organic causes Daytime wetting	Fluid reduction/ night waking	11/30	No details	Military hospital (Turkey)
Wagner	6-16	82%	Daytime wetting	Drugs/alarm in prev. yr	13/49	72% 3x week	Media/consultant referral/school/GP

Forest plot for AvB

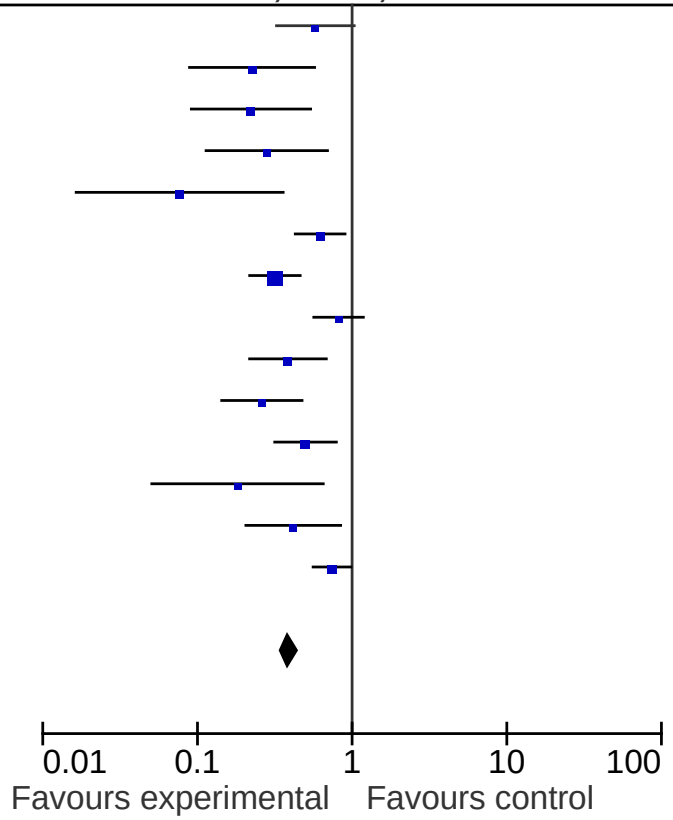
Alarm versus no treatment

Study or Subgroup	Weight	Risk Ratio M-H, Fixed, 95% CI
Bennett 1985	3.7%	0.58 [0.32, 1.03]
Bollard 1981a	6.0%	0.23 [0.09, 0.57]
Bollard 1981b	7.0%	0.22 [0.09, 0.54]
Houts 1986	4.6%	0.28 [0.11, 0.69]
Jehu 1977	7.7%	0.08 [0.02, 0.36]
Lynch 1984	7.2%	0.62 [0.43, 0.90]
Moffat 1987	22.0%	0.32 [0.22, 0.46]
Nawaz 2002	4.3%	0.82 [0.57, 1.18]
Ronen 1992	7.3%	0.39 [0.22, 0.68]
Sacks 1974	4.8%	0.26 [0.14, 0.47]
Sloop 1973	7.7%	0.50 [0.32, 0.79]
Wagner 1982	4.3%	0.18 [0.05, 0.65]
Wagner 1985	4.6%	0.42 [0.21, 0.84]
Werry 1965	8.8%	0.74 [0.56, 0.98]
Total (95% CI)	**100.0%**	**0.39 [0.33, 0.45]**

Total events

Heterogeneity: $Chi^2 = 56.57$, df = 13 ($P < 0.00001$); $I^2 = 77\%$

Test for overall effect: $Z = 12.04$ ($P < 0.00001$)



Forest plot for AvC

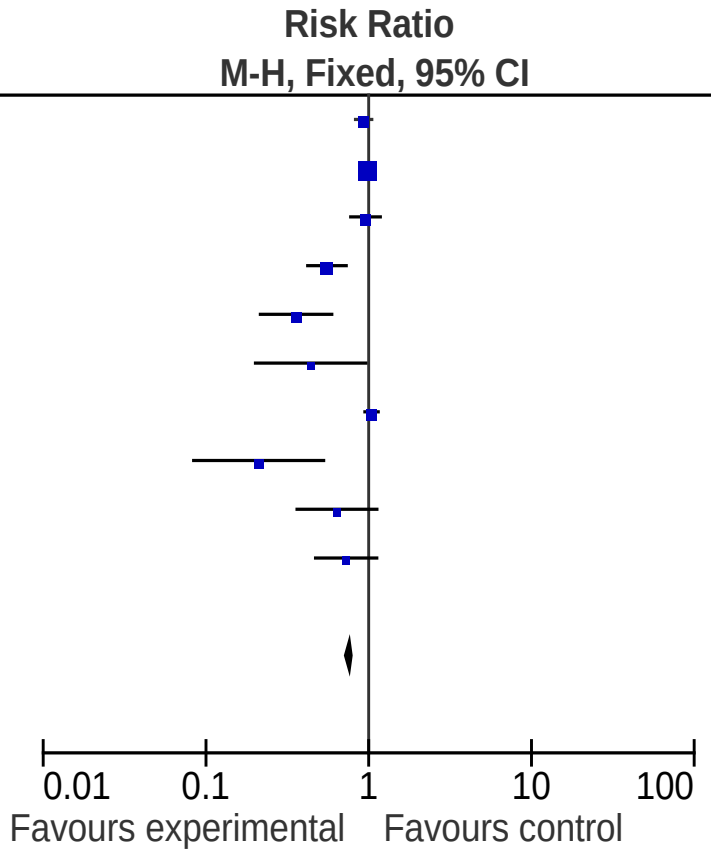
Imipramine versus no treatment

Study or Subgroup	Weight	Risk Ratio M-H, Fixed, 95% CI
Agarwala 1965	10.1%	0.93 [0.83, 1.05]
Forsythe 1969	28.3%	0.99 [0.95, 1.02]
Hodes 1973	10.6%	0.96 [0.77, 1.18]
Khorana 1972	13.0%	0.55 [0.42, 0.73]
Manhas 1967	9.2%	0.36 [0.22, 0.59]
Poussaint 1965	3.3%	0.44 [0.20, 0.96]
Schroder 1971	10.2%	1.04 [0.95, 1.15]
Smellie 1976	7.0%	0.21 [0.08, 0.53]
Tahmaz 2000	4.7%	0.64 [0.36, 1.13]
Wagner 1982b	3.8%	0.73 [0.47, 1.12]
Total (95% CI)	**100.0%**	**0.77 [0.72, 0.83]**

Total events

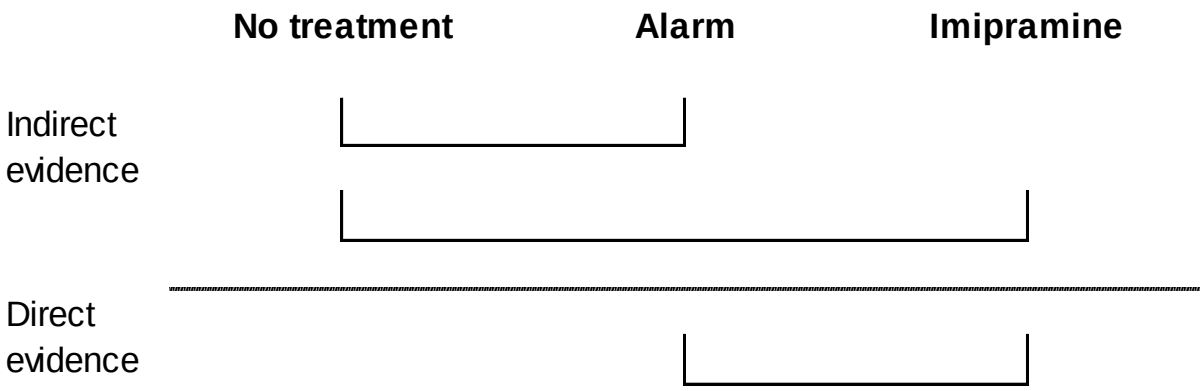
Heterogeneity: $Chi^2 = 269.99$, df = 9 ($P < 0.00001$); $I^2 = 97\%$

Test for overall effect: $Z = 6.97$ ($P < 0.00001$)

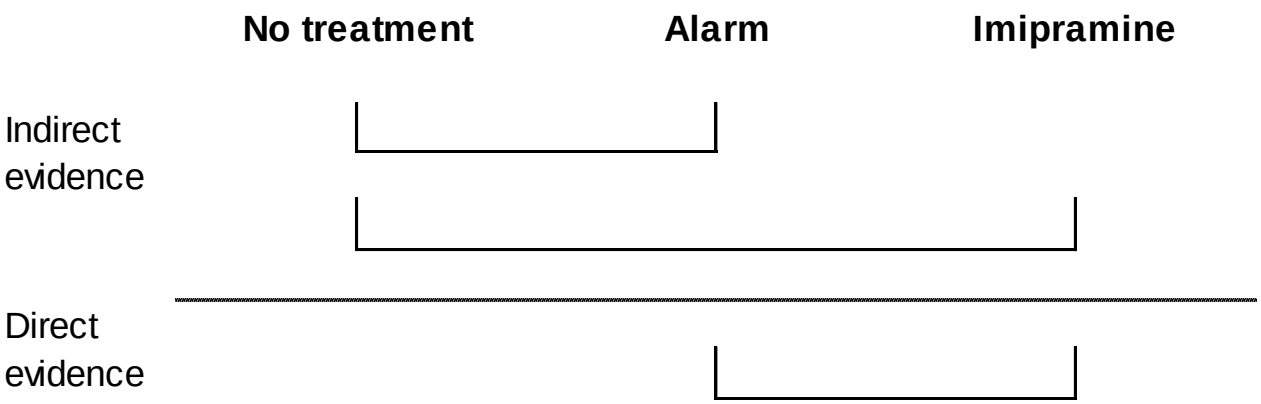


Example evidence structure #2

- Another common evidence structure is where we have **some** direct evidence on the relevant treatment comparisons (active vs active) but on its own its insufficient.



Evidence base: 3 treatment options; 3 comparisons



Summary of results from 3 enuresis meta-analyses

Comparison	n/ N active	n/ N no.treat	Relative Risk	CIs
Alarm vs no treatment	107/ 316	250/ 260	0.39	(0.33 to 0.46)
Imipramine vs no treatment	314/ 400	391/ 403	0.95	(0.87 to 0.99)
Alarm vs imipramine	61/105	82/103	0.77	(0.64 to 0.93)

Multiple-treatments meta-analysis

Combine direct and indirect evidence. Also known as:

1) Mixed treatment comparison

2) Network meta-analysis

ALL 3 mean the same thing – **simultaneous** comparison of multiple competing treatments using direct & indirect evidence (usually from RCTs) in a single analysis.

SAME assumption as made for indirect comparison alone: the consistency assumption.

Combining direct and indirect evidence

Simple approach to pooling direct and indirect evidence on lrr_{BC}

1. lrr_{BC}^{direct}

2. $lrr_{BC}^{indirect}$

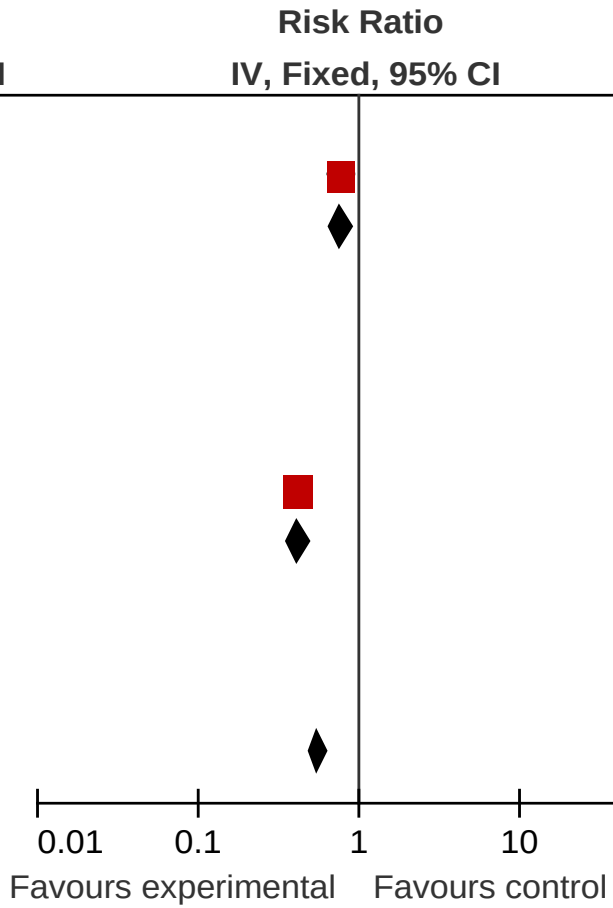
3. $$lrr_{BC}^{MTM} = \frac{(w^{direct}\ lrr_{BC}^{direct}) + (w^{indirect}\ lrr_{BC}^{indirect})}{(w^{direct} + w^{indirect})}$$

$$w = 1/se(BC_i)^2$$

Indirect evidence given less weight than direct evidence

Discussion of indirect and direct estimates

Study or Subgroup	log[Risk Ratio]	SE	Weight	Risk Ratio IV, Fixed, 95% CI
2.2.1 Direct				
Direct B vs C	-0.2571	0.095	47.9%	0.77 [0.64, 0.93]
Subtotal (95% CI)			**47.9%**	**0.77 [0.64, 0.93]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 2.71 (P = 0.007)				
2.2.2 Indirect				
Indirect B vs C	-0.87	0.091	52.1%	0.42 [0.35, 0.50]
Subtotal (95% CI)			**52.1%**	**0.42 [0.35, 0.50]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 9.56 (P < 0.00001)				
Total (95% CI)			**100.0%**	**0.56 [0.49, 0.64]**
Heterogeneity: Chi² = 21.71, df = 1 (P < 0.00001); I² = 95%				
Test for overall effect: Z = 8.78 (P < 0.00001)				



The **big** assumption

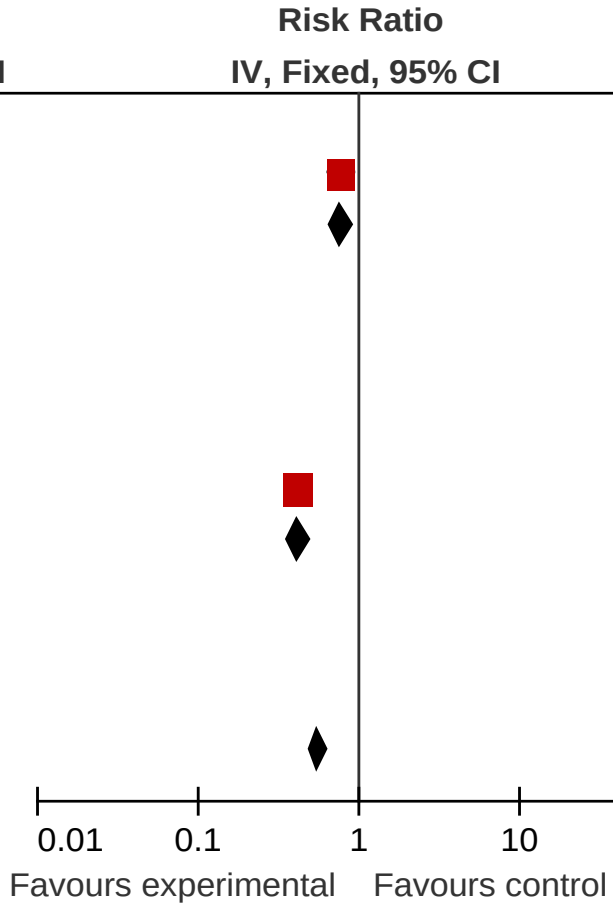
IC and MTM assume that the "Direct" and "Indirect" evidence estimate the same parameter, i.e. are CONSISTENT.

That the Treatment effect μ_{BC} estimated by the BC trials, would be the same as the treatment effect estimated by the AC and AB trials (if they had included B and C arms).

Nearly all the doubts about IC and MTM are doubts about this assumption.

Discussion of indirect and direct estimates

Study or Subgroup	log[Risk Ratio]	SE	Weight	Risk Ratio IV, Fixed, 95% CI
2.2.1 Direct				
Direct B vs C	-0.2571	0.095	47.9%	0.77 [0.64, 0.93]
Subtotal (95% CI)			**47.9%**	**0.77 [0.64, 0.93]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 2.71 (P = 0.007)				
2.2.2 Indirect				
Indirect B vs C	-0.87	0.091	52.1%	0.42 [0.35, 0.50]
Subtotal (95% CI)			**52.1%**	**0.42 [0.35, 0.50]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 9.56 (P < 0.00001)				
Total (95% CI)			**100.0%**	**0.56 [0.49, 0.64]**
Heterogeneity: Chi² = 21.71, df = 1 (P < 0.00001); I² = 95%				
Test for overall effect: Z = 8.78 (P < 0.00001)				



Bucher approach to checking consistency

The difference ω between direct LRR_{BC} and indirect LRR_{BC}

$\hat{\omega}$ = -0.257 - -0.87 = 0.61

To calculate the standard error of the difference we sum the SE from the direct and indirect log risk ratios

$$SE(\Delta) = \sqrt{SE(LLR^{Direct})^2 + SE(LRR^{Indirect})^2}$$

$$= \sqrt{0.095^2 + 0.091^2} \quad = 0.13$$

Bucher approach to checking consistency

Calculate confidence intervals & p-values for : $\hat{\omega}$

95% CI = $\hat{\omega} \pm (1.96*SE)$ = exp [0.36] to exp [0.86]

= **1.43 to 2.37**

z-score = $\frac{\hat{\omega}}{SE(\hat{\omega})}$ = 4.64 p-value = <0.000002

Limitations of simple approach

Straightforward & conceptually intuitive

- Extension of pairwise meta-analysis
- Checking consistency of evidence

BUT it is very LIMITED:

- Pool separately for each treatment comparison (separate meta-analyses)

What happens when

Treatments	4	5	6	7	8	9	10	11
Pairwise	6	10	15	21	28	36	45	55
Indirect	12	30	60	105	168	252	360	495

Methods for larger networks

- Simultaneous comparison of multiple treatments can only be done in a SINGLE MTM analysis
  - Using frequentist or Bayesian approach
  - Provides effect estimates for ALL treatment comparisons regardless of whether they have been directly compared.
  - combining direct and indirect evidence can increase the precision of effect estimates.

- Need to determine which treatment is BEST??
  - Bayesian framework (Markov chain Monte Carlo method
    - Ranking of treatments using simulation approach
    - Estimates probability each treatment is the best.