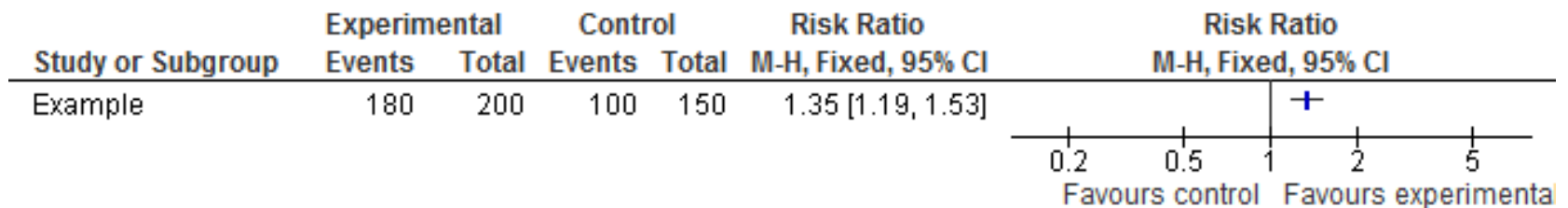# Reviews of conditions with low event rates

Matthew Grainge and Jo Leonardi-Bee
Statistical Editors for CSG

Introduction

- CSG reviews frequently have common outcomes (>10%)
- Statisticians recommend using risk ratio
  ◦ Odds ratio over estimates treatment effect

Study or Subgroup	Experimental Events	Total	Control Events	Total	Risk Ratio M-H, Fixed, 95% CI
Example	180	200	100	150	1.35 [1.19, 1.53]

Risk Ratio M-H, Fixed, 95% CI

0.2 0.5 1 2 5
Favours control   Favours experimental

Study or Subgroup	Experimental Events	Total	Control Events	Total	Odds Ratio M-H, Fixed, 95% CI
Example	180	200	100	150	4.50 [2.54, 7.98]

Odds Ratio M-H, Fixed, 95% CI

0.2 0.5 1 2 5
Favours control   Favours experimental

But what about rare outcomes?

- Rare outcomes (<5%)
  - Not a problem if sample size is large enough (>30 participants in total)

- BUT, if sample size is small and rare outcomes, then results can be misleading

	Experimental		Control		Risk Ratio	Risk Ratio
Study or Subgroup	Events	Total	Events	Total	M-H, Fixed, 95% CI	M-H, Fixed, 95% CI
Example	2	4	1	19	9.50 [1.11, 81.12]	

0.01 0.1 1 10 100

Favours control   Favours experimental

What about no outcomes?

- ▸ Mathematically, can't estimate Risk Ratio
  - ◦ Review Manager cheats by adding 0.5 people to each group with and without the outcomes

	Outcome	No outcome	TOTAL
Intervention	0	8	8
Control	5	8	13
TOTAL	5	16	21
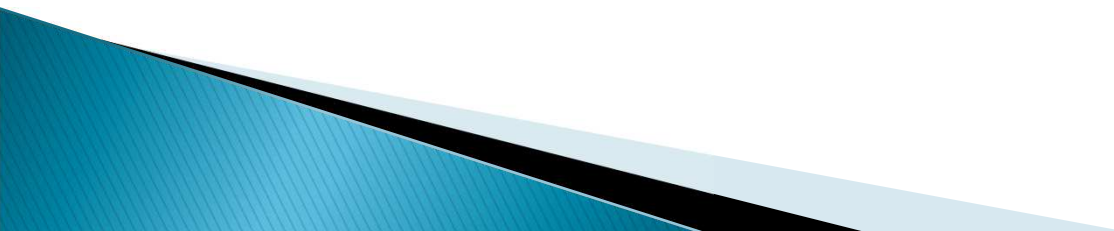
	Outcome	No outcome	TOTAL
Intervention	0.5	8.5	9
Control	5.5	8.5	14
TOTAL	6	17	23

What about confidence intervals?

- Misleading confidence intervals
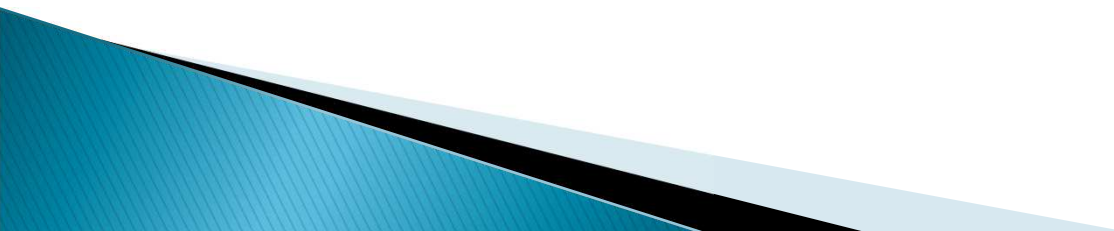 - Assumes data is based on a large sample size

Study or Subgroup	Experimental Events	Total	Control Events	Total	Risk Ratio M-H, Fixed, 95% CI	Risk Ratio M-H, Fixed, 95% CI
Example	2	4	1	19	9.50 [1.11, 81.12]	

0.01 0.1 1 10 100

Favours control   Favours experimental

What does the handbook say?

- **Section 16.9 – Rare events (including zero frequencies)**

- "Many methods of meta-analysis are based on large sample approximations, and are unsuitable when events are rare"

- "The fixed correction (for zero frequencies) … avoids computational errors [but] it has the undesirable effect of biasing study estimates towards no difference and overestimating variances of study estimates…"

Mycosis fungoides

- Most common type of cutaneous T-cell lymphoma
  - Rare malignant, chronic disease (not fungal infection)
  - Uncontrolled growth of blood cells within the skin
  - Several therapies used to induce clinical remission

- Weberschock et al. Interventions for mycosis fungoides. 2012; Issue 9
  - 14 RCTs (675 participants)
  - Majority of results are based on data from one study per treatment

Intervention: Combination therapy

- Active transfer factor with nitrogen mustard
  ◦ Intramuscular injections

- Nitrogen Mustard
  ◦ Topically applied chemotherapy agent

- Comparison: TNM with active transfer factor versus TNM with inactivated transfer factor
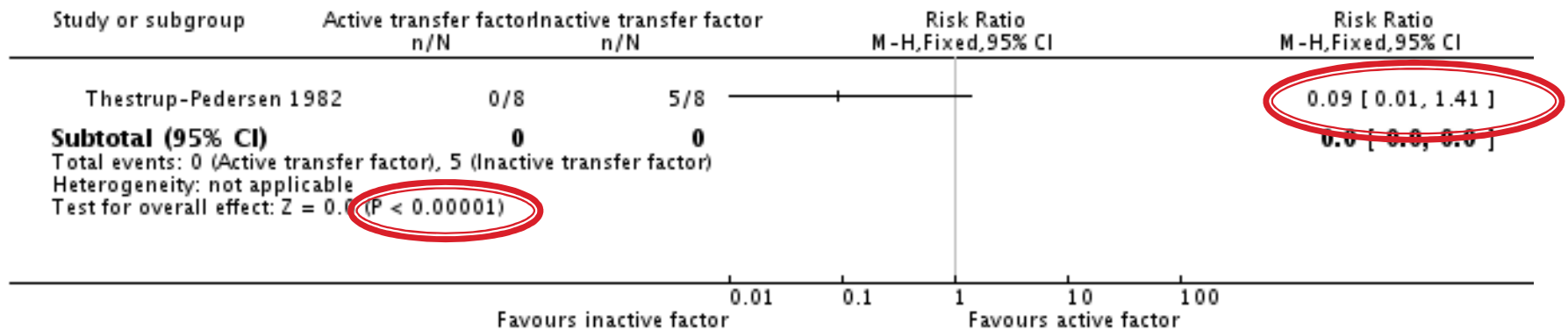
- Outcome: Clearance

Thestrup-Pedersen 1982

- "… differences were seen between the active transfer factor group and the inactivated transfer group for clearance: RR 0.09, 95% CI 0.01 to 1.41, 16 participants, Analysis 1.1"
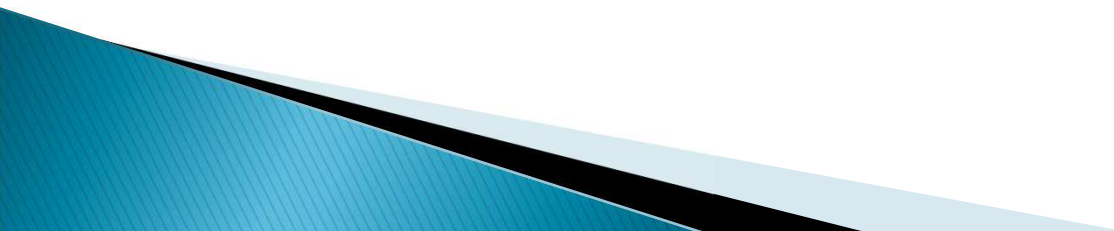
Review: Interventions for mycosis fungoides
Comparison: 11 Topical nitrogen mustard with active transfer factor versus topical nitrogen mustard with inactivated transfer factor
Outcome: 1 Clearance

Study or subgroup	Active transfer factor n/N	Inactive transfer factor n/N	Risk Ratio M-H,Fixed,95% CI	Risk Ratio M-H,Fixed,95% CI
Thestrup-Pedersen 1982	0/8	5/8		0.09 [0.01, 1.41]
Subtotal (95% CI)	**0**	**0**		0.0 [0.0, 0.0]

Total events: 0 (Active transfer factor), 5 (Inactive transfer factor)
Heterogeneity: not applicable
Test for overall effect: Z = 0.0 (P < 0.00001)

0.01   0.1   1   10   100
Favours inactive factor   Favours active factor

N.B: Both these statistics assume large sample sizes

Measures of treatment effect section of the review

- "P values for the subtotal RR were calculated with the Fisher test in order to avoid spurious (non-) significance in studies with small sample sizes or low numbers of events"

- Data: 0/8 v. 5/8
- P=0.03 (from Fisher's Exact test)

Is the treatment effective?

- **P values**
  - From RevMan <0.00001
  - From Fisher's Exact test=0.03
  - Both significant

- **Measure of effect and confidence intervals**
  - Risk Ratio = 0.09
  - 95% Confidence Interval 0.01 to 1.41
  - Not significant

So was this a fluke?

Included study	Outcome of interest	Event rate	RR (95% CI)	Fisher's p-value
Duvic (2001)	Rash	6/29 vs. 1/35	7.24 (0.92, 56.8)	0.04
Wolf (1985)	Fever	5/9 v. 0/9	11.0 (0.70,173.7)	0.03
Stadler (1998)	Adverse events	9/42 v. 2/40	4.29 (0.99,18.6)	0.049
Child (2004)	Improvement	0/8 v. 7/8	0.07 (0.00,1.00)	0.002

Is this phenomena seen in other CSG reviews?

Method 1: Survey of Skin Group full reviews

▸ Random sample of 20 skin group reviews published in the Cochrane Library

▸ Research question: What percent of analyses (Forest plots) are based on?
  ◦ A single study
  ◦ Denominator < 30
  ◦ No. of events < 10

Survey of Skin Group Reviews

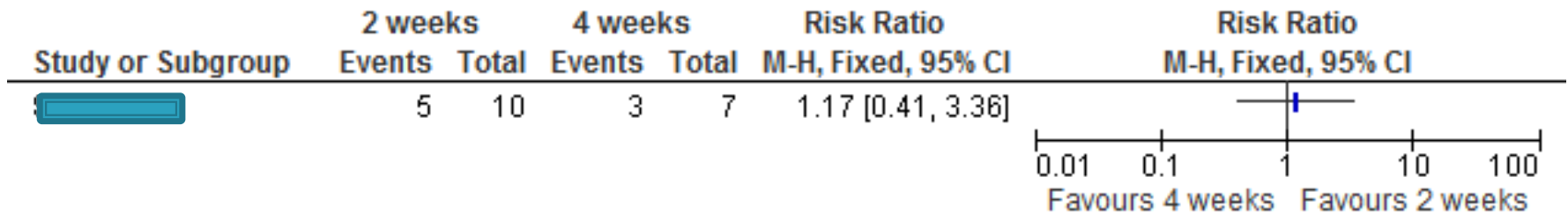|	No. of Forest plots	No. based on 1 study	Denominator* <30	No. with <10 events
All reviews	742 *(0–298)*	603 (81.3) *(0 –267)*	56 *(0 – 12)*	88 *(0 – 22)*
All reviews minus #5	444 *(0–123)*	336 (75.7) *(0 – 71)*	56 *(0 – 12)*	66 *(0– 18)*

* Number of randomised patients in both groups combined

So have small no. comparisons influenced overall findings?

- 20 reviews, 14 contained comparisons with small numbers

Results reported in abstract: 4 reviews reported data based on a single comparison with small numbers
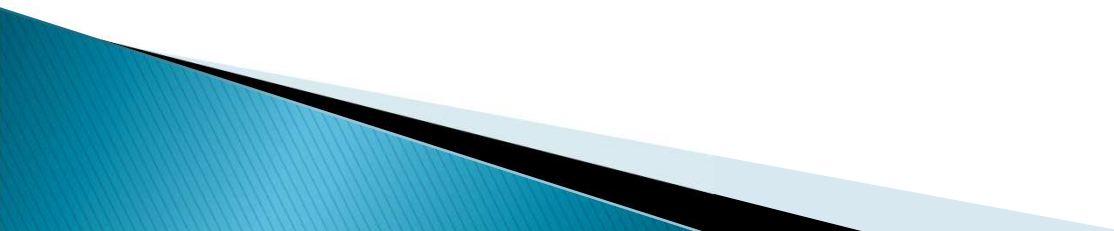
Implications for practice: 2 occasions where implications for practice were influenced by small no. comparisons

2.7 Cured (250 mg/2 weeks versus 250 mg/4 weeks)

Study or Subgroup	2 weeks Events	2 weeks Total	4 weeks Events	4 weeks Total	Risk Ratio M-H, Fixed, 95% CI	Risk Ratio M-H, Fixed, 95% CI
	5	10	3	7	1.17 [0.41, 3.36]	

0.01   0.1   1   10   100
Favours 4 weeks   Favours 2 weeks

Method 2: Email to Statistical Methods Group List

"My query relates to the inconsistency in approaches used to generate the results, where the author is reporting both an exact (Fisher's) p value and asymptotic based 95% confidence intervals for risk ratios from Rev Man. The issue is the discrepancy in the interpretation between the p value (which is less than 0.05) and the confidence intervals (which crosses 1)"
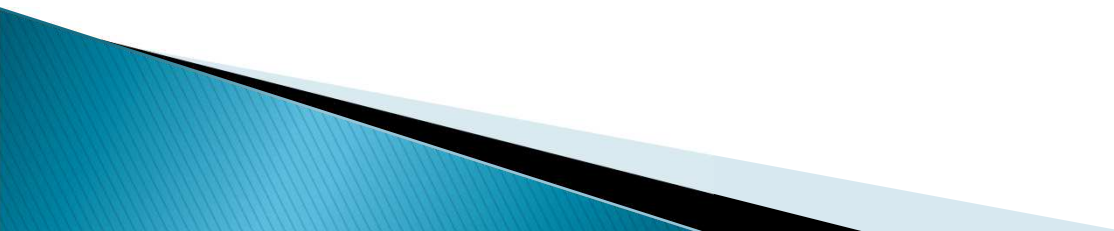
Responses

- Suggested formulae for confidence intervals which work well with small numbers
  - Robert Newcombe's Excel worksheet
  - Miettinen-Nurminen formula

- Philosophical suggestions
  - Should use odds ratios instead of risk ratios if concordance between confidence intervals and p-values is required
  - Warned against of over interpreting borderline results (i.e. where P is close to 0.05)

Comparison of two formulae for Mycosis fungoides review

	Event rates	P	Mantel–Haenszel	Miettinen–Nurminen
Thestrup–Pedersen (1982)	0/8 v. 5/8	0.03	0.09 (0.01, 1.41)	0.00 (0.00,0.61)
Duvic (2001)	6/29 v. 1/35	0.04	7.24 (0.92, 56.8)	7.24 (1.22,45.1)
Wolf (1985)	5/9 v. 0/9	0.03	11.0 (0.70,173.7)	NE (1.59, NE)
Stadler (1998)	9/42 v. 2/40	0.049	4.29 (0.99,18.6)	4.29 (1.13, 17.1)
Child (2004)	0/8 v. 7/8	0.002	0.07 (0.00,1.00)	0.00 (0.00, 0.40)

NE – Not estimable

Solution

- Measures of effect and 95% CIs should be used ideally to inform strength of association and

- Where there are results from individual studies based on small numbers of included participants individual studies are presented narratively and a two sided Fisher's exact p value is used to determine statistical significance

Next steps

- Student project
  - How common is this phenomena in other Cochrane reviews?
  - What methods are used to overcome the phenomena?
- Standard sentence to be added to protocol:
  - "Where results are estimated for individual studies with low numbers of outcomes (<10 in total) or where the total sample size is less than 30 participants, we will report the proportion of outcomes in each treatment group together with a p value from a Fisher's Exact test."
- Meta-analyses which contain 2 or more small studies are beyond the scope of this presentation