Forest Plots:
trying to see the wood, trees, and leaves

Dr Saeed Dastgiri

adapted from:Steff Lewis, Mike Clarke, BMJ 2001;322:1479–80.

Clinicians wishing to quickly answer a clinical question may seek a systematic review, rather than searching for primary articles.

Systematic review and meta analysis?

Quantitative approach for systematically combining results of previous research to arrive at conclusions about the body of research.

What does it mean?

- Quantitative : numbers
- Systematic : methodical
- combining: putting together
- previous research: what's already done
- conclusions: new knowledge

Table 1A: Relevant features of study design to be considered when deciding whether to pool studies in a systematic review

Patients	Interventions	Outcomes	Study methodologies

Table 1B: Relevant features of study design to be considered when deciding whether to pool studies in a systematic review examining the effect of antibiotics in patients with obstructive lung disease

Patients	Interventions	Outcomes	Study methodologies
Patient age Patient sex Type of lung disease (e.g., emphysema, chronic bronchitis)	Same antibiotic in all studies Same class of antibiotic in all studies Comparison of antibiotic with placebo Comparison of one antibiotic with another	Death Peak expiratory flow Forced expiratory volume in the first second	All randomized trials Only blinded randomized trials Cohort studies

TABLE 6. Summary of patient, drug, and size of study for studies included in the meta-analysis

Study/year	Study type	Pathology of mother	No. of cases	Corticosteroid(s)	Dose (prednisolone equivalent)
Popert '62	Cohort	Rheumatoid arthritis, SLE, ankylosing spondylitis, psoriatic arthropathy	22	Prednisolone, cortisone, corticotropin	2.5–27.5 mg/day
Warrell and Taylor '68	Cohort	Asthma, eczema, ulcerative colitis, SLE, urticaria, sarcoidosis	69	Prednisolone	2.5–40 mg/day
Heinonen et al., '77	Cohort	N/A	50,282	Corticosteroid and/or corticotropin	N/A
Mogadam et al., '81	Cohort	Inflammatory bowel disease	521	Corticosteroid or corticosteroid + sulfasalazine	N/A
Mintz et al., '86	Cohort	SLE	204	Prednisone	10 mg/day
Robert et al., '94	Case control	N/A	1,448	Corticosteroids	N/A
Czeizel and Rockenbauer '97	Case control	Asthma, hay fever, rheumatoid arthritis, Addison's disease, subfertility	56,557	Dexamethasone, prednisone, cortisone, betamethasone, methylprednisolone, triamcinolone	5–100 mg/day
Rodriguez-Pinilla and Martinez-Frias, '98	Case control	N/A	12,304	Prednisolone, hydrocortisone, prednisone, triamcinolone	10–30 mg/day
Carmichael and Shaw, '99	Case control	Crohn's disease, asthma, lupus	1,396	Prednisone, cortisone, dexamethazone, triamcinolone	N/A
Park-Wyllie et al., '00 (present study)	Cohort	Crohn's disease, ulcerative colitis, rheumatoid arthritis, SLE, and other	372	Prednisone	5–80 mg/day

SLE, systemic lupus erythematosus; N/A, not available.

Assessment of Quality and Selection of Studies

- Quality varies, therefore
  - Standardized Assessment (?blind*)
  - Group/Rank by quality
- Select a threshold, e.g. all prospective studies with blind reading of reference and index tests.

* assessment of quality blind to study outcome

Assessing a Study of a Test

(Jaeschke et al, JAMA, 1994, 271: 389-91)

- Was an appropriate spectrum of patients included?
  - (Spectrum Bias)
- All patients subjected to a Gold Standard?
  - (Verification Bias)
- Was there an independent, "blind" comparison with a Gold Standard?
  - Observer Bias; Differential Reference Bias
- Methods described so you could repeat test?

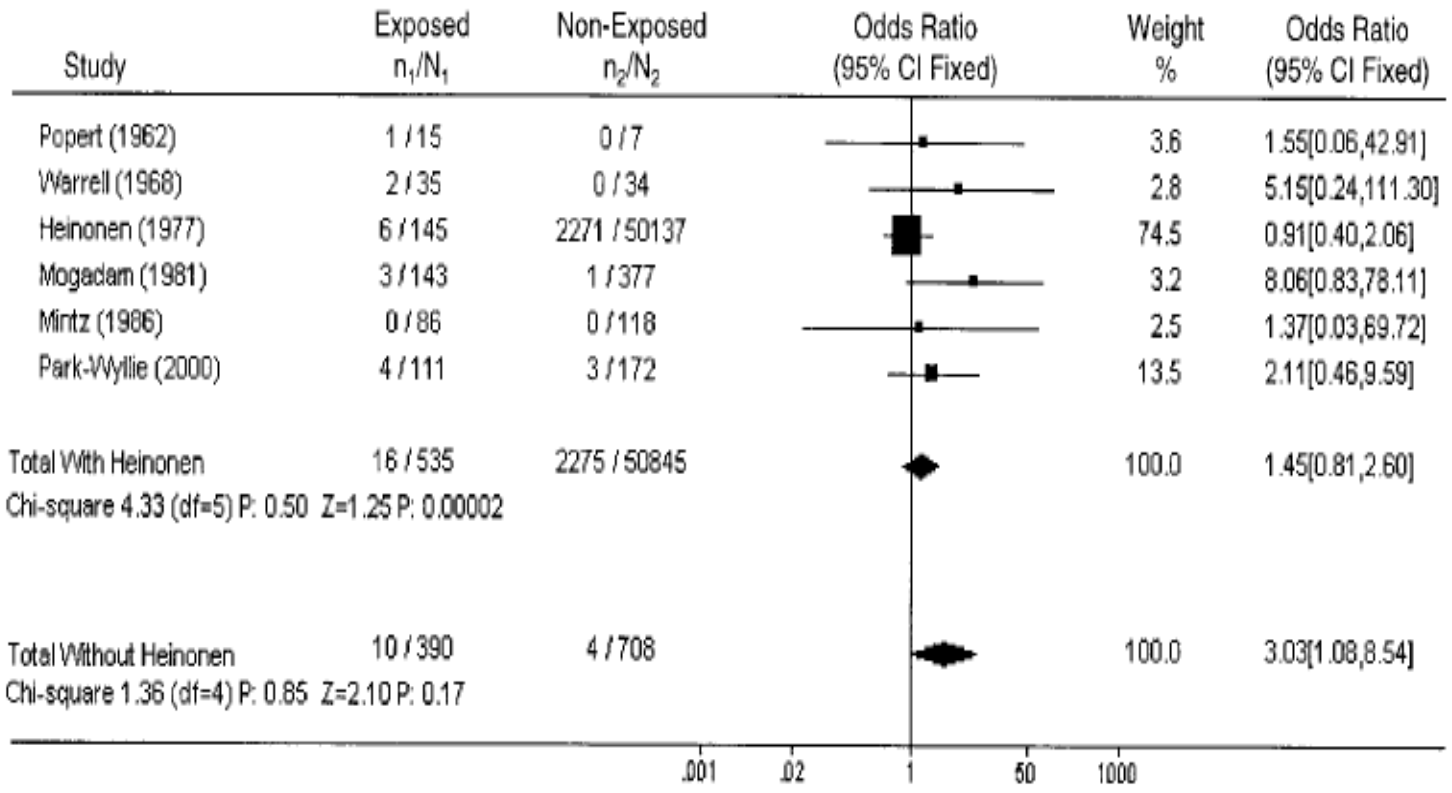


Fig. 1. Individual and cumulative Mantel-Haenszel summary odds ratio for corticosteroid-exposed cohort studies for major malformations with and without the Heinonen et al. ('77) analysis.

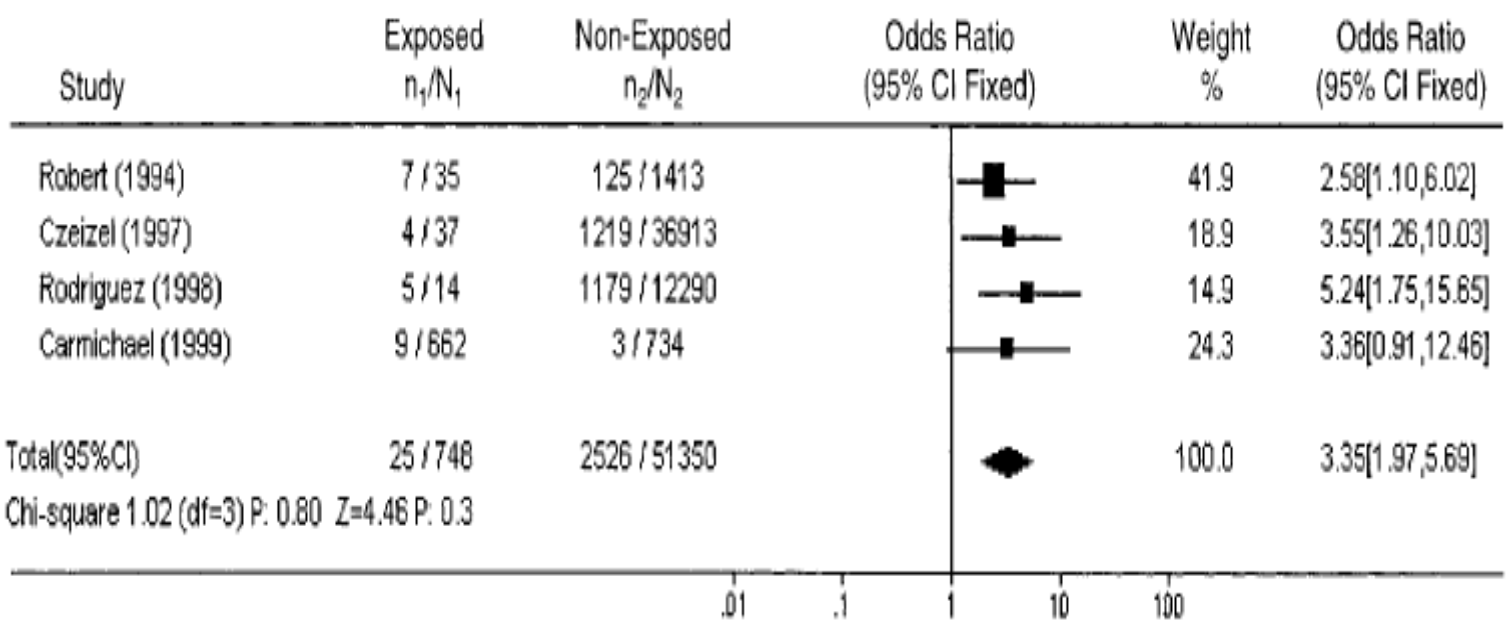


Fig. 2. Individual and cumulative Mantel-Haenszel summary odds ratio for corticosteroid-exposed case-control studies focusing on oral clefts.

Our discussion focuses on the qualitative, rather than the statistical (Cochran's Q test and I^2 statistic), assessment of heterogeneity.

This is an approach to evaluating potentially important differences in the results of individual studies being considered for a meta-analysis. These differences are frequently referred to as heterogeneity.

Weighting studies

- More weight to the studies which give us more information
  - More participants
  - More events
  - More precision

- Weight is proportional to the precision

Does it make sense to combine?

- Do we need studies to be exactly the same?
- When can we say we are measuring the same thing?

Are the studies consistent?

- Are variations in results between studies consistent with chance?

  (Test of homogeneity: has low power)

- If NO, then WHY?
  - Variation in study methods (biases)
  - Variation in intervention
  - Variation in outcome measure (e.g. timing)
  - Variation in population

Two concepts are commonly implied in the assessment of heterogeneity.

The first is an assessment for heterogeneity within 4 key elements of the design of the original studies:

- patients,
- interventions,
- Outcomes, and
- methods.

The second concept relates to assessing heterogeneity among the results of the original studies.

Even if the study designs are similar, the researchers must decide whether it is useful to combine the primary studies' results.

What is the clinician to do when presented with results such as those in these Figures?

- differences in patients (effects may be larger in sicker patients),
- in interventions (larger doses may result in larger effects),
- in outcomes (longer follow-up may diminish the magnitude of effect) and,
- in study design (methodologically weaker studies may generate larger effects).

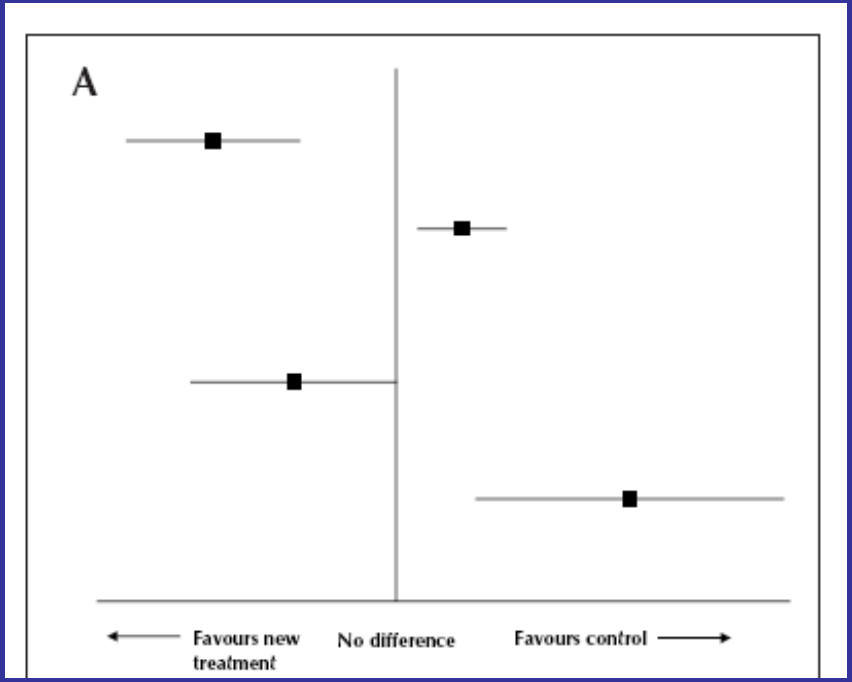
The investigators will then have to examine the extent to which these hypotheses can explain the differences in magnitude of effect across studies. This is called subgroup analyses. This may also be misleading.

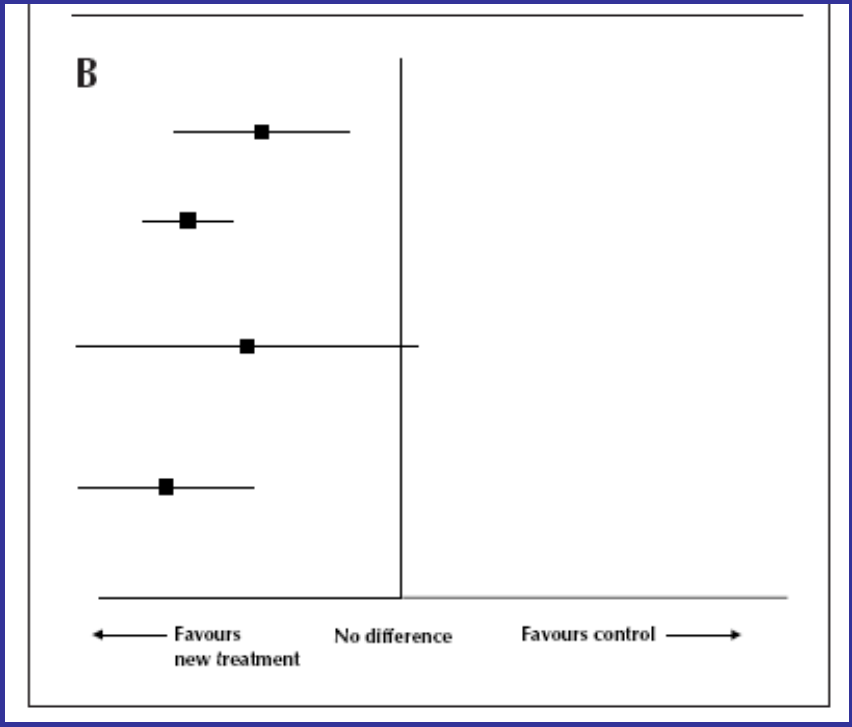
Box 2: Questions to ask when evaluating a subgroup analysis in a meta-analysis[10]

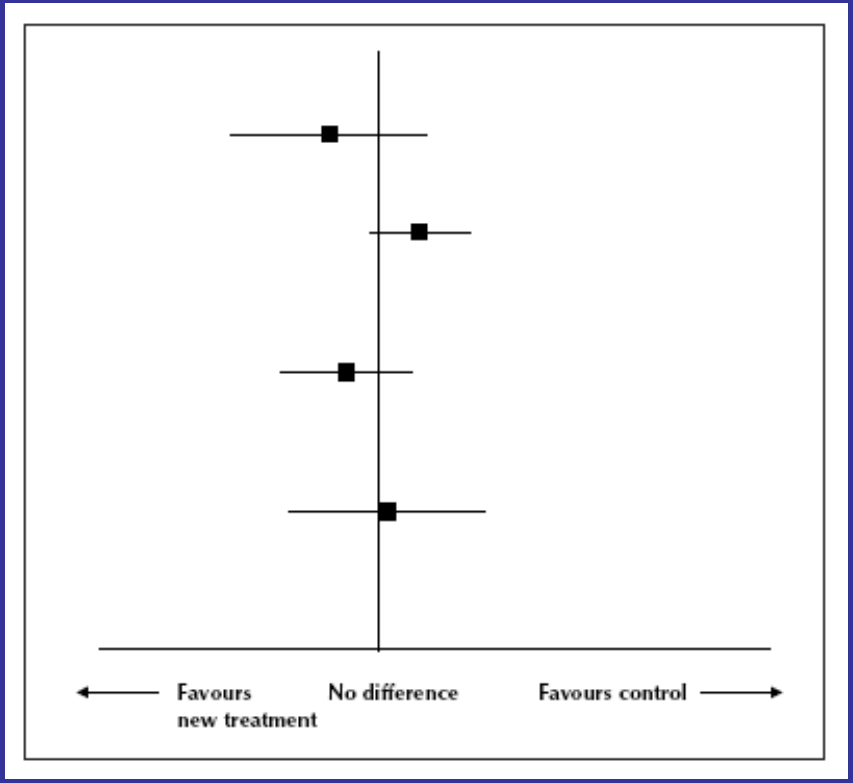
- Was the subgroup comparison based on a within-study, rather than a between-study, comparison?
- Is the magnitude of the difference in effect between subgroups large?
- Is the effect consistent across studies?
- Is the difference in effect statistically significant?
- Was the subgroup analysis planned in advance by the trialists?
- Were many subgroup analyses performed and selectively reported?
- Is the difference in effect in the subgroup supported by a biological hypothesis?

There are two criteria for not combining the results of studies in a meta-analysis:

- Highly disparate point estimates and,
- Confidence intervals with little overlap.


A
Favours new treatment
No difference
Favours control


B
Favours
new treatment
No difference
Favours control


Favours
new treatment
No difference
Favours control

Although statisticians (and statistical software) can calculate 95% confidence intervals, clinicians can readily estimate the upper boundary of confidence intervals for proportions with very small numerators.

Table 1A: Relative risk and relative risk reduction observed in 5 successively larger hypothetical trials

Control event rate	Treatment event rate	Relative risk, %	Relative risk reduction, %*
2/4	1/4	50	50
10/20	5/20	50	50
20/40	10/40	50	50
50/100	25/100	50	50
500/1000	250/1000	50	50

*Calculated as the absolute difference between the control and treatment event rates (expressed as a fraction or a percentage), divided by the control event rate. In the first row in this table, relative risk reduction = (2/4 –1/4) ÷ 2/4 = 1/2 or 50%. If the control event rate were 3/4 and the treatment event rate 1/4, the relative risk reduction would be (3/4 – 1/4) ÷ 3/4 = 2/3. Using percentages for the same example, if the control event rate were 75% and the treatment event rate were 25%, the relative risk reduction would be (75% – 25%) ÷ 75% = 67%.

Table 1B: Confidence intervals (CIs) around the relative risk reduction in 5 successively larger hypothetical trials

Control event rate	Treatment event rate	Relative risk, %	Relative risk reduction, %	CI around relative risk reduction, %	
				Intuitive CI*	Calculated 95% CI*†
2/4	1/4	50	50	–50 to 90	–174 to 92
10/20	5/20	50	50	–20 to 90	–14 to 79.5
20/40	10/40	50	50	0 to 90	9.5 to 73.4
50/100	25/100	50	50	20 to 80	26.8 to 66.4
500/1000	250/1000	50	50	40 to 60	43.5 to 55.9

*Negative values represent an increase in risk relative to control. See text for further explanation.
†Calculated by statistical software.

Table 2: The 3/*n* rule to estimate the upper limit of the 95% confidence interval (CI) for proportions with 0 in the numerator

n	Observed proportion	3/*n*	Upper limit of 95% CI
20	0/20	3/20	0.15 or 15%
100	0/100	3/100	0.03 or 3%
300	0/300	3/300	0.01 or 1%
1000	0/1000	3/1000	0.003 or 0.3%

Table 1: Method for obtaining an approximation of the upper limit of the 95% confidence interval (CI)*

Observed numerator	Numerator for calculating approximate upper limit of 95% CI
0	3
1	5
2	7
3	9
4	10

*For any observed numerator listed in the left hand column, the learner substitutes the numerator in the right hand column. When this value is divided by the number of study subjects, the learner obtains a reasonable approximation of the upper limit of the 95% CI. For example, if the sample size is 15 and the observed numerator is 3, the upper limit of the 95% confidence interval is approximately 9 ÷ 15 = 0.6 or 60%.

Summary points

Forest plots show the information from the individual studies that went into the meta-analysis at a glance

They show the amount of variation between the studies and an estimate of the overall result

Forest plots, in various forms, have been published for about 20 years

During this time, they have been improved, but it is still not easy to draw them in most standard computer packages

References

- Lewis, Mike Clarke. Forest plots: trying to see the wood and the trees, BMJ 2001;322:1479–80.

- Barratt A, Wyer PC, Hatala R, McGinn T, Dans AL, Keitz S, et al. Tips for teachers of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. Available: www.cmaj.ca/ cgi/content/full/171/4/353/DCI.

- Montori VM, Kleinbart J, Newman TB, Keitz S, Wyer PC, Moyer V, et al. Tips for teachers of evidence-based medicine: 2. Confidence intervals and p values. Available: www.cmaj.ca/cgi/content/full/171/6/611/DCI.

- McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, Guyatt G, et al. Tips for teachers of evidence-based medicine: 3. Understanding and calculating kappa. Available: www. cmaj.ca/cgi/content/full/171/11/1369/DC1.

- Hatala R, Keitz S, Wyer P, Guyatt G. Tips for teachers of evidence-based medicine: 4. Assessing heterogeneity of primary studies in systematic reviews and whether to combine their results. Available: www.cmaj.ca/cgi/content/full/172/5/661/DCI.

- Victor M. Montori, Peter Wyer, Thomas B. Newman, Sheri Keitz, Gordon Guyatt, Tips for teachers of evidence-based medicine: 5. The effect of spectrum of disease on the performance of diagnostic tests. Available: www.cmaj.ca/cgi/content/full/173/4/385/DC1.

با تشکر از توجه شما