Interpreting Statistical Results

PAROS Literature Review Workshop
11-12 April 2011, Khoo Teck Puat Hospital, Singapore

Dianne Bautista, PhD
Singapore Clinical Research Institute

Very important!

- Assess bias before interpreting results. Biased studies mislead.
- The treatment effect is worth considering if study is valid
- For RCTs of interventions: adequacy of randomization (sequence generation and allocation concealment), blinding (subjects, care providers, assessors), completeness of follow-up, selective reporting

Example: RCT

Does Enalapril lower mortality compared with Hydralazine + Nitrates in men with congestive heart failure?

- Population: Men with congestive heart failure
- Intervention: Enalapril (ACE inhibitor)
- Comparator: Hydralazine + nitrates (H+N)
- Outcome: Mortality

Group	Deaths	Total	Risk of death
Enalapril	132	403	132/403= 33%
H + N	153	401	153/401= 38%

Hypothesis testing with p-values

- Null hypothesis:
  - There is no difference in the mortality rate
  - p-value = 0.11 > 0.05 (level of significance)
- Conclusion:
  - Do not reject the null hypothesis.
  - There is insufficient evidence to show that Enalaparil reduces mortality compared to H+N among men with congestive heart failure.

Remarks on p-values

- Hypothesis testing using a p-value is a binary (Reject/Do not reject Null) decision.
- Reject Null → "statistically significant"
- p-values do not provide info on direction or size of the treatment effect
- Issue: why make the question of efficacy a dichotomy (Reject/Don't reject) when it may be appropriate to view it as a continuum?

Confidence intervals

- What is the single value most likely to represent the true difference between intervention and control?

  Enalapril: 33%; H+N: 38%

  Absolute risk difference: 33% - 38% = **-5%**

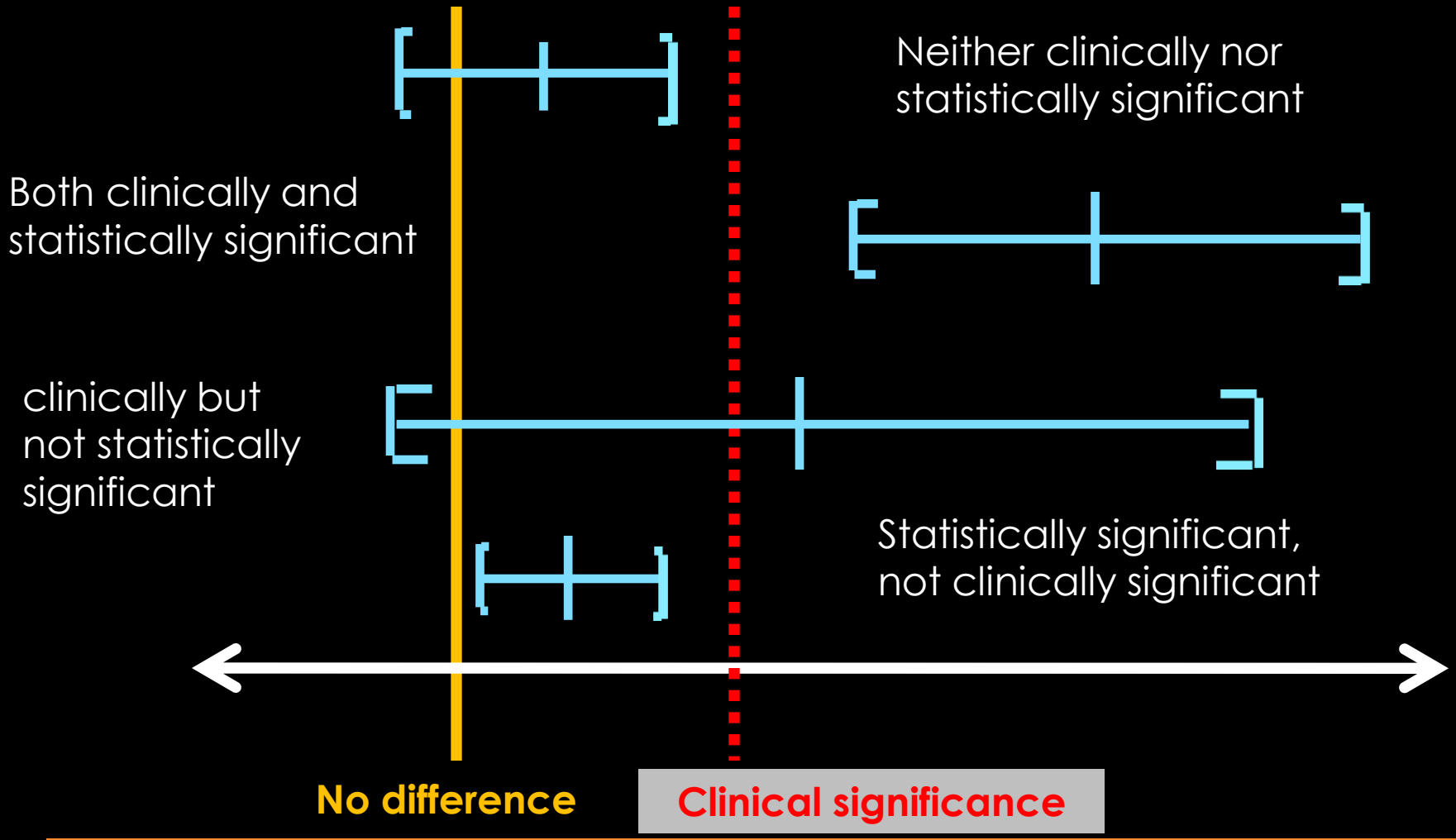
- What is the plausible range of differences within which the difference may lie?

  95% CI: -1.2% to 12%

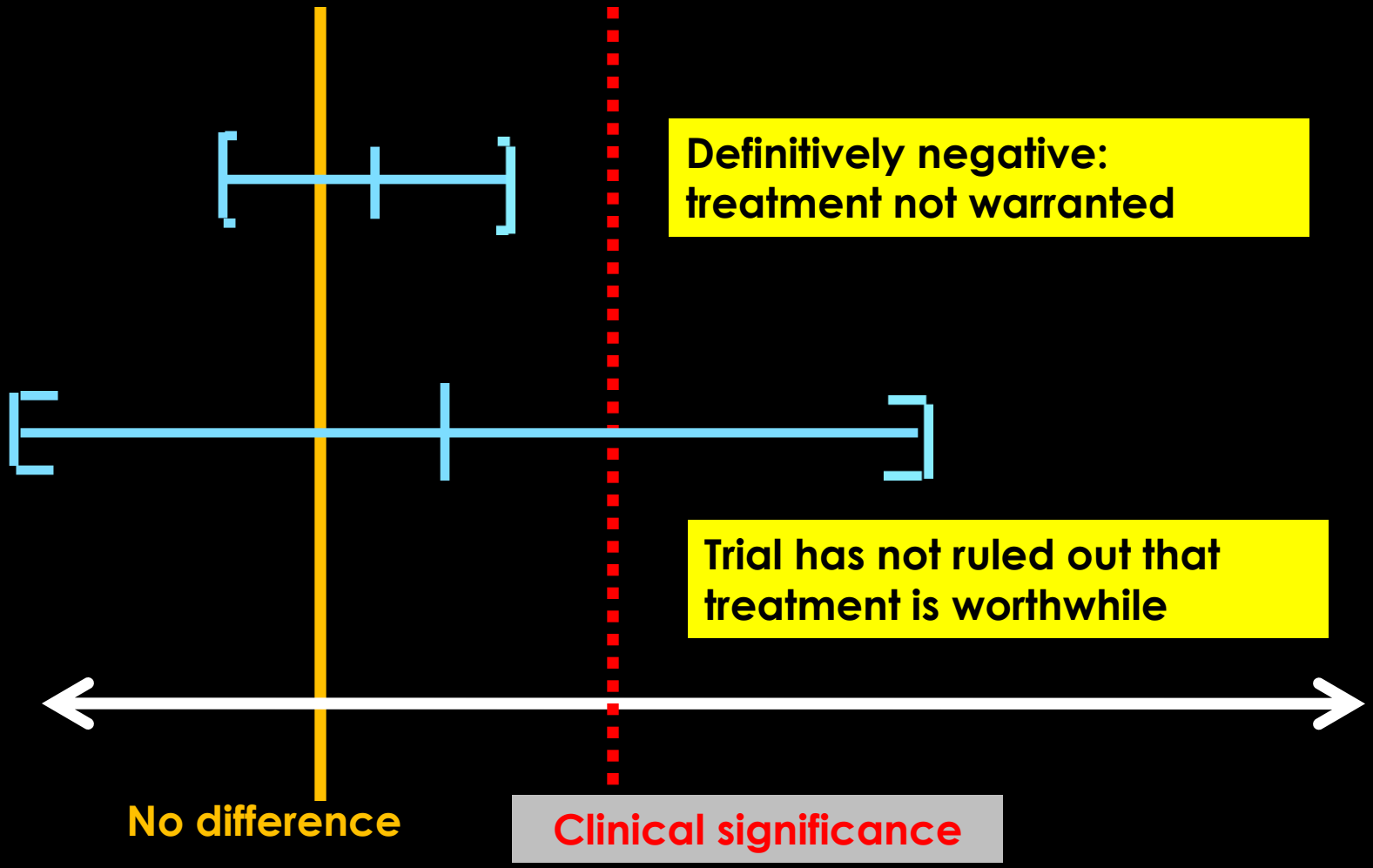
Confidence interval

- Conclusion:
  - Patients offered ACE inhibitors will most likely (but not certainly) die later than patients given H+N.
  - However the size of the difference in expected survival may be trivial or large
  - All else being equal, an ACE inhibitor is the appropriate choice for patients with heart failure, but the evidence is not definitive.

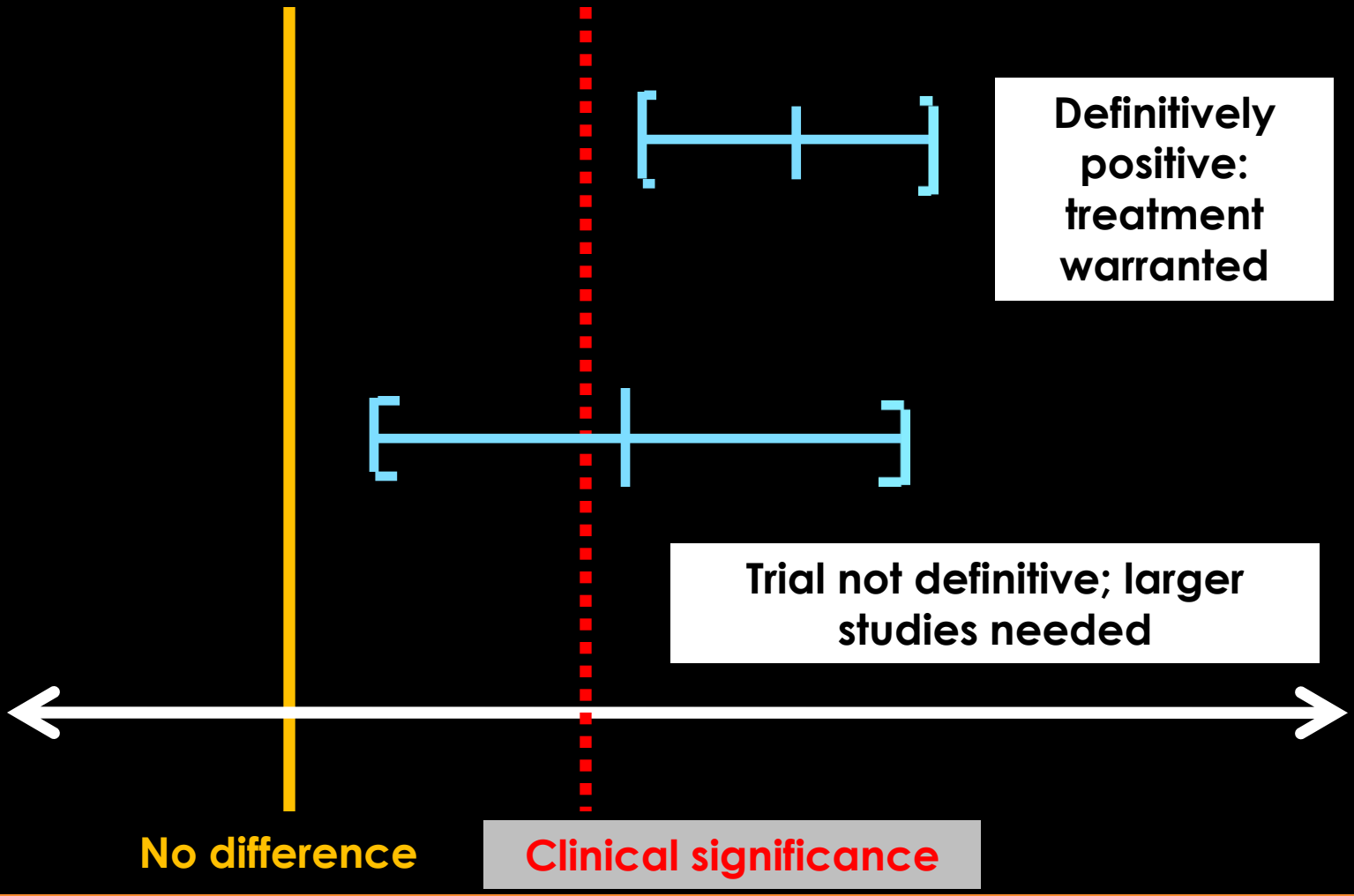
Clinical versus statistical significance



Interpreting "negative" results



Interpreting "positive" results



Assessing Effects of the Treatment

- How can we express the magnitude of the relationship between **I** and **O**?
  - Binary : risk difference, RR, OR, NNT
  - Time-to-event : hazard ratio
  - Continuous : mean difference
- No treatment effect is expressed
  - Difference = 0
  - Ratio = 1

2 x 2 table

- Binary / Dichotomous outcomes
  - We usually compare the number of patients who experienced the "event"
  - Bad outcomes: occurrence of stroke (Y/N), MI, death
  - Good outcomes: resolution of symptoms (Y/N), ulcer healing
  - Even continuous outcomes can be dichotomized: improvement in FEV1 of more than 20% over baseline (Y/N)

Risk Ratio (RR)

Intervention	Outcome, Number of patients		Total patients treated	Risk of death
	Death	Survival		
Int = Ligation	18	46	64	18/64 = 28.1%
Sclerotherapy	29	36	65	29/65 =44.6%
Risk ratio (RR) = 28.1%/ 44.6% = 63%				

Risk ratio also known as relative risk

Interpretation: The risk of death after ligation is about two-thirds as great as the risk of death after sclerotherapy.

Risk Difference (RD)

Intervention	Outcome, Number of patients		Total patients treated	Risk of death
	Death	Survival		
Int = Ligation	18	46	64	18/64 = 28.1%
Sclerotherapy	29	36	65	29/65 =44.6%
Risk difference (RD) = 16.5%, in favor of ligation				

Risk difference is also known as absolute risk reduction or ARR

Interpretation: Treating with ligation rather than sclerotherapy will save the lives of about 16 of 100 patients.

Odds ratio(OR)

Intervention	Outcome, Number of patients		Total patients treated	Odds of death
	Death	Survival		
Int = Ligation	18	46	64	18/46 = 0.39
Sclerotherapy	29	36	65	29/36 =0.80
Odds ratio(OR) = 0.39 / 0.80 = 0.49				

Odds = "piece of the pie" / "rest of the pie"

Interpretation: The odds of death after ligation are half the odds of death after sclerotherapy

RR vs OR vs RD

- Consider two cases with RR=33%
  - Case 1: reduction of risk from 3% to 1%
  - Case 2: reduction of risk 60% to 20%
- But clinical implications may be different if 5% of patients experience side effects
  - Case 1: therapy not worth instituting
  - Case 2: trade-off worthwhile
- Whereas RD gives info on absolute risk, RR and OR do not
- RD and NNT, may be most useful for deciding whether to institute intervention or not

Number needed to treat (NNT)

Intervention	Outcome, Number of patients		Total patients treated	Risk of death
	Death	Survival		
Int = Ligation	18	46	64	18/64 = 28.1%
Sclerotherapy	29	36	65	29/65 =44.6%
NNT = 100/16.5 ≈ 6				

Risk difference is also known as absolute risk reduction or ARD

Interpretation: We need to treat 6 patients with ligation to prevent one death

Meta-analyses

- What is a meta-analysis?
  - ✓ Statistical synthesis of results from a series of studies
  - ✓ Optional part of a systematic review
- Why perform a meta-analysis
  - ✓ To increase power
  - ✓ To improve precision
  - ✓ To answer questions not answered by individual studies
  - ✓ To settle controversies arising from conflicting studies

Framework for synthesis

1. What is the direction of the effect?
2. What is the size of the effect?
3. Is the effect consistent across studies?
4. What is the strength of evidence for the effect?

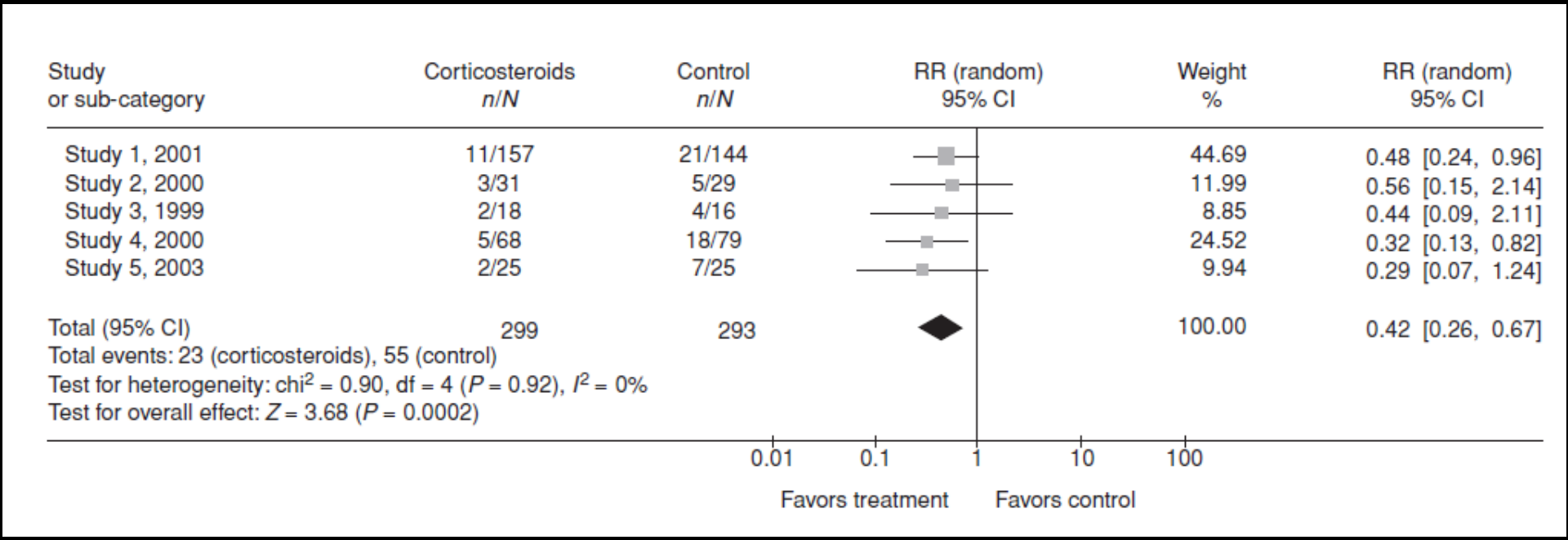
Heterogeneity

- **Clinical heterogeneity**
  - Participants
    - ✓ Age, sex, co-morbidities, disease severity, medication status at start, eligibility criteria, geographical variation
  - Interventions and Comparators
    - ✓ Dose, duration, type of drug, mode of administration, nature of control (none, placebo, standard care)
  - Outcomes
    - ✓ follow-up duration, definition of an event, ways of measuring outcomes

Heterogeneity

- **Methodological heterogeneity**
  - Study design
    - ✓ Randomized vs. non-randomized, parallel group vs. crossover, individual vs. cluster randomized
  - Conduct
    - ✓ Allocation concealment, blind outcome assessment
  - Analysis
    - ✓ ITT vs per protocol, unit of analysis, imputation methods for missing data

Forest plot

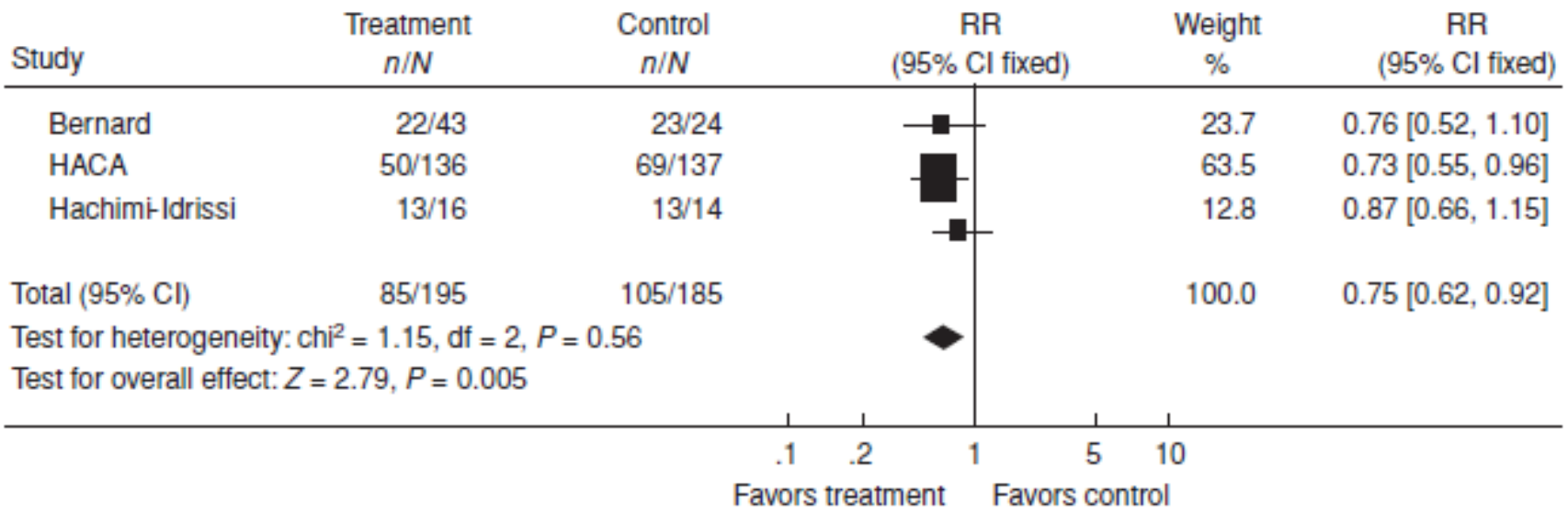

Study
or sub-category
Corticosteroids
n/N
Control
n/N
RR (random)
95% CI
Weight
%
RR (random)
95% CI
Study 1, 2001 11/157 21/144 44.69 0.48 [0.24, 0.96]
Study 2, 2000 3/31 5/29 11.99 0.56 [0.15, 2.14]
Study 3, 1999 2/18 4/16 8.85 0.44 [0.09, 2.11]
Study 4, 2000 5/68 18/79 24.52 0.32 [0.13, 0.82]
Study 5, 2003 2/25 7/25 9.94 0.29 [0.07, 1.24]
Total (95% CI) 299 293 100.00 0.42 [0.26, 0.67]
Total events: 23 (corticosteroids), 55 (control)
Test for heterogeneity: chi² = 0.90, df = 4 (P = 0.92), I² = 0%
Test for overall effect: Z = 3.68 (P = 0.0002)
0.01
0.1
1
10
100
Favors treatment
Favors control

Example

Focused question: In adults who achieve ROSC following cardiac arrest, does the administration of therapeutic hypothermia (intervention) increase the rate of survival to hospital discharge compared to supportive care?

Search strategy:
MEDLINE and EMBASE: (heart arrest OR cardiopulmonary resuscitation) AND (hypothermia, induced OR circulatory arrest, deep hypothermia induced)

Example



Comparison of in-hospital mortality between patients treated with mild hypothermia and control groups in three clinical trials

SINGAPORE CLINICAL RESEARCH INSTITUTE

END

Analysis 1.1. Comparison 1 Exercise versus no intervention - general population, Outcome 1 Anxiety.

Review: Exercise in prevention and treatment of anxiety and depression among children and young people

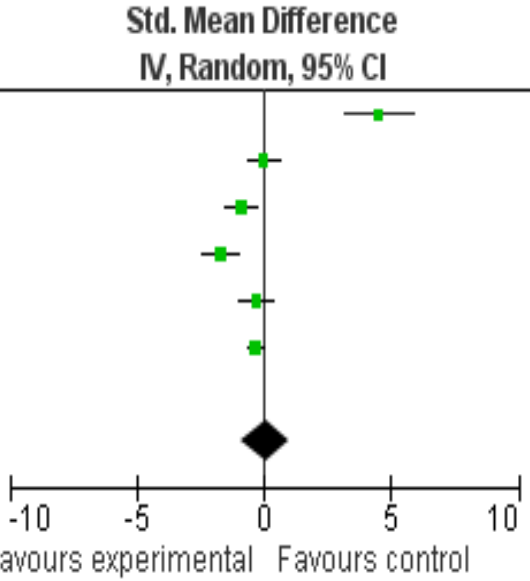
Comparison: 1 Exercise versus no intervention - general population

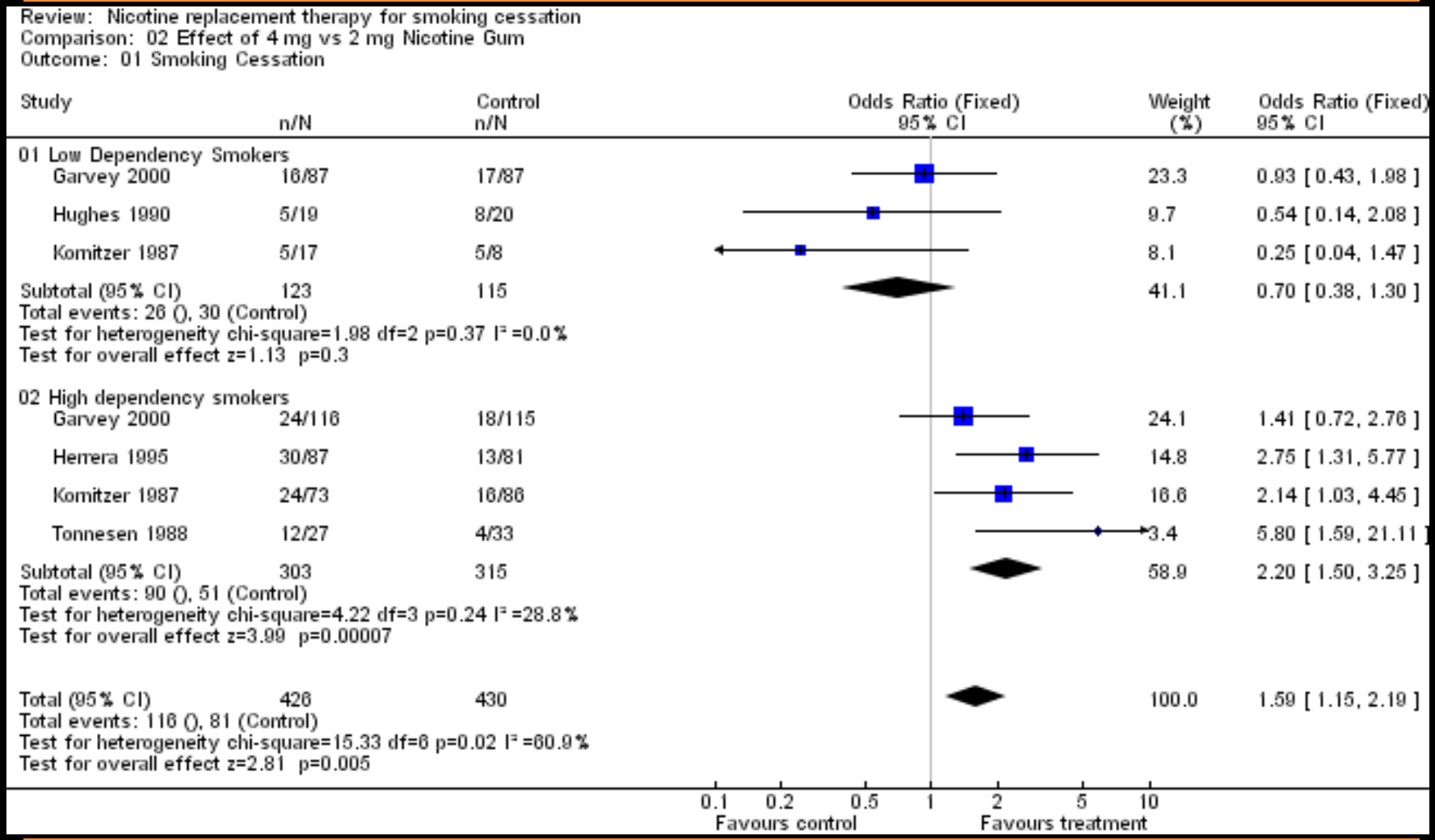
Outcome: 1 Anxiety

	Treatment			Control				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Smith 1983	35.69	8.06	16	0	7.31	16	13.4%	4.52 [3.15, 5.89]
Roth 1987	36.8	8.9	18	37	9.1	18	17.2%	-0.02 [-0.68, 0.63]
Jacobs 1984	30.77	7.68	22	38.23	8.47	22	17.3%	-0.91 [-1.53, -0.28]
Hilyer 1982	29.17	4.39	23	39.1	6.96	20	16.9%	-1.70 [-2.41, -0.99]
Carl 1984	32.12	7.78	15	34.13	4.92	16	16.9%	-0.30 [-1.01, 0.41]
Berger 1988	6.49	6.3	66	8.64	7.37	87	18.3%	-0.31 [-0.63, 0.01]
Total (95% CI)			**160**			**179**	**100.0%**	**0.05 [-0.89, 0.99]**

Heterogeneity: $Tau^2 = 1.22$; $Chi^2 = 66.60$, $df = 5$ ($P < 0.00001$); $I^2 = 92\%$

Test for overall effect: $Z = 0.10$ ($P = 0.92$)




Review: Nicotine replacement therapy for smoking cessation
Comparison: 02 Effect of 4 mg vs 2 mg Nicotine Gum
Outcome: 01 Smoking Cessation
Study
n/N
Control
n/N
Odds Ratio (Fixed) 95% CI
Weight (%)
Odds Ratio (Fixed) 95% CI
01 Low Dependency Smokers
Garvey 2000 16/87 17/87 23.3 0.93 [0.43, 1.98]
Hughes 1990 5/19 8/20 9.7 0.54 [0.14, 2.08]
Komitzer 1987 5/17 5/8 8.1 0.25 [0.04, 1.47]
Subtotal (95% CI) 123 115 41.1 0.70 [0.38, 1.30]
Total events: 26 (), 30 (Control)
Test for heterogeneity chi-square=1.98 df=2 p=0.37 I² =0.0%
Test for overall effect z=1.13 p=0.3
02 High dependency smokers
Garvey 2000 24/116 18/115 24.1 1.41 [0.72, 2.76]
Herrera 1995 30/87 13/81 14.8 2.75 [1.31, 5.77]
Komitzer 1987 24/73 16/86 16.6 2.14 [1.03, 4.45]
Tonnesen 1988 12/27 4/33 3.4 5.80 [1.59, 21.11]
Subtotal (95% CI) 303 315 58.9 2.20 [1.50, 3.25]
Total events: 90 (), 51 (Control)
Test for heterogeneity chi-square=4.22 df=3 p=0.24 I² =28.8%
Test for overall effect z=3.99 p=0.00007
Total (95% CI) 426 430 100.0 1.59 [1.15, 2.19]
Total events: 116 (), 81 (Control)
Test for heterogeneity chi-square=15.33 df=6 p=0.02 I² =60.9%
Test for overall effect z=2.81 p=0.005
0.1 0.2 0.5 1 2 5 10
Favours control
Favours treatment