Meta-analysis

David M. Thompson

Director, Training Unit
Biostatistics and Epidemiology (BERD) Core
OSCTR

SCTR
Oklahoma Shared Clinical & Translational Resources

osctr.ouhsc.edu

Biostatistics, Epidemiology, and Research Design (BERD) Core

http://osctr.ouhsc.edu/biostatisticsepidemiology-core

BERD Training Unit

http://osctr.ouhsc.edu/training-unit

Systematic reviews and meta-analysis

A systematic review
is a review of a ***clearly formulated question***
that uses systematic and explicit methods
to identify, select, and critically appraise relevant research,
and to collect and analyze data from the studies that are included in the review.

Meta-analysis
refers to the use of statistical techniques in a systematic review to analyze, summarize and integrate the results of included studies.

A systematic review may or may not include a meta-analysis..

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097.

PRISMA checklist (Moher et al., 2009)

Objectives, which parallel elements of a clearly formulated clinical question

- **Patient type**
- **Intervention**
- **Comparison**
- **Outcome**
- **Study Design**

Studies or sources of data

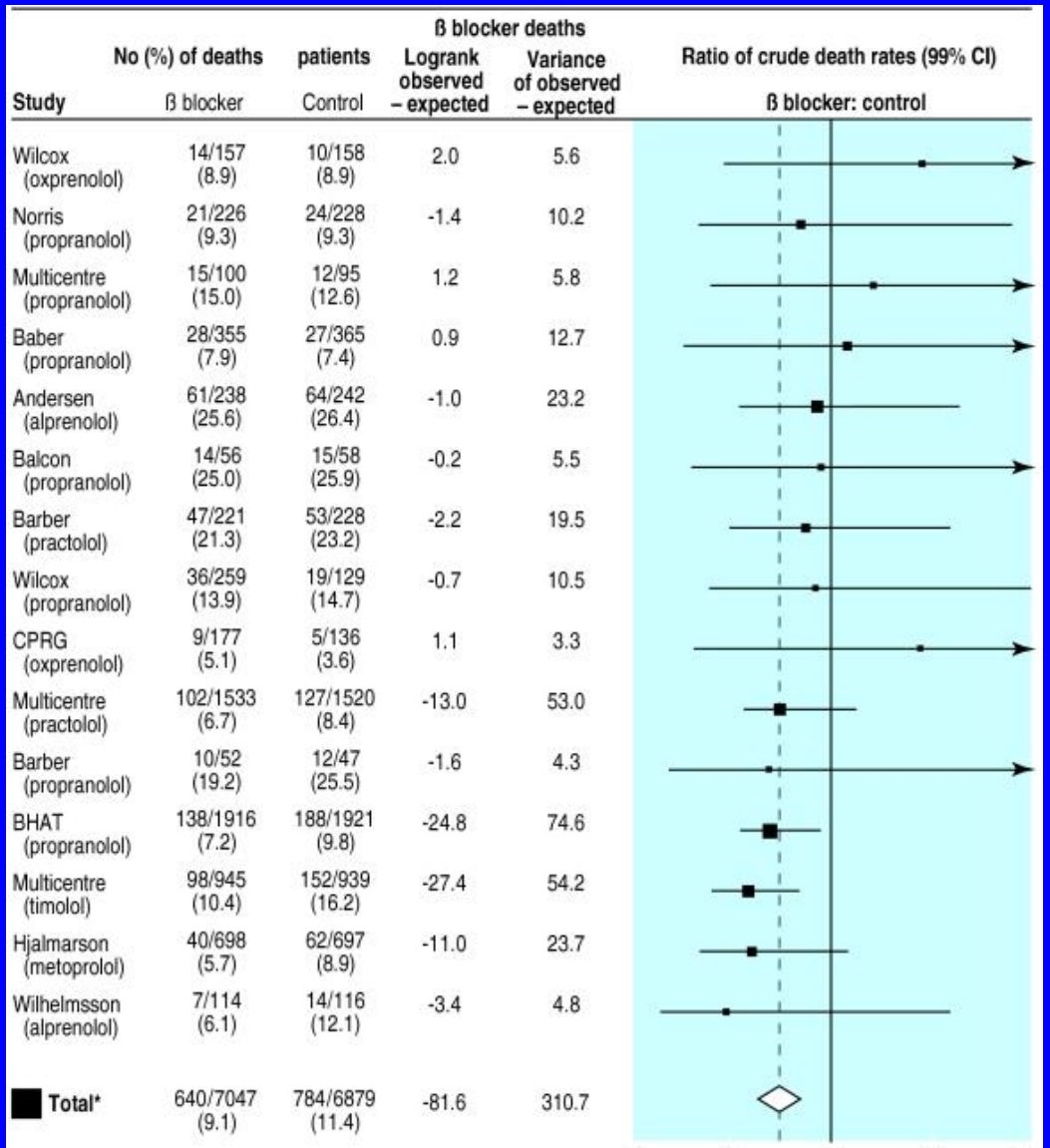
- **data sources and dates searched**
- **replicable electronic search strategy**
- **accounting for "gray literature" and publication bias**

Criteria for inclusion of studies

Quantitative methods

- **Principal summary measures (e.g. relative risk, difference in means)**
- **Results of individual studies, ideally with a forest plot**
- **Methods of combining results, and measure of consistency among studies that shared similar outcome measures, inclusion criteria, type and duration of treatment**
- **Synthesis of results, including confidence interval and measure of consistency**

Results of individual studies reported using “forest plots”



Study	No (%) of deaths ß blocker	patients Control	ß blocker deaths Logrank observed – expected	Variance of observed – expected
Wilcox (oxprenolol)	14/157 (8.9)	10/158 (8.9)	2.0	5.6
Norris (propranolol)	21/226 (9.3)	24/228 (9.3)	-1.4	10.2
Multicentre (propranolol)	15/100 (15.0)	12/95 (12.6)	1.2	5.8
Baber (propranolol)	28/355 (7.9)	27/365 (7.4)	0.9	12.7
Andersen (alprenolol)	61/238 (25.6)	64/242 (26.4)	-1.0	23.2
Balcon (propranolol)	14/56 (25.0)	15/58 (25.9)	-0.2	5.5
Barber (practolol)	47/221 (21.3)	53/228 (23.2)	-2.2	19.5
Wilcox (propranolol)	36/259 (13.9)	19/129 (14.7)	-0.7	10.5
CPRG (oxprenolol)	9/177 (5.1)	5/136 (3.6)	1.1	3.3
Multicentre (practolol)	102/1533 (6.7)	127/1520 (8.4)	-13.0	53.0
Barber (propranolol)	10/52 (19.2)	12/47 (25.5)	-1.6	4.3
BHAT (propranolol)	138/1916 (7.2)	188/1921 (9.8)	-24.8	74.6
Multicentre (timolol)	98/945 (10.4)	152/939 (16.2)	-27.4	54.2
Hjalmarson (metoprolol)	40/698 (5.7)	62/697 (8.9)	-11.0	23.7
Wilhelmsson (alprenolol)	7/114 (6.1)	14/116 (12.1)	-3.4	4.8
Total*	640/7047 (9.1)	784/6879 (11.4)	-81.6	310.7

Ratio of crude death rates (99% CI) ß blocker: control

0 0.5 1.0 1.5 2.0

ß blocker better | ß blocker worse

Reduction 23.1% (SE5.0) P<0.0001

Heterogeneity between 15 trials: $\chi^2 = 13.9$; df = 14; P > 0.1

Treatment effect P < 0.0001

* 95% confidence interval as shown for the odds ratio

Lewis, S., & Clarke, M. (2001). Forest plots: Trying to see the wood and the trees. *BMJ, 322(7300):* 1479–1480.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1120528

Example 1
Fixed effects analysis

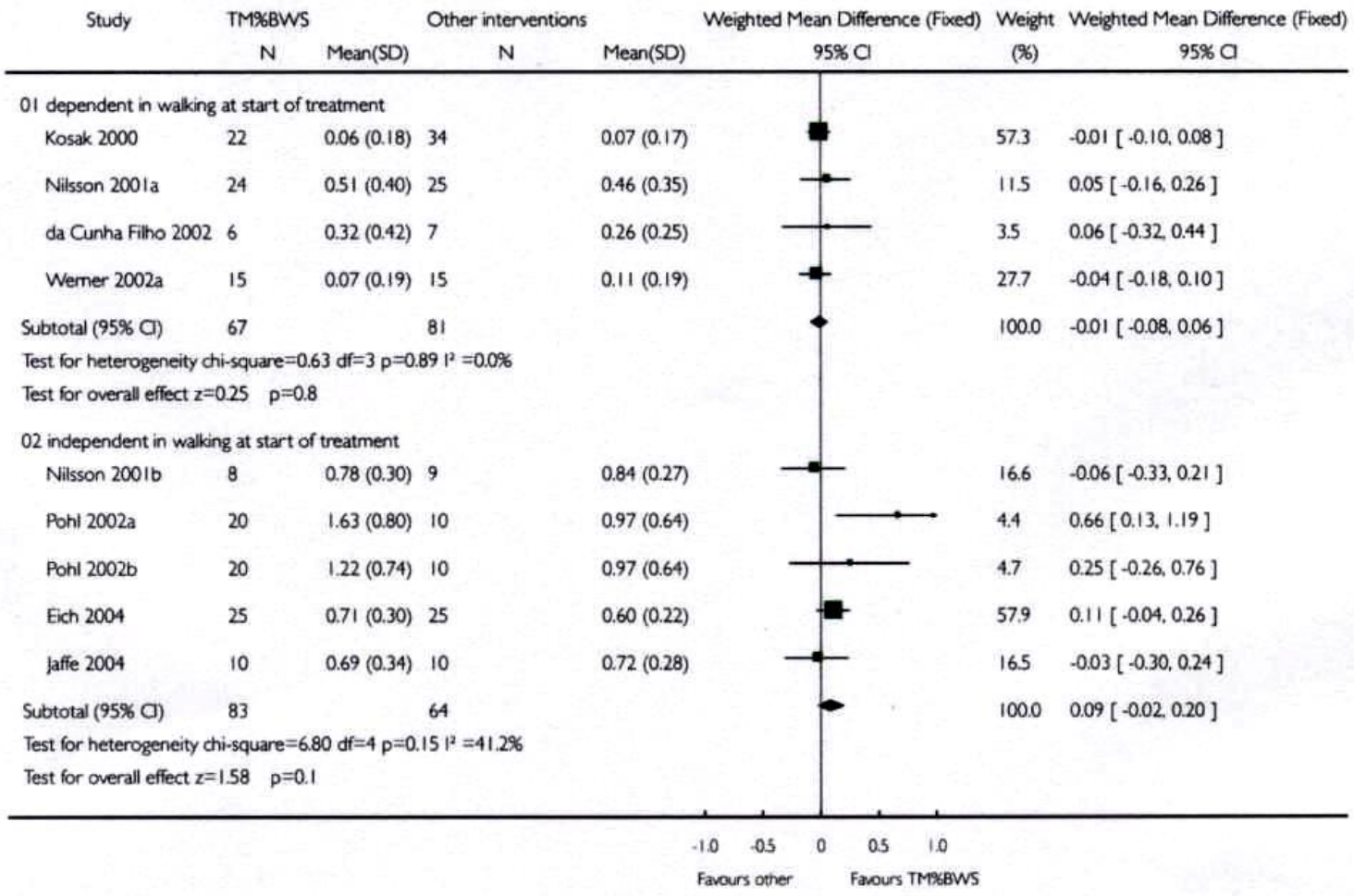
Moseley, A.M., Stark, A., Cameron, I.D., & Pollock, A. (2008). Treadmill training and body weight support for walking after stroke. Cochrane Database of Systematic Reviews, 2, 2008.

Analysis 01.02. Comparison 01 Treadmill and body weight support versus other interventions, Outcome 02 walking speed (m/sec) at end of treatment phase

Review: Treadmill training and body weight support for walking after stroke

Comparison: 01 Treadmill and body weight support versus other interventions

Outcome: 02 walking speed (m/sec) at end of treatment phase



Aggregating (weighting and "pooling") results of several studies

To arrive at overall estimate of outcome, study results are weighted inversely to their variability.

The more precise its estimate, the more heavily a study is weighted.

Weights depend on both sample size and within-sample variability.

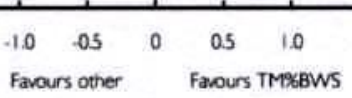
Analysis 01.02. Comparison 01 Treadmill and body weight support versus other interventions, Outcome 02 walking speed (m/sec) at end of treatment phase

Review: Treadmill training and body weight support for walking after stroke

Comparison: 01 Treadmill and body weight support versus other interventions

Outcome: 02 walking speed (m/sec) at end of treatment phase

Study	TM%BWS N	TM%BWS Mean(SD)	Other interventions N	Other interventions Mean(SD)	Weighted Mean Difference (Fixed) 95% CI	Weight (%)	Weighted Mean Difference (Fixed) 95% CI
01 dependent in walking at start of treatment							
Kosak 2000	22	0.06 (0.18)	34	0.07 (0.17)		57.3	-0.01 [-0.10, 0.08]
Nilsson 2001a	24	0.51 (0.40)	25	0.46 (0.35)		11.5	0.05 [-0.16, 0.26]
da Cunha Filho 2002	6	0.32 (0.42)	7	0.26 (0.25)		3.5	0.06 [-0.32, 0.44]
Werner 2002a	15	0.07 (0.19)	15	0.11 (0.19)		27.7	-0.04 [-0.18, 0.10]
Subtotal (95% CI)	67		81			100.0	-0.01 [-0.08, 0.06]
Test for heterogeneity chi-square=0.63 df=3 p=0.89 I^2 =0.0%							
Test for overall effect z=0.25 p=0.8							
02 independent in walking at start of treatment							
Nilsson 2001b	8	0.78 (0.30)	9	0.84 (0.27)		16.6	-0.06 [-0.33, 0.21]
Pohl 2002a	20	1.63 (0.80)	10	0.97 (0.64)		4.4	0.66 [0.13, 1.19]
Pohl 2002b	20	1.22 (0.74)	10	0.97 (0.64)		4.7	0.25 [-0.26, 0.76]
Eich 2004	25	0.71 (0.30)	25	0.60 (0.22)		57.9	0.11 [-0.04, 0.26]
Jaffe 2004	10	0.69 (0.34)	10	0.72 (0.28)		16.5	-0.03 [-0.30, 0.24]
Subtotal (95% CI)	83		64			100.0	0.09 [-0.02, 0.20]
Test for heterogeneity chi-square=6.80 df=4 p=0.15 I^2 =41.2%							
Test for overall effect z=1.58 p=0.1							



Measuring consistency (homogeneity) of studies' results

Individual weights used to calculate Cochran's Q:

$$Q = \Sigma\, w_i \; [\text{outcome of study i} - \textit{overall effect}\,]^2$$

Large values for Q suggest heterogeneity (lack of consistency)

Related statistic: $I^2 = 100\% \times (Q-df)/Q$

percentage of variation among study outcomes due not to chance,

but to heterogeneity among studies.

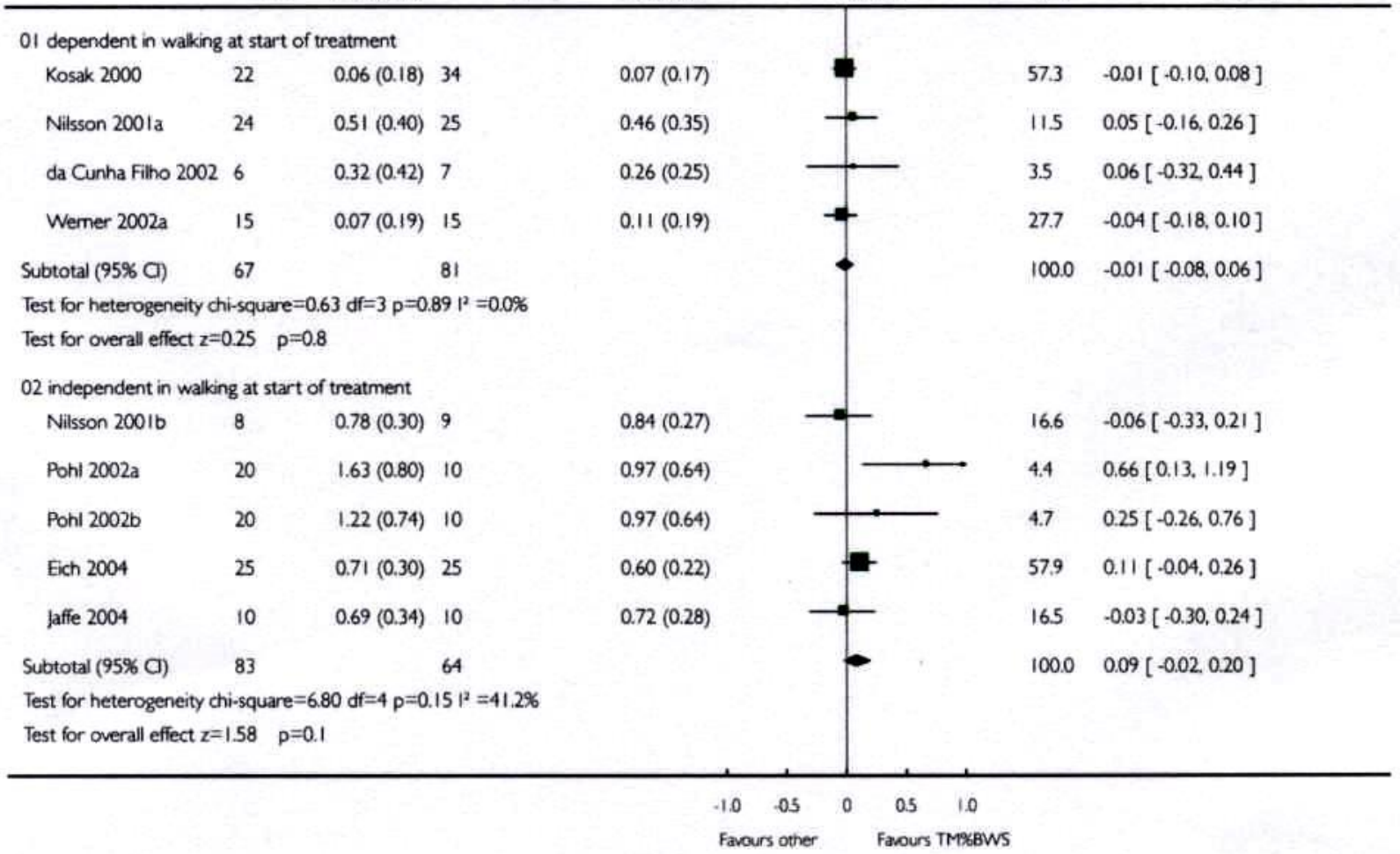
Analysis 01.02. Comparison 01 Treadmill and body weight support versus other interventions, Outcome 02 walking speed (m/sec) at end of treatment phase

Review: Treadmill training and body weight support for walking after stroke

Comparison: 01 Treadmill and body weight support versus other interventions

Outcome: 02 walking speed (m/sec) at end of treatment phase

Study	TM%BWS N	TM%BWS Mean(SD)	Other interventions N	Other interventions Mean(SD)	Weighted Mean Difference (Fixed) 95% CI	Weight (%)	Weighted Mean Difference (Fixed) 95% CI
01 dependent in walking at start of treatment							
Kosak 2000	22	0.06 (0.18)	34	0.07 (0.17)		57.3	-0.01 [-0.10, 0.08]
Nilsson 2001a	24	0.51 (0.40)	25	0.46 (0.35)		11.5	0.05 [-0.16, 0.26]
da Cunha Filho 2002	6	0.32 (0.42)	7	0.26 (0.25)		3.5	0.06 [-0.32, 0.44]
Werner 2002a	15	0.07 (0.19)	15	0.11 (0.19)		27.7	-0.04 [-0.18, 0.10]
Subtotal (95% CI)	67		81			100.0	-0.01 [-0.08, 0.06]
Test for heterogeneity chi-square=0.63 df=3 p=0.89 I^2 =0.0%							
Test for overall effect z=0.25 p=0.8							
02 independent in walking at start of treatment							
Nilsson 2001b	8	0.78 (0.30)	9	0.84 (0.27)		16.6	-0.06 [-0.33, 0.21]
Pohl 2002a	20	1.63 (0.80)	10	0.97 (0.64)		4.4	0.66 [0.13, 1.19]
Pohl 2002b	20	1.22 (0.74)	10	0.97 (0.64)		4.7	0.25 [-0.26, 0.76]
Eich 2004	25	0.71 (0.30)	25	0.60 (0.22)		57.9	0.11 [-0.04, 0.26]
Jaffe 2004	10	0.69 (0.34)	10	0.72 (0.28)		16.5	-0.03 [-0.30, 0.24]
Subtotal (95% CI)	83		64			100.0	0.09 [-0.02, 0.20]
Test for heterogeneity chi-square=6.80 df=4 p=0.15 I^2 =41.2%							
Test for overall effect z=1.58 p=0.1							



Relatively consistent studies are combined using a fixed effects model,

which assumes that each study measures the same outcome,

and that the outcome has a true and fixed value in the population.

Relatively inconsistent (heterogenous) studies can still be combined in a *random effects model*,

which assumes the studies are a random sample from a family of studies that address slightly different questions.

A random effects model estimates the same overall effect as a fixed effects model,

but produces wider confidence intervals, which reflects the underlying studies' heterogeneity.

A family of studies that address "slightly different questions?"

If we conceive of a clinical question as multidimensional:

- **Patient group**
- **Intervention**
- **Comparison**
- **Outcome**

then even if studies address the same outcome, they address different questions if, across studies:

- **patient characteristics vary**
- **interventions are inconsistent**
- **comparison groups are diverse**

Example 2
Random effects analysis

Gibbs, S, & Harvey, I. (2008). Topical treatments for cutaneous warts. Cochrane Database of Systematic Reviews. 2, 2008.

Analysis 15.01. Comparison 15 Aggressive vs gentle cryotherapy, Outcome 01 Cure rate

Review: Topical treatments for cutaneous warts

Comparison: 15 Aggressive vs gentle cryotherapy

Outcome: 01 Cure rate

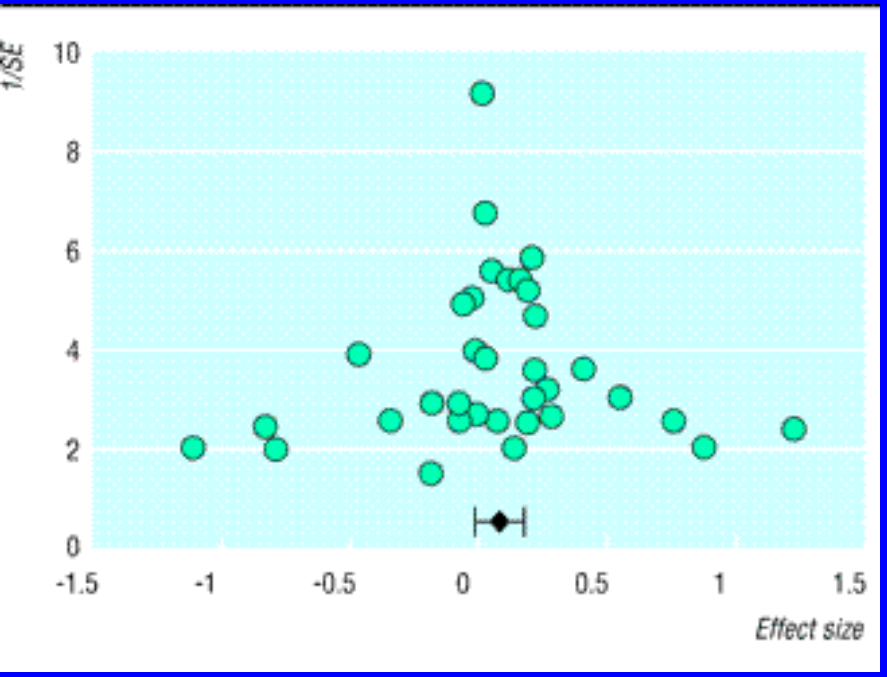
Study	aggressive cryo n/N	'gentle' cryo n/N	Relative Risk (Random) 95% CI	Weight (%)	Relative Risk (Random) 95% CI
Berth-Jones 1994	79/169	57/155		38.2	1.27 [0.98, 1.65]
Connolly 1999	42/71	25/75		34.5	1.77 [1.22, 2.58]
Hansen 1986	24/33	7/27		24.3	2.81 [1.43, 5.49]
Sonnex 1988	14/31	0/31		3.1	29.00 [1.81, 465.72]
Total (95% CI)	304	288		100.0	1.90 [1.15, 3.15]

Total events: 159 (aggressive cryo), 89 ('gentle' cryo)

Test for heterogeneity chi-square=10.76 df=3 p=0.01 I^2 =72.1%

Test for overall effect z=2.50 p=0.01

Funnel plots



Horizontal axis: effect size.

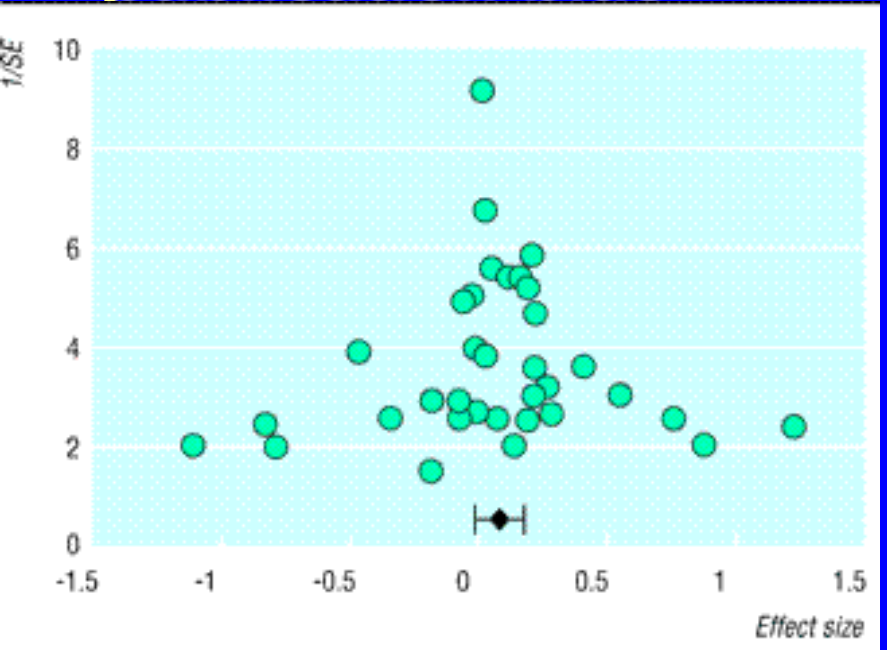
Vert. axis proportional to study size and precision. Less precise studies toward bottom.

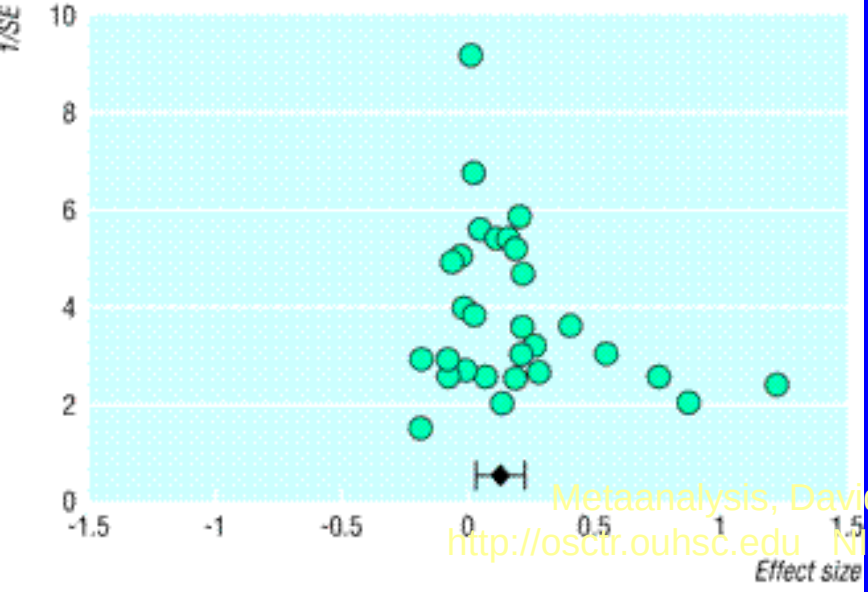
Larger studies (toward top) yield more precise estimates that should approximate true effect size (♦).

Smaller studies (toward bottom) yield less precise, more variable estimates.

Sutton, A.J., Duval, S.J., Tweedie, R.L., Abrams, K.R., & Jones, D.R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *BMJ,*320:1574-1577.

Funnel plots and publication bias





The graph typically resembles an inverted funnel.

Publication bias is suggested if review finds no small and negative studies.

Cochran's Q and I^2 statistics (details)

Measures of consistency vs. heterogeneity among study results

$Q = \Sigma w_i$ [study outcome i - overall effect]^2

a weighted sum
of squared differences
between individual study outcomes
and the overall effect across studies.

Cochran's Q is distributed as a chi-square statistic with k-1 degrees of freedom (where k is number of studies)

The statistic's p-value relates to the null hypothesis that individual study estimates are consistent with one another.

Related statistic: $I^2 = 100\% \times (Q-df)/Q$
percentage of variation across study outcomes due to heterogeneity of studies rather than chance.

Egger test

A test of funnel plot asymmetry

that tests null hypothesis that y-intercept (β_0)=0

in a linear regression model: $y = \beta_0 + \beta_0 x$

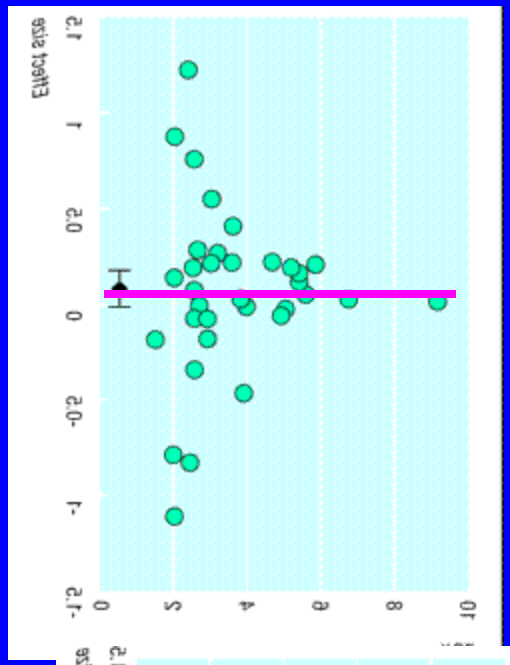
where y is the estimate (or effect size), divided by its standard error

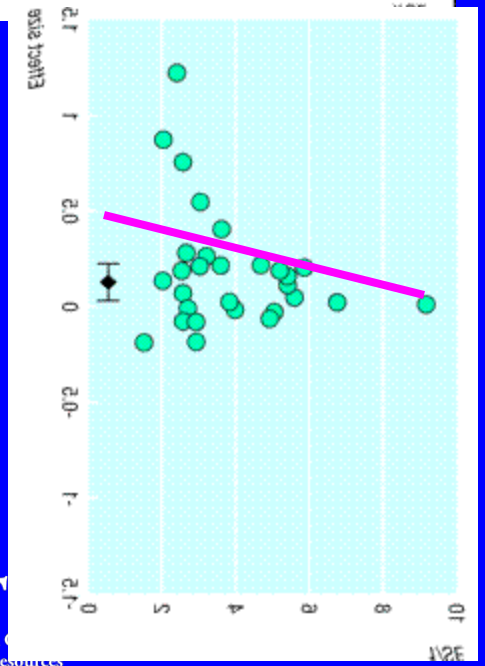
X is precision (reciprocal of the standard error of the estimate).

If $\beta_0 \neq 0$, there is evidence of bias

Test's power to detect bias depends on number of studies (data points in funnel plot)

Egger M, *et al*. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629-634.





Egger essentially flips the funnel plots and calculates a regression line that relates the outcome to the study's precision.

The line's intercept should be zero in the absence of bias.

SCTR
Oklahoma Shared Clinical & Translational Resources

osctr.ouhsc.edu

Biostatistics, Epidemiology, and Research Design (BERD) Core

http://osctr.ouhsc.edu/biostatisticsepidemiology-core

BERD Training Unit

http://osctr.ouhsc.edu/training-unit