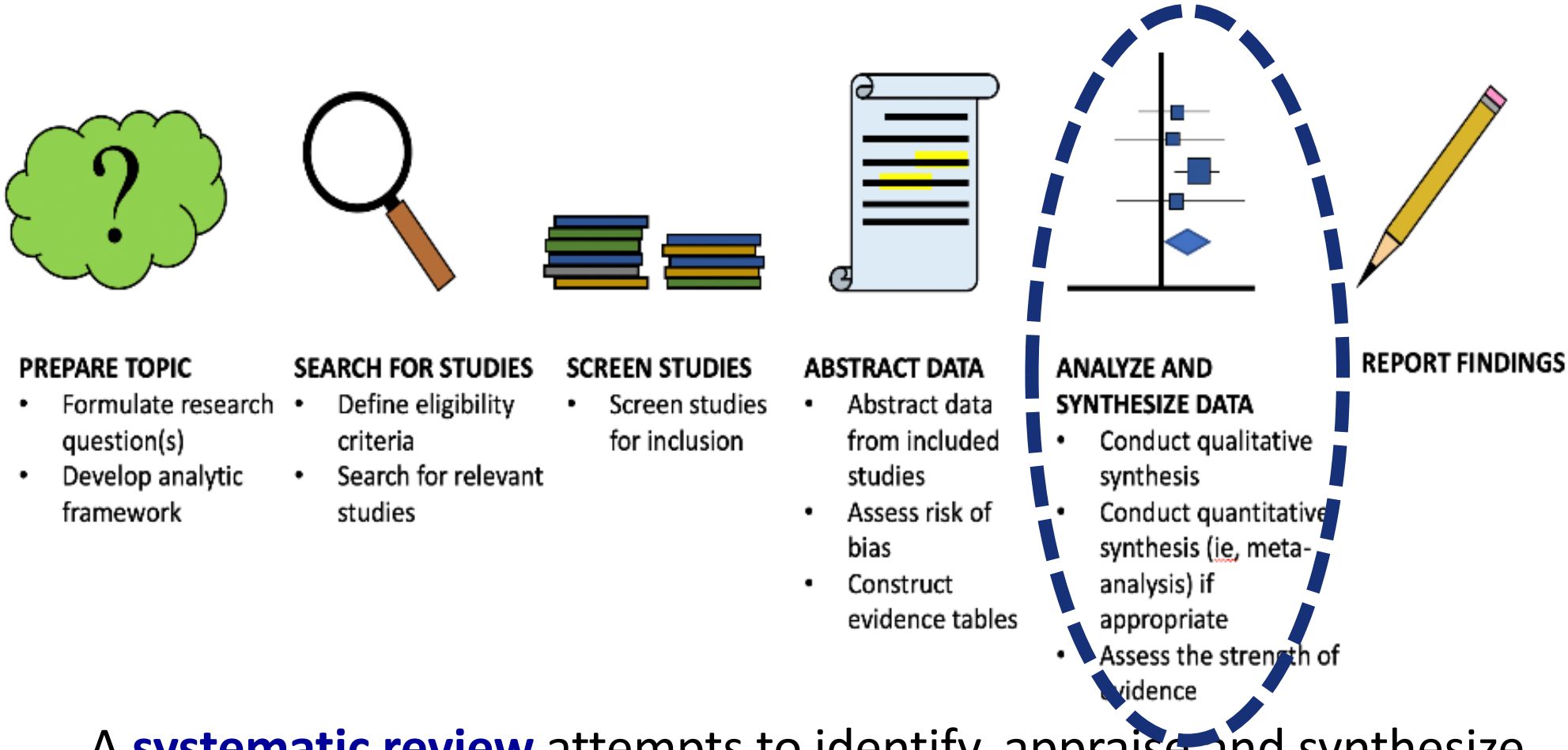
Introduction to Meta-analysis and Heterogeneity

Tianjing Li, MD, MHS, PhD
Associate Professor
June 3, 2022

Department of Ophthalmology
School of Medicine
University of Colorado Anschutz Medical Campus

Steps in Completing a Systematic Review



PREPARE TOPIC
- Formulate research question(s)
- Develop analytic framework

SEARCH FOR STUDIES
- Define eligibility criteria
- Search for relevant studies

SCREEN STUDIES
- Screen studies for inclusion

ABSTRACT DATA
- Abstract data from included studies
- Assess risk of bias
- Construct evidence tables

ANALYZE AND SYNTHESIZE DATA
- Conduct qualitative synthesis
- Conduct quantitative synthesis (ie, meta-analysis) if appropriate
- Assess the strength of evidence

REPORT FINDINGS

A **systematic review** attempts to identify, appraise and synthesize all the empirical evidence that meets pre-specified eligibility criteria to answer a specific research question.

Adapted from Wallace C et al. J Compar Effect Res 2013; 2(3):273-282

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

PLANNING THE ANALYSIS

Introduction

Results of meta-analyses can be very misleading if suitable attention has not been given to:

- the review question
- eligibility criteria
- identification and selection of studies
- collection of appropriate data
- risk of bias in study results

Decide what would be meaningful to analyze

Planning the synthesis

Synthesis is a process of bringing together data from a set of included studies with the aim of drawing conclusions about a body of evidence

- **Qualitative (or narratively):** structured summary (e.g. tabulation), description, and discussion of the study characteristics that may affect the cumulative evidence
- **Quantitative:** involves statistical analysis (meta-analysis)

A general framework for synthesis

Stage 1. At protocol stage

- Set up the comparisons

Stage 2. Summarize the included studies and prepare for synthesis

- Summarize the characteristics of each study.
- Determine which studies are similar enough to be grouped within each comparison.
- Determine what data are available for synthesis.
- Determine if modification to the planned comparisons or outcomes is necessary, noting any deviations from the protocol plans.
- Synthesize the characteristics of the studies contributing to each comparison.

Stage 3. The synthesis itself

- Qualitative synthesis.
- Perform a statistical synthesis if appropriate or provide structured reporting of the effects.
- Interpret the results, including considerations of the direction and size of the effect and certainty of the evidence.

Characteristics of included studies table

Overall study design
Number randomized
Study duration

Inclusion and exclusion criteria
Actual characteristics of participants

All interventions including doses, schedule/frequency, and duration of the interventions

Relevant outcomes, including time of assessment
Number of participants analyzed
Subgroup analyses

Characteristics of included studies [ordered by study ID]

Varshney 2011

Methods	Randomised double-blind, placebo-controlled trial Setting: University health centres, USA
Participants	28 children aged 1 to 16 years from local allergy and immunology clinics or surrounding community physician offices (19 intervention, 9 control). Av. age 7 years Inclusion criteria: clinical history of reaction to peanut within 60min of ingestion; peanut CAP-FEIA >15 kU/L or >7 kU/L if a significant reaction occurred within 6m of enrolment; positive skin prick test Excluded if: history of severe peanut anaphylaxis; moderate to severe persistent asthma; poorly controlled atopic dermatitis; oat allergy; or inability to discontinue antihistamines during OFC or skin testing
Interventions	Oral immunotherapy for approx. 48wks (home dosing/build-up visits approx. 44wks; maintenance phase approx. 4wks; OFC at wk 48). Active = peanut flour; placebo = oat flour *Initial day escalation phase:* Day 1: dosing started at 0.1 mg peanut protein or placebo, approx. doubled every 30min until 6 mg reached/subject had symptoms. Day 2: Dosing began at highest tolerated dose during day 1 *Home dosing:* Children instructed to ingest each dose mixed in a food vehicle daily *Dose escalations:* Every 2 wks, doses increased 50 - 100% until 75 mg dose, 25-33% until 400 mg maintenance dose reached *Maintenance phase:* Ingest 400 mg dose daily for 1m
Outcomes	Double-blind, placebo controlled OFC (increasing doses of peanut or oat flour every 10 to 20mins up to cumulative dose of 5000 mg); adverse events; peanut-specific IgE, IgG, and IgG_4; secreted cytokine assays and T-cell analyses in subset of participants.
Notes	Small sample size Participants followed up after wk 48, investigating long-term immune tolerance (data not yet available)

Characteristics of included studies table

Association between intake of non-sugar sweeteners and health outcomes

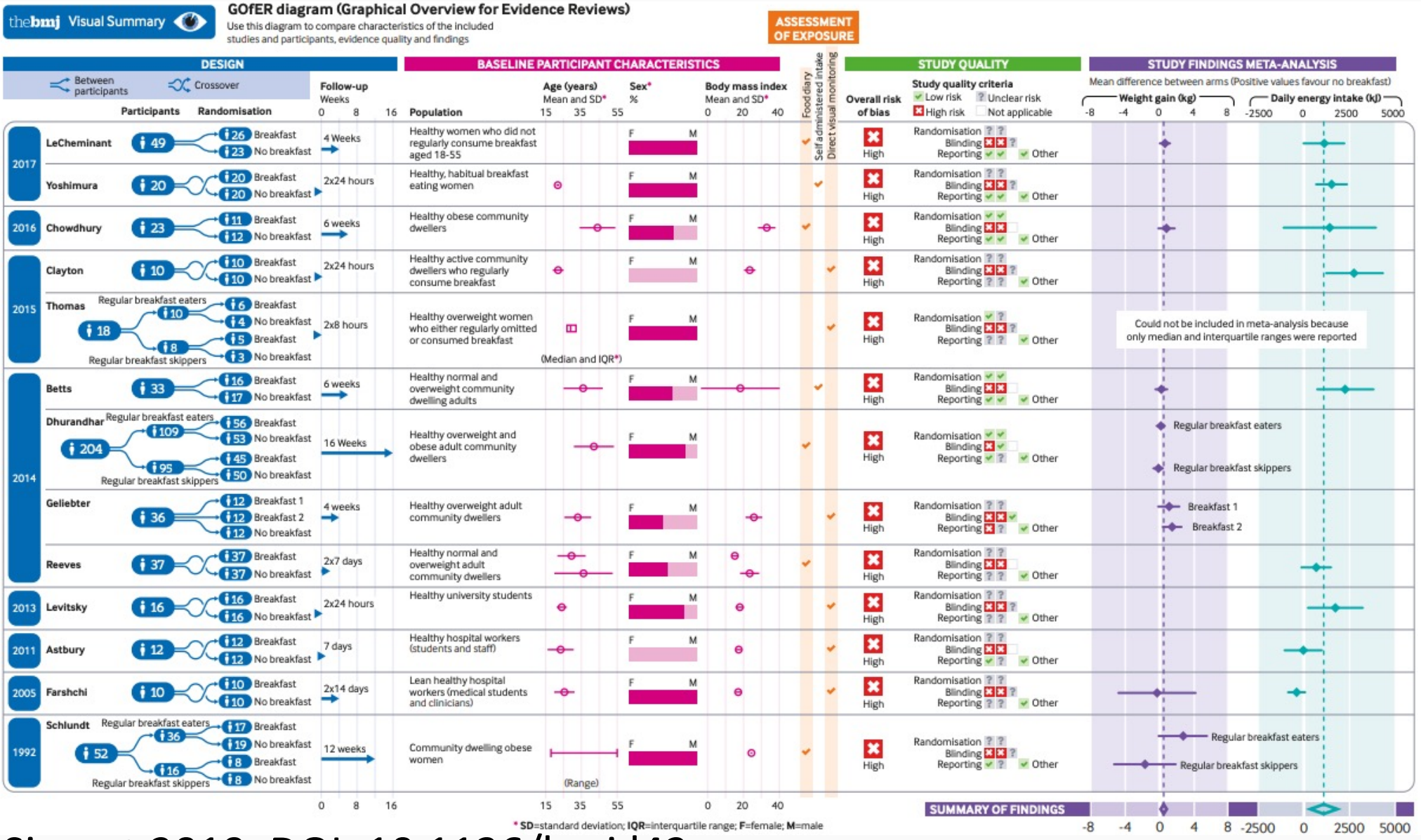
Supplementary file 3: Details of included studies (RCT = randomised, controlled trial; non-RCT = non-randomised controlled trial; AS = artificial sweetener, CVD = cardiovascular disease); *For profit funding includes sponsoring of study material, i.e. intervention substances, as well as financial sponsoring for conducting the study

Study ID	Study design	Population	N Intervention	N Control	N overall / at baseline	For-profit funding*	Intervention	Source	Dose	Duration	Control	Outcomes
1. Baird 2000	RCT	Adults, healthy	77	31	118	yes	Sucralose	Liquid	in ascending dose (125, 250 and 500 mg during week 1-3, 4-7 and 8-12, respectively)	90 days / 12 weeks	Fructose	Body weight, CVD, adverse effects
2. Ballantyne 2011	RCT	Adults, overweight males	20	20	40	yes	Aspartame	Liquid	No information, only no. and kJ of drinks	8 weeks	Sucrose	Body, weight, eating behaviour, preference for sweet taste, mood
3. Blackburn 1997	RCT	Adults, overweight	82	81	186	yes	Aspartame	Mixed	No information	16 weeks	"Avoidance of low energy sweeteners"	Body weight, eating behaviour, preference for sweet taste, adverse effects
4. Bosetti 2009	Case-control study	Adults	-	-	3117	no	Saccharin	Mixed	No information	10-16 years	No intervention	Cancer
5. Cabaniols 2011	Case-control study	Adults	122	122	244	no	Aspartame	Mixed	No information	5 years	No intervention	Cancer
6. Cartwright 1981	Case-control study	Adults probably	-	-	1901	no	Saccharin	Tabletop sweetener	No information	at least 1 year	No intervention	Cancer
7. Colditz 1990	Cohort study	Adults, female	-	-	31940	no	Saccharin	Mixed	0.0 - 5.8, or more than 5.8 g	8 years	No intervention	Body weight
8. deRuyter 2012	RCT	Children, healthy	319	322	641	no	Combination of NSSs (40 mg sucralose, 10 mg acesulfame-K)	Liquid	40 mg sucralose, 10 mg acesulfame-K	18 months	Sucrose	Body weight, eating behaviour
9. Duran Aguero 2014	Cross-sectional study	Children	-	-	1224	no	Aspartame, acesulfame, cyclamate, saccharin	Mixed	No information	No information	No exposure	Body weight

https://www-bmj-com.proxy1.library.jhu.edu/content/bmj/suppl/2019/01/02/bmj.k4718.DC1/toei046063.ww3.pdf

Characteristics of included studies figure

Effect of breakfast on weight and energy intake



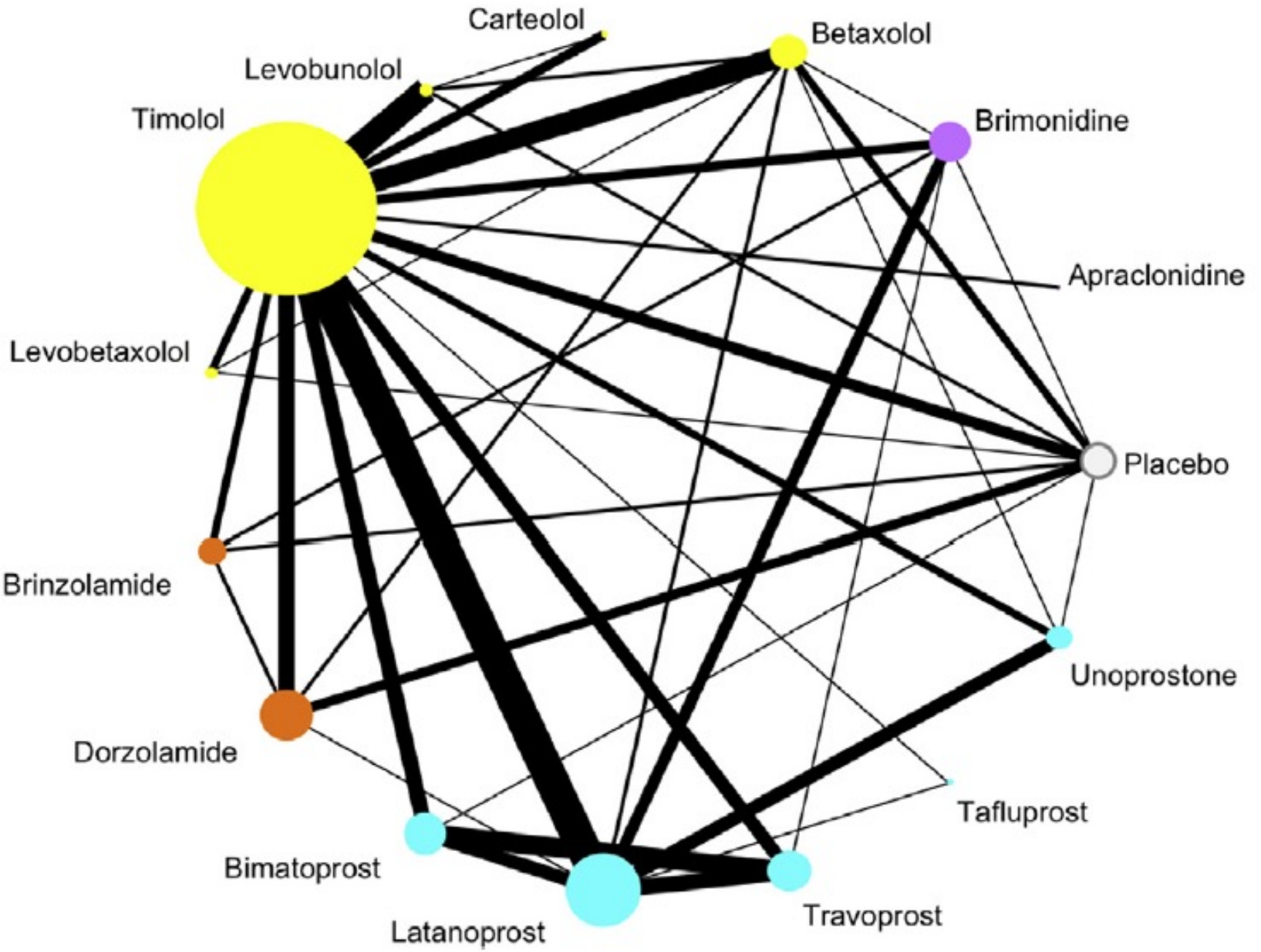
Sievert 2019. DOI: 10.1136/bmj.l42

Analysis plan follows from the aim of the review

- Obtain size of effect from similar studies estimating the same effect
  - e.g. early vs late epidural for women in labor
  - Method: Standard meta-analysis
- Compare multiple interventions for the same condition
  - e.g. medical interventions for glaucoma
  - Method: Network meta-analysis
- Investigate the relationship between the size of an effect and some characteristic(s) of the studies
  - e.g. does prevalence of autistic spectrum disorders depend on time period and location?
  - Method: Meta-regression
- Just test a null hypothesis
  - e.g. organizational change has no effect on mortality
  - Method: Combination of P values [uncommon, in fact]
- Or other aims…

Slide courtesy of Julian Higgins

Compare multiple interventions – network meta-analysis



JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Introduction to Meta-analysis

What is a meta-analysis?

- An optional component of a systematic review

- Original definition
  - "The statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass 1976)

- Alternative definition:
  - "A statistical analysis which combines the results of several independent studies considered by the analyst to be 'combinable'" (Huque 1988)

Combining results in a meta-analysis

- Justification for combining results:
  - Studies are estimating – in whole or in part –a common effect
  - Studies are addressing the same fundamental biological/clinical/mechanistic question

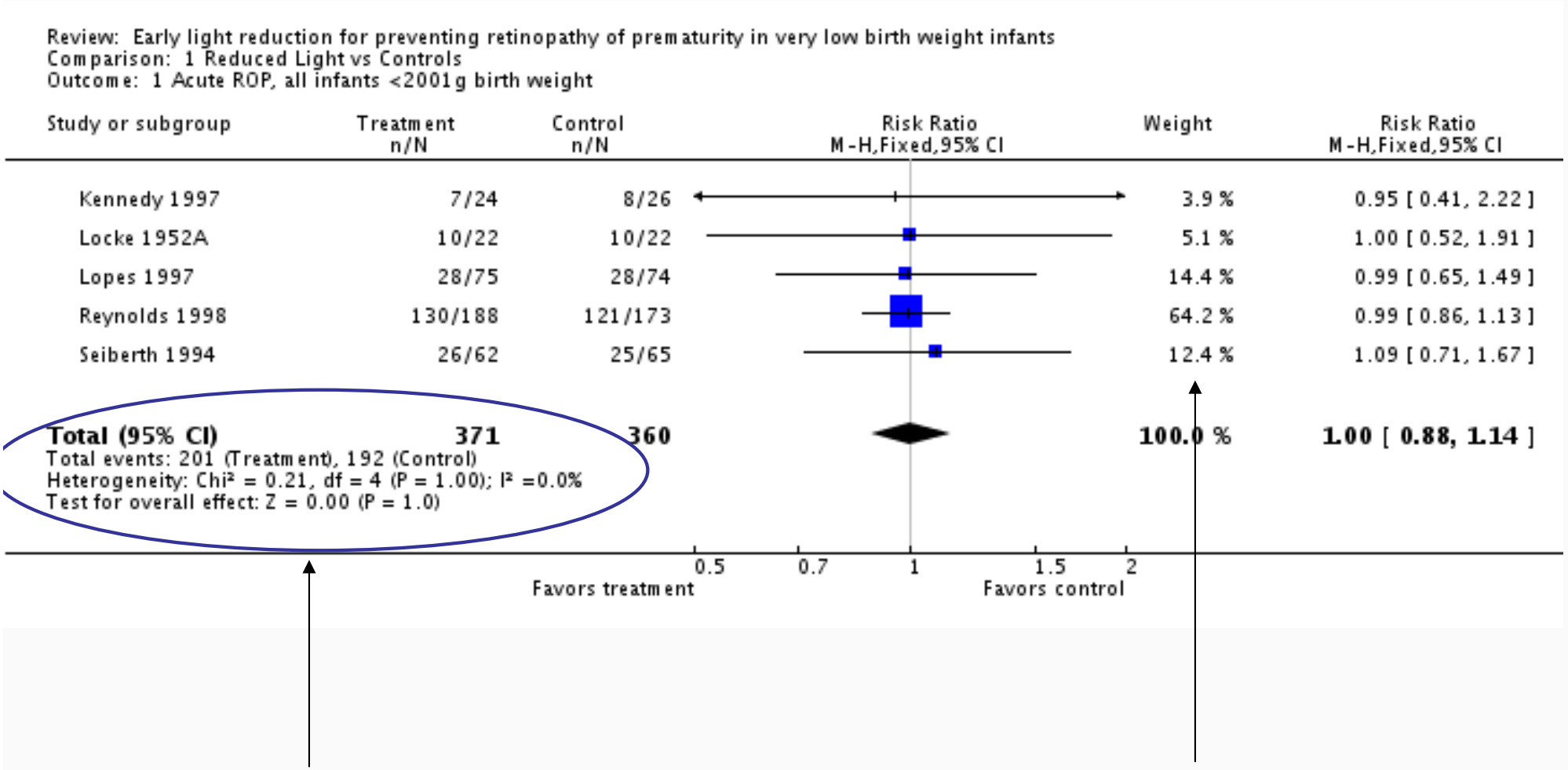
Early light reduction to prevent retinopathy

	Babies	Intervention
Kennedy 1997	<1251g	Goggles on both eyes, <6 hrs to 31 weeks
Locke 1952A	<2001g	Patch on one eye, <24 hrs to dispatch
Lopes 1997	<1600g	Patch on both eyes, birth to 32 weeks
Reynolds 1998	<1251g	97% light reducing goggles from <24 hrs to 31 weeks
Seiberth 1994	<1501g	Patch on both eyes, birth to 35 weeks

OK to combine?

Slide courtesy of Julian Higgins

Forest plot in The Cochrane Library



Further details of meta-analysis
Identification of heterogeneity (see later)

Weights given to each study

Poll Question

Which of the following is a legitimate reason to NOT do a meta-analysis?

- The outcome data needed for the meta-analysis are not available from 9 out of the 15 studies included in the systematic review.
- There are only 2 studies included in the systematic review.
- The studies included in the systematic review enrolled between 127 to 756 patients.
- The studies included in the systematic review do not address the same fundamental research question.

Pooling the raw data

Consider two randomized trials with dichotomous outcome (death)

	Treated (deaths / N)	Controls (deaths / N)
Trial A	40 / 100	50 / 200
Trial B	15 / 200	5 / 100

Slide courtesy of Julian Higgins

Adding up the cells

Consider two randomized trials with dichotomous outcome (death)

	Treated (deaths / N)	Controls (deaths / N)
Trial A	40 / 100	50 / 200
Trial B	15 / 200	5 / 100
Total	55 / 300	55 / 300

Slide courtesy of Julian Higgins

Calculating summary statistics

This is *not* how we do a meta-analysis

	Treated (deaths / N)	Controls (deaths / N)	Relative risk (95% CI)
Trial A	40 / 100	50 / 200	1.6 (1.14, 2.25)
Trial B	15 / 200	5 / 100	1.5 (0.56, 4.01)
Total	55 / 300	55 / 300	1 (0.71 ,1.40)

Slide courtesy of Julian Higgins

De-mystify meta-analysis - step 1

- Each study is summarized by an estimate of effect or association (the 'result'):
  - Risk ratio
  - Odds ratio
  - Mean difference

  - In this way, individual participants are only compared with others in the same study (preserves randomization)

De-mystify meta-analysis – step 2

- The overall measure of effect is a weighted average of results of the individual studies

$$\frac{\sum W_i \times \text{estimate}_i}{\sum W_i} = \frac{\sum W_i Y_i}{\sum W_i}$$

- It seems sensible to give more weight to larger studies
  - The weights are usually chosen to reflect *precision* of the estimate

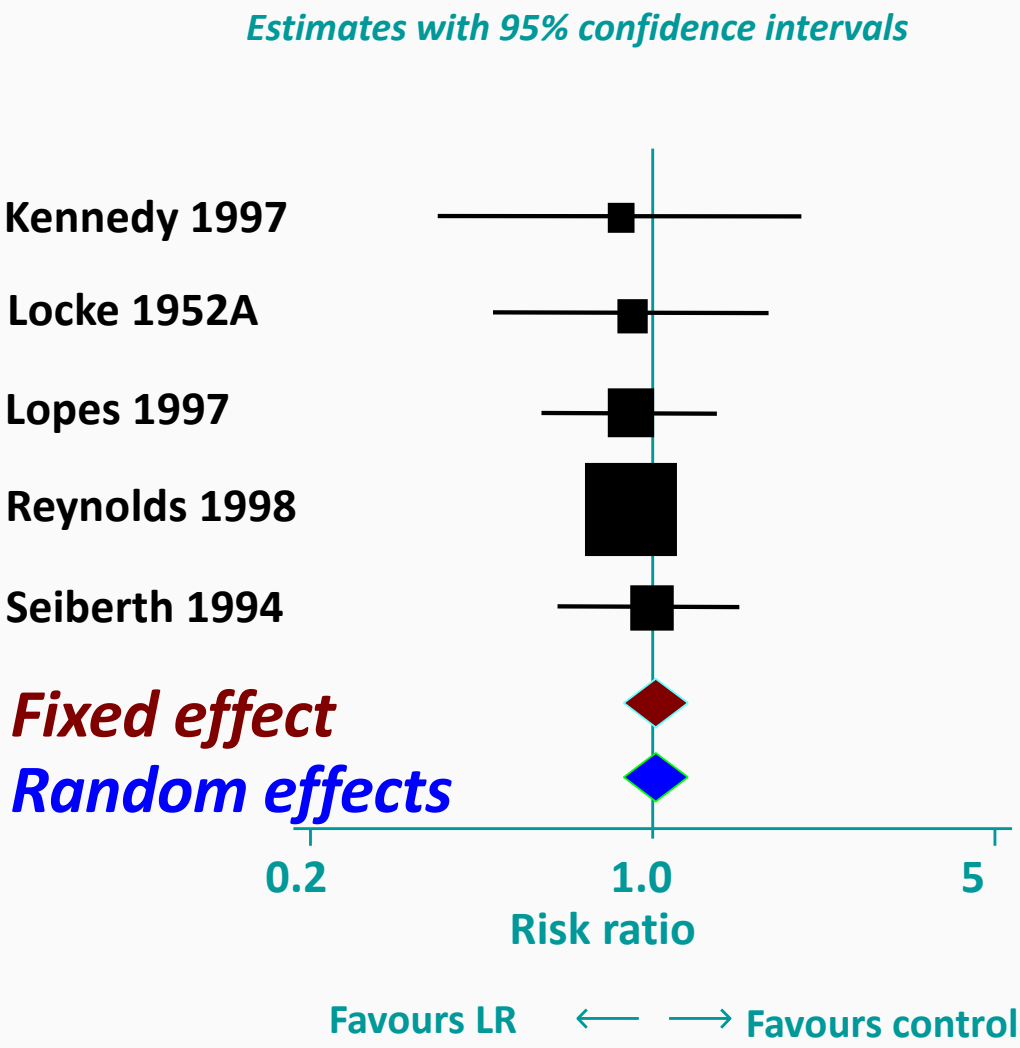
De-mystify meta-analysis – step 3

- Choice of model

The combination of intervention effect estimates across studies may optionally incorporate an assumption that the studies are not all estimating the same intervention effect, but estimate intervention effects that follow a distribution across studies. This is the basis of a random-effects meta-analysis.

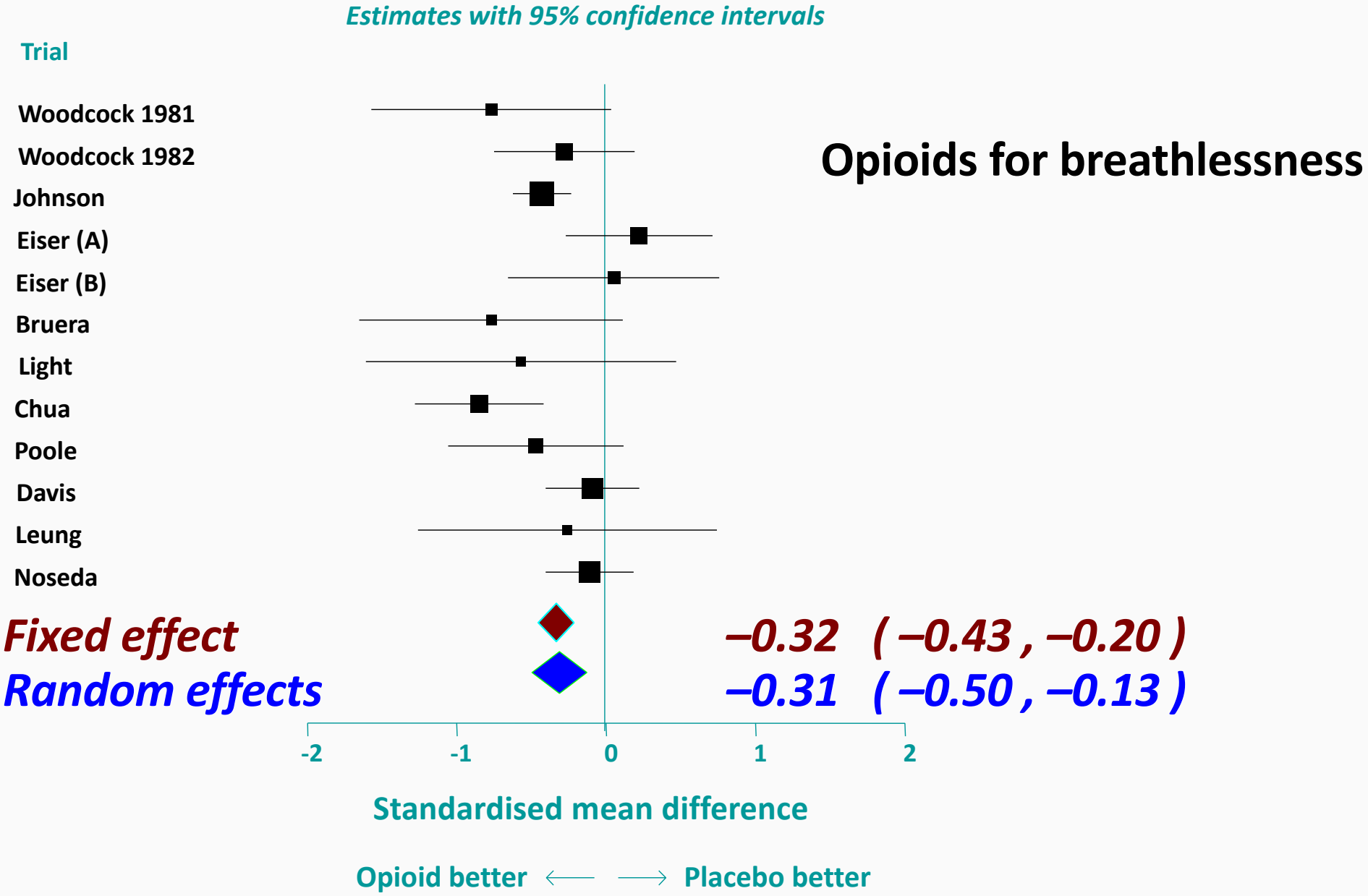
Alternatively, if it is assumed that each study is estimating exactly the same quantity, then a common-effect meta-analysis is performed (often called fixed-effect).

Identical results



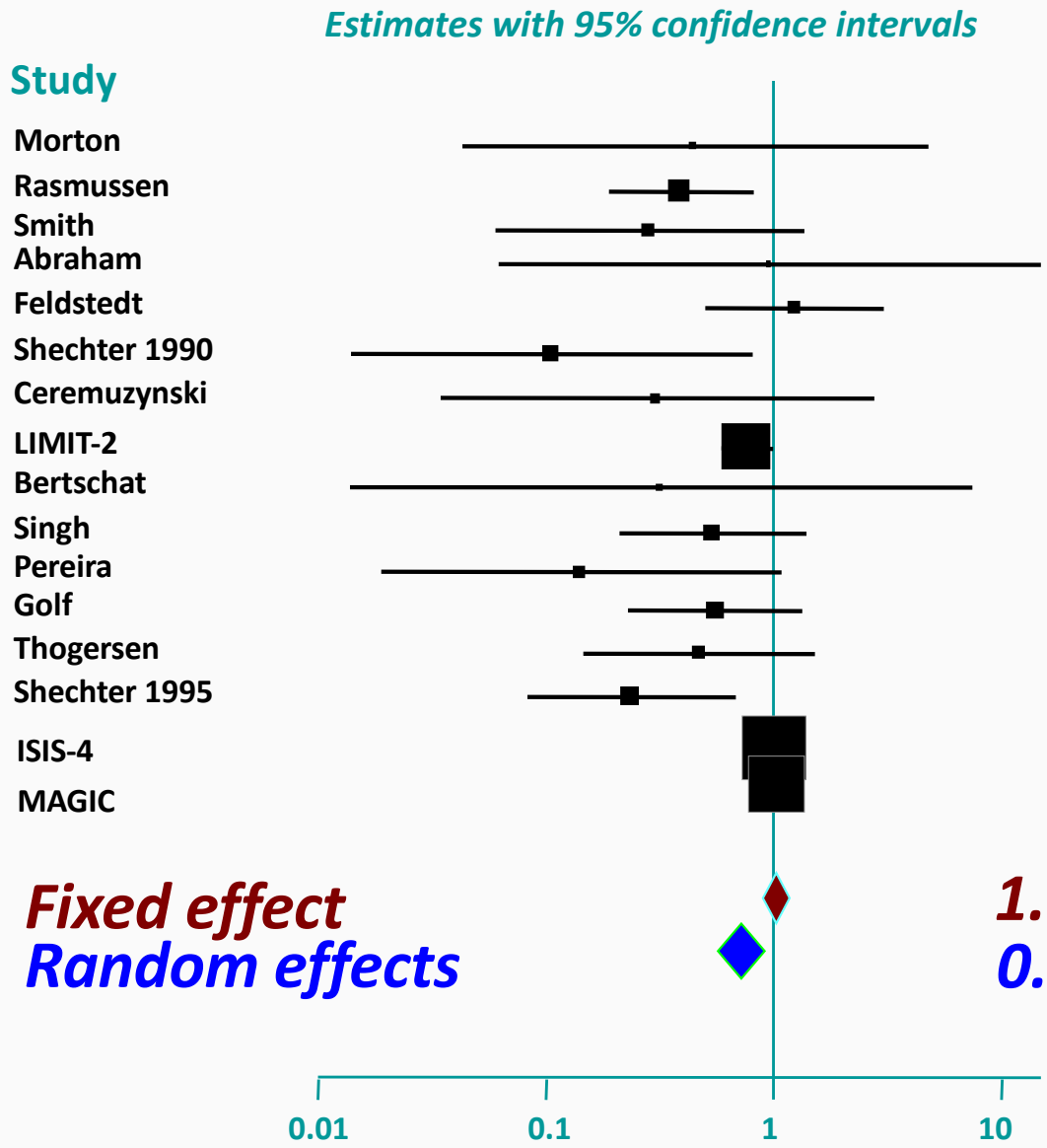
Early light reduction for retinopathy of prematurity

Slightly different results



Opioids for breathlessness

Very different results



Intravenous magnesium for acute myocardial infarction (mortality)

Weights in common and random-effects analyses

- Random-effects meta-analysis will often give a wider confidence interval than common-effect meta-analysis
- Wider confidence intervals for the random-effects estimate reflect greater **uncertainty** because it is assumed that the true effect varies between studies
- If the between-study variance is estimated to be zero, then the common-effect and random-effects estimates will be identical

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Identifying and Quantifying Heterogeneity

What is heterogeneity?

- **Clinical heterogeneity**

  Variability in the participants, interventions (or exposures), and outcomes studied

- **Methodological heterogeneity**

  Variability in the study design and risk of bias

- **Statistical heterogeneity**

  ✓Variability in the intervention effects being evaluated in the different studies

  ✓Is the consequence of clinical and/or methodological heterogeneity

Higgins et al. Chapter 9.5. Heterogeneity. Cochrane Handbook. 2011

Clinical and methodological heterogeneity (Cont'd)

The association between duration of breastfeeding and risk for childhood overweight

- "Our 21 studies were all cohort studies, of which eight were in the US, nine in Europe, and four in Asia, Australia, or the Middle East. The studies analyzed breastfeeding duration ranges from as little as 0-16 weeks to as much as greater than 12 months… Sample sizes ranged from 324 to 117,304. Study dropout rates prior to follow up ranged from 5% to 52%."

Chan G. Class 2011

Clinical and methodological heterogeneity (Cont'd)

Risk of ischemic stroke in people with migraine headaches

- "…as seen in our descriptive tables (Tables 1A, 1B, 2A, 2B, 3A, 3B), there was substantial variation in the sample sizes and characteristics of research subjects across studies. Subjects were drawn from registries, administrative databases, RCT participants, hospitals and the community; these various source populations would be expected to be associated with different baseline risks of stroke. Studies varied in mean age of subjects, which ranged from 15-97 years, but most focused on the 15-50 year old age group. There was also variation in the gender composition of study populations."

Kahn S. Class 2009

Measures of statistical heterogeneity

- Q (weighted sum of squares on a standardized scale) can be calculated to test the null hypothesis that all studies share a common effect
- I^2 is the ***%*** of the variability in effect estimates that is due to heterogeneity rather than sampling error; scale free (ranges from 0%-100%)
- T^2 is the variance of true effects, reflecting the ***amount*** of true heterogeneity; on the same scale as the effects themselves

Borenstein M et al. Chapter 16. Identifying and Quantifying Heterogeneity. Introduction to Meta-analysis. 2009

Interpreting the I^2 statistic

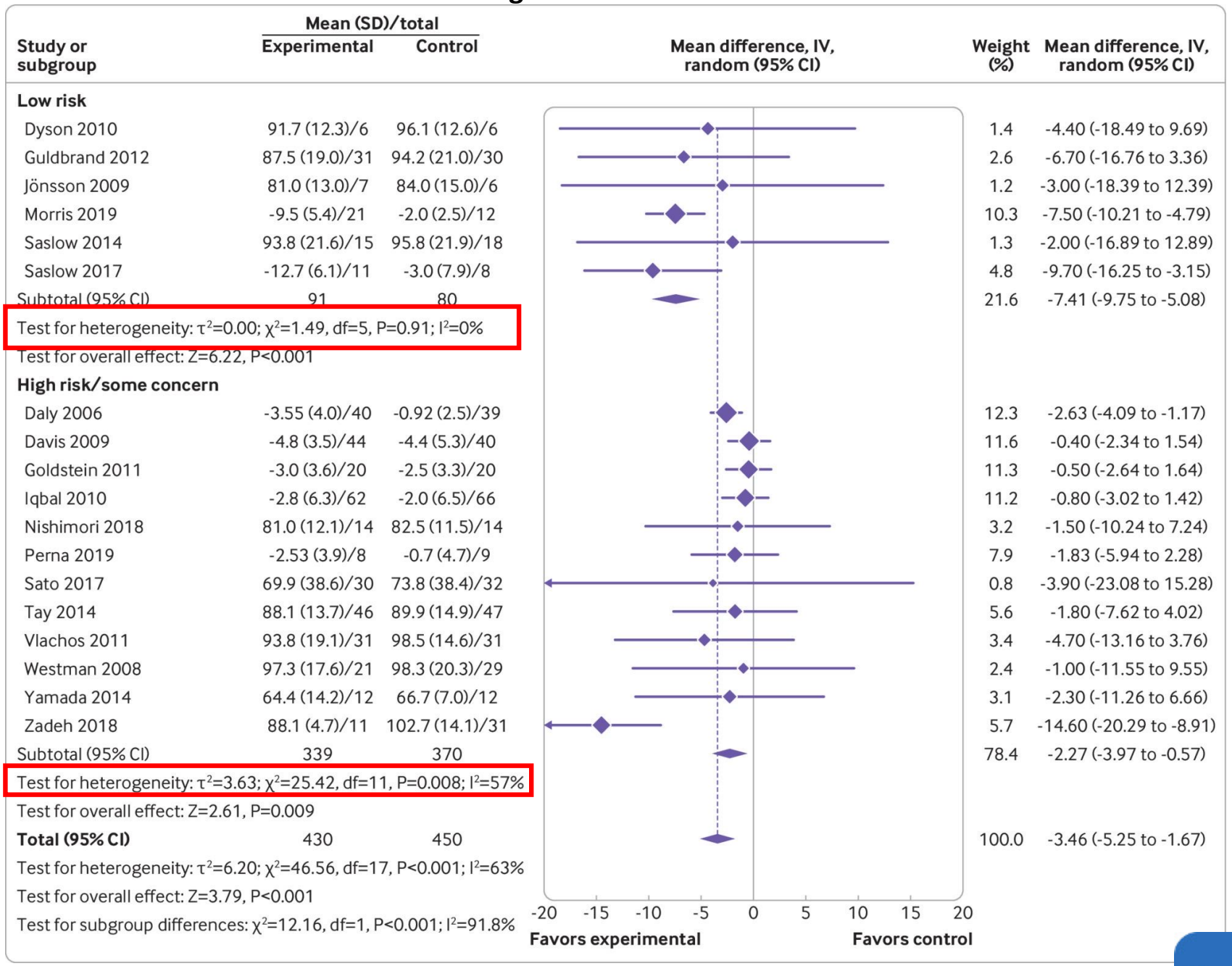
Rough guide to interpretation in the context of meta-analyses of randomized trials:

- 0% to 40%: might not be important
- 30% to 60%: may represent moderate heterogeneity*
- 50% to 90%: may represent substantial heterogeneity*
- 75% to 100%: considerable heterogeneity*

*importance of the observed value of I^2 depends on

1. magnitude and direction of effects, and
2. strength of evidence for heterogeneity (e.g. P value from heterogeneity test, or a confidence interval for I^2)

Weight loss at six months



Joshua Z Goldenberg et al. BMJ 2021;372:bmj.m4743

What to do with heterogeneity?

Options	
• Ignore it	• Don't do that!
• Do not do a meta-analysis	• A systematic review need not contain any meta-analysis
• Check the data	• Incorrect data extraction; data transcription and transformation errors
• Change the effect measure	• Mean difference vs. standardized mean difference; risk ratio vs. risk difference etc.
• Exclude studies	• Exclude outlying studies in a sensitivity analysis
• Encompass it	• Random-effects meta-analysis (but don't say a RE model takes care of it!)
• Explore it	• Subgroup analysis • Meta-regression

Higgins et al. Chapter 9.9.5. Strategies for Addressing Heterogeneity. Cochrane Handbook. 2011

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Subgroup Analysis & Meta-regression

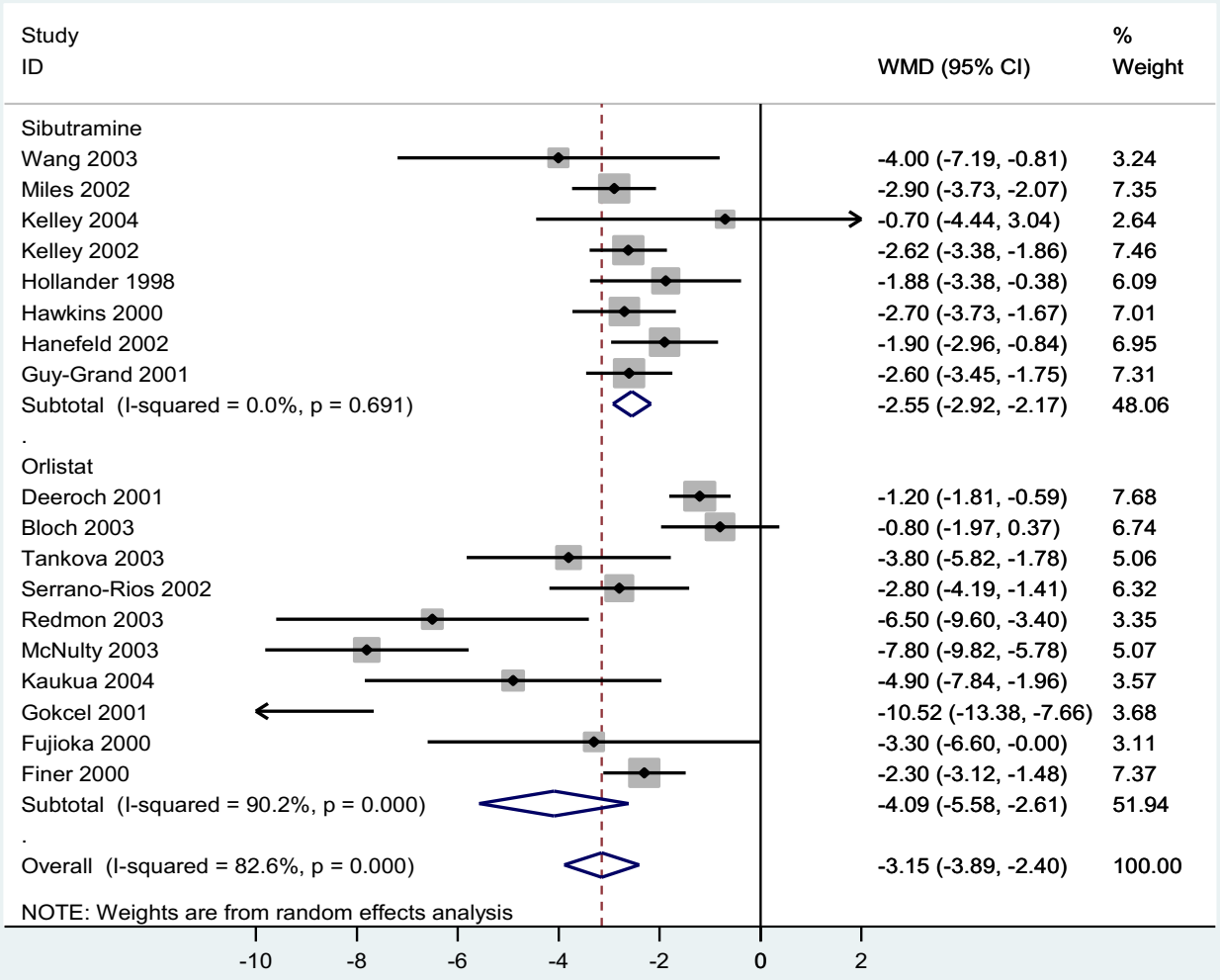
Exploring heterogeneity

- Characteristics of studies may be associated with the size of treatment effect
  - For example,
    - age group of participants
    - setting of study
    - dose of drug
    - risk of bias

- For discrete characteristics, can use **subgroup analyses**
- For discrete or continuous characteristics, can use **meta-regression**

Subgroup analysis

- Divide up the studies
  - e.g. by drug
- Separate meta-analysis for each group of studies



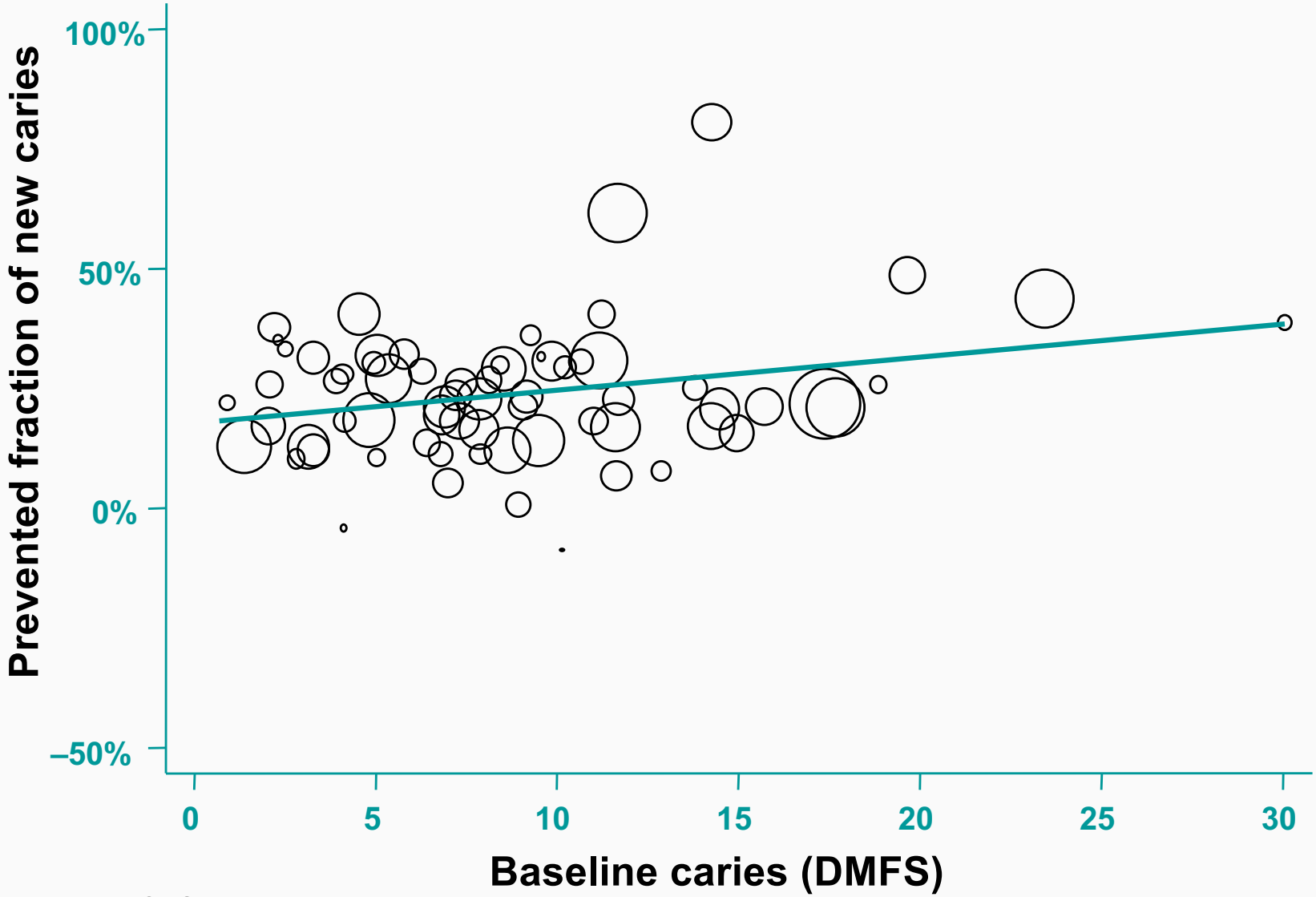
Meta-regression

- **Linear regression** describes a linear relationship between two characteristics
  - model the outcome variable (or dependent variable)
  - using values of the explanatory variable (or independent variable)

- In **meta-regression**:
  - The outcome variable is the effect size (e.g. mean difference or logRR)
  - The explanatory variable is some summary description of the study (e.g. duration or location)

Meta-regression

Does effectiveness of toothpaste depend on baseline population levels of caries?



Source: Marinho et al. CDSR 2003