Introduction to meta-analysis

Steve Simon

Abstract

Meta-analysis is the quantitative pooling of data from multiple studies. The three threats to the validity of a meta-analytic finding are heterogeneity, publication bias, and poor individual study quality. This talk will introduce you to the major design issues that you must address in your research protocol to insure that your meta-analysis will have credibility.

Today's talk is a basic introduction. It will focus on design issues for a meta-analytic study.

Abstract (continued)

You will also learn some of the fundamental graphical and analytic tools used in meta-analysis: the forest plot, Cochran's Q and I-squared, the funnel plot, and the L'Abbe plot. You will compare the results from a fixed effects and a random effects model and understand the choices available for summary statistics. Finally, you will see how to publish your results using the PRISMA guidelines.

It will also cover some of the common statistics and graphics used in a meta-analytic study, and publication standards for a meta-analytic study.

A motivating example
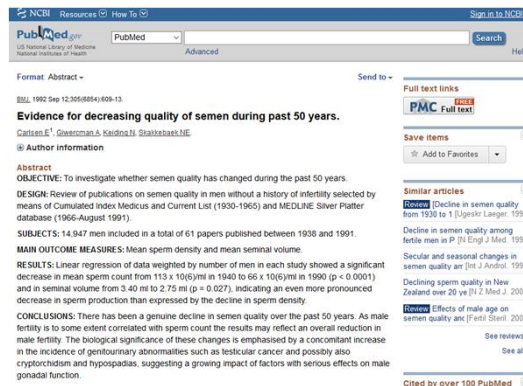
Pubmed abstract of a controversial meta-analysis

In 1992, the British Medical Journal published a controversial meta-analysis. This study (Carlsen 1992) reviewed 61 papers published from 1938 and 1991 and showed that there was a significant decrease in sperm count and in seminal volume over this period of time.
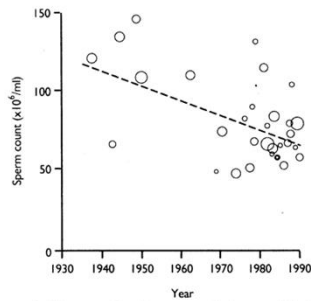
A motivating example

FIG 1—*Linear regression of mean sperm density reported in 61 publications (represented by circles whose area is proportional to the logarithm of the number of subjects in study) each weighted according to number of subjects, 1938-90*

graph showing decline in sperm counts over time

For example, a linear regression model on the pooled data provided an estimated average count of 113 million per ml in 1940 and 66 million per ml in 1990.
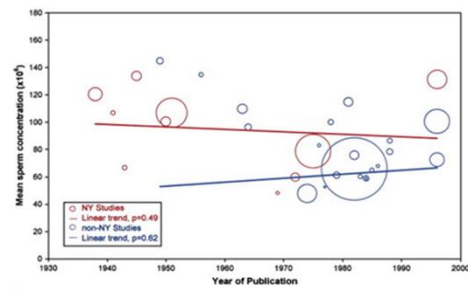
Alternative analysis

Fig. 1.
Re-analysis of data from Carlsen et al shows no decline in sperm counts (*blue regression line*) when data from New York are excluded. Bubble size corresponds to number of men in study. (*From* Saidi J, Chang D, Goluboff E, et al. Declining sperm counts in the United States? A critical review. J Urol 1999;161:460–2; with permission.)

Alternative display of semen studies

Several researchers (Olsen 1995; Fisch 1996) noted heterogeneity in this meta-analysis, a mixing of apples and oranges. Studies before 1970 were dominated by studies in the United States and particularly studies in New York. Studies after 1970 included many other locations including Third World countries.

Additional sources of variation

– The source of patients
  • Sperm donor clinics,
  • Fertility work-ups
  • In vitro fertilization clinics
– Request for minimum abstinence time.
– Tobacco and marijauna use

There was also substantial variation from study to study in the source of patients. Sperm donor clinics typically ask donors to prove they are fertile (e.g., having fathered at least two children). A fertility work-up, on the other hand, is done in settings where fertility is questionable, and would not require such a condition. An in vitro clinic might have men with normal fertility, if the issue of infertility were restricted to the female side. But often a poor sperm count is a contributing factor to female infertility.

Another thing that varied from study to study were absitinence requirements. Some studies asked the men to abstain from sex for two days prior to providing a sample and some didn't. Abstinent men would tend to have higher sperm counts than non-abstinent men.

If patients who used tobacco or marijuana were excluded from some studies, but not from others, this could also affect sperm counts.

Illustration of strengths and weaknesses of meta-analysis

- Weakness
  - Mixing North American apples with third world oranges.
- Strengths
  - Objective process.
  - Ability to re-analyze.

This meta-analysis and the subsequent criticisms illustrate, at the same time, the greatest weakness and the greatest strength of meta-analysis.

Apples and oranges

Here are the main threats to the validity of a meta-analysis:

- Did you mix apples and oranges? (heterogeneity)
- Did you leave some apples on the tree? (publication bias)
- Did the pile of apples amount to more than just a hill of beans? (no practical significance)
- Were all of the apples rotten? (poor study quality)

One of the seven habits of Steven Covey is "Begin with the end in mind." If you read the criticisms of meta-analysis, they tend to fall into four broad categories. You should design your study and analyze it keeping these criticisms in mind. You want to minimize the number of objections to your work.

Did you mix apples and oranges?

Meta-analysis: a multi-center clinical trial where each center uses a different protocol.

How do clinical trials differ?

- How the patient population was selected.
- How the intervention was administered.
- How the controls were selected/treated.
- How the effectiveness of the intervention was measured.

There are many ways that one clinical trial can differ from another. One might study a severely ill group of patients and another might study a mildly ill group. The intervention studied could vary in the frequency and duration of the intervention. The controls could be given a placebo or the best available alternative. The outcome used to measure the intervention could be something as severe and final as death or it could be a surrogate marker (e.g., CD4 cell counts in an AIDS trial).

A little bit of heterogeneity is actually quite good. If a new therapy is shown to be effective across a range of patient populations using a variety of different outcomes, you have a robust result. Mixing apples and oranges is okay, it gives you fruit salad. But you're not supposed to mix apples and onions.

You can examine heterogeneity using

– the forest plot

– L'Abbe plot

– Cochran's Q

– I-squared

– sensitivity/subgroup analysis

– meta regression

You can assess how much heterogeneity there is among your studies using graphical approaches, such as the forest plot and the L'Abbe plot, with numerical measures like Cochran's Q or I-squared, or through sensitivity and subgroup analyses or through meta regression.

You'll see how to calculate these quantities in just a bit.

Did you leave some apples on the tree?

Publication bias is difficult to assess and difficult to control for. You should

- have a comprehensive search protocol
  - non-Medline indexed journals
  - conference presentations
  - clinical trial registries
- assess publication bias using a funnel plot.

It is very important to try to get every study conducted in the area you're studying. Unpublished results and more likely to have negative results. It was originally thought to be because journals would preferentially publish only positive studies, but there is some evidence that authors self-censor the negative studies, especially negative studies with small sample sizes.

You need to try hard to find studies that are hard to find.

The funnel plot is a graphical method commonly used to identify whether publication bias has occured. We'll talk about it in a bit.

Did the pile of apples amount to more than just a hill of beans?

Very few meta-analytic studies address practical significance

- Summary measures in meta-analysis are unitless.
- Translate your findings to a meaningful scale.

A neglected issue in meta-analysis is the practical interpretation of the results. You need to assess more than just the overall statistical significance of your meta-analysis. You need to consider the scientific or practical significance as well.

The unitless quantities often used in meta-analysis make assessment of practical significance difficult. We'll address that issue in more detail in just a little bit.

Were all of the apples rotten?

Meta-analysis cannot "make a silk purse out of a sow's ear"

Quality scores
- Jadad
- PEDro

Meta-analysis cannot remove the biases and imprecision associated with poor research methodologies. If all of the studies have "issues" then a meta-analysis will amplify those issues.

You can assess quality issues by limiting studies based on scoring systems like Jadad or PEDro or by weighting studies based on these scores. We'll talk about this a bit later.

1.Was the study described as randomized (this includes words such as randomly, random, and randomisation)? (+1 Point)

1.Was the method used to generate the sequence of randomisation described and appropriate (table of random numbers, computer-generated, etc)? (+1 Point)

1.Was the study described as double blind? (+1 Point)

1.Was the method of double blinding described and appropriate (identical placebo, active placebo, dummy, etc)? (+1 Point)

1.Was there a description of withdrawals and dropouts? (+1 Point)

1.Deduct one point if the method used to generate the sequence of randomisation was described and it was inappropriate (patients were allocated alternately, or according to date of birth, hospital number, etc)

1.Deduct one point if the study was described as double blind but the method of blinding was inappropriate (e.g., comparison of tablet vs. injection with no double dummy).

1.eligibility criteria were specified no, yes where:

1.subjects were randomly allocated to groups (in a crossover study, subjects were randomly allocated an order in which treatments were received) no, yes where:

1.allocation was concealed no, yes where:

1.the groups were similar at baseline regarding the most important prognostic indicators no, yes where:

1.there was blinding of all subjects no, yes where:

1.there was blinding of all therapists who administered the therapy no, yes where:

1.there was blinding of all assessors who measured at least one key outcome no, yes where:

1.measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups no, yes where:

1.all subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome was analysed by "intention to treat" no, yes where:

1.the results of between-group statistical comparisons are reported for at least one key outcome no, yes where:

1.the study provides both point measures and measures of variability for at

least one key outcome no, yes where:

Design of a meta-analytic study

Detailed protocol
- – Get help from a librarian
- – Search strategy
- – Inclusion/exclusion criteria
- – Process for extracting numerical results

A detailed protocol is a must for meta-analysis. You have to provide a level of detail comparable to a clinical trial. A librarian is especially helpful here.

Medline is a comprehensive database of research publications in medical journals. It's a great place to start, but you should also search in databases other than Medline, such as Embase and CINAHL.

Look through presentation abstracts at major conferences. Also look for clinical trials that have been registered in a clinical trial database.

Make sure you pay to get translations for relevant articles published in languages other than English.

Detail exactly which studies do and do not qualify for inclusion in your meta-analysis. Have two independent raters evaluate each candidate article against your inclusion/exclusion criteria.

Also, detail what results you are going to extract from each article. Have two independent parties do the extraction and then compare.

Data analysis

– Pick a summary measure

– Forest plot,

– Cochran's Q and I-squared,

– Funnel plot

– L'Abbe plot.

– Fixed versus random effects

– Meta regression

Let's talk about data analysis. This talk will cover several different statistical measures and several different plots. You'll also see controversies about fixed versus random effects analysis.

BCG description

"Results from 13 studies examining the effectiveness of the Bacillus Calmette-Guerin (BCG) vaccine against tuberculosis."

Available at https://rdrr.io/cran/metafor/man/dat.colditz1994.html

BCG data, first three columns

```
##   trial               author year
## 1     1              Aronson 1948
## 2     2      Ferguson & Simes 1949
## 3     3        Rosenthal et al 1960
## 4     4      Hart & Sutherland 1977
## 5     5 Frimodt-Moller et al 1973
## 6     6        Stein & Aronson 1953
```

BCG data, last six columns

```
##    tpos  tneg cpos  cneg ablat     alloc
## 1    4   119   11   128    44    random
## 2    6   300   29   274    55    random
## 3    3   228   11   209    42    random
## 4   62 13536  248 12619    52    random
## 5   33  5036   47  5761    13 alternate
## 6  180  1361  372  1079    44 alternate
```

LOS description

Results from 9 studies on the length of the hospital stay of stroke patients under specialized care and under conventional/routine (non-specialist) care.

LOS data, first two columns

```
##   study             source
## 1     1          Edinburgh
## 2     2      Orpington-Mild
## 3     3  Orpington-Moderate
## 4     4    Orpington-Severe
## 5     5       Montreal-Home
## 6     6   Montreal-Transfer
```

LOS data, last six columns

```
##    n1i m1i sd1i n2i m2i sd2i
## 1 155  55   47 156  75   64
## 2  31  27    7  32  29    4
## 3  75  64   17  71 119   29
## 4  18  66   20  18 137   48
## 5   8  14    8  13  18   11
## 6  57  19    7  52  18    4
```

The standardized mean difference (SMD)

For a continuous outcome, the standardized mean difference is computed as

$$\hat{\theta}_i = \frac{\overline{X}_T - \overline{X}_C}{Estimated\ Standard\ Deviation}$$

or

$$\hat{\theta}_i = \frac{\overline{X}_C - \overline{X}_T}{Estimated\ Standard\ Deviation}$$

For continuous outcomes, the most common summary measure is the standardized mean difference. It is the difference in means divided by an estimate of the standard deviation (not the standard error) of an individual patient.

Which way do you subtract? It depends on whether a large value is good or a small value is good. The Cochrane Collaboration suggests subtracting in such a way that negative values suggest that the new treatment is better and positive values suggest that the control treatment is better.

Estimating the standard deviation

Different estimated standard deviations
- Cohen's d (pooled standard deviation)
- Hedge's g (bias correction)
- Adjustments for heteroscedascity or pairing

There are several ways to estimate the standard deviation. With the exception of pairing, where the estimated standard deviation could be substantially lower, these approaches are not all that different.

Cohen's d

Uses a pooled standard deviation.

$$S_p = \sqrt{\frac{n_T S_T^2 + n_C S_C^2}{n_T + n_C}}$$

The simplest standardized mean difference is Cohen's d. This is the famous Jacob Cohen who wrote "Statistical Power Analysis for the Behavioral Sciences" back in 1977 and invented to concept of effect size.

Cohen's d uses a pooled standard deviation in the denominator.

Hedge's g

Pooled standard deviation adjusted by a bias correction factor

$$J \approx 1 - \frac{3}{4df - 1}$$

Hedge's g multiples the pooled standard deviation by a bias correction factor. The exact form of the correction factor is messy, so sometimes an approximate formula based on the pooled degrees of freedom is used in its place.

Both the exact and the approximate bias adjustment formulas are less than one, making Hedge's g smaller than Cohen's d. But for large sample sizes, this adjustment becomes trivial.

Adjusting for heteroscedascity

Two choices for heteroscedascity within a study

$$\sqrt{\frac{S_T^2 + S_C^2}{2}}$$

or

$$S_C$$

If there is heterogeneity within a study, you can use the square root of the average the two variances or the square root of the control group variance.

Adjusting for pairing

For pairing use

$$S_D$$

if it is available, or

$$\sqrt{S_T^2 + S_C^2 - 2\hat{\rho}S_T S_C}$$

if it is not.

For paired data, use the standard deviation of the differences, if it reported, or estimate the standard deviation of the differences using a reasonable guess as to the correlation.

Analyzing proportions

    – Odds ratio

    – Relative risk

    – Risk difference

Odds ratios and relative risk are always analyzed and displayed on a log scale.

When you are comparing proportions rather than means, you have three ways of summarizing your data. You can use the odds ratio, the relative risk, or the risk difference in proportions.

Titanic data

	Survived	Died
Male	142	709
Female	308	154

Mortality versus gender table

You already have seen the various summary measures used for comparing two proportions, but let me review them anyway.

This is table from the Titanic, listing passengers only (not crew members) by gender and by whether they lived or died. Kate Winslet is in the lower left corner of women who survived and Leonardo DiCaprio, sadly, is in the upper right corner of men who died.

Odds ratio

	Survived	Died	odds
Male	142	709	709/142=4.993
Female	308	154	154/308=0.5
			odds ratio = 9.986

Odds ratio calculation

The ratio of deaths to survivors is about 5 to 1 among men and exactly 1 to 2 among women. So the odds ratio is approximately 10.

Relative risk

	Survived	Died	P[Death]
Male	142	709	709/(709+142)=0.833
Female	308	154	154/(154+308)=0.333
			Relative risk = 2.502

Relative risk calculation

You could also look at the probability of death. It is 83% among men and 33% among women. That ratio, approximately 2.5 is the relative risk.

Risk difference

	Survived	Died	P[Death]
Male	142	709	709/(709+142)=0.833
Female	308	154	154/(154+308)=0.333
			Risk difference = 0.500

Risk difference calculation

Rather than dividing one ratio by another, you could subtract one from the other. You get a value of 0.5, which is the risk difference or the absolute risk.

Which measure to choose? There are practical reasons to prefer the relative risk or the risk difference to the odds ratio, but the odds ratio has an important advantage. It is not artificially constrained by any bound.

The other measures will have an upper bound depending on what the control group is. It is hard to get a relative risk of 2.0 or a risk difference of 0.5 when the proportion in your control group is 0.8.

The artifical constraints of the relative risk and risk difference will sometimes create unwanted heterogeneity in your meta-analysis.

Standardized mean differences for los data

```
##   n1i m1i sd1i n2i m2i sd2i       yi      vi
## 1 155  55   47 156  75   64 -0.3552 0.0131
## 2  31  27    7  32  29    4 -0.3479 0.0645
## 3  75  64   17  71 119   29 -2.3176 0.0458
## 4  18  66   20  18 137   48 -1.8880 0.1606
## 5   8  14    8  13  18   11 -0.3840 0.2054
## 6  57  19    7  52  18    4  0.1721 0.0369
```

Odds ratios for the BCG data

```
##   tpos  tneg cpos  cneg       yi     vi
## 1    4   119   11   128 -0.9387 0.3571
## 2    6   300   29   274 -1.6662 0.2081
## 3    3   228   11   209 -1.3863 0.4334
## 4   62 13536  248 12619 -1.4564 0.0203
## 5   33  5036   47  5761 -0.2191 0.0520
## 6  180  1361  372  1079 -0.9581 0.0099
```

Fixed effects model

$$V(\hat{\theta}_i) \approx \frac{1}{n_T} + \frac{1}{n_C}$$

$$\approx \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}$$

$$w_i = \frac{1}{V(\hat{\theta}_i)}$$

Once you have your standardized mean differences, you can compute an overall estimate, but you need to account for the fact that some estimates are more precise than others.

The variance of the standardized mean difference is approximated by a simple function of the sample sizes in the treatment and control for the standardized mean difference and by a simple function of the four cell counts for the log odds ratio. You should calculate a weight equal to the inverse of the variance to insure that the studies with the largest sample sizes (and thus the greatest precision) get weighted more heavily.

Fixed effects model (cont'd)

$$\hat{\theta} = \frac{\Sigma w_i \hat{\theta}_i}{\Sigma w_i}$$

$$V(\hat{\theta}) = \frac{1}{\Sigma w_i}$$

$$\frac{\hat{\theta}^2}{V(\hat{\theta})} \sim \chi^2(1)$$

The overall estimate, theta-hat is a weighted average of the individual study estimates. If you take the square of the overall estimate and divide by the estimated variance of the overall estimate, you get a statistic for testing the hypothesis that the overall mean difference is zero.

Compare this test statistic to a chi-squared distribution with 1 degree of freedom.

Cochrane's Q

Test of homogeneity

$$Q = \Sigma w_i (\hat{\theta}_i - \hat{\theta})^2$$

$$Q \sim \chi^2 (k - 1)$$

$$I^2 = 100 \frac{Q - (k - 1)}{Q}$$

Heterogeneity will manifest itself with a greater degree of variation from one study to another than the amount you'd expect due to the sample sizes in each study.

If the value of Q is close to its degrees of freedom (k-1), then you have evidence that the studies are homogenous. If the value of Q is much larger than the degrees of freedom, then you have evidence of heterogeneity.

I-squared is a measure of the degree of heterogeneity that ranges between 0 and 100.

Random effects model

Rough approximation

$$\hat{\tau}^2 = max\left(0, Q - (k - 1)\right)$$

$$W_i^* = \frac{1}{V(\hat{\theta}_i) + \hat{\tau}^2}$$

You can fit a random effects model using standard software like PROC MIXED in SAS. The random effects model assumes that each study has an extra source of random variation, caused by the numerous small differences that occur from one study to another. The weights in a random effects meta-analysis are closer to one another, much closer to one another when the variance componet (tau-hat squared) is large.

Random effects model for LOS data

```
Model Results:

estimate      se     zval     pval     ci.lb
ci.ub
 -0.5371  0.3087  -1.7401  0.0818  -1.1421
0.0679  .
```

Random effects model for BCG data

```
Model Results:

estimate       se      zval     pval     ci.lb
ci.ub
 -0.7452  0.1860  -4.0057  <.0001  -1.1098  -
0.3806  ***
```
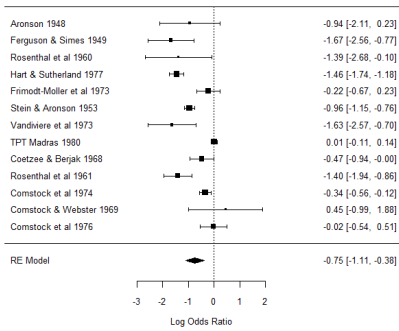
Test of homogeneity for the LOS data

```
tau^2 (estimated amount of total heterogeneity):
0.7908 (SE = 0.4281)

tau (square root of estimated tau^2 value):
0.8893

I^2 (total heterogeneity / total variability):
95.49%

H^2 (total variability / sampling variability):
22.20


Test for Heterogeneity:
Q(df = 8) = 123.7293, p-val < .0001
```

Test of homogeneity for the BCG data

```
tau^2 (estimated amount of total heterogeneity):
0.3378 (SE = 0.1784)
tau (square root of estimated tau^2 value):
0.5812
I^2 (total heterogeneity / total variability):
92.07%
H^2 (total variability / sampling variability):
12.61

Test for Heterogeneity:
Q(df = 12) = 163.1649, p-val < .0001
```
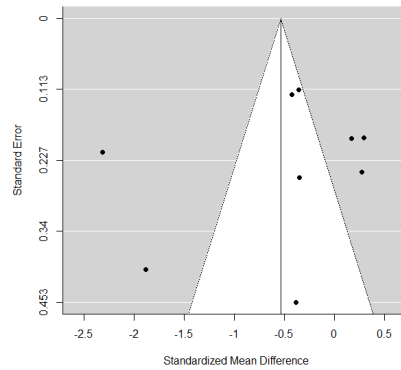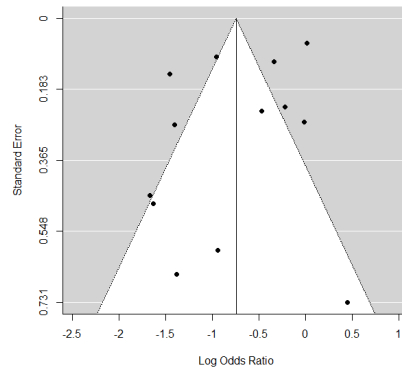
Forest plot for the LOS data

Edinburgh	-0.36 [-0.58, -0.13]
Orpington-Mild	-0.35 [-0.85, 0.15]
Orpington-Moderate	-2.32 [-2.74, -1.90]
Orpington-Severe	-1.89 [-2.67, -1.10]
Montreal-Home	-0.38 [-1.27, 0.50]
Montreal-Transfer	0.17 [-0.20, 0.55]
Newcastle	0.27 [-0.21, 0.75]
Umea	-0.42 [-0.66, -0.19]
Uppsala	0.29 [-0.08, 0.66]
RE Model	-0.54 [-1.14, 0.07]

Standardized Mean Difference
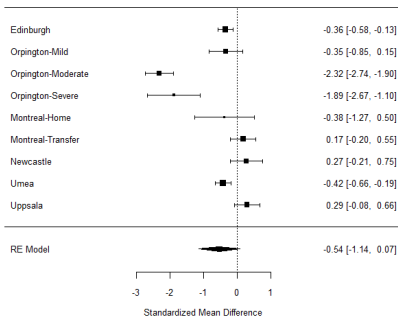
Forest plot for the BCG data

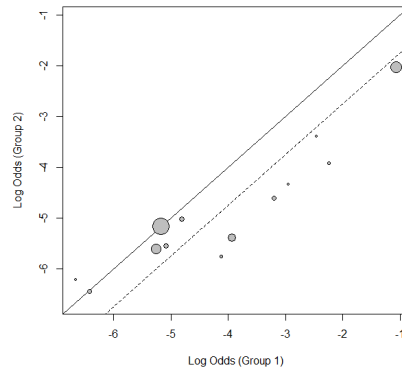# Funnel plot for the LOS data

Funnel plot for the BCG data

L'Abbe plot
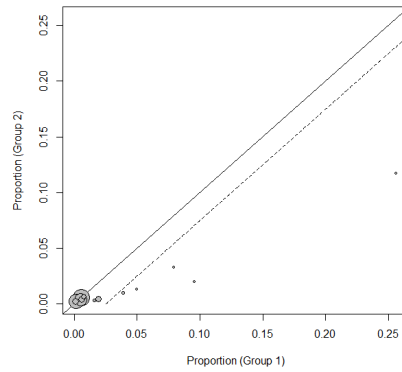
L'Abbe plot on log odds scale

Meta regression

```
Test of Moderators (coefficient(s) 2:3):
QM(df = 2) = 16.2533, p-val = 0.0003


Model Results:


         estimate        se     zval     pval
ci.lb    ci.ub
intrcpt  -10.5347  27.3739  -0.3848  0.7004  -
64.1865  43.1172
ablat     -0.0288   0.0095  -3.0311  0.0024   -
0.0475  -0.0102   **
year       0.0055   0.0138   0.3949  0.6929   -
0.0216   0.0325
```

Change in heterogeneity for meta regression model

```
tau^2 (estimated amount of residual
heterogeneity):     0.0913 (SE = 0.0745)

tau (square root of estimated tau^2 value):
0.3022

I^2 (residual heterogeneity / unaccounted
variability): 67.29%

H^2 (unaccounted variability / sampling
variability):   3.06

R^2 (amount of heterogeneity accounted for):
72.96%


Test for Residual Heterogeneity:
QE(df = 10) = 25.0121, p-val = 0.0053
```
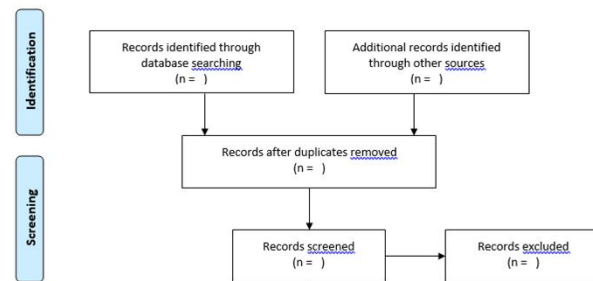
Publication guidelines

– PRISMA (2009)
  - Preferred Reporting Items for Systematic Reviews and Meta-Analyses
– QUOROM (1996)
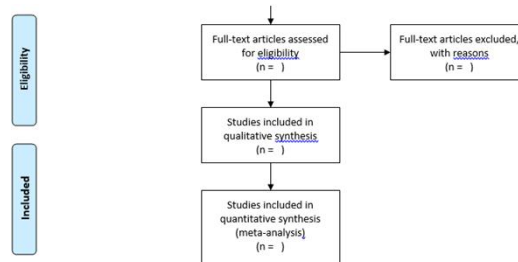 - QUality Of Reporting Of Meta-analysis

PRISMA flow diagram

PRISMA 2009 Flow Diagram

Identification

Screening

Records identified through
database searching
(n =)

Additional records identified
through other sources
(n =)

Records after duplicates removed
(n =)

Records screened
(n =)

Records excluded
(n =)

Top half of PRISMA flow diagram

PRISMA flow diagram

Bottom half of PRISMA flow diagram

PRISMA checklist, part 1

Section/topic	#	Checklist item
TITLE		
Title	1	Identify the report as a systematic review, meta-analysis, or both.
ABSTRACT		
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.
INTRODUCTION		
Rationale	3	Describe the rationale for the review in the context of what is already known.
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).
METHODS		
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.

Checklist for PRISMA

PRISMA checklist, part 2

Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.

Checklist for PRISMA

PRISMA checklist, part 3

Section/topic	#	Checklist item
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.
RESULTS		
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.

Checklist for PRISMA

PRISMA checklist, part 4

Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).
DISCUSSION		
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.
FUNDING		
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.

Checklist for PRISMA

Conclusion

– Design of meta-analysis
  • Include a librarian
  • Aggressive search strategy
– Statistical summary
  • Random versus fixed effects versus meta regression
  • Forest, funnel, L'Abbe plots
  • Cochrane's Q and I-squared
– Reporting
  • PRISMA flow diagram and checklist