MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

HEADS PhD PROGRAMME IN HEALTH DATA SCIENCE

DECIDE

Introduction to Health Interventions, Policy and Services

Coordinator:

Luís Filipe Azevedo

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

HEADS PhD PROGRAMME IN HEALTH DATA SCIENCE

DECIDE

Introduction to Health Interventions, Policy and Services

Methods in Evidence Synthesis – Part II

Authors

Luís Filipe Azevedo

Bernardo Sousa Pinto

HEADS

Summary

- Meta-analysis
- Heterogeneity: assessment and exploration
- Publication biases
- Protocol and dissemination of results
- GRADE approach

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

HEADS

Evidence synthesis

- Background
- Research question
- Primary studies search and selection
- Data extraction and processing
- Quality assessment
- Data analysis, heterogeneity and publication biases
- Protocol and dissemination of results

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

Systematic reviews and meta-analyses

- Qualitative synthesis
  - General and methodological characteristics of included studies
  - Quality criteria
  - Obtained results
- Quantitative synthesis
  - Identification and obtention of effect measures
  - Obtention of measures assessing the precision of estimates (standard-errors or confidence intervals)
  - Quantitative synthesis – meta-analysis
  - Heterogeneity assessment

Meta-analysis

Statistical analysis

- Effect measures
  - Identification of the effect measure (e.g., assessing frequency, association…) capable of quantiatively providing – for each primary study – the response to the defined research question
    - For example, in a randomised controlled trial, several quantiative measures – including the risk difference or the relative risk, among other commonly used effect measures - can be chosen to assess the efficacy of the intervention.

Statistical analysis

- Types of effect measures (EF)
  - Continuous variables
    - Difference in means (DM)
    - Standardized mean difference (SMD)
  - Binary variables
    - Proportions – Prevalence (P), cumulative incidence (CI)
    - Rates – Incidence rate (IR)
    - Risk difference – Absolute risk reduction (RD)
    - Relative risk – risks ratio (RR)
    - Odds ratio (OR)
  - Other EF
    - Correlation coefficient
    - Hazard ratio (HR) and other time-to-event measures
    - Number-needed-to-treat (NNT)

Prevalence:

$$P = \frac{n}{N} \qquad SE_P = \sqrt{\frac{P(1-P)}{N}} \qquad 95\%CI = P \pm 1.96\sqrt{\frac{P(1-P)}{N}}$$

Cumulative incidence:

$$CI = \frac{I}{N} \qquad SE_{CI} = \sqrt{\frac{CI(1-CI)}{N}} \qquad 95\%CI = CI \pm 1.96\sqrt{\frac{CI(1-CI)}{N}}$$

Incidence rate:

$$IR = \frac{I}{PT} \qquad SE_{IR} = \frac{\sqrt{I}}{PT} \qquad 95\%CI = IR \pm 1.96\frac{\sqrt{I}}{PT} = \frac{I \pm 1.96\sqrt{I}}{PT}$$

Risk difference:

$$RD = CI_E - CI_U = \frac{a}{a+b} - \frac{c}{c+d} \qquad SE_{RD} = \sqrt{\frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}}$$

$$95\%CI = RD \pm 1.96\sqrt{\frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}}$$

Relative risk (risk ratio):

$$RR = \frac{CI_E}{CI_U} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} \qquad SE_{\ln RD} = \sqrt{\frac{b}{a(a+b)} + \frac{d}{c(c+d)}}$$

$$95\%CI = e^{\ln RR \pm 1.96\sqrt{\frac{b}{a(a+b)} + \frac{d}{c(c+d)}}}$$

Incidence rates difference:

$$IRD = IR_E - IR_U = \frac{I_E}{PT_E} - \frac{I_{NE}}{PT_{NE}} \qquad SE_{IRD} = \sqrt{\frac{I_E}{PT_E^2} + \frac{I_{NE}}{PT_U^2}}$$

$$95\%CI = IRD \pm 1.96\sqrt{\frac{I_E}{PT_E^2} + \frac{I_{NE}}{PT_U^2}}$$

Incidence rates ratio:

$$IRR = \frac{IR_E}{IR_U} = \frac{\frac{I_E}{PT_E}}{\frac{I_{NE}}{PT_{NE}}} \qquad SE_{\ln IRR} = \sqrt{\frac{1}{I_E} + \frac{1}{I_{NE}}}$$

$$95\%CI = e^{\ln IRR \pm 1.96\sqrt{\frac{1}{I_E} + \frac{1}{I_{NE}}}}$$

Odds ratio:

$$OR = \frac{ad}{bc} \qquad SE_{\ln OR} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

$$95\%CI = e^{\ln OR \pm 1.96\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$

Difference in means:

$$DM = \bar{X}_{G1} - \bar{X}_{G2} \qquad SE_{DM} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

Standardized mean difference (SMD, Cohen's d):

$$d = \frac{\bar{X}_{G1} - \bar{X}_{G2}}{s_{pooled}} \qquad SE_d = \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2} + \frac{d^2}{2 \cdot (n_1 + n_2)}}$$

$$s_{pooled} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}} \qquad d' = d\left[1 - \frac{3}{4(n_1 + n_2) - 9}\right]$$

Pearson correlation coefficient (r):

$$Z_r = \frac{1}{2}\ln\left[\frac{1 + r}{1 - r}\right] \Leftrightarrow r = \frac{e^{2Z_r} - 1}{e^{2Z_r} + 1} \qquad SE_d = \sqrt{\frac{1}{n - 3}}$$

r – Pearson correlation coefficient
Z_r – Fisher transformation of the correlation coefficient

HEADS

Statistical analysis

- Principles for pooling effect measures
  - Transformation (to a scale in which the asymptotic normality of the effects estimators can be assumed; e.g., logarithmic transformation of effect measures expressed by ratios – odds ratios, relative risks…)
  - Obtention of weighted pooled effects measures
  - Weighting of each individual effect measure by the inverse variance of the effect estimator

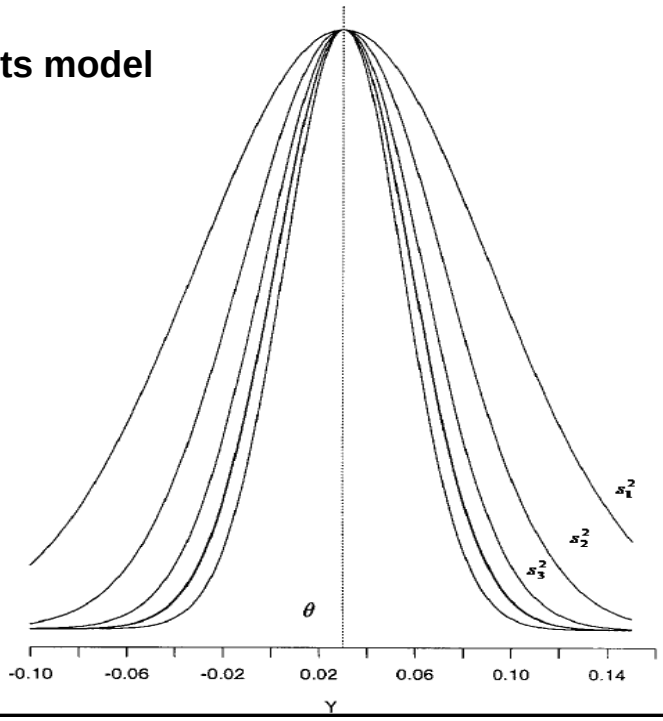
HEADS

Statistical analysis

- Quantitative synthesis – meta-analysis
  - **Fixed effects model** (assumes that all studies were obtained from the same population; weighting given by $w_i=1/s_i^2$)
  - **Random effects model** (assumes that all studies were obtained from diferente populations, estimating an overall measure that takes into account inter-studies variability; weighting given by $w_i=1/[s_i^2 + \tau^2]$)

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

Fixed effects model



Fixed effects model

$$Y_i \overset{indep}{\sim} N\left(\theta, s_i^2\right) \quad for \quad i = 1,2,...,k$$

$$\ell\left(\theta | Y_i\right) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi \cdot s_i^2}} e^{-\frac{1}{2}\left(\frac{Y_i - \theta}{s_i}\right)^2}$$

$$\hat{\theta}_{EMV} = \frac{\sum_{i=1}^{k}\left[\frac{Y_i}{s_i^2}\right]}{\sum_{i=1}^{k}\left[\frac{1}{s_i^2}\right]} = \frac{\sum_{i=1}^{k} Y_i W_i}{\sum_{i=1}^{k} W_i}$$

$$W_i = \frac{1}{s_i^2}$$

Test for effects homogeneity

$$H_0 : \theta = \theta_1 = \theta_2 = ... = \theta_k$$

H_1 : At least one of the θ_i is different from the others

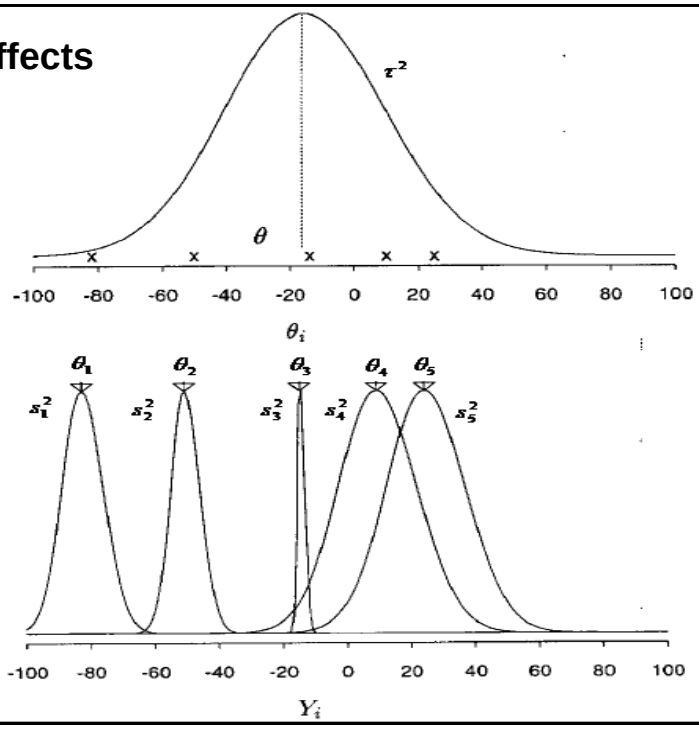
Under the null hypothesis validity, and for large enough samples, the statistic:

$$Q_W = \sum_{i=1}^{k} W_i \left(Y_i - \hat{\theta}_{EMV}\right)^2$$

follows an asymptotic distribution of chi-square with k-1 degrees of freedom.

$$\hat{\theta}_{EMV} = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \qquad W_i = \frac{1}{s_i^2}$$

Random effects model



MEDCIDS
DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE

Random effects model

$$Y_i \left| \theta_i, s_i^2 \right. \overset{indep}{\sim} N\left(\theta_i, s_i^2\right) \quad for \quad i = 1,2,\ldots,k$$

$$\theta_i \left| \theta, \tau^2 \right. \overset{indep}{\sim} N\left(\theta, \tau^2\right) \quad for \quad i = 1,2,\ldots,k$$

$$Y_i \left| s_i^2, \theta_i, \theta, \tau^2 \right. \overset{indep}{\sim} N\left(\theta, s_i^2 + \tau^2\right) \quad for \quad i = 1,2,\ldots,k$$

$$\theta_i \left| \mathbf{y}, \theta, \tau^2 \right. \sim N\left(B_i\theta + (1 - B_i)Y_i \ , \ s_i^2(1 - B_i)\right) \quad for \quad i = 1,2,\ldots,k$$

$$\hat{\theta}_{EMV} = \frac{\sum_{i=1}^{k}\left[\frac{Y_i}{s_i^2 + \tau^2}\right]}{\sum_{i=1}^{k}\left[\frac{1}{s_i^2 + \tau^2}\right]} = \frac{\sum_{i=1}^{k}\left[Y_i W_i(\tau)\right]}{\sum_{i=1}^{k}\left[W_i(\tau)\right]}$$

$$B_i = \frac{s_i^2}{s_i^2 + \tau^2} \qquad W_i(\tau) = \frac{1}{s_i^2 + \tau^2}$$

Random effects model

DerSimonian and Laird method

$$\hat{\tau}^2_{DL} = \max\left\{0, \frac{Q_W - (k-1)}{\sum_{i=1}^{k} W_i - \frac{\sum_{i=1}^{k} W_i^2}{\sum_{i=1}^{k} W_i}_{i}}\right\}$$

$$\hat{\theta}_{DL} = \frac{\sum_{i=1}^{k}\left[Y_i W_i(\hat{\tau}_{DL})\right]}{\sum_{i=1}^{k}\left[W_i(\hat{\tau}_{DL})\right]} \quad ; \quad W_i(\hat{\tau}_{DL}) = \frac{1}{s_i^2 + \hat{\tau}^2_{DL}}$$

MEDCIDS
DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE

HEADS
PhD PROGRAMME IN HEALTH DATA SCIENCE

Data analysis

- Forest Plots

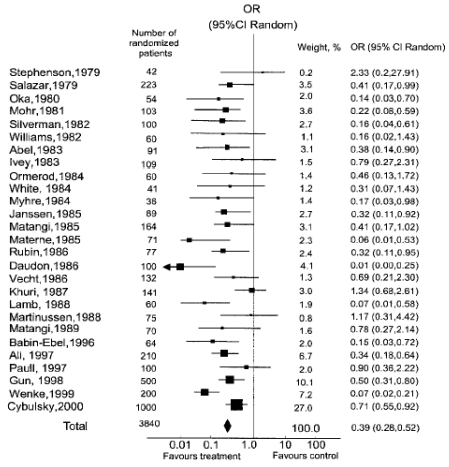


Figure 1. β-Blockers versus placebo or no treatment for the prevention of AF. Test for heterogeneity P=0.00001. Test for overall effect P<0.00001.

MEDCIDS
DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE
U.PORTO
CINTESIS
Health. Research.

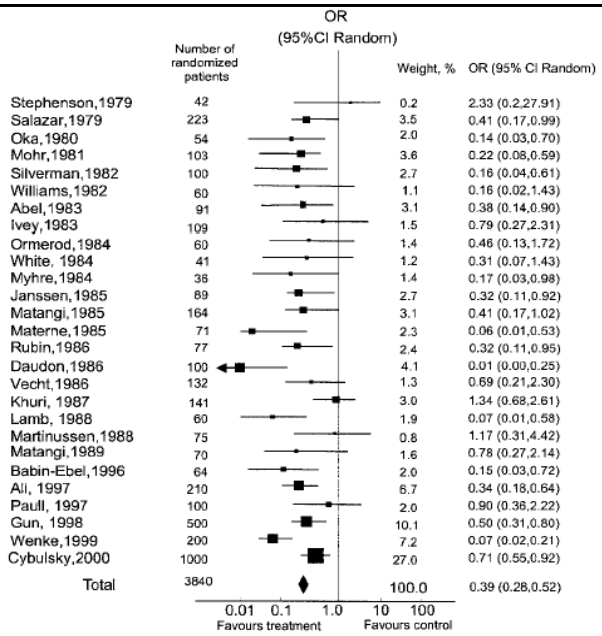
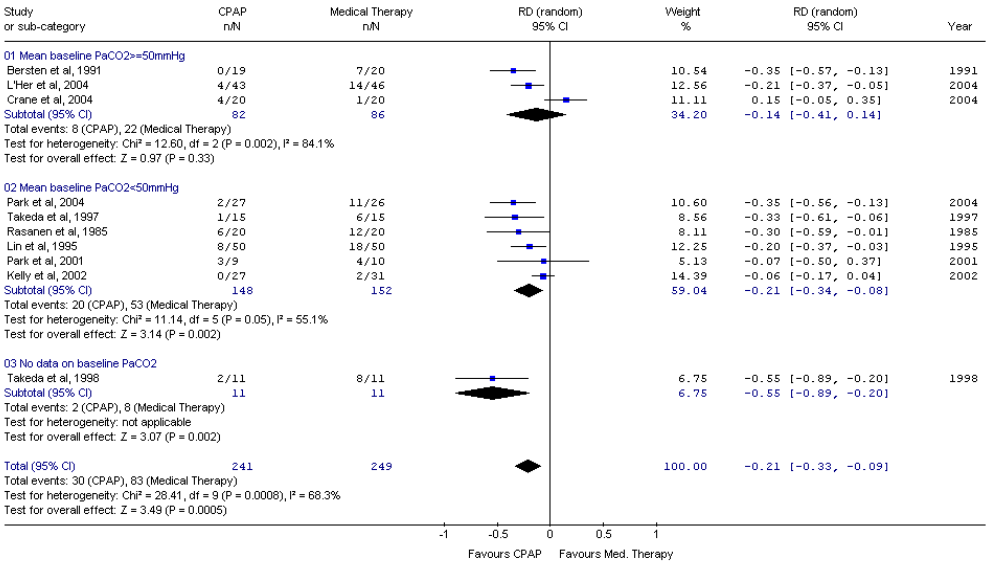


Figure 1. β-Blockers versus placebo or no treatment for the prevention of AF. Test for heterogeneity P=0.00001. Test for overall effect P<0.00001.



Heterogeneity – assessment and exploration

HEADS

Heterogeneity

- Definition
  - Differences in results between studies that cannot be explained by sampling variability and random error
- Recommendations
  - Assessment of the magnitude of heterogeneity
  - Identification of the causes/sources of heterogeneity
  - Exploration of the causes/sources of heterogeneity
  - Adaptation of data analysis methods
- Sources of heterogeneity
  - Clinical heterogeneity
  - Methodological heterogeneity

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

HEADS

Heterogeneity

- Assessing heterogeneity
  - Chi-square test – Cochran Q statistic
  - Q/GL(>1 possible heterogeneity; <1 improbable heterogeneity)
  - I^2statistics =[(Est.Q - GL)/Est.Q] ×100 %
  - Forest plot and L'Abbé plot
- Exploring the causes of heterogeneity
  - Adoption of a random effects model (E.g., DerSimonian & Laird random effects model)
  - Subgroup and sensitivity analyses
  - Change of scale
  - Meta-regression

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

Test for effects homogeneity

$$H_0 : \theta = \theta_1 = \theta_2 = ... = \theta_k$$

$$H_1 : \text{At least one of the } \theta_i \text{ is different from the others}$$

Under the null hypothesis validity, and for large enough samples, the statistic:

$$Q_W = \sum_{i=1}^{k} W_i \left(Y_i - \hat{\theta}_{EMV} \right)^2$$

follows an asymptotic distribution of chi-square with k-1 degrees of freedom.

$$\hat{\theta}_{EMV} = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \qquad W_i = \frac{1}{s_i^2}$$

Heterogeneity – Other statistics

Higgins&Thompson H statistic: $H = \sqrt{Q/(k-1)}$

The statistic I^2 $\quad I^2 = \dfrac{Q-(k-1)}{Q} \times 100\%$

HEADS

Heterogeneity

- Subgroup analysis
  - Identification of variables potentially explaining heterogeneity
  - Analysis within each subgroup according those variables
  - Use of hypothesis tests to test for subgroup differences (i.e., differences between subgroups).

Heterogeneity

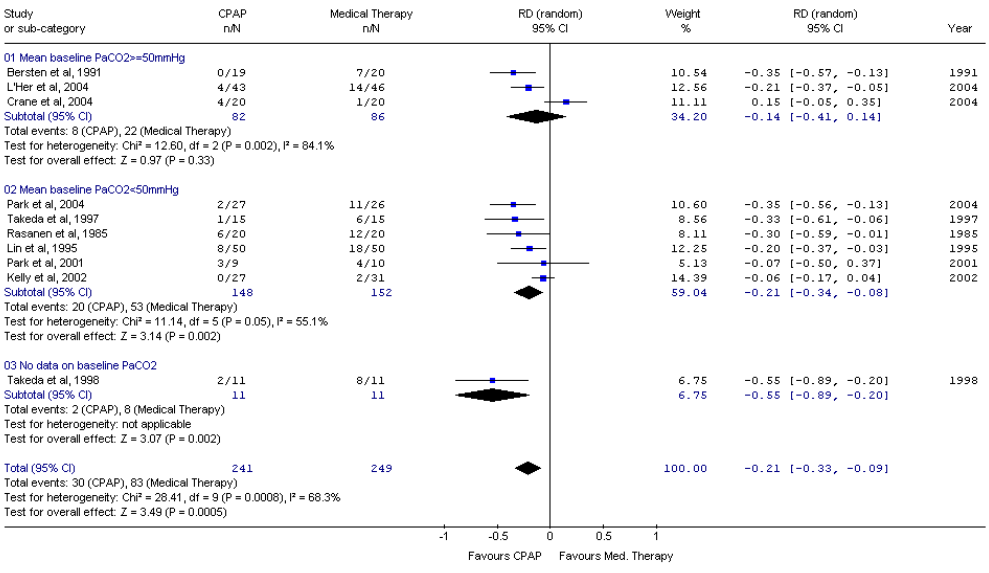
- Subgroup analysis

Test for comparison of subgroups

The test is valid for all methods except the Mantel-Haenszel methods for binary data. The Q statistic defined by either Q_{IV} or Q_{Peto} is calculated separately for each of the S subgroups and for the totality of studies, yielding statistics $Q_1, \ldots, Q_S$ and Q_{tot}. The test statistic is given by

$$Q_{\text{int}} = Q_{\text{tot}} - \sum_{j=1}^{S} Q_j .$$

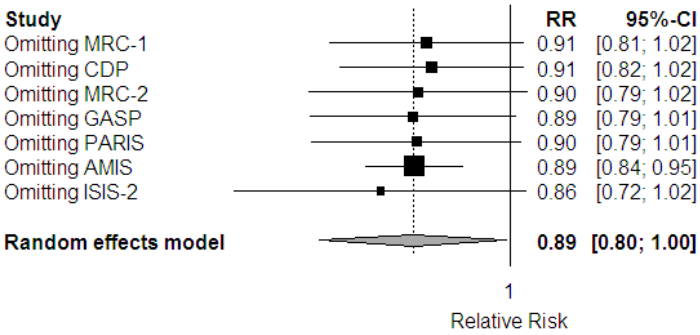
Under the null hypothesis that there are no differences in treatment effect among subgroups this follows a chi-squared distribution with $S-1$ degrees of freedom (where S is the number of subgroups).



Heterogeneity

- Sensitivity analysis
  - Analysis of the meta-analysis results when each study is removed one by one (leave-one-out sensitivity analysis), or when subgroups of studies are removed
  - Assessment of the pooled statistics and of the heterogeneity

Meta-analysis on Aspirin in Preventing Death after Myocardial Infarction



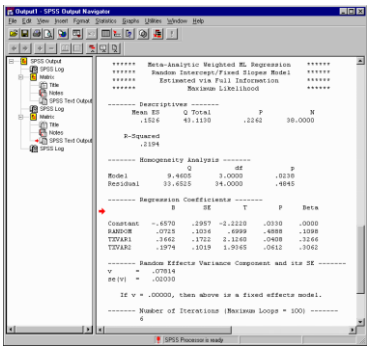
MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

HEADS

Heterogeneity

- Meta-regression

$$y_i \sim \mathrm{N}(\alpha + \beta x_i, v_i + \tau^2).$$



MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

Publication biases

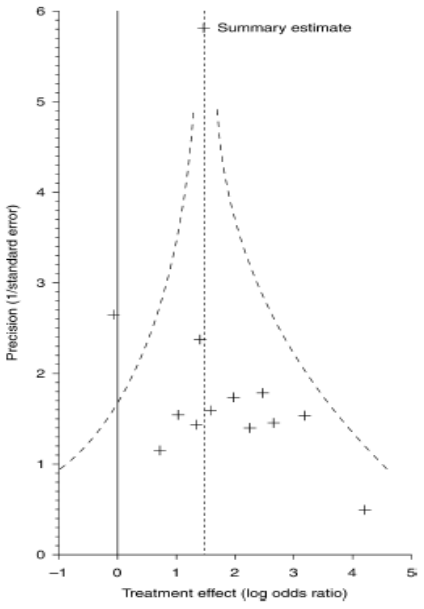
Table 10.1.a Definitions of some types of reporting biases

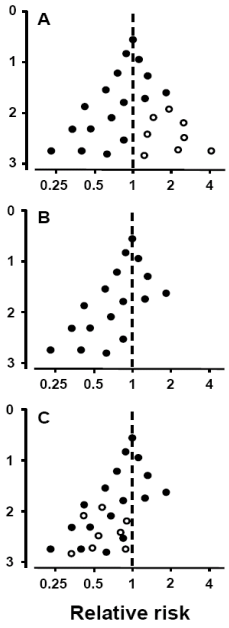
Type of reporting bias	Definition
Publication bias	The *publication* or *non-publication* of research findings, depending on the nature and direction of the results
Time lag bias	The *rapid* or *delayed* publication of research findings, depending on the nature and direction of the results
Multiple (duplicate) publication bias	The *multiple* or *singular* publication of research findings, depending on the nature and direction of the results
Location bias	The publication of research findings in journals with different *ease of access* or *levels of indexing* in standard databases, depending on the nature and direction of results.
Citation bias	The *citation* or *non-citation* of research findings, depending on the nature and direction of the results
Language bias	The publication of research findings *in a particular language*, depending on the nature and direction of the results
Outcome reporting bias	The *selective reporting* of some outcomes but not others, depending on the nature and direction of the results

HEADS

Publication biases

- Importance of a systematic and comprehensive search
- Analysis methods
  - Funnel plot
  - Rank correlation test of Begg
  - Others





Note: Dashed outer lines show boundaries of an 'ideal' funnel; if there is no heterogeneity the points are distributed evenly on either side of the summary estimate.

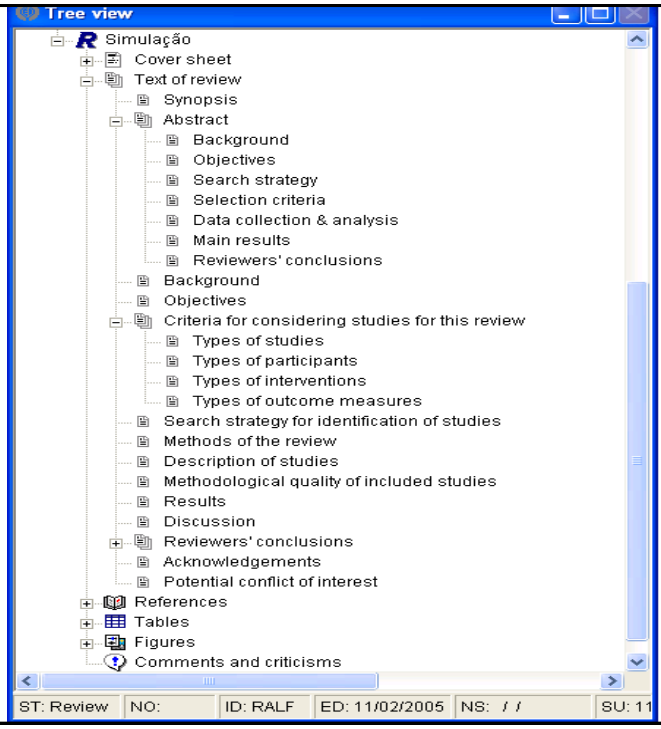
Publication bias

- Funnel plot asymetry

Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315: 629–634.

1. **Selection biases**:
   - Publication bias:
     - Delayed publication (also known as 'time-lag' or 'pipeline') bias.
     - Location biases:
       - Language bias;
       - Citation bias;
       - Multiple publication bias.
   - Selective outcome reporting.
2. **Poor methodological quality leading to spuriously inflated effects in smaller studies**:
   - Poor methodological design;
   - Inadequate analysis;
   - Fraud.
3. **True heterogeneity**:
   - Size of effect differs according to study size (for example, due to differences in the intensity of interventions or differences in underlying risk between studies of different sizes).
4. **Artefactual**:
   - In some circumstances (see Section 10.4.3), sampling variation can lead to an association between the intervention effect and its standard error.
5. **Chance**.

Protocol and dissemination of results



Box 2.3.a Sections of a protocol for a Cochrane review

Title*
Protocol information:
- Authors*
- Contact person*
- Dates
- What's new
- History

The protocol:
- Background*
- Objectives*
- Methods:
  - Criteria for selecting studies for this review:
    - Types of studies*
    - Types of participants*
    - Types of interventions*
    - Types of outcome measures*
  - Search methods for identification of studies*
  - Data collection and analysis*
- Acknowledgements
- References:
  - Other references:
    - Additional references
    - Other published versions of this review
- Tables and figures:
  - Additional tables
  - Figures

Supplementary information:
- Appendices
- Feedback:
  - Title
  - Summary
  - Reply
  - Contributors

About the article:
- Contributions of authors
- Declarations of interest*
- Sources of support:
  - Internal sources
  - External sources
- Published notes

Box 2.3.b Sections of a Cochrane review

Title*
Review information:
- Authors*
- Contact person*
- Dates*
- What's new
- History

Abstract:
- Background*
- Objectives*
- Search strategy*
- Data collection and analysis*
- Results*
- Authors' conclusions*

Plain language summary:
- Plain language title*
- Summary text*

The review:
- Background*
- Objectives*
- Methods:
  - Criteria for selecting studies for this review:
    - Types of studies*
    - Types of participants*
    - Types of interventions*
    - Types of outcome measures*
  - Search methods for identification of studies*
  - Data collection and analysis*
- Results:
  - Description of studies*
  - Risk of bias in included studies*
  - Effects of interventions*
- Discussion*
- Authors' conclusions:
  - Implication for practice*
  - Implication for research*
- Acknowledgements
- References:
  - References to studies:
    - Included studies
    - Excluded studies
    - Studies awaiting classification
    - Ongoing studies

Box 2.3.b Sections of a Cochrane review (cont.)

- Other references:
  - Additional references
  - Other published versions of this review
- Tables and figures:
  - Characteristics of studies:
    - Characteristics of included studies (*includes 'Risk of bias' tables*)
    - Characteristics of excluded studies
    - Characteristics of studies awaiting assessment
    - Characteristics of ongoing studies
  - 'Summary of findings' tables
  - Additional tables
  - Figures
- **Supplementary information:**
  - Data and analyses
  - Appendices
  - Feedback:
    - Title
    - Summary
    - Reply
    - Contributors
- **About the article:**
  - Contributions of authors
  - Declarations of interest*
  - Differences between protocol and review
  - Sources of support:
    - Internal sources
    - External sources
  - Published notes

HEADS

Dissemination of results

- Guidelines for redacting systematic reviews and meta-analyses
  - PRISMA
  - MOOSE
- Work base
  - Cochrane Collaboration structure for results dissemination: Available through RevMan

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement

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.
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.
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.

GRADE approach

HEADS

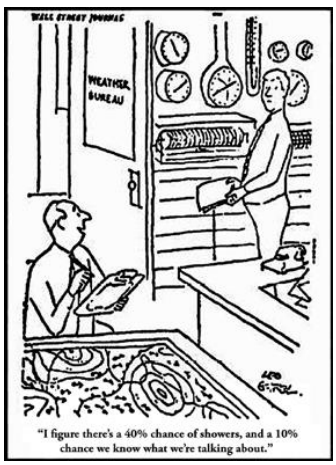
Grading of Recommendations Assessment, Development and Evaluation

GRADE

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

HEADS

Why bother about grading evidence?



People draw conclusions about
- Quality of evidence
- Strength of recommendations

Systematic, explicit approaches help
- Facilitate critical appraisal
- Protect against errors
- Resolve disagreements
- Communicate information

GRADE Working Group (2004)

(1) Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490–1494. (2) Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–926.

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.

GRADE

Grading of **R**ecommendations **A**ssessment, **D**evelopment and **E**valuation

System (and common language) that

- was developed and updated by GRADE Working Group
- has been endorsed by large number of organisations
- expresses degree of confidence one can place in **quality of evidence** and **strength of recommendation**

System for assessing quality of evidence of a body of evidence based on

- study design
- criteria for downgrading/upgrading

GRADE Working Group (2004); Guyatt et al (2008)

(1) Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490–1494. (2) Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–926.

GRADE quality of evidence

Definition: The degree of confidence in an estimate of effect.

High (++++)	We are very confident that the true effect lies close to that of the estimate of the effect.
Moderate (+++)	We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low (++)	Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
Very low (+)	We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of the effect.

(1) Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490–1494. (2) Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–926.

Determinants of quality of evidence (1)

Study design:

- Bodies of RCTs start as high
- All other study designs start as low

Upgrading/downgrading criteria:

- Five factors can decrease quality of evidence
- Three factors can increase quality of evidence

RCTs versus observational study designs usually run through assessment separately

Determinants of quality of evidence (2)

Quality	Study design	Lower quality if:	Higher quality if:
High (++++)	Randomised controlled trials	Risk of bias (-1, -2) Inconsistency (-1, -2) Indirectness (-1, -2) Imprecision (-1, -2) Publication bias (-1, -2)	Large effect (+1, +2) Dose-response (+1) Direction of residual confounding and biases (+1)
Moderate (+++)			
Low (++)	Observational studies		
Very low (+)	Case reports, expert opinion, modelling		

Risk of bias (-/--)

- Risk of bias:
  Limitations in study design and execution with relevance to given outcome, based on quality appraisal of individual studies

- Indicators of risk of bias
  - Moderate or high risk of bias across most studies
  - etc.

- Lower quality of evidence
  - -1 if serious limitations in study design and execution
  - -2 if very serious limitations in study design and execution

Risk of bias (-/--)

Limitations in RCTs
lack of concealment
intention to treat principle violated
inadequate blinding
loss to follow-up
early stopping for benefit
selective outcome reporting

Inconsistency (-/--)

- Consistency:
  Similarity of estimates of effect across studies

- Indicators of inconsistency
  - Differences in direction of effect
  - Variation in size of effect
  - Large I^2 value
  - etc.
- Distinguish between
  - **explained** heterogeneity (e.g. population, intervention, outcome)
  - **unexplained** heterogeneity

- Lower quality of evidence
  - -1 if large unexplained inconsistency
  - -2 if very large unexplained inconsistency

Indirectness (-/--)

- Directness:
  Extent to which populations, interventions, comparisons and outcomes are similar to those of interest

- Indicators of indirectness
  - Very different populations (e.g. age, sex, illness)
  - Surrogate outcomes
  - No direct comparisons
  - etc.

- Lower quality of evidence
  - -1 if serious uncertainty about directness
  - -2 if very serious uncertainty about directness

Imprecision (-/--)

- Precision:
  Is a consequence of sample size and number of events

- Indicators of imprecision
  - Small population (sparse data)
  - Small number of events
  - Wide confidence intervals around pooled effect (e.g. including RR=1)
  - etc.

- Lower quality of evidence
  - -1 if imprecise or sparse data
  - -2 if very imprecise or sparse data

Publication bias (-/--)

- Publication bias:
  Systematic under- or overestimate of effect due to selective publication of studies

- Indicators of publication bias
  - Small studies
  - Industry-sponsored studies
  - Asymmetric funnel plot
  - etc.

- Lower quality of evidence
  - -1 if publication bias is strongly suspected
  - -2 if publication bias is very strongly suspected

Three factors for upgrading

- Large or very large effects are less likely to be spurious
  - +1 if relative risk reduction ≤ 0.5 or risk ratio ≥ 2
  - +2 if relative risk reduction ≤ 0.8 or risk ratio ≥ 5
- Evidence of dose-response gradient
  - +1 if dose-response gradient observed
- If effect observed: All plausible residual confounding and biases would have reduced effect
  If no effect observed: All plausible residual confounding and biases would have increased the effect
  - +1 if appropriate direction of residual confounding and biases

HEADS

GRADE strength of recommendation

Definition: The degree of confidence that desirable effects of adherence to a recommendation outweigh undesirable effects.

- **Strong** recommendation:
  The panel is confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects.
- **Weak/ conditional/ discretionary** recommendation:
  The panel concludes that the desirable effects of adherence to a recommendation probably outweigh the undesirable effects, but is not confident.

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health.Research.

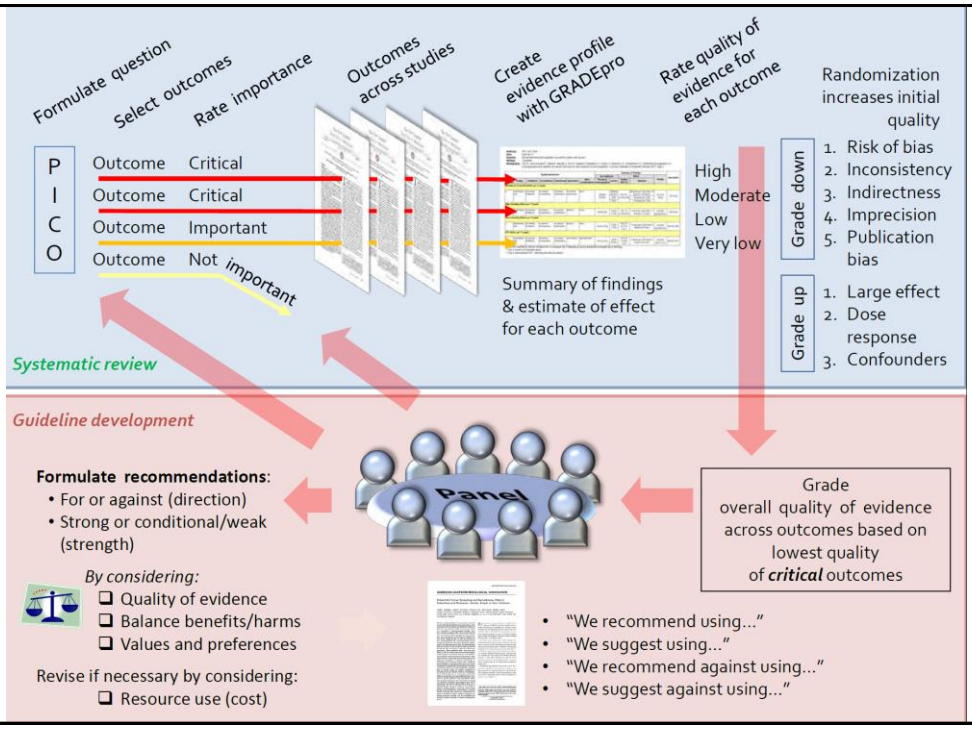
(1) Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490–1494. (2) Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–926.

HEADS

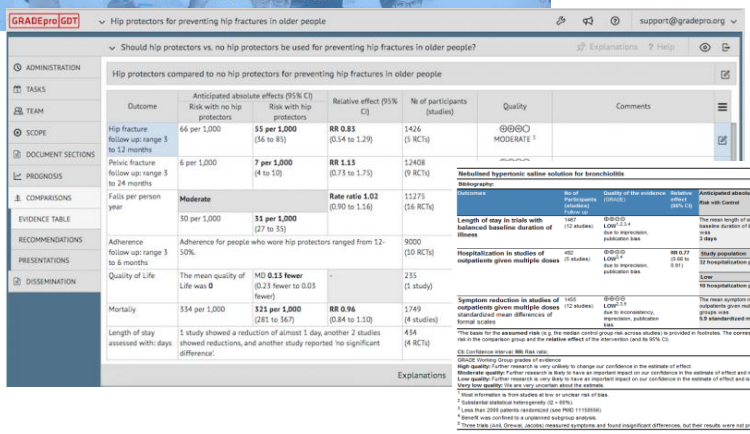
Low quality of evidence, strong recommendation?

- Clear distinction between
  - quality of evidence, driven by confidence in body of evidence
  - strength of recommendation, driven by confidence in body of evidence plus other considerations
- All combinations possible, e.g.:
  - High quality of evidence -> weak recommendation
  - Very low quality of evidence -> strong recommendation
- Paradigmatic situations for strong recommendations in the absence of high-quality evidence (Alexander et al. 2016)

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE U.PORTO CINTESIS Health. Research.



https://gradepro.org/



HEADS

Bibliography

- **Higgins J, Thomas J, Ed. Cochrane handbook for systematic reviews of interventions. 2019.**
  - Core methods – Chapters 6, 9-10 and 13-15.

- **Recommended articles**
  - As available in *Moodle*

MEDCIDS DEPARTAMENTO DE MEDICINA DA COMUNIDADE, INFORMAÇÃO E DECISÃO EM SAÚDE
U.PORTO
CINTESIS Health.Research.