

## Incorporating study quality into meta-analyses

Adrian Barnett

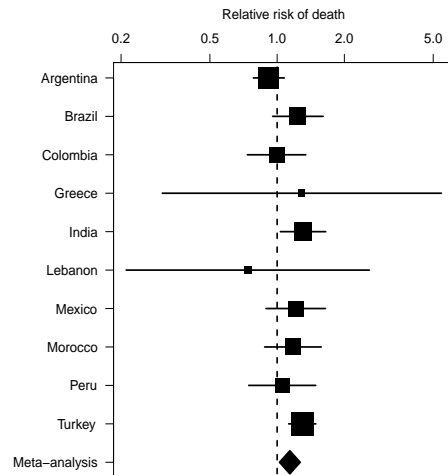
School of Public Health  
Queensland University of Technology

25 June 2010

## A perfect meta-analysis

Almost no heterogeneity and a statistically significant mean

- ▶ Forest plot: squares proportional to study weight (created using *rmeta* in *R*)
- ▶ Risk of in-hospital death after a ventilator acquired pneumonia
- ▶ 10 countries
- ▶ Mean relative risk = 1.14 (95% CI: 1.02, 1.27) (using random effects)
- ▶ Heterogeneity variance = 0.009 ( $p = 0.134$ )
- ▶ Same protocol in each country



## Preamble

Plan:

- ▶ Heterogeneity
- ▶ Assessing quality
- ▶ Meta-regression

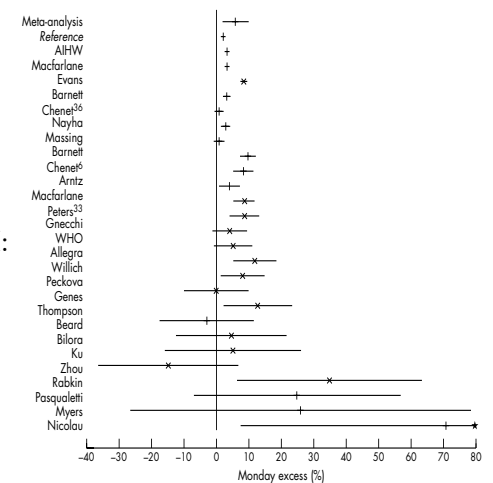
Two principles:

- ▶ Distinction between quality of **individual studies** and the quality of **meta-analyses**
- ▶ Higher quality  $\Rightarrow$  Greater accuracy
- ▶ My examples are from **observational, cohort** and **diagnostic** studies (rather than RCTs)

## A less than perfect meta-analysis

Huge heterogeneity

- ▶ Excess coronary events on Mondays (Barnett & Dobson 2005)
- ▶ 28 studies
- ▶ Mean excess (%) = 5.9 (95% CI: 2.0, 9.9) (using random effects)
- ▶ Estimated heterogeneity:  $I^2 = 87.3\%$  (Higgins & Thompson 2002)
- ▶ Meaningless mean?



## Sources of heterogeneity

- ▶ Sources due to **diversity**:
  - ▶ Populations (e.g., countries, age groups)
  - ▶ Interventions (e.g., dose, treatment length)
  - ▶ Inclusion/Exclusion criteria (e.g., more severe patients)
- ▶ Sources due to **bias** (“any process tending to produce results that differ systematically from the true values”, Jüni et al 2009)
  - ▶ Design (e.g., inadequate follow-up of lost patients)
  - ▶ Conduct (e.g., poor randomisation)
- ▶ **Publication bias**: “the phenomenon where uninteresting findings get left in a desk drawer unpublished forever”, Ben Goldacre
- ▶ Monday effect: publication bias + data quality bias (higher data quality associated with smaller Monday effect) + alcohol effect

## Heterogeneity can be useful

- ▶ Chance to uncover more about the treatment action (e.g., dose-response)
- ▶ Collect key information on each study
- ▶ Use this information to explain heterogeneity (e.g., latitude predicts the effects of heat waves—Hajat & Kosatsky 2009)
- ▶ If we can't explain the heterogeneity then:
  - ▶ Random effects (admit failure) + stress caution
  - ▶ *or* Do not do a meta-analysis
- ▶ Always create a forest plot

## Collecting the key information

- ▶ Think of all the things that may effect quality, e.g.,
  - ▶ Randomised vs observational
  - ▶ Prospective vs retrospective data collection
  - ▶ Cohort vs case-control
  - ▶ Outcome assessment (e.g., objectively assessed vs self-report)
  - ▶ Study funding
- ▶ Think of all the things that may effect the treatment difference, e.g.,
  - ▶ Source of patients (e.g., hospital vs GP)
  - ▶ Dose
  - ▶ Location (latitude for heat wave example)
- ▶ Created *a priori* to avoid “observer bias”
- ▶ Make a score sheet for each study

## QUADAS

Quality assessment of studies of diagnostic accuracy (Whiting et al 2003)

- ▶ 14 items scored as ‘Yes’, ‘No’, ‘Unclear’
  - ▶ Was the spectrum of patients representative of the patients who will receive the test in practice? **Yes**
  - ▶ Were selection criteria clearly described? **Unclear**
  - ▶ Is the reference standard likely to correctly classify the target condition? **No**
  - ▶ Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests? **Yes**
  - ▶ Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis? **Yes**
  - ▶ Did patients receive the same reference standard regardless of the index test result? **No**
  - ▶ Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)? **No**
  - ▶ Was the execution of the index test described in sufficient detail to permit replication of the test? **Yes**
  - ▶ Was the execution of the reference standard described in sufficient detail to permit its replication? **Yes**
  - ▶ Were the index test results interpreted without knowledge of the results of the reference standard? **No**
  - ▶ Were the reference standard results interpreted without knowledge of the results of the index test? **Yes**
  - ▶ Were the same clinical data available when test results were interpreted as would be available when the test is used in practice? **No**
  - ▶ Were uninterpretable/intermediate test results reported? **No**
  - ▶ Were withdrawals from the study explained? **Yes**

▶ Score = 7 yeses out of 14 = 50%

## Other meta-analysis guidelines/checklists

- ▶ PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses (Moher et al 2009); 27 items
- ▶ PRISMA came out of QUOROM (quality of reporting of meta-analysis)
- ▶ MOOSE: Meta-analysis of observational studies in epidemiology (Stroup et al 2000); 35 items
- ▶ Worth reading guidelines, even when submitting single study papers
- ▶ Design your own checklist (or augment an existing list) with those items that have special significance in your field
- ▶ MAGIC: Meta-analysis: Getting it correct (Barnett et al 2010)

## Collecting the key information

- ▶ Assessment can be **subjective** ⇒ use an independent researcher who is **blind** to your hypotheses
- ▶ But they may be less qualified—so the quality of the quality review is compromised!
- ▶ Assessed by two raters, disagreements examined using a third researcher (Kappa statistic for rater agreement)
- ▶ Do a pilot sample first (3–6 papers; try to include low/high quality studies)
- ▶ What if something isn't mentioned? E.g., no mention of whether doctors who assessed the scans were blind to treatment group
- ▶ Assume that the quality was inadequate unless the information to the contrary is provided (the “guilty until proved innocent” approach—Jüni et al 2001).

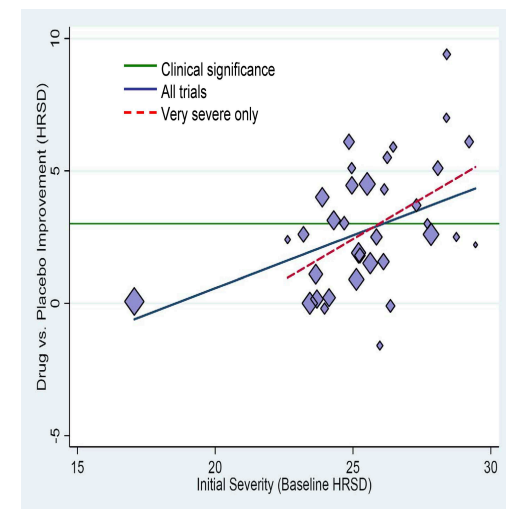
## What to do with study quality?

- ▶ Always report the quality of the studies (include results for each study in an appendix)
- ▶ **Weight** studies according to total quality score? Not recommended as poor studies still have some influence
- ▶ Only use those studies over some quality threshold. Becomes very dependent on the threshold
- ▶ **Plot** study results against quality scores
- ▶ **Stratify** on important variables instead (e.g., separate meta-analyses for cohort and case-control studies)
- ▶ Stratification decisions should be made *a priori*
- ▶ Leave-one-out meta-analysis can also be useful
- ▶ Cochrane Reviewers' Handbook: “None of the currently available scales for measuring the validity or ‘quality’ of trials can be recommended without reservation”

## Meta-regression

Using regression to explain heterogeneity

- ▶ Mean Drug–Placebo difference as a function of initial severity (Kirsch et al 2009)
- ▶ Studies of antidepressants
- ▶ Regression with weights based on study size
- ▶ Greater treatment effect in studies with more severe depression (greater placebo effect)



## Want to improve the quality of your meta-analysis?

- ▶ Then use a statistician...or an epidemiologist
- ▶ Delgado et al (2001) found that meta-analyses with a statistician/epidemiologist were more likely to:
  - ▶ Have a sample size calculation (Odds ratio = 1.5)
  - ▶ Have study start and end dates (OR  $\approx$  2.1)
  - ▶ Use an objectively assessed outcome (OR = 2.4)
  - ▶ Have an intention to treat analysis (OR = 2.0)
- ▶ 178 out of 467 studies had a stat/epi person (38%)
- ▶ **Big assumption:** Is the quality of the *report* of a published study related to its true quality?







## Summary

- ▶ Meta-analyses are a great way of increasing statistical power, and should give more **convincing evidence** than individual studies
- ▶ There's a lot of rubbish out there!
- ▶ "If the 'raw material' is flawed then the conclusions of systematic reviews cannot be trusted." Jüni et al (2001)
- ▶ Some type of quality review of studies is recommended as part of a meta-analysis
- ▶ But it can be time consuming, and may not justify the effort
- ▶ Less agreement on what to do with the results of the review
- ▶ Informal analysis (e.g., plots) favoured over formal analysis (e.g., weighted regression)
- ▶ Important decisions should be made *a priori*

## References

-  P Alderson, S Green, and J Higgins.  
*Cochrane Reviewers' Handbook 4.2.2.*  
John Wiley & Sons, Ltd, Chichester, UK, 2004.
-  A G Barnett and A J Dobson.  
The excess in cardiovascular events on Mondays: a meta-analysis and prospective study.  
*Journal of Epidemiology and Community Health*, 59:109–114, 2005.
-  M Delgado-Rodriguez, M Ruiz-Canela, J De Irala-Estevez, J Llorca, and M A Martinez-Gonzalez.  
Participation of epidemiologists and/or biostatisticians and methodological quality of published controlled clinical trials.  
*Journal of Epidemiology and Community Health*, 55(8):569–572, 2001.
-  P. P. Glasziou and S. L. Sanders.  
Investigating causes of heterogeneity in systematic reviews.  
*Statistics in Medicine*, 21(11):1503–1511, 2002.
-  Shakoor Hajat and Tom Kosatsky.  
Heat-related mortality: a review and exploration of heterogeneity.  
*Journal of Epidemiology and Community Health*, pages –, 2009.
-  P. T. Higgins Julian and G. Thompson Simon.  
Quantifying heterogeneity in a meta-analysis.  
*Statistics in Medicine*, 21(11):1539–1558, 2002.
-  P. T. Higgins Julian, G. Thompson Simon, and J. Spiegelhalter David.  
A re-evaluation of random-effects meta-analysis.  
*Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1):137–159, 2009.

## References

-  Peter Jüni, Anne Witschi, Ralph Bloch, and Matthias Egger.  
The Hazards of Scoring the Quality of Clinical Trials for Meta-analysis.  
*JAMA*, 282(11):1054–1060, 1999.
-  Irving Kirsch, Brett J Deacon, and et al.  
Initial severity and antidepressant benefits: A meta-analysis of data submitted to the food and drug administration.  
*PLoS Med*, 5(2):e45, 2008.
-  Thomas Lumley.  
*rmeta: Meta-analysis*, 2009.  
R package version 2.16.
-  David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and for the PRISMA Group.  
Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.  
*BMJ*, 339(jul21\_1):b2535, 2009.
-  D F Stroup, et al, and for the Meta-analysis Of Observational Studies in Epidemiology Group.  
Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting.  
*JAMA*, 283(15):2008–2012, 2000.
-  Penny Whiting, Anne Rutjes, Johannes Reitsma, Patrick Bossuyt, and Jos Kleijnen.  
The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews.  
*BMC Medical Research Methodology*, 3(1):25, 2003.