UNIVERSITY OF BIRMINGHAM

Empirical evidence concerning non-randomized studies

Jon Deeks and the BRANDO team

UNIVERSITY OF BIRMINGHAM

Some background ...

- RCTs regarded as a 'gold standard', principally because randomisation reduces the potential for *selection bias.*

- Many healthcare interventions will only ever be evaluated in non-randomised studies (NRS), and it is important to know how biased they may be.

- Estimates of bias also required for down-weighting processes in evidence syntheses, and to assess the need for RCTs

To understand risks of including NRS in Cochrane Reviews

- Need - the degree and consistency of bias in non-randomised studies (NRS) of healthcare interventions

- Method – comparison of results of RCTs with those of NRS that address the same healthcare questions, synthesised across many topics

8 Previous evaluations (+ 1 update)

Author (year)	RCTs/NRSs	Topics
Sacks (1982)	50/56	6
Kunz 1 (1998)	122/152	11
Kunz 2 (2002)	263/246	23
Britton (1998)	46/41	18
MacLehose (2000)	31/68	14 (38 outcomes)
Benson (2000)	85/53	19
Concato (2000)	55/44	5
Ioannidis (2001)	240/168	45
Lipsey (1996,2001)	Not stated	76

Previous evaluations ...

- Kunz and Oxman, BMJ 1998; 1185-90.
  - Compared RCTs with non-randomised trials for 8 interventions.
  - *"on average non-randomised trials result in overestimates of effect. This bias, however, can go in either direction, can reverse the direction of effect, or can mask an effect"*

- Britten, McKee, Black et al. HTA 1998; 2(13)
  - Compared individual RCTs with non-randomised comparable groups for 14 interventions.
  - *"results from RCTs and non-randomised studies do not inevitably differ … the effect of adjustment for baseline differences between groups in non-randomised studies is inconsistent".*

Previous evaluations ...

- Benson and Hartz. NEJM 2000; 342: 1878-86.
  - Compared 19 interventions with RCTs and observation studies of unspecified design (13 in common with Britten).

  - *"observational studies and randomised controlled trials usually produce similar results …. Our results suggest that observational studies usually do provide valid information".*

  - *"for most treatment comparisons there were insufficient data to exclude the possibility of clinically important differences between the two types of study"*

UNIVERSITY OF BIRMINGHAM

Previous evaluations ...

- Concato, Shah, Horwitz. NEJM 2000; 342: 1887-92.
  - Data from 5 MAs combining RCTs with concurrent cohort or case-control studies
  - "*the summary results of RCTs and observational studies were remarkably similar… the observational studies had less variability in point estimates than the RCTs*"

UNIVERSITY OF BIRMINGHAM

Previous evaluations ...

- MacLehose, Reeves, Harvey et al. HTA 2000: 4(34).
  - 38 comparisons (including Kunz's and Britten's) were graded as "fair" or "unfair" according to the similarity of population and methods used to control for biases.
  - *"Previous comparisons may have over-emphasised the differences between randomised and non-randomised evidence due to poor quality of most non-randomised evidence".*

UNIVERSITY OF BIRMINGHAM

The evidence …

- Report mixed findings
- Substantial overlap of data
- Possibly explained by differences in methods
- Critique according to first principles

UNIVERSITY OF BIRMINGHAM

Confounding

- Mixed methodological quality of RCTs
  - for example, with or without blinded outcome assessment

- Clinical heterogeneity
  - different forms of the intervention
  - recruiting different patient groups
  - using different outcome definitions.

Selection bias

- 6/8 used secondary research studies
  - comparisons of RCTs and NRS for a single topic
  - systematic reviews that included both RCTs and NRS.
  - cannot be regarded as representative samples.

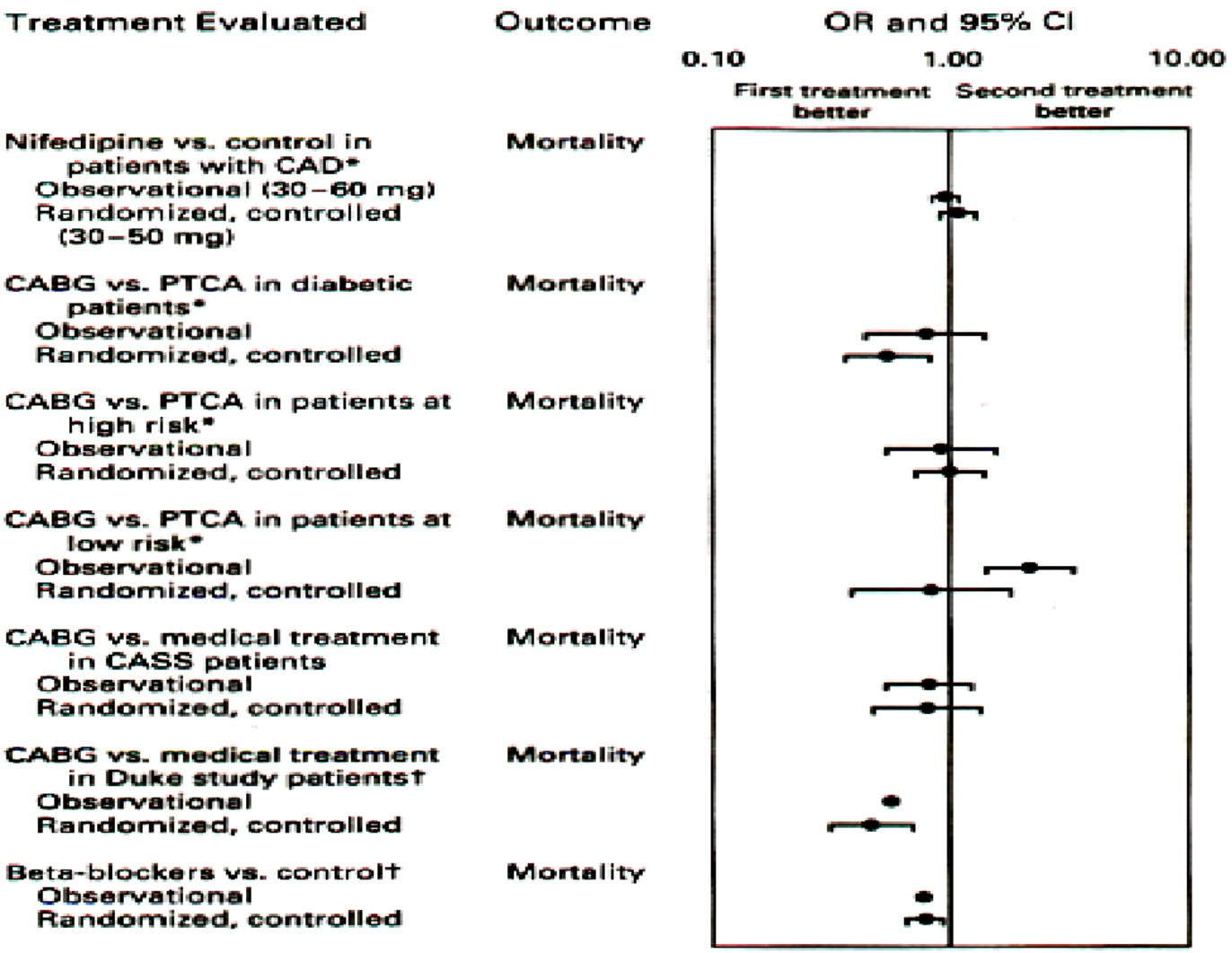
- Two undertook their own primary searches of the literature.

Statistical methods

- Mixed statistical methods
  - Vote counting of statistically significant positive findings
  - Consideration of overlapping confidence intervals of estimates
  - Testing the statistical significance of differences in individual topics

- None provided estimate of average bias
- Low power in within topic analyses

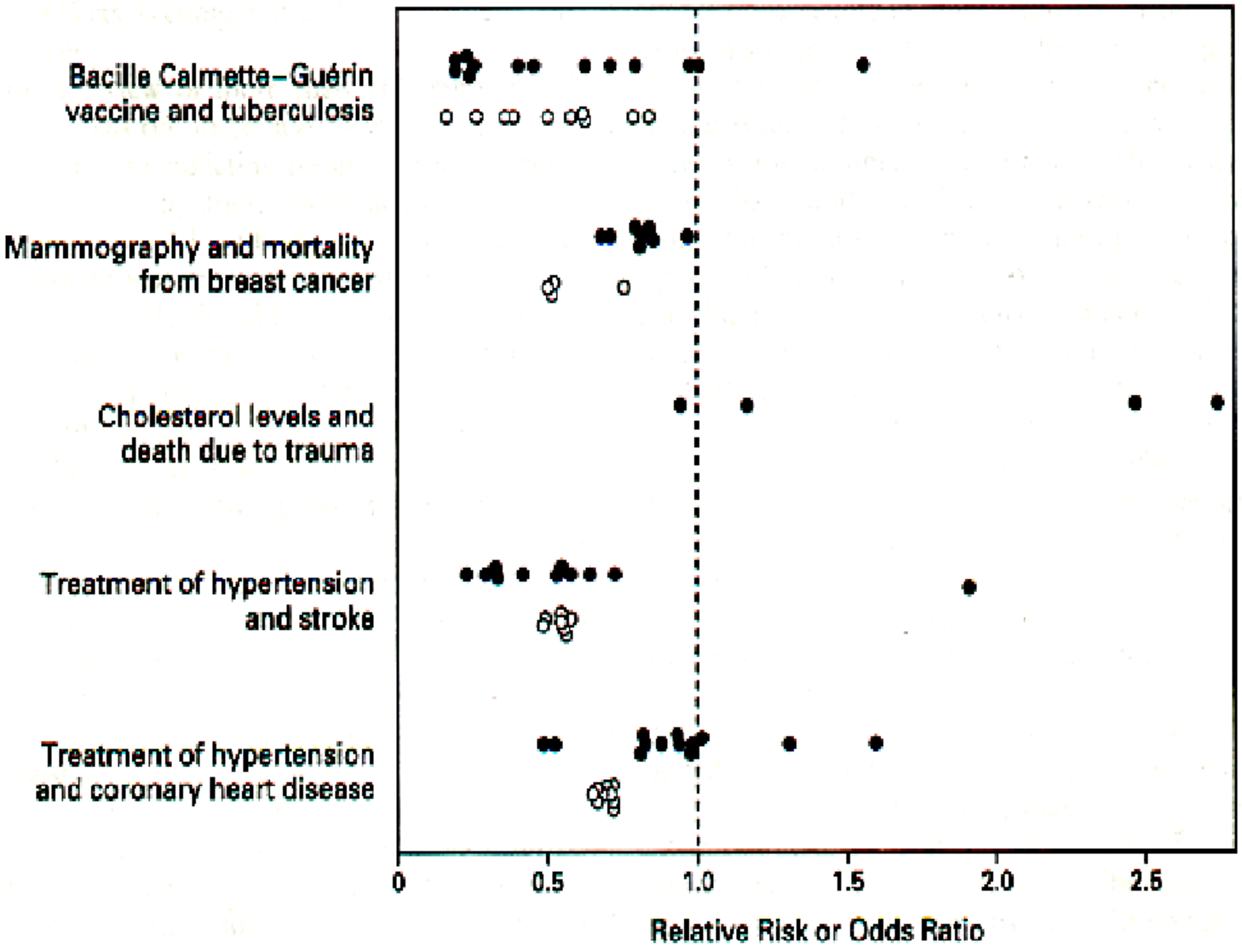
Benson and Hartz NEJM 2000; 342: 1878-86.



Statistical methods for heterogeneity

- Two studies assessed whether there are differences in the variability of NRS and RCT results
  - One compared the spread of results
  - One made a comparison of statistically significant heterogeneity

Concato, Shah, Horwitz NEJM 2000; 342: 1887-92.



In summary ...

- These comparisons have reported mixed results and have:
  1. been underpowered to detect differences
  2. rarely estimated differences
  3. poor criteria for declaring results to be “the same” or to be “different”
  4. are potentially confounded by known and unknown between-study differences in the populations and interventions
  5. are potentially confounded by other differences in study methodology beyond the allocation methods
  6. not considered sample size when assessing variability
  7. are suspect to publication and selection biases

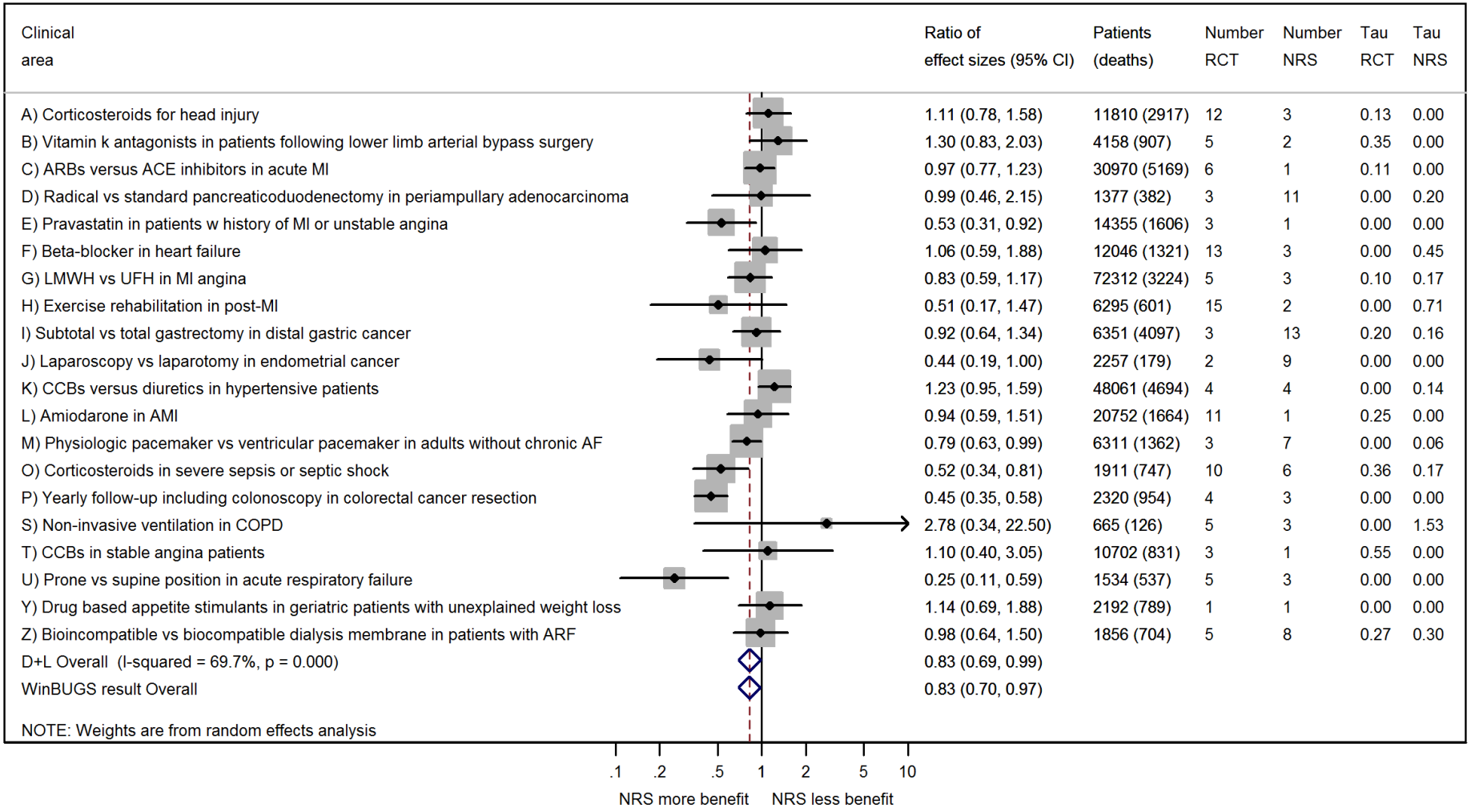
BRANDO NRS study

- Randomly selecting topics to avoid selection bias
- Restricting outcomes to “all cause mortality” reducing heterogeneity and minimising outcome assessment bias
- Standardised assessments of NRS and RCT for each topic reducing heterogeneity
- Quantification of effect sizes and proper quantitative synthesis across topics
- Investigating impact of other methodological issues in sensitivity analysis

PICOs included

TOPIC ID	TOPIC	Source of RCTs
A	Corticosteroids for head injury	Cochrane
B	Vitamin k antagonists in patients following lower limb arterial bypass surgery	ISI
C	Angiotensin II receptor antagonists versus ACE inhibitors in acute MI patients	Full search
D	Radical pancreaticoduodenectomy vs standard pacnreaticoduodenectomy in patients w periampullary adenocarcinoma	ISI
E	Pravastatin vs placebo in patients w history of MI or unstable angina and increased plaque cholesterol	ISI
F	Beta-blocker in heart failure patients	ISI
G	LMWH vs UFH in MI/ angina patients	Cochrane
H	Exercise rehabilitation in post-MI patients	Cochrane
I	Subtotal vs total gastrectomy (curative) in distal gastric cancer patients	Full search
J	Laparoscopy vs laparotomy in endometrial cancer patients	Full search
K	Calcium channel blockers versus diuretics in high risk hypertensives	ISI
L	Amiodarone in AMI patients	Full search
M	Physiologic pacemaker vs ventricular pacemaker in adults w symptomatic bradycardia without chronic atrial fibrillation	Cochrane
O	Corticosteroids in adult patients with severe sepsis or septic shock	Cochrane
P	Yearly follow-up including colonoscopy in patients undergoing curative resection for colorectal cancer	Cochrane
S	Non-invasive ventilation in chronic obstructive pulmonary disease	Cochrane
T	Calcium channel blockers vs no CCB in stable angina patients	Full search
U	Prone vs supine position in adult acute respiratory failure	Full search
Y	Drug based appetite stimulants in geriatric patients with weight loss with no apparent cause	Full search
Z	Bioincompatible (Cuprophan) dialysis membrane versus biocompatible (polymethyl methacrylate) dialysis membrane in	Cochrane
TOTAL		

All Data


Clinical area
Ratio of effect sizes (95% CI)
Patients (deaths)
Number RCT
Number NRS
Tau RCT
Tau NRS
A) Corticosteroids for head injury 1.11 (0.78, 1.58) 11810 (2917) 12 3 0.13 0.00
B) Vitamin k antagonists in patients following lower limb arterial bypass surgery 1.30 (0.83, 2.03) 4158 (907) 5 2 0.35 0.00
C) ARBs versus ACE inhibitors in acute MI 0.97 (0.77, 1.23) 30970 (5169) 6 1 0.11 0.00
D) Radical vs standard pancreaticoduodenectomy in periampullary adenocarcinoma 0.99 (0.46, 2.15) 1377 (382) 3 11 0.00 0.20
E) Pravastatin in patients w history of MI or unstable angina 0.53 (0.31, 0.92) 14355 (1606) 3 1 0.00 0.00
F) Beta-blocker in heart failure 1.06 (0.59, 1.88) 12046 (1321) 13 3 0.00 0.45
G) LMWH vs UFH in MI angina 0.83 (0.59, 1.17) 72312 (3224) 5 3 0.10 0.17
H) Exercise rehabilitation in post-MI 0.51 (0.17, 1.47) 6295 (601) 15 2 0.00 0.71
I) Subtotal vs total gastrectomy in distal gastric cancer 0.92 (0.64, 1.34) 6351 (4097) 3 13 0.20 0.16
J) Laparoscopy vs laparotomy in endometrial cancer 0.44 (0.19, 1.00) 2257 (179) 2 9 0.00 0.00
K) CCBs versus diuretics in hypertensive patients 1.23 (0.95, 1.59) 48061 (4694) 4 4 0.00 0.14
L) Amiodarone in AMI 0.94 (0.59, 1.51) 20752 (1664) 11 1 0.25 0.00
M) Physiologic pacemaker vs ventricular pacemaker in adults without chronic AF 0.79 (0.63, 0.99) 6311 (1362) 3 7 0.00 0.06
O) Corticosteroids in severe sepsis or septic shock 0.52 (0.34, 0.81) 1911 (747) 10 6 0.36 0.17
P) Yearly follow-up including colonoscopy in colorectal cancer resection 0.45 (0.35, 0.58) 2320 (954) 4 3 0.00 0.00
S) Non-invasive ventilation in COPD 2.78 (0.34, 22.50) 665 (126) 5 3 0.00 1.53
T) CCBs in stable angina patients 1.10 (0.40, 3.05) 10702 (831) 3 1 0.55 0.00
U) Prone vs supine position in acute respiratory failure 0.25 (0.11, 0.59) 1534 (537) 5 3 0.00 0.00
Y) Drug based appetite stimulants in geriatric patients with unexplained weight loss 1.14 (0.69, 1.88) 2192 (789) 1 1 0.00 0.00
Z) Bioincompatible vs biocompatible dialysis membrane in patients with ARF 0.98 (0.64, 1.50) 1856 (704) 5 8 0.27 0.30
D+L Overall (I-squared = 69.7%, p = 0.000) 0.83 (0.69, 0.99)
WinBUGS result Overall 0.83 (0.70, 0.97)
NOTE: Weights are from random effects analysis
.1
.2
.5
1
2
5
10
NRS more benefit
NRS less benefit

Our own limitations…

- Sampling biased to inclusion of topics with many trials and where trials have multiple reports
- Many topics excluded as no NRS identified
- Potential publication bias, searches only of Medline

- Poor quality methods and reporting of NRS (poor quality methods and reporting of RCTs shown not to impact)
- Inadequate power to investigate effect of NRS study features.

Summary

- Systematic reviews of empirical studies are important to assess the evidence base
- Substantial data overlap may preclude meta-analysis
- Critical analysis of methods of the empirical studies is essential.
  - Assessment may need to be derived from first principles
  - Empirical assessments are typically retrospective and observational so may be prone to many biases!