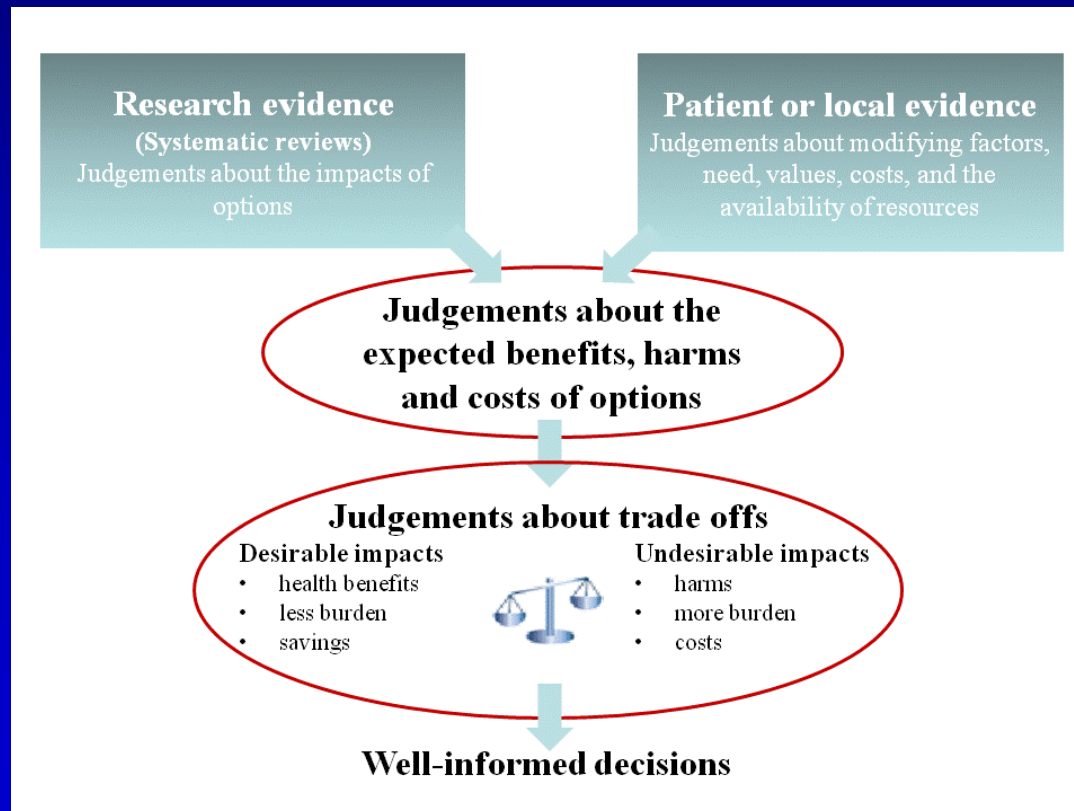


Plan

- evidence and decision-making
- GRADE background
- GRADE approach to evaluating quality of evidence
- GRADE approach to summarizing evidence

What is the role of evidence in policy?

- The role of evidence is to inform policy
- Evidence is essential, but not sufficient
- Judgements are needed, including judgements about confidence (the quality of the evidence), what to expect in a specific setting, and trade-offs



Dilemma: proliferation of systems

Solution: common international grading system?

- GRADE (*Grades of recommendation, assessment, development and evaluation*)
- international group
 - Australian NMRC, SIGN, USPSTF, WHO, NICE, Oxford CEBM, CDC, CC
- ~ 25 meetings over last ten years
 - (~10 - 50 attendants)

GRADE Uptake

Agencia sanitaria regionale, Bologna, Italia
Agency for Health Care Research and Quality (AHRQ)
Allergic Rhinitis and Group - Independent Expert Panel
American Association for the study of liver diseases
American College of Cardiology Foundation
American College of Chest Physicians
American College of Emergency Physicians
American College of Physicians
American Endocrine Society
American Gastroenterology Association
American Society for Colposcopy and Cervical Pathology
American Society of Gastrointestinal Endoscopy
American Society of Interventional Pain Physicians
American Thoracic Society (ATS)
Austrian Ludwig Boltzmann Institute for HTA
BMJ Clinical Evidence
British Medical Journal
Canadian Agency for Drugs and Technology in Health
Canadian Cardiovascular Society
Canadian Society of Nephrology
Canadian Task Force on Preventive Health Care
Centers for Disease Control
Cochrane Collaboration
Critical Ultrasound Journal
Dutch Institute for Healthcare Improvement CBO
EBM Guidelines Finland
Emergency Medical Services for Children National Resource Center
European Association for the Study of the Liver
European Monitoring Centre for Drugs and Drug Addicaton
European Respiratory Society
European Society of Thoracic Surgeons

Evidence-based Nursing Sudtirool, Alta Adiga, Italy
Finnish Office of Health Technology Assessment
German Agency for Quality in Medicine
Health Inspectorate for Scotland
Infectious Disease Society of America
Institute for Clinical Systems Improvement
Japanese Society of Oral and Maxillofacial Radiology
Joslin Diabetes Center
Journal of Infection in Developing Countries
Kaiser Permanente
Kidney Disease International Guidelines Organization
National and Gulf Centre for Evidence-based Medicine
National Institute for Clinical Excellence (NICE)
National Kidney Foundation
Norwegian Knowledge Centre for the Health Services
Ontario MOH Medical Advisory Secretariat
Panama and Costa Rica National Clinical Guidelines Program
Polish Institute for EBM
Scottish Intercollegiate Guideline Network (SIGN)
Society of Critical Care Medicine
Society of Pediatric Endocrinology
Society of Vascular Surgery
Spanish Society of Family Practice (SEMFYC)
Stop TB Diagnostic Working Group
Surviving sepsis campaign
Swedish Council on Technology Assessment in Health Care
Swedish National Board of Health and Welfare
University of Pennsylvania Health System for EB Practice
UpToDate
WINFOCUS
World Allergy Organization
World Health Organization (WHO)

What are we grading?

- two components
- quality of body of evidence
 - extent to which confidence in estimate of effect adequate to support decision
 - high, moderate, low, very low
- strength of recommendation
 - strong and weak

Determinants of quality

- RCTs start high
- observational studies start low
- what can lower quality?
 - detailed design and execution
 - inconsistency
 - indirectness
 - imprecision
 - reporting bias

Risk of Bias

- well established
 - concealment
 - intention to treat principle observed
 - blinding
 - completeness of follow-up
- more recent
 - selective outcome reporting bias

Consistency of results

- if inconsistency, look for explanation
 - patients, intervention, outcome, methods
- judgment of consistency
 - variation in size of effect
 - overlap in confidence intervals
 - statistical significance of heterogeneity
 - I^2

Relative Risk with 95% CI for Vitamin D Non-vertebral Fractures

Chapuy et al, (1994) 0.79 (0.69, 0.92)

Lips et al, (1996) 1.10 (0.87, 1.39)

Dawson-Hughes et al, (1997) 0.46 (0.24, 0.88)

Pfeifer et al, (2000) 0.48 (0.13, 1.78)

Meyer et al, (2002) 0.92 (0.68, 1.24)

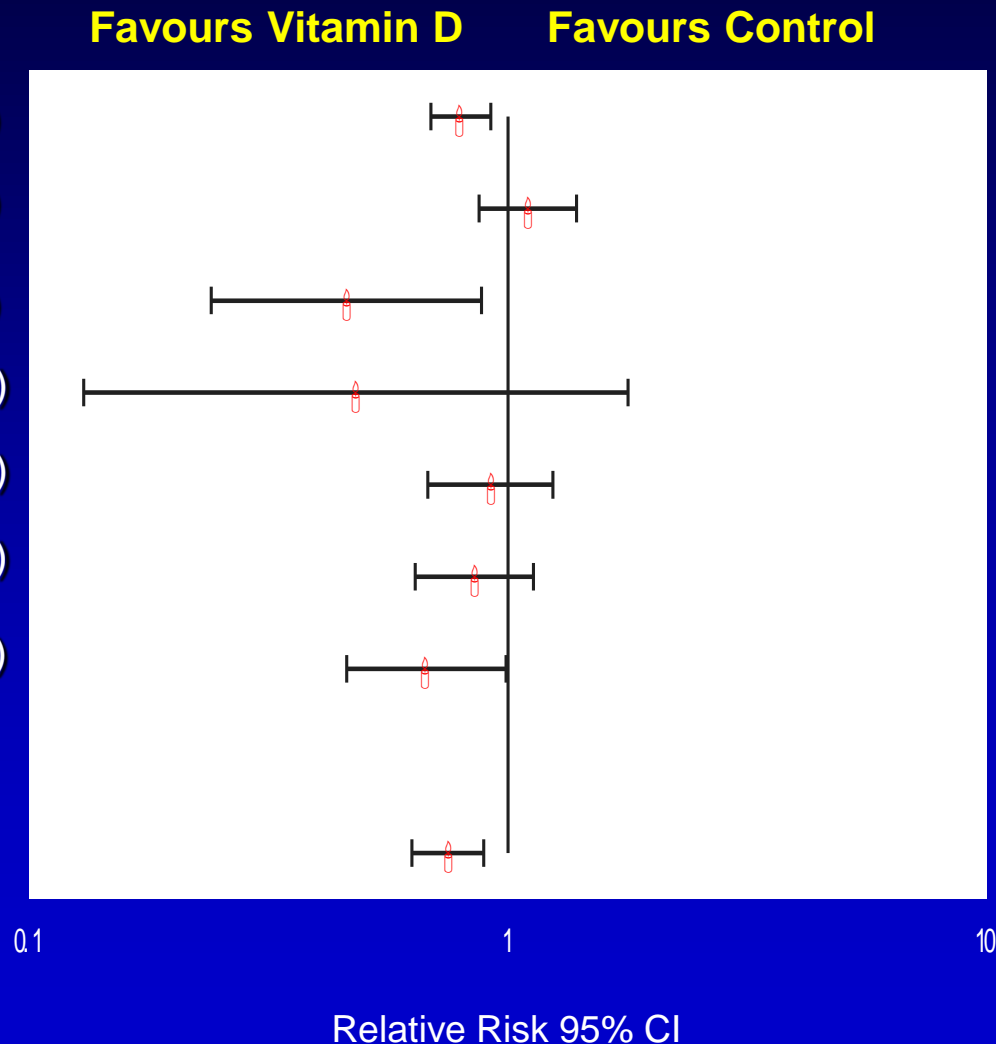
Chapuy et al, (2002) 0.85 (0.64, 1.13)

Trivedi et al, (2003) 0.67 (0.46, 0.99)

Pooled Random Effect Model

0.82 (0.69 to 0.98)

p= 0.05 for heterogeneity, I²=53%

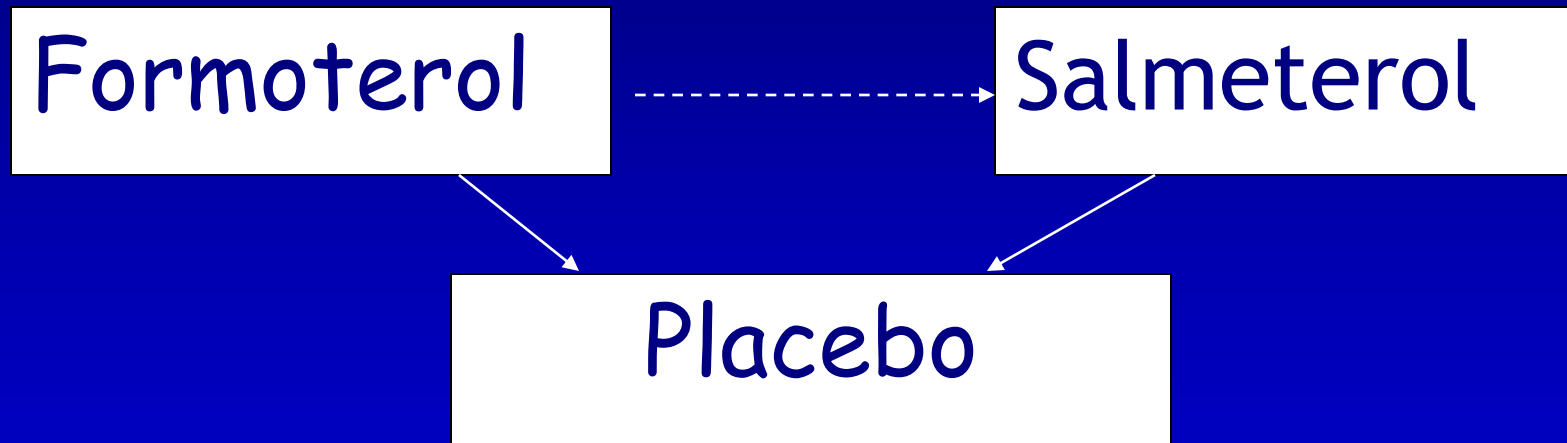


Quality judgments: Directness

- populations
 - older, sicker or more co-morbidity
- interventions
 - warfarin in trials vs clinical practice
- outcomes
 - important versus surrogate outcomes
 - glucose control versus CV events

Directness

interested in A versus B
available data A vs C, B vs C



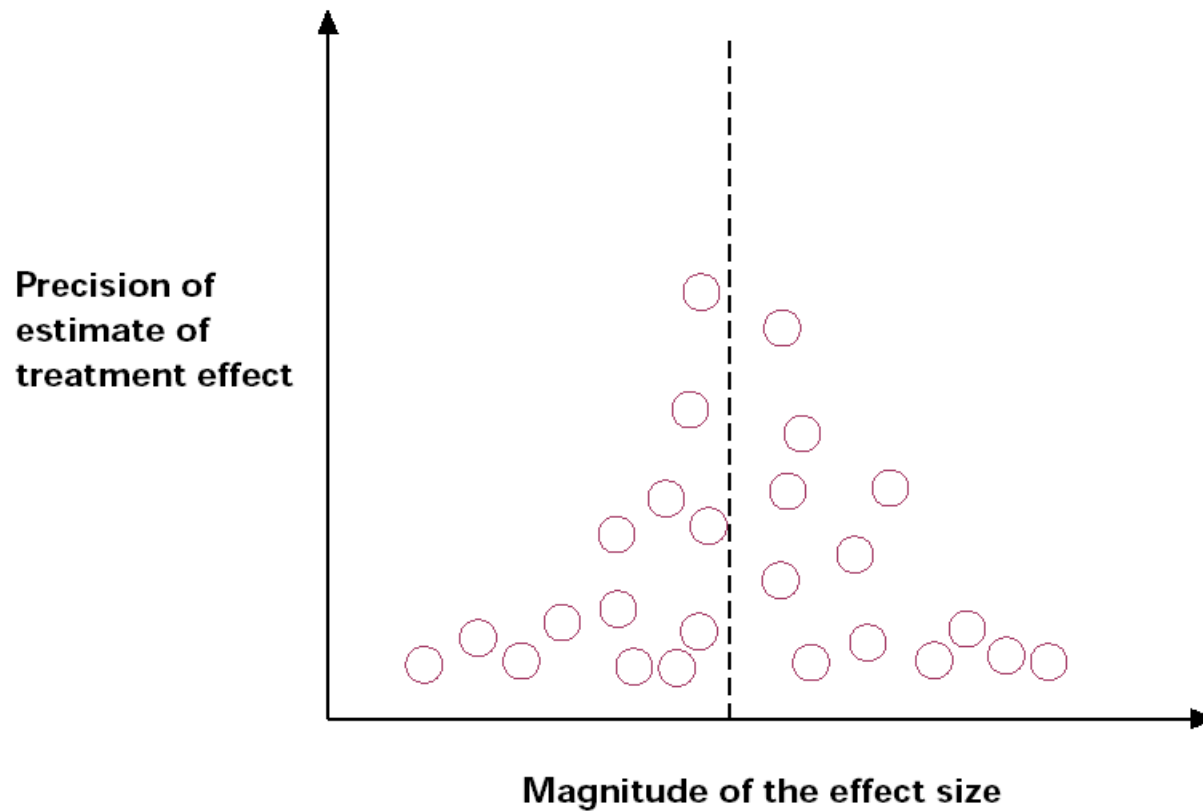
Imprecision

- small sample size
 - small number of events
- wide confidence intervals
 - uncertainty about magnitude of effect

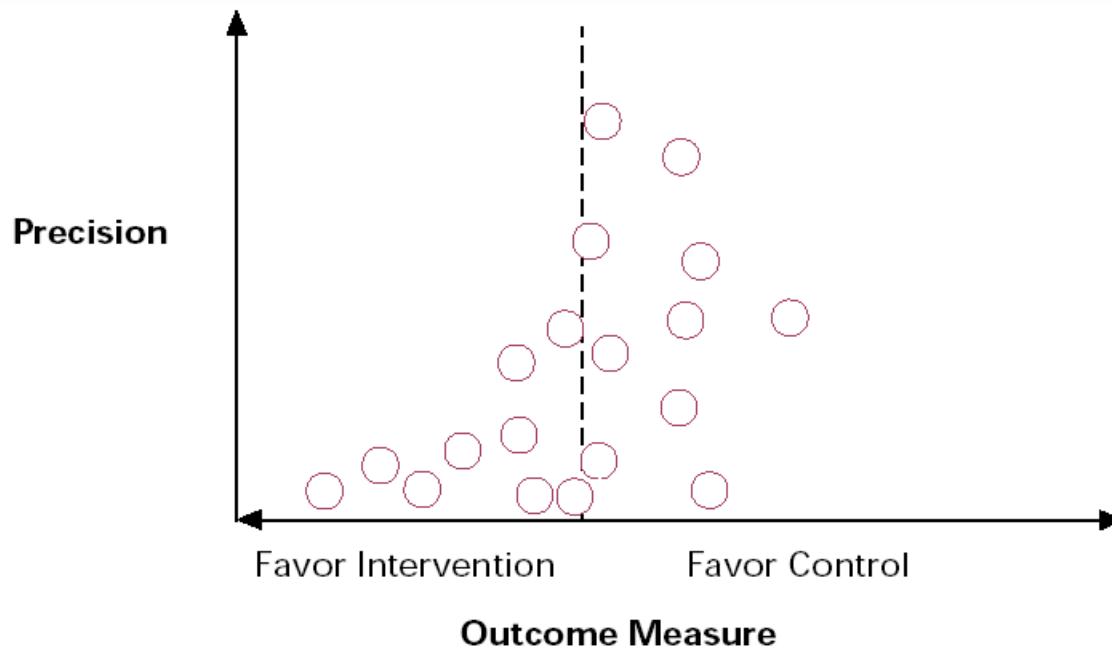
Publication bias

- high likelihood could lower quality
- when to suspect
 - number of small studies
 - industry sponsored

Funnel Plot



Publication Bias



Funnel Plot

Fish oil on mortality

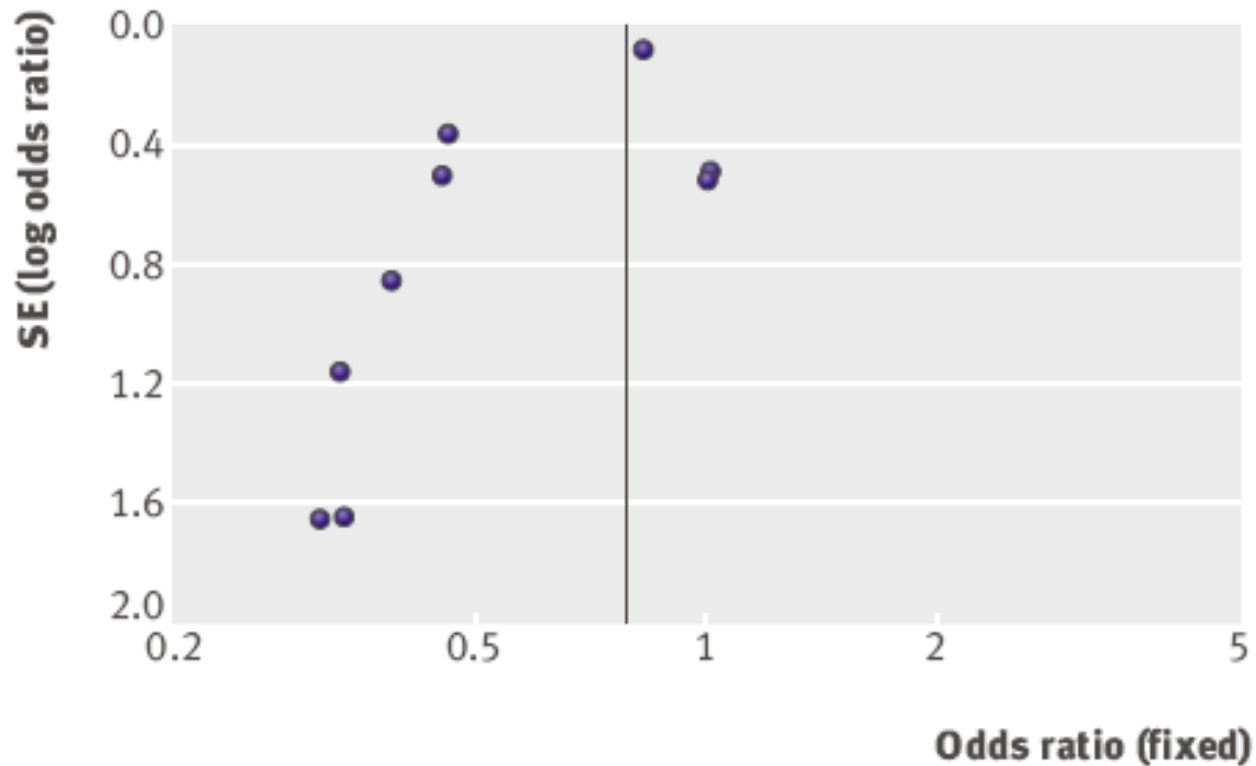


Fig 4 | Funnel plot for assessment of publication bias for death from cardiac causes in 11 included studies reporting data on this outcome

What can raise quality?

- large magnitude can rate up one level
 - very large two levels
- common criteria
 - everyone used to do badly
 - almost everyone does well
 - quick action
- hip replacement for hip osteoarthritis
- mechanical ventilation in respiratory failure

Quality assessment criteria

Study Design	Quality of Evidence	Lower if	Higher if
Randomised trial →	High	Risk of bias -1 Serious -2 Very serious	Large effect +1 Large +2 Very large
	Moderate	Inconsistency -1 Serious -2 Very serious	Dose response +1 Evidence of a gradient
Observational study →	Low	Indirectness -1 Serious -2 Very serious	All plausible confounding +1 Would reduce a demonstrated effect or
	Very low	Imprecision -1 Serious -2 Very serious	+1 Would suggest a spurious effect when results show no effect
		Publication bias -1 Likely -2 Very likely	

Beta blockers in non-cardiac surgery

Quality Assessment							Summary of Findings		
							Quality	Relative Effect (95% CI)	Absolute risk difference
Outcome	Number of participants (studies)	Risk of Bias	Consistency	Directness	Precision	Publication Bias			
Myocardial infarction	10,125 (9)	No serious limitations	No serious imitations	No serious limitations	No serious limitations	Not detected	High	0.71 (0.57 to 0.86)	1.5% fewer (0.7% fewer to 2.1% fewer)
Mortality	10,205 (7)	No serious limitations	Possibly inconsistent	No serious limitations	Imprecise	Not detected	Moderate or low	1.23 (0.98 – 1.55)	0.5% more (0.1% fewer to 1.3% more)
Stroke	10,889 (5)	No serious limitaions	No serious limitations	No serious limitations	Possible imprecision	Not detected	High	2.21 (1.37 – 3.55)	0.5% more (0.2% more to 1.3% more0

Conclusion

- in deciding on essential medicines, policy-makers need summaries of evidence including quality
- GRADE
 - simple, transparent, systematic
 - increasing wide adoption