Holger Schünemann, MD, PhD

Chair, Department of Clinical Epidemiology & Biostatistics
Michael Gent Chair in Healthcare Research
McMaster University, Hamilton, Canada

Montreal, February 12, 2010
McGill University, Montreal Chest Institute

GRADE

schuneh@mcmaster.ca

Content

- Background about GRADE

- GRADE approach
  - Quality of evidence
  - Strength of recommendations

Evidence based clinical decisions

Clinical state and circumstances

Patient values and preferences

Expertise

Research evidence

Haynes et al. 2002

Confidence in evidence

- There always is evidence
  - "When there is a question there is evidence"
- Better research ⟹ greater confidence in the evidence and decisions
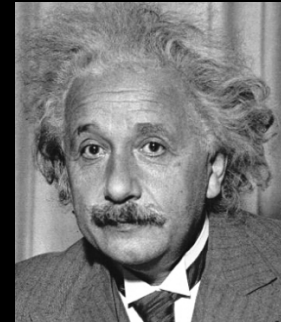
Hierarchy of evidence
based on quality

STUDY DESIGN

- Randomized Controlled Trials

- Cohort Studies and Case Control Studies

- Case Reports and Case Series, Non-systematic observations

- Expert Opinion

BIAS
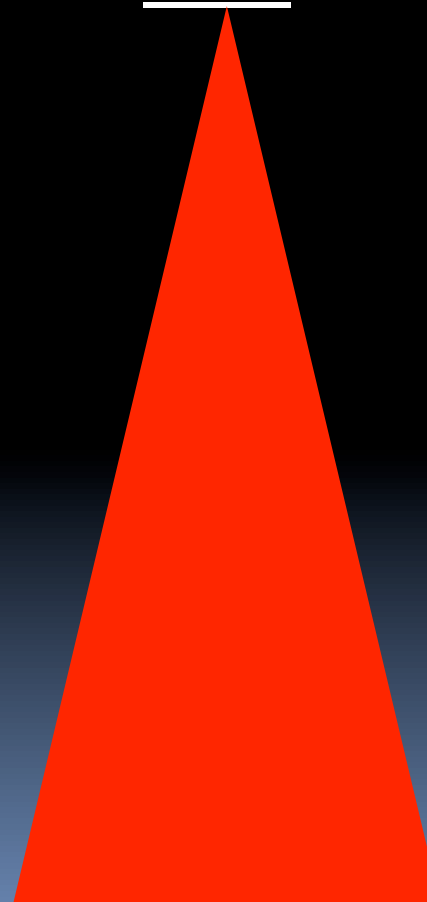
Can you explain the following?

- Concealment of randomization
- Blinding (who is blinded in a double blinded trial?)
- Confounding, effect modification & ext. validity
- Intention to treat analysis and its correct application
- P-values and confidence intervals

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

McMaster
University

Inspiring Innovation and Discovery

Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomised controlled trials

Simple hierarchies are (too) simplistic

STUDY DESIGN

- Randomized Controlled Trials

- Cohort Studies and Case Control Studies

- Case Reports and Case Series, Non-systematic observations

  Expert Opinion

BIAS

Expert Opinion

Schünemann & Bone, 2003
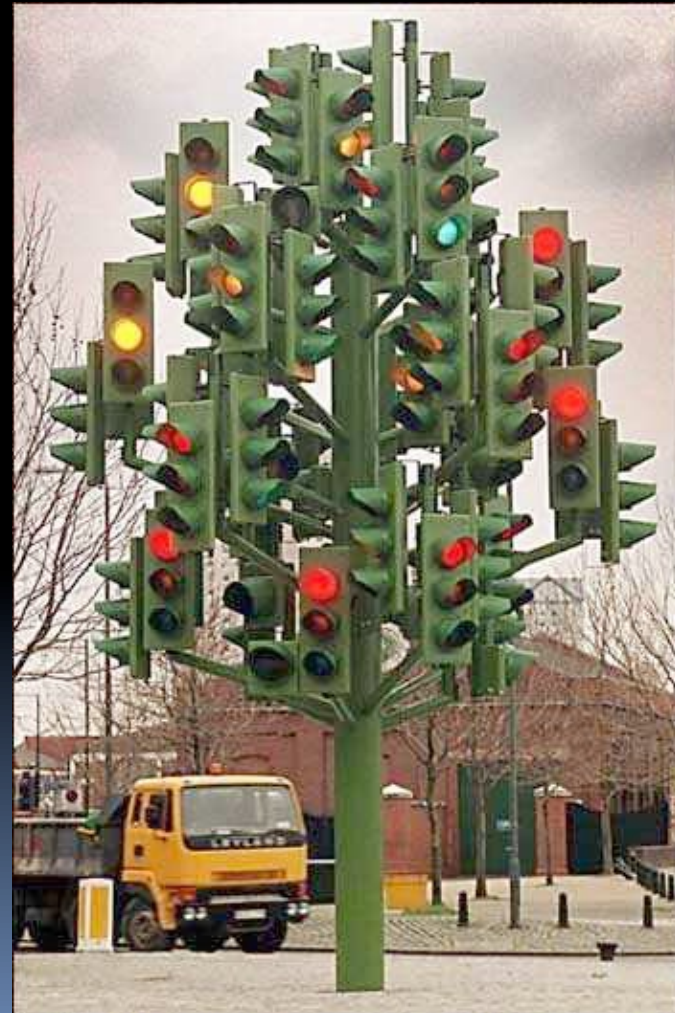
Which hierarchy?

Recommendation for use of oral anticoagulation in patients with atrial fibrillation and rheumatic mitral valve disease

Evidence	Recommendation	Organization
▪ B	Class I	➤ AHA
▪ A	1	➤ ACCP
▪ IV	C	➤ SIGN

What to do?

Grades of **R**ecommendation **A**ssessment, **D**evelopment

and **E**valuation

GRADE
WORKING GROUP

- Since 2000
- Guideline developers, methodologists & clinicians from around the world

McMaster University
Inspiring Innovation and Discovery

RATING QUALITY OF EVIDENCE AND STRENGTH OF RECOMMENDATIONS

GRADE

common, transparent and sensible

GRADE: an emerging consensus on rating quality of evidence and strength of recommendations

Guidelines are inconsistent in how they rate the quality of evidence and the strength of recommendations. This article explores the advantages of the GRADE system, which is increasingly being adopted by organisations worldwide

BMJ | 26 APRIL 2008 | VOLUME 336

CMAJ 2003, BMJ 2004, BMC 2004, BMC 2005, AJRCCM 2006, Chest 2006, BMJ 2008

GRADE Uptake

- World Health Organization
- Allergic Rhinitis in Asthma Guidelines (ARIA)
- American Thoracic Society
- American College of Physicians
- European Respiratory Society
- European Society of Thoracic Surgeons
- British Medical Journal
- Infectious Disease Society of America
- American College of Chest Physicians
- UpToDate
- National Institutes of Health and Clinical Excellence (NICE)
- Scottish Intercollegiate Guideline Network (SIGN)
- Cochrane Collaboration
- Infectious Disease Society of America
- Clinical Evidence
- Agency for Health Care Research and Quality (AHRQ)
- Partner of **GIN**
- Over 40 major organizations
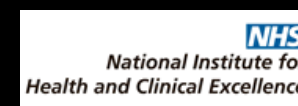
Guideline development Process

McMaster University
Inspiring Innovation and Discovery

Prioritise Problems, establish panel, questions
↓

Systematic Review
↓

Evidence Profile
↓

Relative importance of outcomes
↓

Overall quality of evidence
↓

GRADE

Benefit – downside evaluation
↓

Strength of recommendation
↓

Implementation and evaluation of guidelines

The GRADE approach

Clear separation of 2 issues:

1) 4 categories of quality of evidence: ⊕⊕⊕⊕ (High), ⊕⊕⊕◯(Moderate), ⊕⊕◯◯(Low), ⊕◯◯◯(Very low)?

- methodological quality of evidence
- likelihood of bias
- by outcome and across outcomes

2) Recommendation: 2 grades – weak/conditional or strong (for or against)?

- Quality of evidence only one factor
- Balance of benefits and downsides, values and preferences, resource use

GRADE Quality of Evidence

In the context of a systematic review

- The quality of evidence reflects the extent to which we are confident that an estimate of effect is correct.

In the context of making recommendations

- The quality of evidence reflects the extent to which our confidence in an estimate of the effect is adequate to support a particular recommendation.

Figure 1. Belief and confidence: a two-dimensional weather report. (Reprinted by permission from the Wall Street Journal).

Likelihood of and confidence in an outcome

Determinants of quality

- RCTs start ⊕⊕⊕⊕ (high)

- observational studies start at ⊕⊕○○ (low)

- 5 factors that can lower quality
  1. limitations in detailed design and execution *(risk of bias criteria)*
  2. Inconsistency *(or heterogeneity)*
  3. Indirectness *(PICO and applicability)*
  4. Imprecision *(number of events and confidence intervals)*
  5. Publication bias
  3 factors can increase quality
  1. large magnitude of effect
  2. all plausible residual confounding may be working to reduce the demonstrated effect or increase the effect if no effect was observed
  3. dose-response gradient

1. Design and Execution/Risk of Bias

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

Design and Execution/RoB

Regular treatment with salmeterol for chronic asthma: serious adverse events (Review)

Cates CJ, Cates MJ

Figure 4. Risk of bias summary: review authors' judgments about each risk of bias item for each included study.
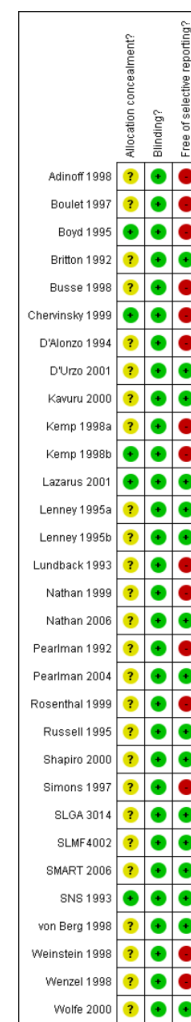
Design and Execution

Regular treatment with salmeterol for chronic asthma: serious adverse events (Review)

Cates CJ, Cates MJ

Figure 4. Risk of bias summary: review authors' judgments about each risk of bias item for each included study.

	Allocation concealment?	Blinding?	Free of selective reporting?
Adinoff 1998	?	+	−
Boulet 1997	?	+	−
Boyd 1995	+	+	−
Britton 1992	?	+	+

Design and Execution

Figure 3. Risk of bias graph: review authors' judgments about each risk of bias item presented as percentages across all included studies.
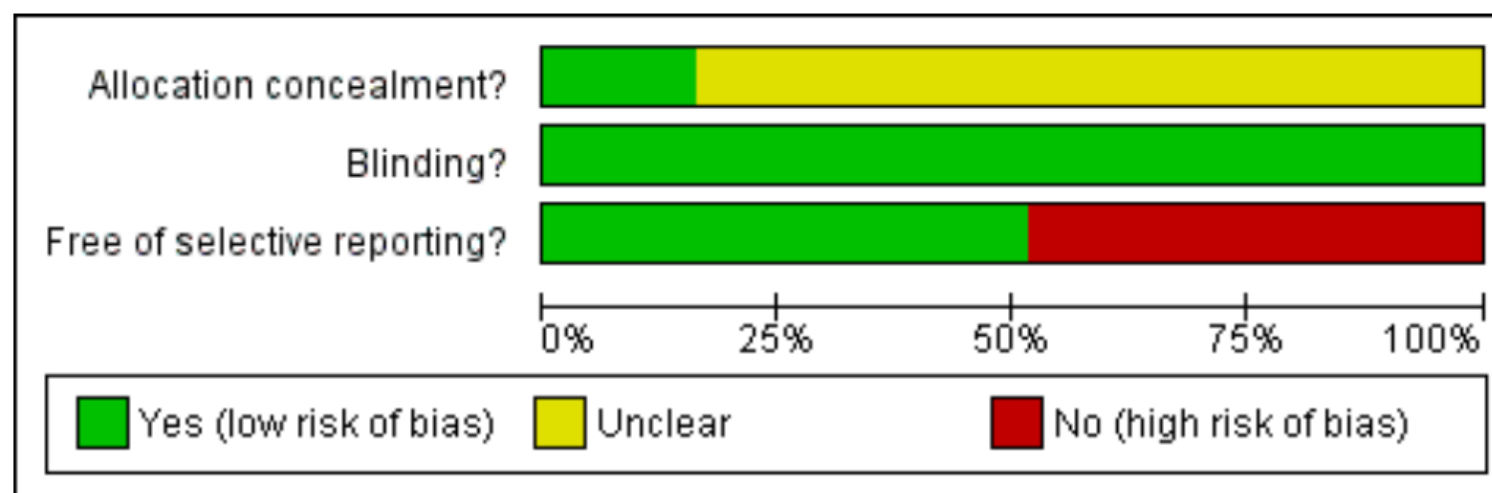
Overall judgment required

Who believes the risk of bias is of concern?

Yes

No

Don't know or undecided

Detailed study design and execution

Mortality, cancer and anticoagulation

Akl E, Barba M, Rohilla S, Terrenato I, Sperati F, Schünemann HJ. "*Anticoagulation for the long term treatment of venous thromboembolism in patients with cancer*". Cochrane Database Syst Rev. 2008 Apr 16;(2):CD006650.

Five trials

Analysis 01.01. Comparison 01 Heparin vs placebo, Outcome 01 Mortality over duration of study

Review: Parenteral anticoagulation for prolonging survival in patients with cancer who have no other indication for anticoagulation

Comparison: 01 Heparin vs placebo

Outcome: 01 Mortality over duration of study

Study	Heparin N	Control N	log [Hazard Ratio] (SE)	Hazard Ratio (Random) 95% CI	Weight (%)	Hazard Ratio (Random) 95% CI
01 SCLC						
Altinbas 2004	42	42	-0.65 (0.23)		10.8	0.52 [0.33, 0.82]
Lebeau 1994	138	139	-0.33 (0.12)		23.7	0.72 [0.56, 0.91]
Subtotal (95% CI)					34.5	0.65 [0.49, 0.87]
Test for heterogeneity chi-square=1.48 df=1 p=0.22 I²=32.4%						
Test for overall effect z=2.93 p=0.003						
02 Advanced cancer						
Kakkar 2004	190	184	-0.24 (0.11)		25.9	0.79 [0.63, 0.98]
Klerk 2005	148	154	-0.28 (0.11)		25.5	0.75 [0.60, 0.94]
Sideras 2006	68	69	0.14 (0.19)		14.1	1.15 [0.79, 1.68]
Subtotal (95% CI)					65.5	0.84 [0.68, 1.03]
Test for heterogeneity chi-square=3.81 df=2 p=0.15 I²=47.5%						
Test for overall effect z=1.68 p=0.09						
Total (95% CI)					100.0	0.77 [0.65, 0.91]
Test for heterogeneity chi-square=7.63 df=4 p=0.11 I²=47.5%						
Test for overall effect z=3.01 p=0.003						

0.2   0.5   1   2   5

Favours heparin   Favours control

Who believes the risk of bias is of concern?

Yes

No

Don't know or undecided

2. Inconsistency of results (Heterogeneity)

- if inconsistency, look for explanation
  - patients, intervention, outcome
- unexplained inconsistency downgrade quality

Heparin or vitamin K antagonists for survival in patients with cancer

Akl E, Barba M, Rohilla S, Terrenato I, Sperati F, Schünemann HJ. "*Anticoagulation for the long term treatment of venous thromboembolism in patients with cancer*". Cochrane Database Syst Rev. 2008 Apr 16;(2):CD006650.
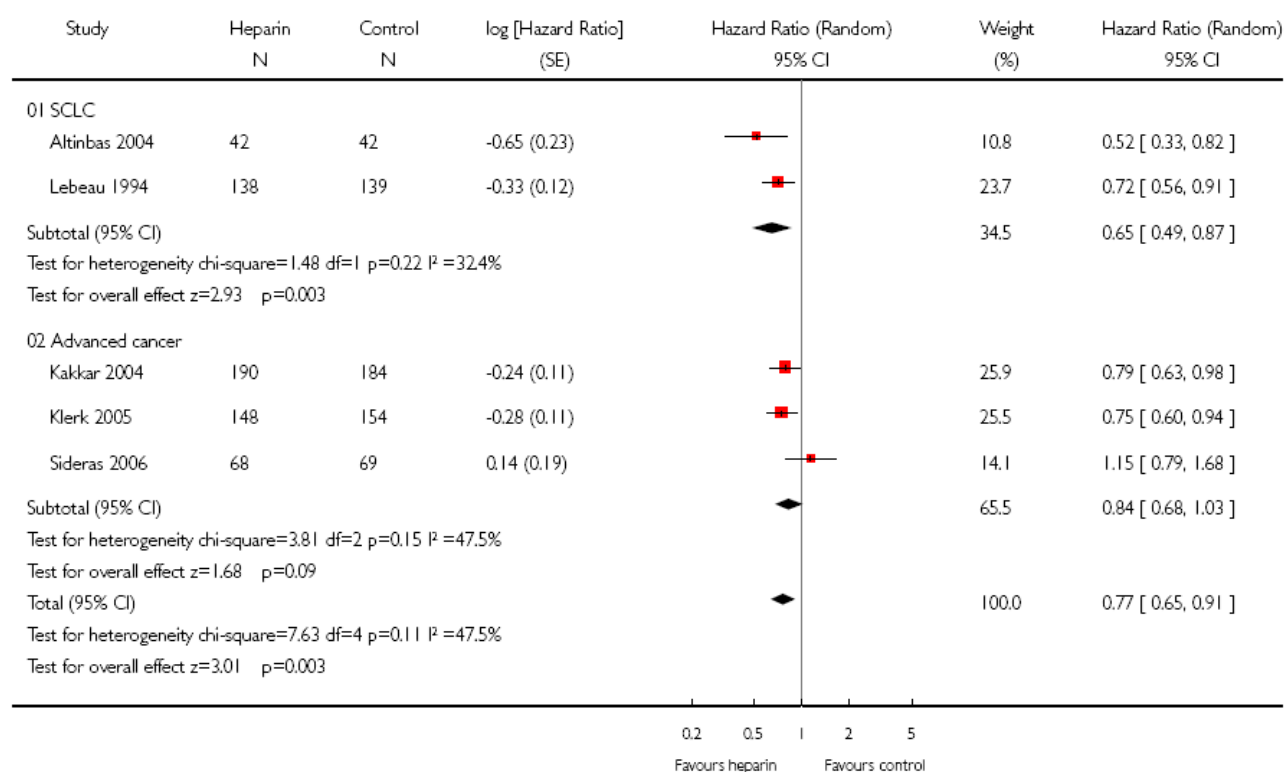
Non-steroidal drug use and risk of pancreatic cancer

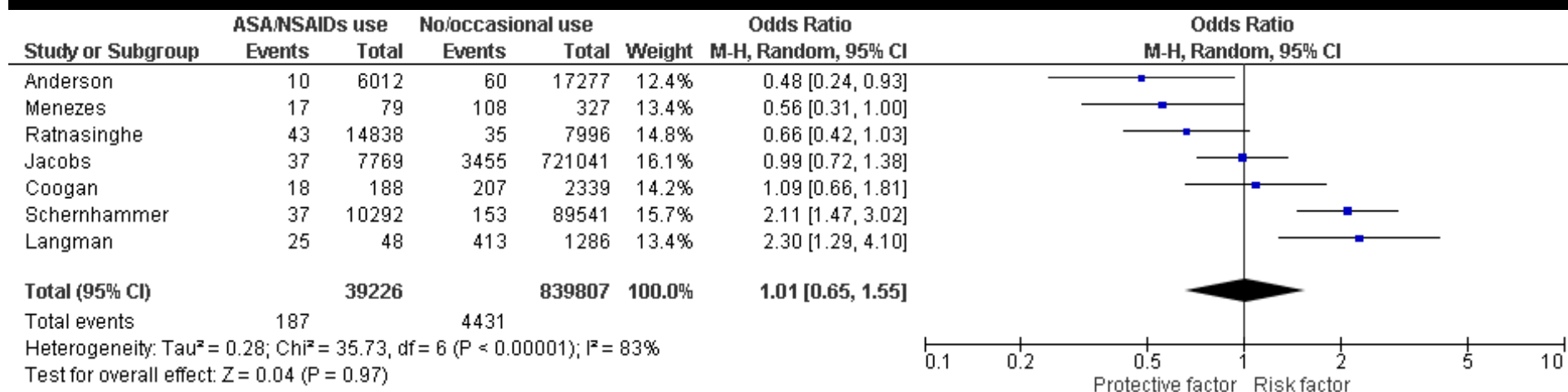| Study or Subgroup	ASA/NSAIDs use Events	Total	No/occasional use Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
Anderson	10	6012	60	17277	12.4%	0.48 [0.24, 0.93]
Menezes	17	79	108	327	13.4%	0.56 [0.31, 1.00]
Ratnasinghe	43	14838	35	7996	14.8%	0.66 [0.42, 1.03]
Jacobs	37	7769	3455	721041	16.1%	0.99 [0.72, 1.38]
Coogan	18	188	207	2339	14.2%	1.09 [0.66, 1.81]
Schernhammer	37	10292	153	89541	15.7%	2.11 [1.47, 3.02]
Langman	25	48	413	1286	13.4%	2.30 [1.29, 4.10]
Total (95% CI)		39226		839807	100.0%	1.01 [0.65, 1.55]
Total events	187		4431			

Heterogeneity: Tau² = 0.28; Chi² = 35.73, df = 6 (P < 0.00001); I² = 83%
Test for overall effect: Z = 0.04 (P = 0.97)

Capurso G, Schünemann HJ, Terrenato I, Moretti A, Koch M, Muti P, Capurso L, Delle Fave G.
Meta-analysis: the use of non-steroidal anti-inflammatory drugs and pancreatic cancer risk for different exposure categories.
Aliment Pharmacol Ther. 2007 Oct 15;26(8):1089-99.

Who believes there is important inconsistency (rather than random error)?
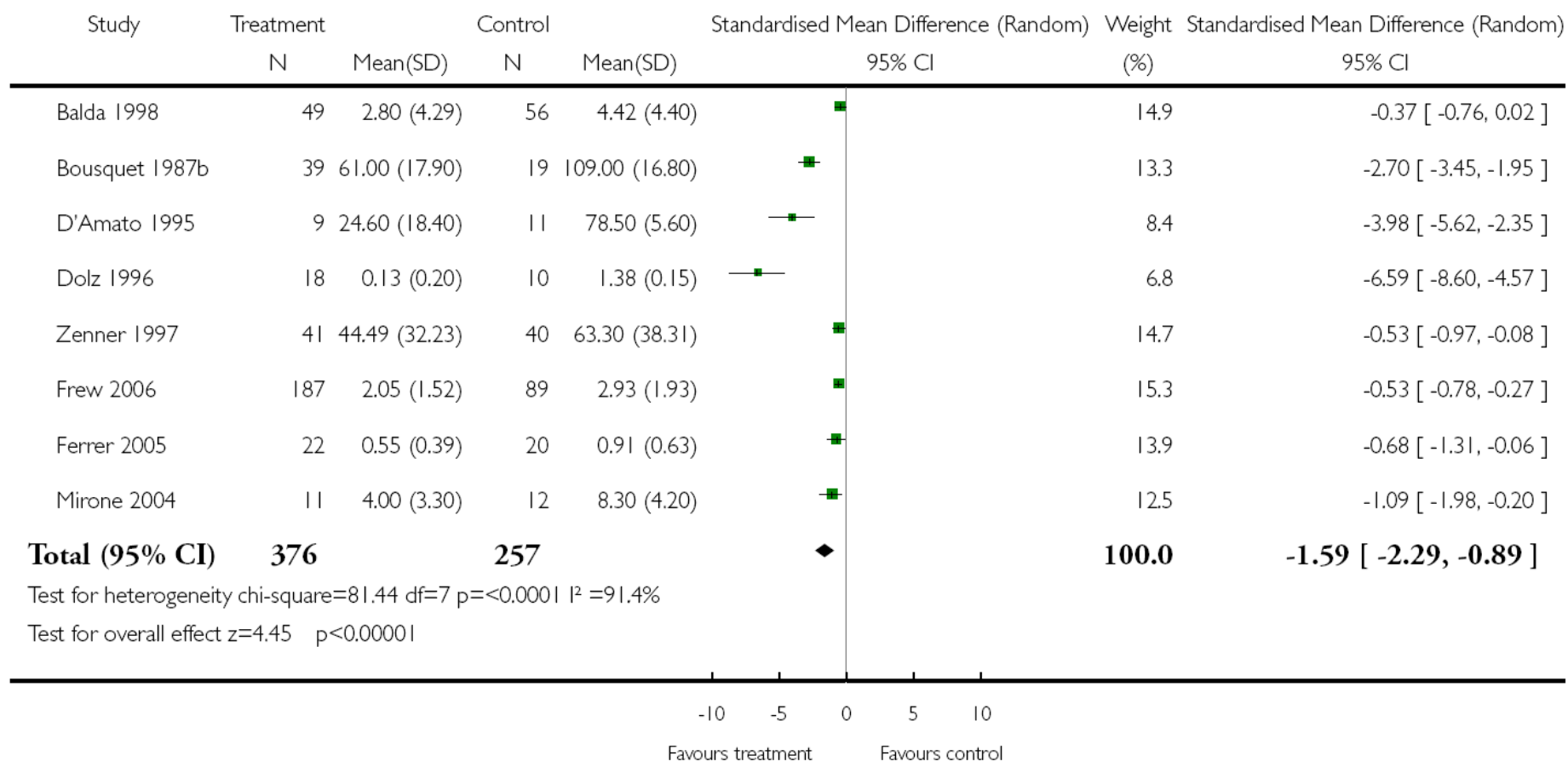
Yes

No

Don't know or undecided

Analysis 01.04.   Comparison 01 Active versus placebo, Outcome 04 Nasal symptom

Review: Allergen injection immunotherapy for seasonal allergic rhinitis

Comparison: 01 Active versus placebo

Outcome: 04 Nasal symptom

Study	Treatment		Control		Standardised Mean Difference (Random)	Weight	Standardised Mean Difference (Random)
	N	Mean(SD)	N	Mean(SD)	95% CI	(%)	95% CI
Balda 1998	49	2.80 (4.29)	56	4.42 (4.40)		14.9	-0.37 [-0.76, 0.02]
Bousquet 1987b	39	61.00 (17.90)	19	109.00 (16.80)		13.3	-2.70 [-3.45, -1.95]
D'Amato 1995	9	24.60 (18.40)	11	78.50 (5.60)		8.4	-3.98 [-5.62, -2.35]
Dolz 1996	18	0.13 (0.20)	10	1.38 (0.15)		6.8	-6.59 [-8.60, -4.57]
Zenner 1997	41	44.49 (32.23)	40	63.30 (38.31)		14.7	-0.53 [-0.97, -0.08]
Frew 2006	187	2.05 (1.52)	89	2.93 (1.93)		15.3	-0.53 [-0.78, -0.27]
Ferrer 2005	22	0.55 (0.39)	20	0.91 (0.63)		13.9	-0.68 [-1.31, -0.06]
Mirone 2004	11	4.00 (3.30)	12	8.30 (4.20)		12.5	-1.09 [-1.98, -0.20]
Total (95% CI)	376		257			100.0	-1.59 [-2.29, -0.89]

Test for heterogeneity chi-square=81.44 df=7 p=<0.0001 I² =91.4%

Test for overall effect z=4.45   p<0.00001

-10   -5   0   5   10

Favours treatment   Favours control

Inconsistency when 1 study?

- Do not downgrade

3. Directness of Evidence

- differences in
  - populations/patients (mild versus severe COPD, older, sicker or more co-morbidity)
  - interventions (all inhaled steroids, new vs. old)
  - outcomes (important vs. surrogate; long-term health-related quality of life, short –term functional capacity, laboratory exercise, spirometry)
- indirect comparisons
  - interested in A versus B
  - have A versus C and B versus C
  - formoterol versus salmeterol versus tiotropium

Indirect comparison

Source of indirectness	Question of interest	Example
Indirect comparison	Early emergency department systemic corticosteroids to treat acute exacerbations in adult patients with asthma	Both oral and intravenous routes are effective but there is no direct comparison of these two routes of administration in adults.

Difference in populations

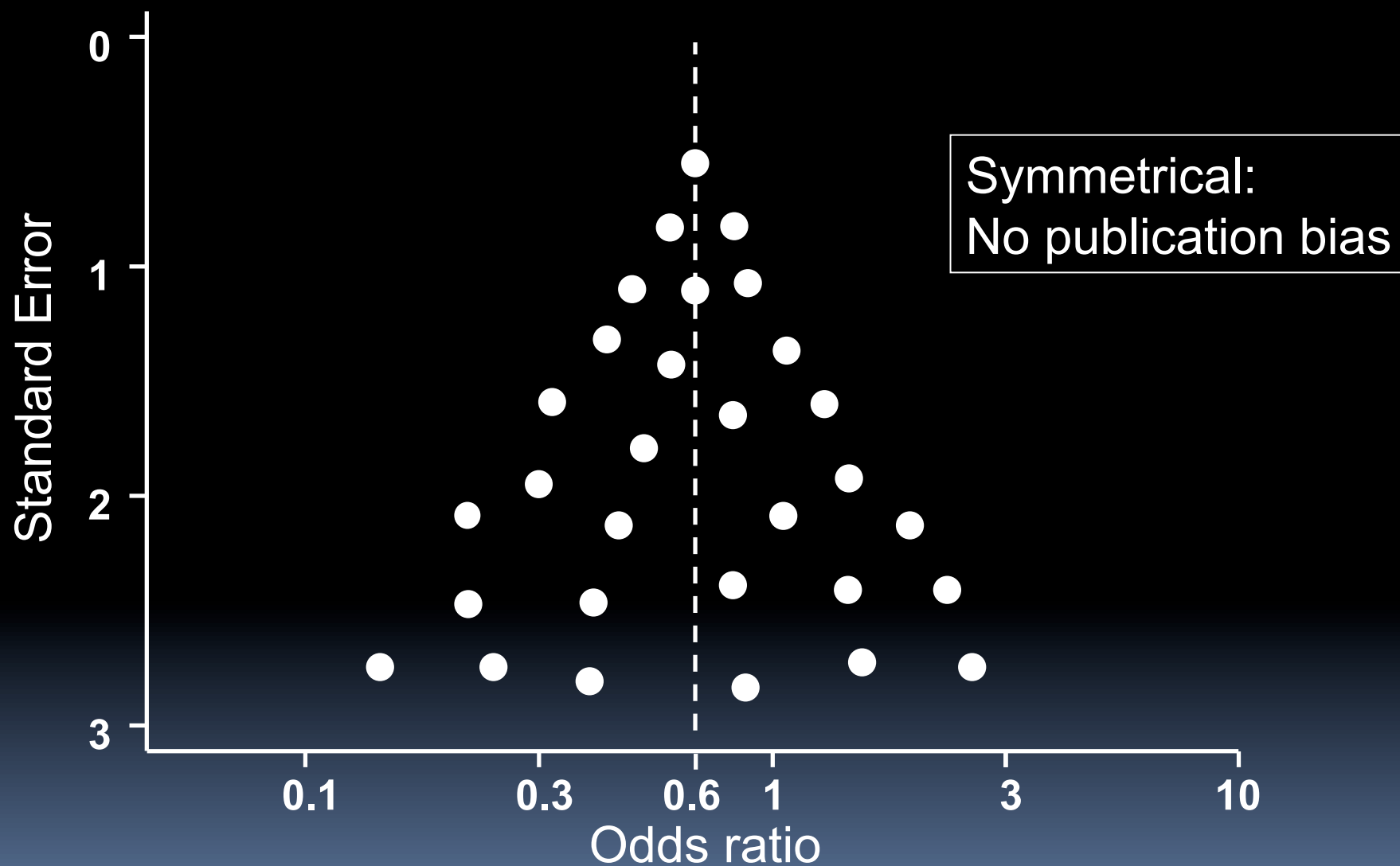| Source of indirectness	Question of interest	Example
Differences in populations	Anti-leukotrienes plus inhaled glucocorticosteroids vs. inhaled glucocorticosteroids alone to prevent asthma exacerbations and nighttime symptoms in patients with chronic asthma and allergic rhinitis.	Trials that measured asthma exacerbations and nighttime symptoms did not include patients with allergic rhinitis.

Differences in intervention

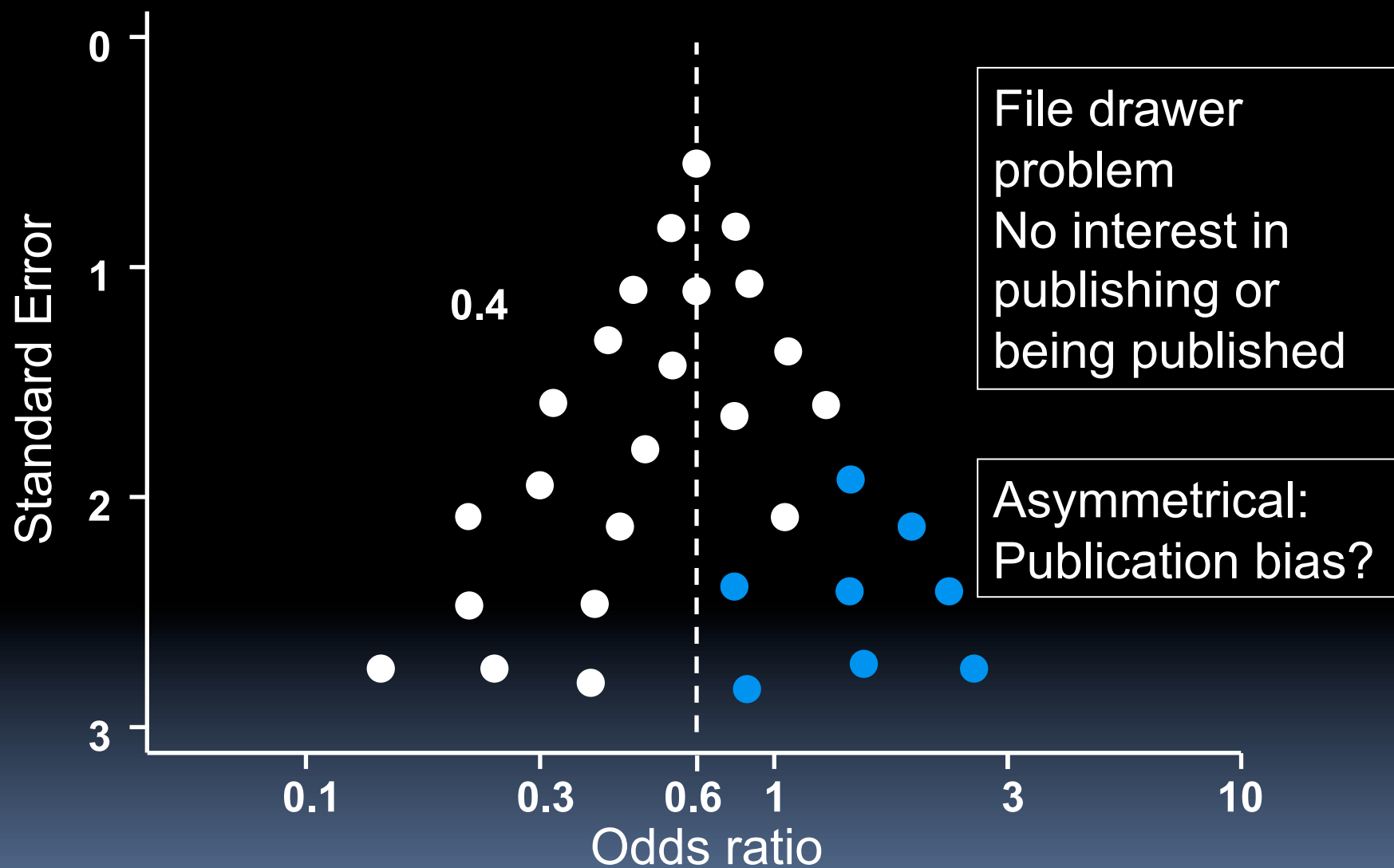| Source of indirectness	Question of interest	Example
Differences in intervention	Avoidance of pet allergens in non-allergic infants or preschool children to prevent development of allergy.	Available studies used multifaceted interventions directed at multiple potential risk factors in addition to pet avoidance.

Differences in outcomes

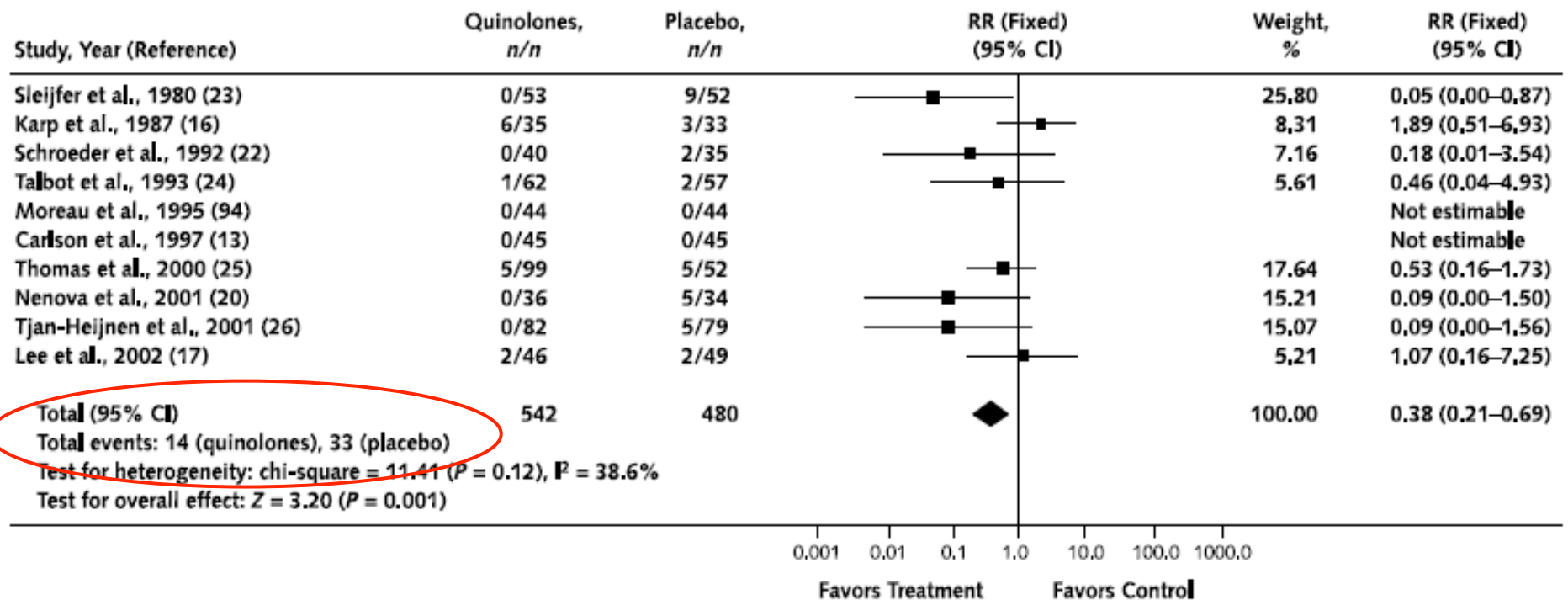| Source of indirectness	Question of interest	Example
Differences in outcomes of interest	Intranasal glucocorticosteroids vs. oral H_1-antihistamines in children with seasonal allergic rhinitis	In the available study parents were rating the symptoms and quality of life of their teenage children, instead the children themselves

Funnel plot

Symmetrical:
No publication bias

Funnel plot

Standard Error (y-axis): 0, 1, 2, 3

Odds ratio (x-axis): 0.1, 0.3, 0.6, 1, 3, 10

0.4

File drawer problem
No interest in publishing or being published
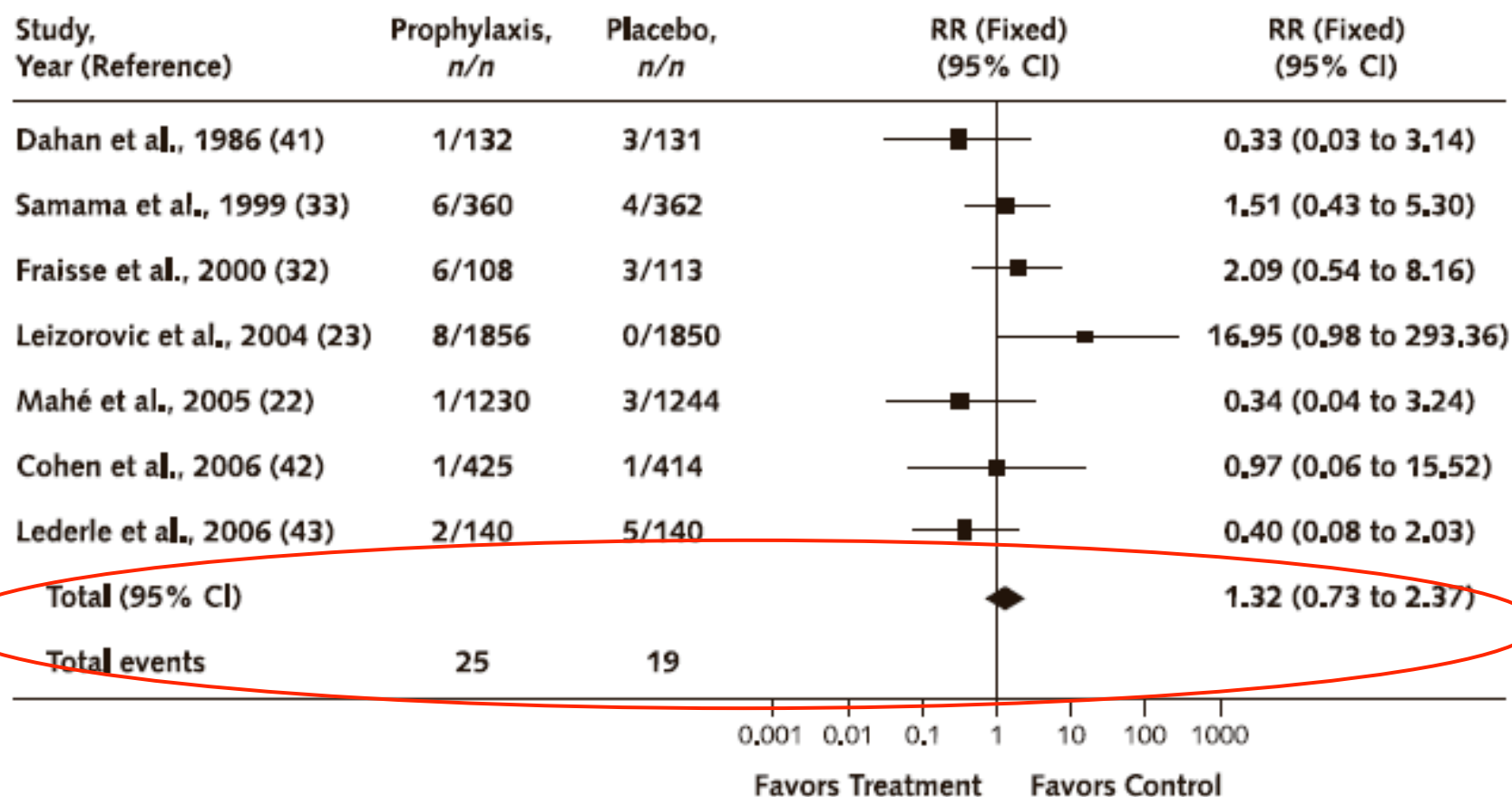
Asymmetrical: Publication bias?

5. Imprecision

- small sample size
  - small number of events

- wide confidence intervals
  - uncertainty about magnitude of effect

- extent to which confidence in estimate of effect adequate to support decision

Fluoroquinolone prophylaxis in neutropenia: infection-related mortality

Study, Year (Reference)	Quinolones, n/n	Placebo, n/n	RR (Fixed) (95% CI)	Weight, %	RR (Fixed) (95% CI)
Sleijfer et al., 1980 (23)	0/53	9/52		25.80	0.05 (0.00–0.87)
Karp et al., 1987 (16)	6/35	3/33		8.31	1.89 (0.51–6.93)
Schroeder et al., 1992 (22)	0/40	2/35		7.16	0.18 (0.01–3.54)
Talbot et al., 1993 (24)	1/62	2/57		5.61	0.46 (0.04–4.93)
Moreau et al., 1995 (94)	0/44	0/44			Not estimable
Carlson et al., 1997 (13)	0/45	0/45			Not estimable
Thomas et al., 2000 (25)	5/99	5/52		17.64	0.53 (0.16–1.73)
Nenova et al., 2001 (20)	0/36	5/34		15.21	0.09 (0.00–1.50)
Tjan-Heijnen et al., 2001 (26)	0/82	5/79		15.07	0.09 (0.00–1.56)
Lee et al., 2002 (17)	2/46	2/49		5.21	1.07 (0.16–7.25)
Total (95% CI)	542	480		100.00	0.38 (0.21–0.69)

Total events: 14 (quinolones), 33 (placebo)
Test for heterogeneity: chi-square = 11.41 (P = 0.12), I² = 38.6%
Test for overall effect: Z = 3.20 (P = 0.001)

0.001 0.01 0.1 1.0 10.0 100.0 1000.0

Favors Treatment   Favors Control

Example: Bleeding in the hospital

Study, Year (Reference)	Prophylaxis, n/n	Placebo, n/n	RR (Fixed) (95% CI)	RR (Fixed) (95% CI)
Dahan et al., 1986 (41)	1/132	3/131		0.33 (0.03 to 3.14)
Samama et al., 1999 (33)	6/360	4/362		1.51 (0.43 to 5.30)
Fraisse et al., 2000 (32)	6/108	3/113		2.09 (0.54 to 8.16)
Leizorovic et al., 2004 (23)	8/1856	0/1850		16.95 (0.98 to 293.36)
Mahé et al., 2005 (22)	1/1230	3/1244		0.34 (0.04 to 3.24)
Cohen et al., 2006 (42)	1/425	1/414		0.97 (0.06 to 15.52)
Lederle et al., 2006 (43)	2/140	5/140		0.40 (0.08 to 2.03)
Total (95% CI)				1.32 (0.73 to 2.37)
Total events	25	19		

0.001 0.01 0.1 1 10 100 1000

Favors Treatment   Favors Control

Dentali et al. Ann Int Med, 2007

What can raise quality?

1. large magnitude can upgrade (RRR 50%)
   - very large two levels (RRR 80%)
   - common criteria
     - everyone used to do badly
     - almost everyone does well
   - oral anticoagulation for mechanical heart valves
   - insulin for diabetic ketoacidosis
   - hip replacement for severe osteoarthritis
   - parachutes to prevent death when jumping from airplanes

What can raise quality?

2. dose response relation

- (higher INR – increased bleeding)
- childhood lymphoblastic leukemia
    - risk for CNS malignancies 15 years after cranial irradiation
    - no radiation: 1% (95% CI 0% to 2.1%)
    - 12 Gy: 1.6% (95% CI 0% to 3.4%)
    - 18 Gy: 3.3% (95% CI 0.9% to 5.6%)

3. all plausible confounding may be working to reduce the demonstrated effect or increase the effect if no effect was observed

All plausible confounding
would result in an underestimate of the treatment effect

- Higher death rates in private for-profit versus private not-for-profit hospitals
  - patients in the not-for-profit hospitals likely sicker than those in the for-profit hospitals
  - for-profit hospitals are likely to admit a larger proportion of well-insured patients than not-for-profit hospitals (and thus have more resources with a spill over effect)

All plausible confounding
would result in an overestimate of effect

- Hypoglycaemic drug phenformin causes lactic acidosis
- The related agent metformin is under suspicion for the same toxicity.
- Large observational studies have failed to demonstrate an association
  - Clinicians would be more alert to lactic acidosis in the presence of the agent

Quality assessment criteria

Quality of evidence	Study design	Lower if	Higher if
High	Randomised trials	**Study quality**: Serious limitations Very serious limitations Important **inconsistency** **Directness**: Some uncertainty Major uncertainty **Sparse or imprecise data** High probability of **publication bias**	**Strong association**: Strong, no plausible confounders Very strong, no major threats to validity Evidence of a **Dose response gradient** All plausible **confounders** would have reduced the effect
Moderate			
Low	Observational studies		
Very low			

Evidence Profiles/Summaries

Question 42 [profile 2]
Date: 2007-08-27
Question: Should ketotifen be used for long-term control of asthma and wheeze in children?
Bibliography: 1. Bassler D., Mitra A., Ducharme F.M., Forster J., Schwarzer G. Ketotifen alone or as additional medication for long-term control of asthma and wheeze in children. Cochrane database of systematic reviews (Online), 2004:CD001384.

No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	ketotifen	control	Relative (95% CI)	Absolute	Quality	Importance
Asthma symptoms (follow-up 10 to 12 weeks; Better indicated by less)												
4	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	very serious[3,4]	none[5,6]	72	76	-	SMD -0.49 (-0.16 to -0.82)	⊕◯◯◯ VERY LOW	CRITICAL
Asthma exacerbations (follow-up 12 weeks)												
2	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2,7]	serious[4]	none[6]	10/105	32/104	RR 0.31 (0.19 to 0.59)	213 fewer per 1000 (from 126 fewer to 249 fewer)	⊕⊕◯◯ LOW	CRITICAL
Use of oral glucocorticosteroid (follow-up 10 to 20 weeks)												
4	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	serious[4]	none[5,6]	21/156	73/150	RR 0.28 (0.13 to 0.58)	351 fewer per 1000 (from 205 fewer to 424 fewer)	⊕⊕◯◯ LOW	CRITICAL
Efficacy assessed either by participants or parents (follow-up 12 to 26 weeks)												
7	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	serious[8]	none[6]	101/301	143/298	RR 0.71 (0.52 to 0.96)	139 fewer per 1000 (from 19 fewer to 230 fewer)	⊕⊕◯◯ LOW	CRITICAL
Efficacy evaluated by physicians (follow-up 10 to 26 weeks)												
10	randomised trial	serious[9]	serious[10]	very serious[2,11]	no serious imprecision	reporting bias[12]	113/310	188/315	RR 0.6 (0.46 to 0.79)	239 fewer per 1000 (from 125 fewer to 322 fewer)	⊕◯◯◯ VERY LOW	IMPORTANT
Reduction in the use of bronchodilators (follow-up 12 to 16 weeks)												
12	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	serious[4]	none[6]	56/76	21/73	RR 2.39 (1.64 to 3.48)	400 more per 1000 (from 184 more to 714 more)	⊕⊕◯◯ LOW	IMPORTANT
Sedation (follow-up 10 to 26 weeks)												
7	randomised trial	no serious limitations[1]	no serious inconsistency	no serious indirectness	very serious[4,13]	none[6]	45/218	26/221	RR 1.69 (1.11 to 2.59)[14]	81 more per 1000 (from 13 more to 188 more)	⊕⊕◯◯ LOW	CRITICAL
Weight gain (follow-up 10 to 26 weeks)												
5	randomised trial	no serious limitations[1]	no serious inconsistency	no serious indirectness	very serious[13,15]	none[6,16]	38/142	24/141	RR 1.42 (1.02 to 1.99)	71 more per 1000 (from 3 more to 168 more)	⊕⊕◯◯ LOW	IMPORTANT
Withdrawal from study due to side effects (follow-up 10 to 16 weeks)												
3	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	very serious[4,17]	none[6,16]	5/129	3/109	RR 1.22 (0.3 to 4.92)	6 more per 1000 (from 20 fewer to 110 more)	⊕◯◯◯ VERY LOW	CRITICAL

[1] Most trials have been published before 1990, when reporting of methods were not as stringent as they are now, which may lead to inadequate reporting of good methods rather than bad methods per se.
[2] Inhaled corticosteroids were allowed as additional intervention in eight trials. There was not enough information in the studies to assess the effect of ketotifen as an add-on therapy to inhaled corticosteroids that are the mainstay of therapy of asthma today.
[3] Results include small or large effect.
[4] Very small trials with few events.
[5] Only four trials reported that outcome, but we did not downgrade since we already downgraded for very serious imprecision.

Evidence Profiles/Summaries

Quality assessment						
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations
Asthma symptoms (follow-up 10 to 12 weeks; Better indicated by less)						
4	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	very serious[3,4]	none[5,6]
Asthma exacerbations (follow-up 12 weeks)						
2	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2,7]	serious[4]	none[6]
Use of oral glucocorticosteroid (follow-up 10 to 20 weeks)						
4	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	serious[4]	none[5,6]
Efficacy assessed either by participants or parents (follow-up 12 to 26 weeks)						
7	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	serious[8]	none[6]
Efficacy evaluated by physicians (follow-up 10 to 26 weeks)						
10	randomised trial	serious[9]	serious[10]	very serious[2,11]	no serious imprecision	reporting bias[12]
Reduction in the use of bronchodilators (follow-up 12 to 16 weeks)						
12	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	serious[4]	none[6]
Sedation (follow-up 10 to 26 weeks)						
7	randomised trial	no serious limitations[1]	no serious inconsistency	no serious indirectness	very serious[4,13]	none[6]
Weight gain (follow-up 10 to 26 weeks)						
5	randomised trial	no serious limitations[1]	no serious inconsistency	no serious indirectness	very serious[13,15]	none[6,16]
Withdrawal from study due to side effects (follow-up 10 to 16 weeks)						
3	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	very serious[4,17]	none[6,16]

Evidence Profiles/Summaries

Question 42 [profile 2]
Date: 2007-08-27
Question: Should ketotifen be used for long-term control of asthma and wheeze in children?
Bibliography: 1. Bassler D., Mitra A., Ducharme F.M., Forster J., Schwarzer G. Ketotifen alone or as additional medication for long-term control of asthma and wheeze in children. Cochrane database of systematic reviews (Online), 2004:CD001384.

No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	ketotifen	control	Relative (95% CI)	Absolute	Quality	Importance
Asthma symptoms (follow-up 10 to 12 weeks; Better indicated by less)												
4	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	very serious[3,4]	none[5,6]	72	76	-	SMD -0.49 (-0.16 to -0.82)	⊕OOO VERY LOW	CRITICAL
Asthma exacerbations (follow-up 12 weeks)												
2	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2,7]	serious[4]	none[6]	10/105	32/104	RR 0.31 (0.19 to 0.59)	213 fewer per 1000 (from 126 fewer to 249 fewer)	⊕⊕OO LOW	CRITICAL
Use of oral glucocorticosteroid (follow-up 10 to 20 weeks)												
4	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	serious[4]	none[5,6]	21/156	73/150	RR 0.28 (0.13 to 0.58)	351 fewer per 1000 (from 205 fewer to 424 fewer)	⊕⊕OO LOW	CRITICAL
Efficacy assessed either by participants or parents (follow-up 12 to 26 weeks)												
7	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	serious[8]	none[6]	101/301	143/298	RR 0.71 (0.52 to 0.96)	139 fewer per 1000 (from 19 fewer to 230 fewer)	⊕⊕OO LOW	CRITICAL
Efficacy evaluated by physicians (follow-up 10 to 26 weeks)												
10	randomised trial	serious[9]	serious[10]	very serious[2,11]	no serious imprecision	reporting bias[12]	113/310	188/315	RR 0.6 (0.46 to 0.79)	239 fewer per 1000 (from 125 fewer to 322 fewer)	⊕OOO VERY LOW	IMPORTANT
Reduction in the use of bronchodilators (follow-up 12 to 16 weeks)												
12	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	serious[4]	none[6]	56/76	21/73	RR 2.39 (1.64 to 3.48)	400 more per 1000 (from 184 more to 714 more)	⊕⊕OO LOW	IMPORTANT
Sedation (follow-up 10 to 26 weeks)												
7	randomised trial	no serious limitations[1]	no serious inconsistency	no serious indirectness	very serious[4,13]	none[6]	45/218	26/221	RR 1.69 (1.11 to 2.59)[14]	81 more per 1000 (from 13 more to 188 more)	⊕⊕OO LOW	CRITICAL
Weight gain (follow-up 10 to 26 weeks)												
5	randomised trial	no serious limitations[1]	no serious inconsistency	no serious indirectness	very serious[13,15]	none[6,16]	38/142	24/141	RR 1.42 (1.02 to 1.99)	71 more per 1000 (from 3 more to 168 more)	⊕⊕OO LOW	IMPORTANT
Withdrawal from study due to side effects (follow-up 10 to 16 weeks)												
3	randomised trial	no serious limitations[1]	no serious inconsistency	serious[2]	very serious[4,17]	none[6,16]	5/129	3/109	RR 1.22 (0.3 to 4.92)	6 more per 1000 (from 20 fewer to 110 more)	⊕OOO VERY LOW	CRITICAL

Table header (spanning): Quality assessment | Summary of findings — No of patients (ketotifen, control), Effect (Relative (95% CI), Absolute), Quality, Importance

[1] Most trials have been published before 1990, when reporting of methods were not as stringent as they are now, which may lead to inadequate reporting of good methods rather than bad methods per se.
[2] Inhaled corticosteroids were allowed as additional intervention in eight trials. There was not enough information in the studies to assess the effect of ketotifen as an add-on therapy to inhaled corticosteroids that are the mainstay of therapy of asthma today.
[3] Results include small or large effect.
[4] Very small trials with few events.
[5] Only four trials reported that outcome, but we did not downgrade since we already downgraded for very serious imprecision.

Evidence Profiles/Summaries

No of patients		Effect		Quality	Importance
ketotifen	control	Relative (95% CI)	Absolute		
72	76	-	SMD -0.49 (-0.16 to -0.82)	⊕○○○ VERY LOW	CRITICAL
10/105	32/104	RR 0.31 (0.19 to 0.59)	213 fewer per 1000 (from 126 fewer to 249 fewer)	⊕⊕○○ LOW	CRITICAL
21/156	73/150	RR 0.28 (0.13 to 0.58)	351 fewer per 1000 (from 205 fewer to 424 fewer)	⊕⊕○○ LOW	CRITICAL
101/301	143/298	RR 0.71 (0.52 to 0.96)	139 fewer per 1000 (from 19 fewer to 230 fewer)	⊕⊕○○ LOW	CRITICAL
113/310	188/315	RR 0.6 (0.46 to 0.79)	239 fewer per 1000 (from 125 fewer to 322 fewer)	⊕○○○ VERY LOW	IMPORTANT
56/76	21/73	RR 2.39 (1.64 to 3.48)	400 more per 1000 (from 184 more to 714 more)	⊕⊕○○ LOW	IMPORTANT
45/218	26/221	RR 1.69 (1.11 to 2.59)[14]	81 more per 1000 (from 13 more to 188 more)	⊕⊕○○ LOW	CRITICAL
38/142	24/141	RR 1.42 (1.02 to 1.99)	71 more per 1000 (from 3 more to 168 more)	⊕⊕○○ LOW	IMPORTANT
5/129	3/109	RR 1.22 (0.3 to 4.92)	6 more per 1000 (from 20 fewer to 110 more)	⊕○○○ VERY LOW	CRITICAL

Summary of findings

GRADE Evidence Profiles

Should subcutaneous specific immunotherapy be used for treatment of allergic rhinitis in children without concomitant asthma?

Author(s): JLB & HJS
Date: 2007-08-08
Question: Should subcutaneous immunotherapy be used in children with allergic rhinitis?
Settings:
Bibliography: Fontana V.J., Holt L.E., Jr., Mainland D. Effectiveness of hyposensitization therapy in ragweed hay-fever in children. JAMA, 1966;195:985-992. Niggemann B., Jacobsen L., Dreborg S., et al. Five-year follow-up on the PAT study: specific immunotherapy and long-term prevention of asthma in children. Allergy, 2006;61:855-859. Calderon M., Alves B., Jacobson M., Hurwitz B., Sheikh A., Durham S. Allergen injection immunotherapy for seasonal allergic rhinitis. Cochrane Database Syst Rev, 2007:CD001936.

No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	subcutaneous immunotherapy	control	Relative (95% CI)	Absolute	Quality	Importance
Quality assessment							**Summary of findings**					
							No of patients		**Effect**			
nasal symptoms (follow-up 3 to 5 years; range of scores: 0-0; Better indicated by less)												
2	randomised trial	serious[1]	serious[2]	no serious indirectness	serious[3]	none	93	84	-	not pooled[4]	⊕000 VERY LOW	CRITICAL
development of asthma (follow-up 5 years)												
1	randomised trial	serious[5]	no serious inconsistency	no serious indirectness	serious[6]	none	15/75	29/67	RR 0.46 (0.27 to 0.77)	234 fewer per 1000 (from 100 fewer to 316 fewer)	⊕⊕00 LOW	CRITICAL
non-life threatening systemic adverse events												
13	randomised trial	no serious limitations	no serious inconsistency	serious[7]	no serious imprecision	none	43/615[8]	3/463[8]	not pooled	not pooled	⊕⊕⊕0 MODERATE	CRITICAL
anaphylactic shock												
9	randomised trial	no serious limitations	no serious inconsistency	serious[7]	serious[9]	none	0/417[8]	0/303[8]	not pooled	not pooled	⊕⊕00 LOW	CRITICAL
adrenaline use for systemic reaction												
13	randomised trial	no serious limitations	no serious inconsistency	serious[7]	no serious imprecision	none	0/0[10]	0/0[11]	not pooled	not pooled	⊕⊕⊕0 MODERATE	CRITICAL

[1] One old small trial and a subgroup analysis of one small recent trial.
[2] One trial showed benefit while the other did not.
[3] One trial found no difference between the SCIT and placebo groups (improvement in 11/25 and 11/26 children, respectively) and the other found improvement in symptom score of 14 mm on a 100 mm visual analog scale.
[4] Reporting of symptoms did not alow for meta-analysis. One trial found no difference between the treated and placebo groups (improvement in 11/25 and 11/26 children, respectively). Second found that the symptoms scores measured on a visual analog scale improved more in the SCIT group compared to placebo (-21.5 mm vs -7.4 mm).
[5] post hoc subgroup analysis
[6] small trial with small number of events
[7] extrapolated from trials in adults
[8] Most studies reported number of adverse events, rather than the number of participants in which one or more adverse events were observed.
[9] very small number of events
[10] 19 events of 14,085 injections
[11] 1 event in 8278 injections

Self management for patients with chronic obstructive pulmonary disease

Patient or population: patients with chronic obstructive pulmonary disease
Settings: primary care, community, outpatient
Intervention: self management[1]
Comparison: usual care

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk usual care	Corresponding risk self management				
Quality of Life St George's Respiratory Questionnaire. Scale from: 0 to 100. (follow-up: 3 to 12 months)	The mean quality of life ranged across control groups from 38 to 60 points	The mean quality of Life in the intervention groups was **2.58 lower** (5.14 to 0.02 lower)		698 (7)	⊕⊕⊕O moderate[2]	Lower score indicates better quality of life. A change of less than 4 points is not shown to be important to patients.
Dyspnoea Borg Scale. Scale from: 0 to 10. (follow-up: 3 to 6 months)	The mean dyspnoea ranged across control groups from **1.2 to 4.1 points**	The mean dyspnoea in the intervention groups was **0.53 lower** (0.96 to 0.1 lower)		144 (2)	⊕⊕OO low[3,4]	Lower score indicates improvement
Number and severity of exacerbations[5]	See comment	See comment	Not estimable[5]	591 (3)	See comment	Effect is uncertain
Respiratory-related hospital admissions (follow-up: 3 to 12 months)	Low risk population[6]		OR 0.64 (0.47 to 0.89)	966 (8)	⊕⊕⊕O moderate[7]	
	10 per 100	7 per 100 (5 to 9)				
	High risk population[6]					
	50 per 100	39 per 100 (32 to 47)				
Emergency department visits for lung diseases (follow-up: 6 to 12 months)	The mean emergency department visits for lung diseases ranged across control groups from 0.2 to 0.7 visits per person per year	The mean emergency department visits for lung diseases in the intervention groups was **0.1 higher** (0.2 lower to 0.3 higher)		328 (4)	⊕⊕⊕O moderate[4]	
Doctor and nurse visits (follow-up: 6 to 12 months)	The mean doctor and nurse visits ranged across control groups from 1 to 5 vists per person per year	The mean doctor and nurse visits in the intervention groups was **0.02 higher** (1 lower to 1 higher)		629 (8)	⊕⊕⊕O moderate[8]	

*The basis for the assumed risk (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval; OR: Odds ratio;

How to create a summary of findings table or evidence profile

- GRADEpro – software to create SoF
  - Available from Cochrane IMS website
- Import data from RevMan 5 into GRADEpro
- Create table – author makes suggestions about information to present and GRADEs the evidence
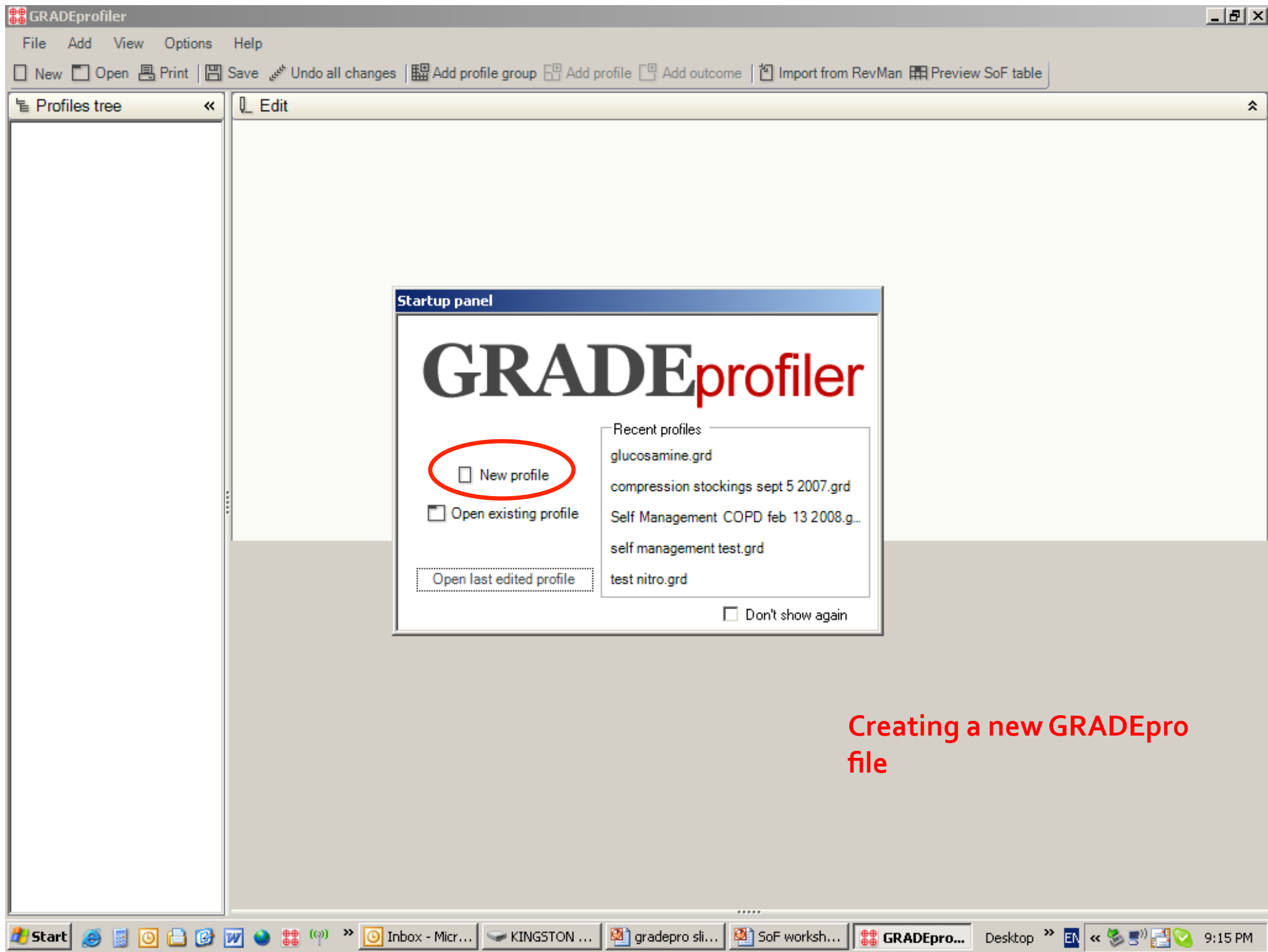
Creating a new GRADEpro file

Content

- Background about GRADE

- GRADE approach
  - Quality of evidence
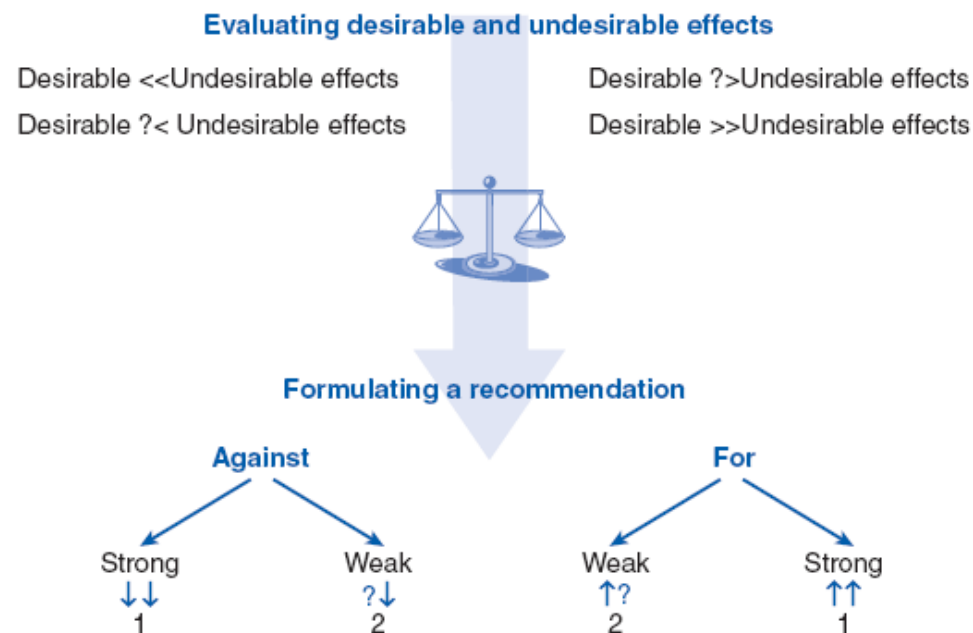 - Strength of recommendations

Strength of recommendation

- "The strength of a recommendation reflects the extent to which we can, across the range of patients for whom the recommendations are intended, be confident that desirable effects of a management strategy outweigh undesirable effects."
- Strong or weak/conditional

Quality of evidence & strength of recommendation

- Linked but no automatism
- Other factors beyond the quality of evidence influence our confidence that adherence to a recommendation causes more benefit than harm
- Systems/approaches failed to make this explicit
- GRADE separates quality of evidence from strength of recommendation

Developing recommendations

Strength of Recommendations

Evaluating desirable and undesirable effects

Desirable <<Undesirable effects

Desirable ?< Undesirable effects

Desirable ?>Undesirable effects

Desirable >>Undesirable effects

Formulating a recommendation

Against		For	
Strong ↓↓ 1	Weak ?↓ 2	Weak ↑? 2	Strong ↑↑ 1

The figure describes the balance between important benefits and downsides relate to a recommendation. The process begins by evaluating whether desirable effects outweigh undesirable effects or vice versa. Moving on to making a recommendation requires a decision: if the balance is clear, a strong recommendation for or against an action follows (<< and >> denote a clear balance). If the balance is not clear, a weak recommendation for or against an action follows (?< and ?> denote a balance that is not clear). Widely differing values (the importance or preference patients assign to a certainhealth state) can also lead to a less clear balance of benefits versus downsides.

Implications of a strong recommendation

- **Patients**: Most people in your situation would want the recommended course of action and only a small proportion would not
- **Clinicians**: Most patients should receive the recommended course of action
- **Policy makers**: The recommendation can be adapted as a policy in most situations

Implications of a weak/conditional recommendation

- **Patients**: The majority of people in your situation would want the recommended course of action, but many would not
- **Clinicians**: Be prepared to help patients to make a decision that is consistent with their own values
- **Policy makers**: There is a need for substantial debate and involvement of stakeholders

Factors determining strength of recommendation

Factors that can strengthen a recommendation	Comment
Quality of the evidence	The higher the quality of evidence, the more likely is a strong recommendation.
Balance between desirable and undesirable effects	The larger the difference between the desirable and undesirable consequences, the more likely a strong recommendation warranted. The smaller the net benefit and the lower the certainty for that benefit, the more likely is a weak recommendation.
Values and preferences	The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely weak recommendation warranted.
Costs (resource allocation)	The higher the costs of an intervention – that is, the more resources consumed – the less likely is a strong recommendation warranted

Current state of recommendations

INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS 78 (2009) 354–363

journal homepage: www.intl.elsevierhealth.com/journals/ijmi

ELSEVIER

The Yale Guideline Recommendation Corpus: A representative sample of the knowledge content of guidelines

Tamseela Hussain*, George Michel, Richard N. Shiffman

Yale Center for Medical Informatics, Yale University School of Medicine, New Haven, CT, United States

Current state of recommendations

- Reviewed 7527 recommendations
  - 1275 randomly selected
- Inconsistency across/within
- 31.6% did not recommendations clearly
  - Most of them not written as executable actions
- 52.7% did not indicated strength

Challenges in wording recommendations

- Need to express (two) levels
  - Strong vs weak/conditional
- Need to express direction
- Differences across languages

	Wording 1	Wording 2	Wording 3
Strong recommendation for	We recommend…	Clinicians should…	We recommend…
Weak recommendation for	We suggest	Clinicians might…	We conditionally recommend…
Weak recommendation against	We suggest...not	Clinicians might not…	We conditionally recommend...not
Strong recommendation against	We recommend …not	Clinicians should not…	We recommend …not

- Need codes (letters, symbols, numbers)

Date: 2009-12-01
Question: Should **extensively hydrolysed formula** vs **soy formula** be used in children with cow's milk allergy?
Bibliography: 1. Agostoni C., Fiocchi A., Riva E., Terracciano L., Sarratud T., Martelli A., Lodi F., D'Auria E., Zuccotti G., Giovannini M. Growth of infants with IgE-mediated cow's milk allergy fed different formulas in the complementary feeding period. Pediatric Allergy & Immunology, 2007;18:599-606. 2. Klemola T., Vanto T., Juntunen-Backman K., Kalimo K., Korpela R., Varjonen E. Allergy to soy formula and to extensively hydrolyzed whey formula in infants with cow's milk allergy: a prospective, randomized study with a follow-up to the age of 2 years. Journal of Pediatrics, 2002;140:219-224.

McMaster University
Inspiring Innovation and Discovery

No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other consideration	extensively hydrolysed formula	soy formula	Relative (95% CI)	Absolute	Quality	Importance
			Quality assessment				No of patients		Effect		Quality	Importance
Severe symptoms of CMA (severe laryngeal edema, severe asthma, anaphylaxis) (follow-up 12 and 24 months)												
2	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	0/125	0/117	Not estimable[1]	Not estimable[1]	⊕⊕OO LOW	CRITICAL
Allergic reaction to formula (follow-up 12 and 24 months)												
2	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	2/125 (1.6%)	13/117 (11.1%)	RR 0.18 (0.05 to 0.71)	91 fewer per 1000 (from 32 fewer to 106 fewer)	⊕⊕OO LOW	CRITICAL
Moderate symptoms of CMA (mild laryngeal edema or mild asthma)												
2	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	0/125	0/117	Not estimable[1]	Not estimable[1]	⊕⊕OO LOW	CRITICAL
Enteropathy or enterocolitis/proctocolitis (follow-up 12 and 24 months)												
2	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	0/125	0/117	Not estimable[1]	Not estimable[1]	⊕⊕OO LOW	CRITICAL
Failure to thrive (measured as: length for age z-score) (follow-up 12 months; Better indicated by higher values)												
1	randomised trials	serious[2]	no serious inconsistency	serious[4]	serious[5]	none	31	32	-	MD 0.27 higher (0.19 lower to 0.73 higher)	⊕OOO VERY LOW	CRITICAL
Failure to thrive (measured as: weight for age z-score) (follow-up 12 months; Better indicated by higher values)												
1	randomised trials	serious[2]	no serious inconsistency	serious[6]	serious[5]	none	31	32	-	MD 0.23 higher (0.01 to 0.45 higher)	⊕OOO VERY LOW	CRITICAL

[1] No events in both studies. [2] Allocation concealment was not reported and studies were not blinded. One study reported the results of per protocol analysis only.

[3] Only 15 events. [4] There is uncertainty to what extent a length for age z-score reflects a change in growth that would have an important consequence for a patient.

[5] Only 62 children. [6] There is uncertainty to what extent a weight for age z-score reflects a change in growth that would have an important consequence for a patient.

[7] Allocation concealment was not reported and studies were not blinded. In one study outcome was measured only in patients who developed symptoms. [8] One additional study (Salpietro 2005) included children with cow's milk allergy (23%) or intolerance and reported a relative risk of secondary sensitization to extensively hydrolysed casein formula compared to soy formula of 1.33 (95% CI: 0.37 to 4.82).

[9] It is uncertain how important is sensitization alone. [10] Only 11 events. [11] Only 4 events.

Example recommendation from the cow milk allergy guidelines

In children with IgE-mediated cow's milk allergy, we suggest extensively hydrolysed milk formula rather than soy formula (conditional recommendation | ⊕○○○/very low quality evidence).

Underlying values and preferences

This recommendation places a relatively high value on avoiding adverse reactions to soy formula, and a relatively low value on an inferior acceptance of the extensively hydrolysed formula and resource utilization. In settings where relative importance of resource expenditure is lower an alternative choice may be equally reasonable.

Relevant healthcare question?

Clinical question:

Population: Avian Flu/influenza A (H5N1) patients

Intervention: Oseltamivir (or Zanamivir)

Comparison: No pharmacological intervention

Outcomes: Mortality, hospitalizations, resource use, adverse outcomes, antimicrobial resistance

Example: Oseltamivir for Avian Flu

Recommendation: In patients with confirmed or strongly suspected infection with avian influenza A (H5N1) virus, clinicians should administer oseltamivir treatment as soon as possible (strong recommendation, very low quality evidence).

Values and Preferences

Remarks: This recommendation places a high value on the prevention of death in an illness with a high case fatality. It places relatively low values on adverse reactions, the development of resistance and costs of treatment.

Schunemann et al., The Lancet ID, 2007

Other explanations

Remarks: Despite the lack of controlled treatment data for H5N1, this is a strong recommendation, in part, because there is a lack of known effective alternative pharmacological interventions at this time.

The panel voted on whether this recommendation should be strong or weak and there was one abstention and one dissenting vote.

Schunemann et al., The Lancet ID, 2007

Formulate question   **Select outcomes**   **Rate importance**   Outcomes across studies   *Create evidence profile with GRADEpro*   Rate quality of evidence for each outcome   RCT start high, obs. data start low

P	Outcome	Critical
I	Outcome	Critical
C	Outcome	Important
O	Outcome	Not important

High
Moderate
Low
Very low

Summary of findings & estimate of effect for each outcome

Grade down
1. Risk of bias
2. Inconsistency
3. Indirectness
4. Imprecision
5. Publication bias

Grade up
1. Large effect
2. Dose response
3. Confounders

Systematic review

Guideline development

Formulate recommendations:
• For or against (direction)
• Strong or weak (strength)

By considering:
☐ Quality of evidence
☐ Balance benefits/harms
☐ Values and preferences

Revise if necessary by considering:
☐ Resource use (cost)

Panel

Rate **overall quality of evidence** across outcomes based on lowest quality of *critical* outcomes

AMERICAN GASTROENTEROLOGICAL ASSOCIATION

Colorectal Cancer Screening and Surveillance: Clinical Guidelines and Rationale—Update Based on New Evidence

• "We recommend using…"
• "We suggest using…"
• "We recommend against using…"
• "We suggest against using…"

GRADE – conclusions

- Do we need grading?
  - Impossible to train *all* practitioners in research methods and the related complexity quickly
- GRADE: Unifying system to evaluate quality of evidence & strength of recommendations
- Clear separation of 2 issues
  - Quality of evidence – 4 levels
  - Recommendations – conditional or strong
- Transparent
- Systematic by and across outcomes

Thank you!