Holger Schünemann

Director, Cochrane Canada

Chair and Professor

Dept. of Health Research Methods, Evidence & Impact*

Professor of Medicine | McMaster University

*formerly „Clinical Epidemiology and Biostatistics"

@schunemann_mac @CochraneCanada

GRADE Training

Cochrane Canada

McMaster University
HEALTH RESEARCH METHODS, EVIDENCE, AND IMPACT

Disclosures

Cochrane Canada Director

GRADE working group Co-chair

No direct financial COI

Views expressed my own

Today

Before Coffee

- Considering and understanding PICOs
- Your SoF Table
- Understand GRADEpro
- Assessing Evidence – Understand GRADE

Coffee to lunch

- Assessing the certainty – Do GRADE

Lunch to afternoon break

- Complete your SoF Table

A clinically sensible question

Population: In patients with (lung) cancer, what is the impact of

Intervention: heparin
(comparison) compared with no heparin

Outcomes: on the risk for venous thromboembolism, death, bleeding, burden…?

PICO

Questions

Should be

practice

NOT

evidence driven

Good questions…

Questions you have when trying to decide what to prescribe/recommend to your patient

Questions you have when trying to decide what to provide in your country/region/ clinic

What should you do with the person in front of you?

Outcomes

Should be

importance driven

NOT

evidence driven

Cuello-Garcia *et al. World Allergy Organization Journal* (2016) 9:10
DOI 10.1186/s40413-016-0102-7

World Allergy Organization Journal

POSITION ARTICLE AND GUIDELINES Open Access

World Allergy Organization-McMaster University Guidelines for Allergic Disease Prevention (GLAD-P): Prebiotics

Carlos A. Cuello-Garcia[1,2], Alessandro Fiocchi[3†], Ruby Pawankar[4†], Juan José Yepes-Nuñez[1,5], Gian Paolo Morgano[1], Yuan Zhang[1], Kangmo Ahn[6], Suleiman Al-Hammadi[7], Arnav Agarwal[8], Shreyas Gandhi[8], Kirsten Beyer[9], Wesley Burks[10], Giorgio W. Canonica[11], Motohiro Ebisawa[12], Rose Kamenwa[13], Bee Wah Lee[14], Haiqi Li[15], Susan Prescott[16], John J. Riva[1,17], Lanny Rosenwasser[18], Hugh Sampson[19], Michael Spigler[20], Luigi Terracciano[21], Andrea Vereda[22], Susan Waserman[23], Holger J. Schünemann[1,23*] and Jan L. Brożek[1,23]

* Correspondence: holger.schunemann@mcmaster.ca
†Equal contributors
[1]Department of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, ON, Canada

3. Should prebiotics vs. no prebiotics be used in infants?

Depends on the outcomes

For the question comparing cryotherapy performed by a doctor or a non-physician.

Please rate the outcome on a scale from 1 (least important) to 9 (most critical).
Note: You can rate multiple outcomes with the same number.

	1	2	3	4	5	6	7	8	9
cervical cancer incidence	○	○	○	○	○	○	○	○	○
recurrence of CIN/cure rates	○	○	○	○	○	○	○	○	○
retreatment rates	○	○	○	○	○	○	○	○	○
major bleeding *(requiring hospital admission or blood transmission with no long term sequelae)*	○	○	○	○	○	○	○	○	○
minor bleeding (e.g. spotting)	○	○	○	○	○	○	○	○	○
major infection *(requiring hospital admission and antibiotics, no long term sequelae)*	○	○	○	○	○	○	○	○	○
minor infection *(requiring outpatient treatment only)*	○	○	○	○	○	○	○	○	○
all severe adverse events *(including major bleeding, major infections, etc.)*	○	○	○	○	○	○	○	○	○
all minor adverse events *(including minor bleeding, minor infection, discharge, flushing, feeling faint, etc.)*	○	○	○	○	○	○	○	○	○
pain *(requiring local treatment)*	○	○	○	○	○	○	○	○	○
feeling faint	○	○	○	○	○	○	○	○	○
flushing	○	○	○	○	○	○	○	○	○
discharge	○	○	○	○	○	○	○	○	○
resource use *(including cost, human resources and length of stay)*	○	○	○	○	○	○	○	○	○
acceptability to providers *(please clarify acceptability)*	○	○	○	○	○	○	○	○	○
acceptability to women *(please clarify acceptability)*	○	○	○	○	○	○	○	○	○
referrals after cryotherapy for complications or follow-up treatment	○	○	○	○	○	○	○	○	○
HIV transmission *(HIV infection, HIV shedding)*	○	○	○	○	○	○	○	○	○

List of outcomes can be long...

Not everything that is measured is important and not everything that is important is measured

That may be particularly true for outcomes in nutrition research

- Lab values – patient important outcomes?
- Vitamin D levels, carotenoid levels, etc.

Choose outcomes

Establish methods for rating the relative importance of outcomes:

- Consultation with consumers and stakeholders (e.g., survey)
- Systematic review of consumers and stakeholders' views
- Input of panelists (including consumers and stakeholders): informal vs. formal and structured

Choosing outcomes

Generate a list of outcomes (from literature, from the panel, from clinical experts, from patient groups)

Issues

Surrogate outcomes

- Outcomes that are relatively infrequent or occur over a long period of time
- Use substitutes or surrogates
- May not be important to decision making

e.g. IgE levels, biopsy gastrointestinal tract vs diarrhoea, weight loss

Other issues

“we will not find any data for ....”

Do not exclude outcomes for which you think there will not be data

Do not let little data influence the ranking of the outcome (if it’s critical, it is critical)

Choosing outcomes

Generate a list of outcomes (from literature, from the panel, from clinical experts, from patient groups)

Ask panel to rank the outcomes by importance (anonymous)

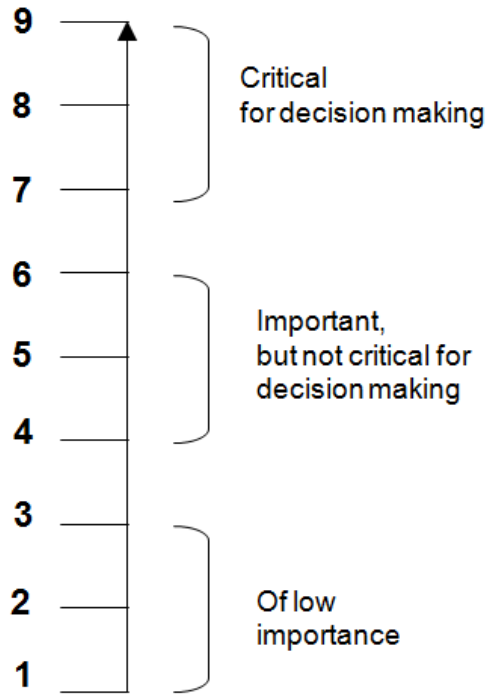
Approach to outcome determination

Desirable outcomes

- Lower mortality
- reduced hospital stay
- Pulmonary embolism

Undesirable outcomes

- Adverse reactions
- Major bleeding



→Consider desirable and undesirable outcomes explicitly

Not more than ~7 outcomes (SoF Tables)

TASK

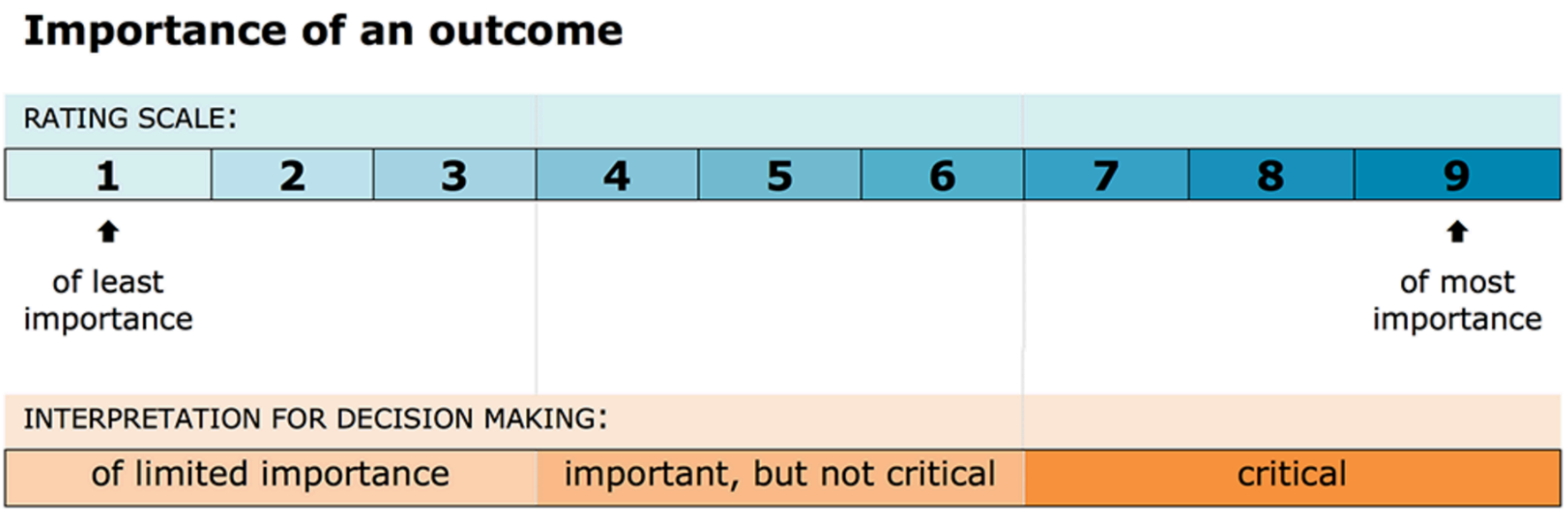
Rate the relative importance of the outcome for decision-making (i.e. formulating a recommendation) of each outcome on a scale from 1 to 9. The meaning of the ratings are:

1–3 are of limited or no importance for decision-making

4–6 are important, but not critical for decision-making

7–9 are critical for decision-making.

Again, you can use the same rating for different outcomes more than once.



Click Next to proceed to the Outcome Importance Rating.

RATING SCALE:

1	2	3	4	5	6	7	8	9

of **least** importance

of **most** importance

of limited importance for making a decision (not included in evidence profile)	important, but not critical for making a decision (included in evidence profile)	Critical for making a decision (included in evidence profile)

Outcome	Your rating of importance (1 to 9)	Group rating of importance (1 to 9)	Include in Evidence profile	
			Yes	No
			Yes	No
			Yes	No
			Yes	No
			Yes	No
			Yes	No
			Yes	No
			Yes	No
			Yes	No
			Yes	No
			Yes	No

3. Outcomes for treatments

1. OUTCOMES FOR TREATMENT OPTIONS (cold knife conization, cryotherapy and LEEP)

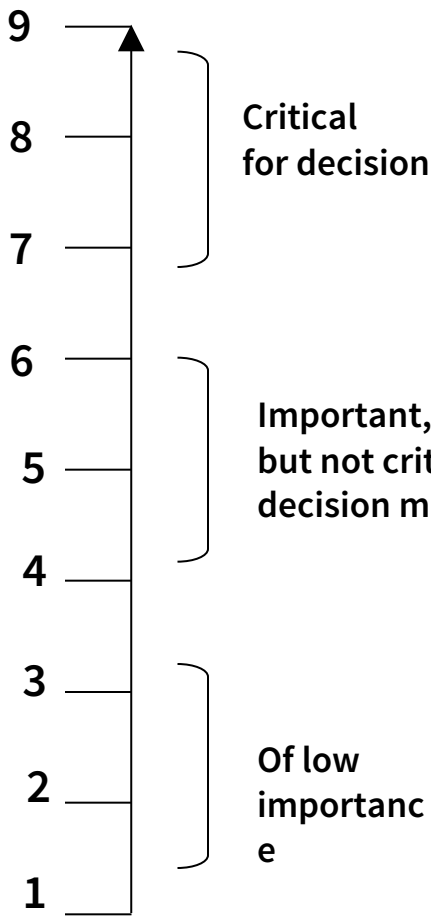
Choose the most important outcomes for decision making: Consider outcomes that might be important to someone making a decision to use or not to use the treatment (make sure to rank benefits and adverse effects)

Rate the relative importance of each outcome on a 9 point scale ranging from 1 (not important) to 9 (critical). You can use the same rating several times (i.e. same number for more than one outcome).
1 – 3 not important,
4 – 6 important, but not critical for making a decision
7 – 9 critical for making a decision

	1 (not important)	2	3	4	5	6	7	8	9 (critical)
Residual/recurrent CIN 2,3 (after 6, 12 months and 24 months)									
HPV negative (after 6, 12 and 24 months)									
Pelvic Inflammatory Disease									
Major infections *(requiring hospital admission and antibiotics)*									
Premature delivery									
infertility									
maternal death									
fetal/neonatal spontaneous abortions									
Major Bleeding *(requires hospitalization/blood transfusion)*									
Minor bleeding *(requires packing or suturing)*									
Damage to other organs/other surgery required – such as injury to bladder or urethra									

Comments or other outcomes (indicate not important, important, critical)

How do you rate outcomes?



- Survey the guideline panel about important outcomes
- Rate on a scale of 1 (least critical) to 9 (critical to decision making)
- Critical and important outcomes will be discussed when making recommendations

Choosing outcomes

Generate a list of outcomes (from literature, from the panel, from clinical experts, from patient groups)

Ask panel to rate the outcomes by importance (anonymous)

Calculate the mean or median rating for each outcome (between 1-9)

Identify outcomes with large variability in rating, discuss these with panel

OUTCOMES	IMPORTANCE
resource use (including cost, human resources and length of stay)	8.00
CIN 2-3	7.68
cervical carcinoma incidence	7.53
acceptability to women (e.g. satisfaction with process or provider, DOES NOT include incidence of adverse events)	7.53
referrals after cryotherapy for complications or follow-up treatment	7.53
acceptability to providers	7.42
HIV transmission (HIV acquisition, HIV shedding)	7.42
all severe adverse events (i.e. a composite outcome including major bleeding, major infections, etc.)	7.37
major infection (requiring hospital admission and antibiotics)	7.05
major bleeding (requiring hospital admission or blood transmission)	7.00
Mortality	6.53
Fertility (e.g. conception)	5.95
CIN (1 or 2-3)	5.58
Spontaneous abortion	5.47
pain (requiring local treatment)	5.11
Maternal morbidity	4.95
minor infection (requiring outpatient treatment only)	4.26

Answer Options	1 (least important)	2	3	4	5	6	7	8	9 (critical)	Rating Average
cervical carcinoma incidence	0	1	0	1	2	1	1	3	10	7.53
CIN (1 or 2-3)	2	0	4	1	3	1	2	2	2	3.98

Choosing outcomes

Generate a list of outcomes (from literature, from the panel, from clinical experts, from patient groups)

Ask panel to rate the outcomes by importance (anonymous)

Calculate the mean or median rating for each outcome (between 1-9)

Identify outcomes with large variability in rating, discuss these with panel

Obtain agreement on rating of outcomes

Critical and important outcomes are included

Risks of not ranking outcomes

Confusion: guideline panels typically cannot balance more than 5 to 7 outcomes

++ evidence retrieval of outcomes that will not be important in decision making (e.g. minor side effects)

Surprises and risk for conflicts of interest to occur

A sensible health care question

Population:	**In infants,**
Intervention:	**what is the impact of using prebiotics compared with**
(comparison)	**not using prebiotics on**
Outcomes:	**eczema (general), allergic rhinitis, asthma, food allergy, any allergy, adverse events, nutritional status**

PICO

Summary of findings:

Prebiotics compared to no prebiotics for prevention of allergies

Patient or population: prevention of allergies (160315)
Setting: ambulatory care
Intervention: prebiotics
Comparison: no prebiotics

Outcome № of participants (studies)	Relative effect (95% CI)	Anticipated absolute effects (95% CI)			Quality	What happens /comments
		Without prebiotics	**With prebiotics**	**Difference**		
Eczema (general) assessed with: clinical criteria and/or parent definition follow up: range 3 to 24 months to № of participants: 2030 (6 RCTs)	**RR 0.68** (0.40 to 1.15)	18.7%	**12.7%** (7.5 to 21.5)	**6.0% fewer** (11.2 fewer to 2.8 more)	⨁⨁ LOW [a,b,c,d]	
Allergic rhinitis assessed with: clinical criteria № of participants: (0 studies)	not estimable	0.0%	**0.0%** (0.0 to 0.0)	**0.0% fewer** (0 fewer to 0 fewer)	- [d]	None of the studies assessed allergic rhinitis symptoms as an outcome
Asthma assessed with: "recurrent wheezing", or self-reported or proxy reported asthma follow up: range 18 to 24 months to № of participants: 249 (2 RCTs)	**RR 0.37** (0.17 to 0.80)	17.4%	**6.4%** (3.0 to 13.9)	**10.9% fewer** (14.4 fewer to 3.5 fewer)	⨁ VERY LOW [c,e,f]	

[Intervention Review]

Hip protectors for preventing hip fractures in older people

Nancy Santesso[1], Alonso Carrasco-Labra[2], Romina Brignardello-Petersen[2]

[1]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada. [2]Evidence Based Dentistry Unit, Faculty of Dentistry, University of Chile, Santiago, Chile

Contact address: Nancy Santesso, Department of Clinical Epidemiology and Biostatistics, McMaster University, 1200 Main Street West, Hamilton, Ontario, L8N 3Z5, Canada. santesna@mcmaster.ca.

Editorial group: Cochrane Bone, Joint and Muscle Trauma Group.
Publication status and date: New search for studies and content updated (conclusions changed), published in Issue 3, 2014.
Review content assessed as up-to-date: 18 June 2013.

Citation: Santesso N, Carrasco-Labra A, Brignardello-Petersen R. Hip protectors for preventing hip fractures in older people. *Cochrane Database of Systematic Reviews* 2014, Issue 3. Art. No.: CD001255. DOI: 10.1002/14651858.CD001255.pub5.

Systematic review for group work

What are the effects of wearing hip protectors for older adults living in institutional settings?

Soft hip protector

Hard hip protector

What is the impact of wearing hip protectors for older adults (living in institutional settings)?

Outcomes:

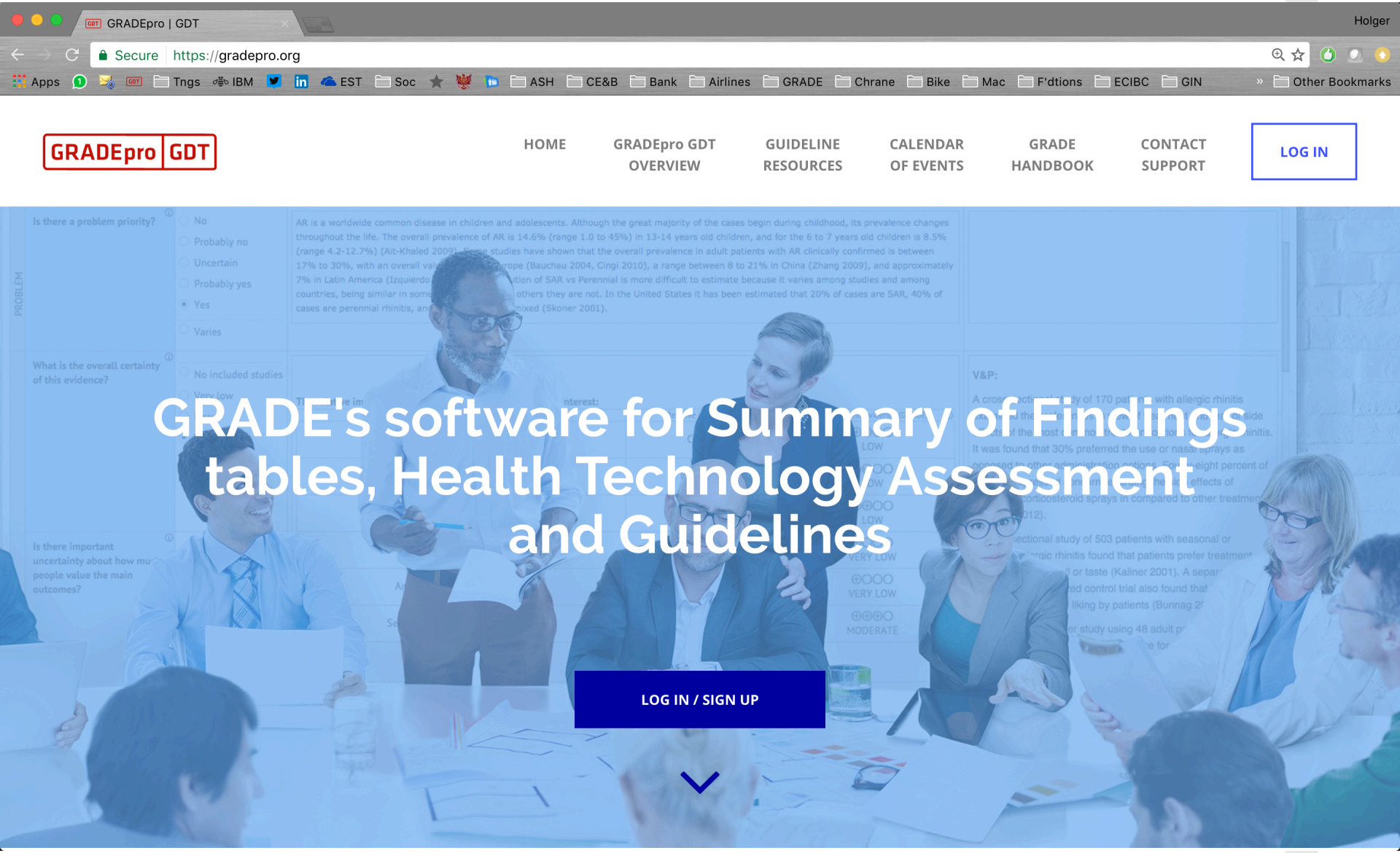
…

Introduction to GRADEpro

Creating a project

Inserting your question

Importing


GRADEpro | GDT
Holger
Secure https://gradepro.org
Apps Tngs IBM EST Soc ASH CE&B Bank Airlines GRADE Chrane Bike Mac F'dtions ECIBC GIN Other Bookmarks
GRADEpro GDT
HOME
GRADEpro GDT OVERVIEW
GUIDELINE RESOURCES
CALENDAR OF EVENTS
GRADE HANDBOOK
CONTACT SUPPORT
LOG IN
GRADE's software for Summary of Findings tables, Health Technology Assessment and Guidelines
LOG IN / SIGN UP

A systematic review of the literature on the treatment of pityriasis rubra pilaris type 1 with TNF-antagonists

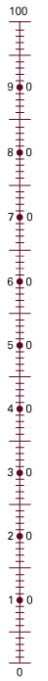
G. Petrof,* N. Almaani, C.B. Archer

St John's Institute of Dermatology,

*Correspondence: G. Petrof. E-

How confident are you in the estimates of effect?

About here?



← 100% confident

About here?

← 0% confident

Abstract

Background Adult pityrias
and histological parallels with pso
tumour necrosis factor (TNF) antagonists.

Objectives Our objective was to systematically review the literature for eviden
the treatment of adult PRP.

Methods We performed a systematic search of the Cochrane library, EMBASE
We defined diagnosis of PRP, classified clinical response and whether th
antagonists. We also reviewed disease, treatment duration and follow up.

Results Sixteen articles were selected for detailed review. From these, 12 art
criteria and were included in the systematic review. The authors identified tw
archive. A total of 15 evaluable cases were included for analysis. Twelve show
TNF-antagonists with a mean time to maximal response of 5 months. In 10 of t
attributable to TNF antagonist therapy.

Conclusion These data indicate that TNF-antagonists may be of value in treating adult type 1 PRP refractory to other systemic agents but selective reporting bias, together with the lack of standard diagnostic criteria and established spontaneous resolution in PRP, prevent any firm recommendations on their place in management.

Received: 10 November 2011; Accepted: 13 January 2012

GRADE

- **Certainty of evidence**
  - Involves assessing evidence transparently
  - Confidence in an estimate of effect, association?
  - Starts with single studies
  - Ends with a body of evidence by outcome and a recommendation
- **Recommendations**
  - Involves making judgments and decisions transparent, rating evidence
  - Evidence to Decision (EtD) frameworks
  - Comprehensive list of criteria that influence a recommendation
  - Clearly developed & formulated action message
    - (strong or conditional/weak recommendations for or against an option)

Formulate question

Select outcomes

Rate importance

Outcomes across studies

P
I/E
C
O

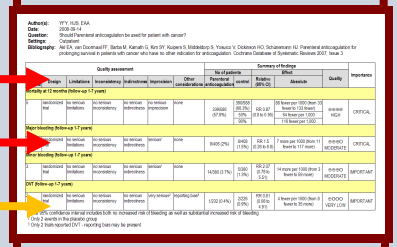
Outcome Critical

Outcome Critical

Outcome Important

Outcome Not important

Create **evidence profile/SoF** Table with GRADEpro



Summary of findings & estimate of effect for each outcome

Rate quality of evidence for each outcome

High

Moderate

Low

Very low

Randomization raises initial quality
RCTs: high
Observational: low

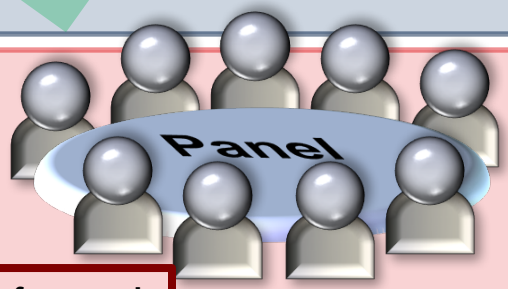
Grade down

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

Grade up

1. Large effect
2. Dose response
3. Opposing bias & Confounders

Evidence synthesis (systematic review/HTA)



Recommendation/Decision

Grade recommendations (Evidence to Recommendation)

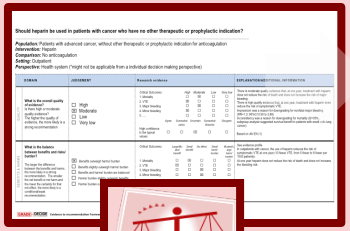
- For or against (direction) ↓↑
- Strong or conditional/weak (strength)

By considering balance of consequences (evidence to recommendations):

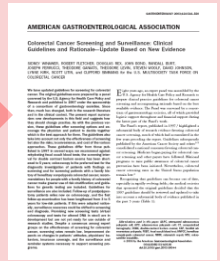
- ❑ Quality of evidence
- ❑ Balance benefits/harms
- ❑ Values and preferences (equity)
- ❑ Resource use (cost, feasibility)
- ❑ Acceptability

EtD framwork

GRADEpro GDT



Guideline



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

Formulate Recommendations (↓↑ | ⊕…)

"The panel recommends that ….should..."

"The panel suggests that ….should..."

"The panel suggests to **not** ..."

"The panel recommends to **not**..."

Transparency, clear, actionable

Assessing the certainty (quality) of evidence

Drawing conclusions about the certainty of the evidence?

Are the studies well done? **Risk of bias**

Are the results consistent across studies? **Inconsistency**

How directly relate the results to my question? **Indirectness**

Is this effect size precise? **Imprecision**

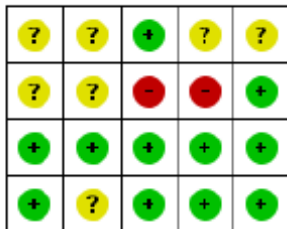
Are these all of the studies? **Publication bias**

Plus additional factors for observational studies

Dose response, size of effect, confounding

When interpreting results, consider…

Risk of bias

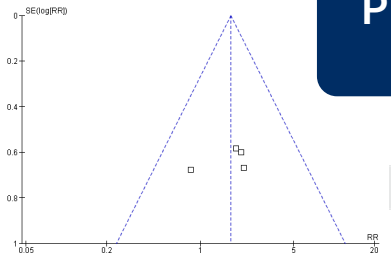


Agnelli 2009

Methods	Randomized clinical trial
Participants	1150 patients with metastatic or locally advanced lung, breast, gastrointestinal (stomach, colon-rectum, pancreas), ovarian or head and neck cancer undergoing chemotherapy.
Interventions	Subcutaneous low molecular weight heparin (nadroparin calcium, one injection/day) vs. placebo for the overall duration of chemotherapy or up to a maximum of 4 months
Outcomes	• Survival, (4 months and 12 months follow-up) • Response to chemotherapy (4 months follow-up) • For patients with central venous catheters (CVC), complications of possible thrombotic origin, such as malfunction or requirement of CVC removal (4 months follow-up) • Superficial thrombophlebitis of lower limbs (4 months follow-up) • Asymptomatic thromboembolic events diagnosed during tests performed for other purposes (4 months follow-up) • Safety (major bleeding, minor bleeding, other adverse events) (4 months follow-up)

Indirectness

Publication bias



Study or Subgroup	Caffeinated coffee Events	Total	Decaffeinated coffee Events	Total	Weight	Risk Ratio M, Fixed, 95% CI
Amore-Coffea 2000	2	31	10	34	6.6%	0.22 [0.05, 0.92]
Deliciozza 2004	10	40	9	40	21.9%	1.11 [0.51, 2.44]
Mama-Kaffa 1999	12	53	9	61	22.2%	1.53 [0.70, 3.35]
Morrocona 1998	3	15	1	17	2.9%	3.40 [0.39, 29.31]
Norscafe 1998	19	68	9	64	26.4%	1.99 [0.97, 4.07]
Oohlahlazza 1998	4	35	2	37	5.1%	2.11 [0.41, 10.83]
Piazza-Allerta 2003	8	35	6	37	14.9%	1.41 [0.54, 3.65]
Total (95% CI)		**277**		**290**	**100.0%**	**1.38 [0.96, 2.00]**
Total events	58		46			
Heterogeneity: Chi² = 8.58, df = 6 (P = 0.20); I² = 30%						
Test for overall effect: Z = 1.73 (P = 0.08)						

Risk Ratio M, Fixed, 95% CI

0.02 0.1 1 10 50

…rs caffeine Favours decaf

Imprecision

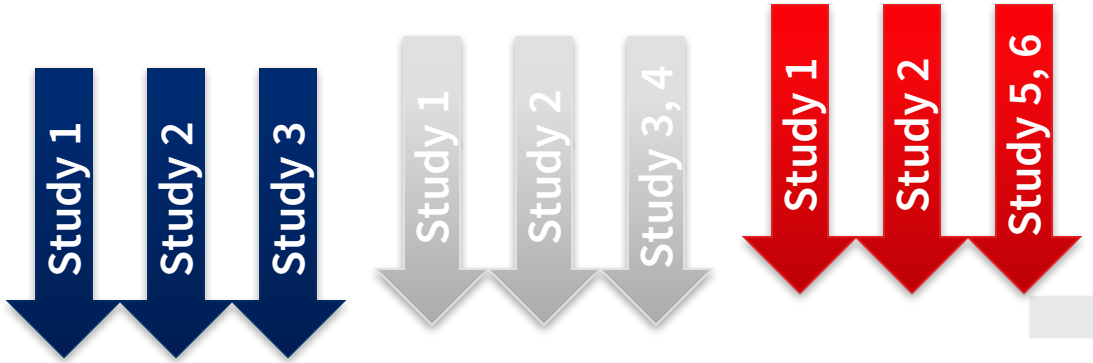
Inconsistency

Effect size and direction

GRADE criteria	Rating (circle one)	Footnotes (explain reasons for downgrading)	Quality of the evidence (Circle one)
Outcome:			
Risk of Bias *(use the Risk of Bias tables and figures)*	No serious (-1) very serious (-2)		⊕⊕⊕⊕ High ⊕⊕⊕◯ Moderate ⊕⊕◯◯ Low ⊕◯◯◯ Very Low
Inconsistency	No serious (-1) very serious (-2)		
Indirectness	No serious (-1) very serious (-2)		
Imprecision	No serious (-1) very serious (-2)		
Publication Bias	Undetected Strongly suspected (-1)		
Other (upgrading factors, circle all that apply)	Large effect (+1 or +2) Dose response (+1 or +2) No Plausible confounding (+1 or +2)		

1. Design and Execution/Risk of Bias

Limitation in observational studies	Explanations
Failure to develop and apply appropriate eligibility criteria (inclusion of control population)	• under- or over-matching in case-control studies • selection of exposed and unexposed in cohort studies from different populations
Flawed measurement of both exposure and outcome	• differences in measurement of exposure (e.g. recall bias in case-control studies) • differential surveillance for outcome in exposed and unexposed in cohort studies
Failure to adequately control confounding	• failure of accurate measurement of all known prognostic factors • failure to match for prognostic factors and/or adjustment in statistical analysis
Incomplete or inadequately short follow-up	



Domain	Outcome 1 (e.g. mortality)	Outcome 2 (e.g. VTE)	Outcome 3 (e.g. bleeding)	
Bias due to confounding	Low risk	Low risk	Low risk	Pre or at intervention
Bias in selection of participants into the study	Low risk	Low risk	Low risk	
Bias in classification of interventions	Moderate risk	Moderate risk	Moderate risk	
Bias due to deviations from intended interventions	Low risk	Low risk	Low risk	Post intervention
Bias due to missing data	Low risk	Serious risk	Serious risk	
Bias in measurement of outcomes	Low risk	Moderate risk	Moderate risk	
Bias in selection of reported results	Low risk	Low risk	Low risk	
Overall bias	Low risk	Moderate risk	Moderate risk	

Determinants of certainty in a body of evidence: GRADE

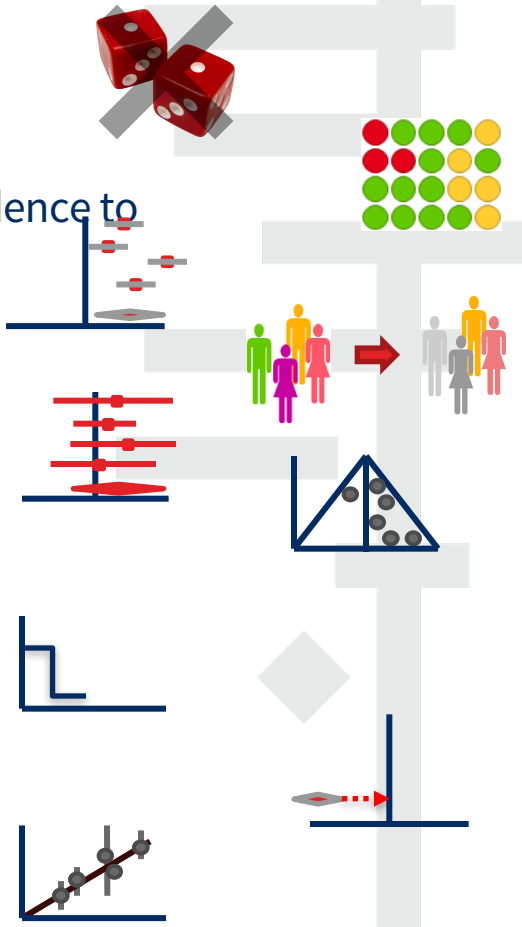
A body of evidence starts as: high | ⊕⊕⊕⊕

5 factors that can lower quality

1. Risk of bias criteria
   - Lack of randomization (observational studies) lowers confidence to low
2. Inconsistency *(or heterogeneity)*
3. Indirectness *(PICO and applicability)*
4. Imprecision
5. Publication bias

3 factors can increase quality

1. large magnitude of effect
2. opposing plausible residual bias or confounding
3. dose-response gradient



Suggested approach

1. Until more experience with ROBINS-I and its use in GRADE is available, those assessing RoB in NRS use a default approach of downgrading evidence from high-certainty to low-certainty due to risk of bias (i.e., as part of risk of bias domain in GRADE) as a result of lack of randomization.

Suggested approach

2. Not downgrading from high to low-certainty, that is by two levels, requires transparent and detailed justification. This should be done for the items of the ROBINS-I tool

- note to avoid confusion that ROBINS-I calls the seven items in the GRADE RoB domain, domains themselves.
- **E.g. residual bias and confounding applies**

1. RCTs Design and Execution/Risk of Bias

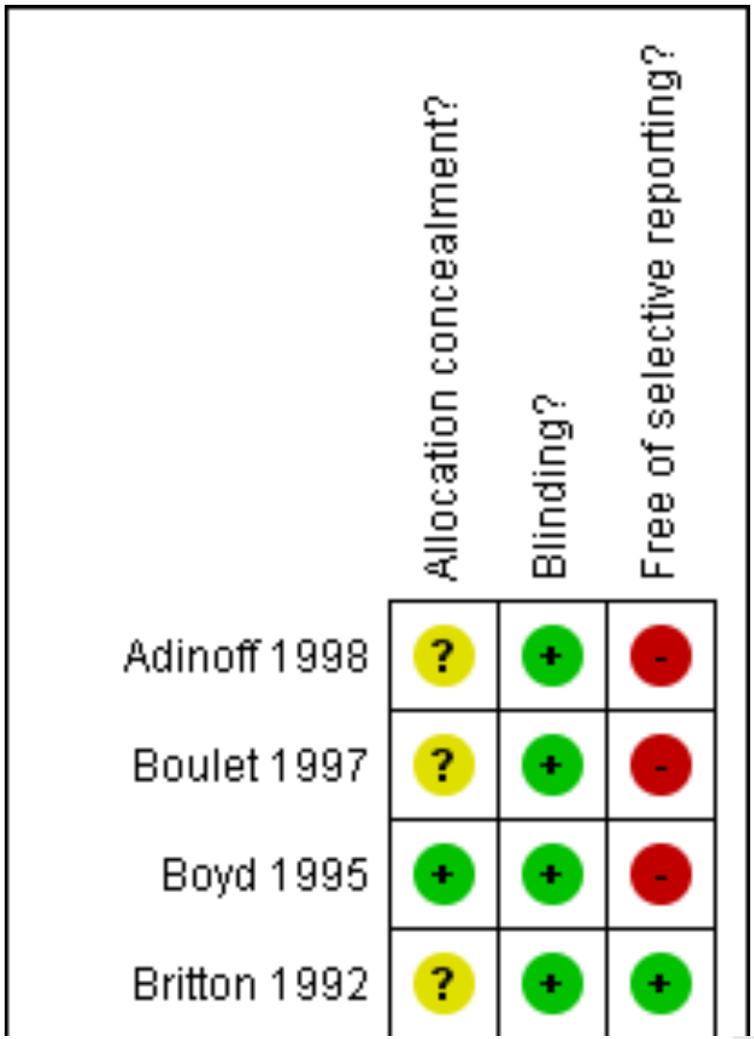
Limitations in RCTs
lack of concealment
intention to treat principle violated
inadequate blinding
loss to follow-up
early stopping for benefit
selective outcome reporting

Design and Execution/RoB

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

Cates CJ, Cates MJ

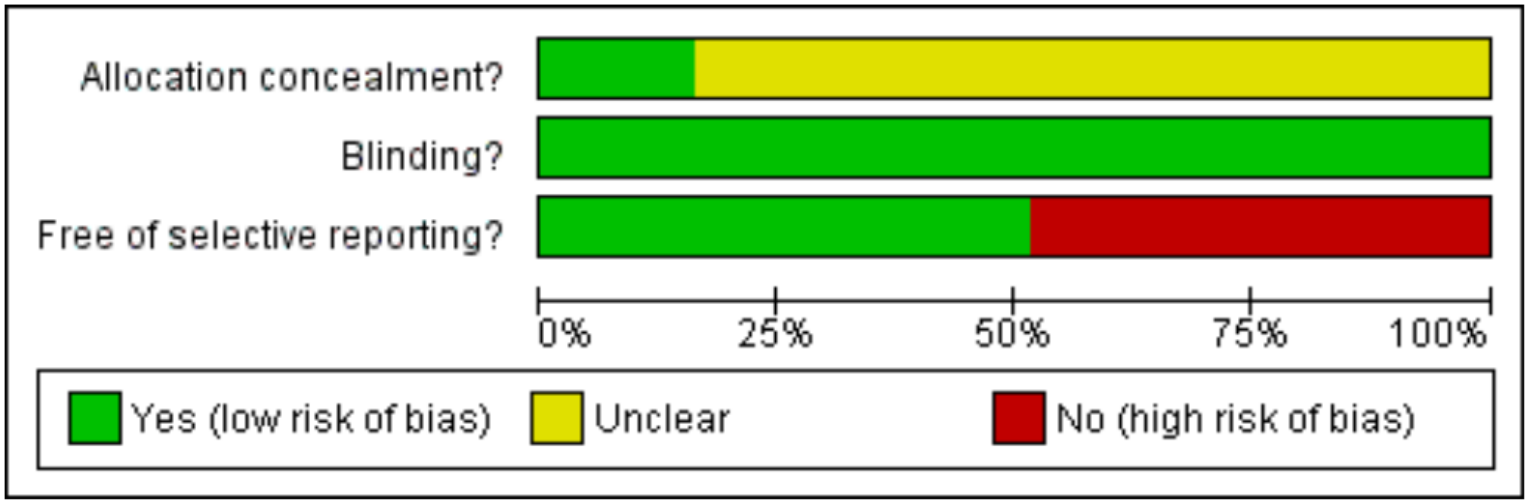
Figure 4. Risk of bias summary: review authors' judgments about each risk of bias item for each included study.



From Cates , CDSR 2008

Design and Execution/RoB

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?

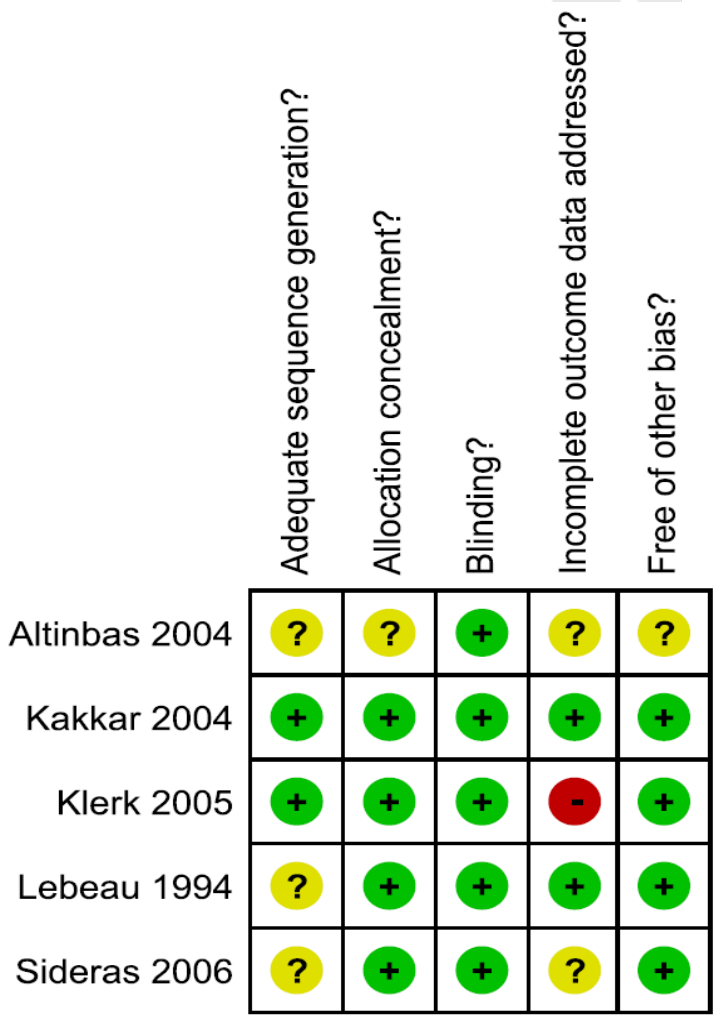
Yes

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

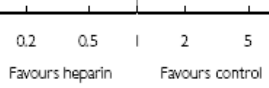
Analysis 01.01. Comparison 01 Heparin vs placebo, Outcome 01 Mortality over duration of study

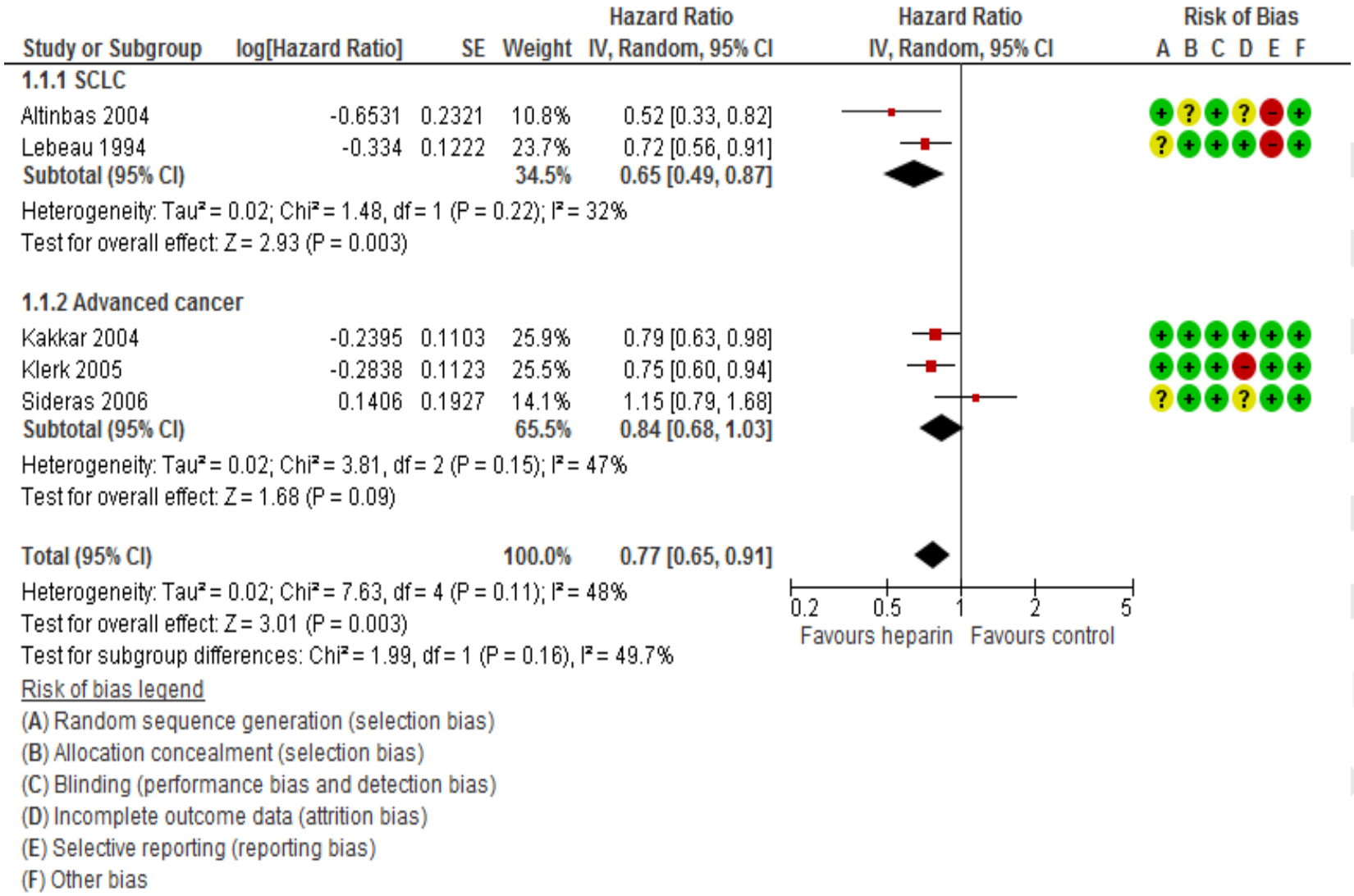
Review: Parenteral anticoagulation for prolonging survival in patients with cancer who have no other indication for anticoagulation

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						





Risk of bias legend

(A) Random sequence generation (selection bias)

(B) Allocation concealment (selection bias)

(C) Blinding (performance bias and detection bias)

(D) Incomplete outcome data (attrition bias)

(E) Selective reporting (reporting bias)

(F) Other bias

Is the risk of bias…

- Not serious
- Serious
- Very Serious

The majority of studies had allocation concealment , and used blinded outcome and adjudication. We did not downgrade although there was some concern about lack of blinding in some studies; the overall risk of bias was felt to be very low.

…that my confidence in the result is reduced?

What if…

Not possible to blind?

Couldn’t pool results?

Abstracts or little information about risk of bias?

2. Imprecision

Small sample size

- small number of events

Wide confidence intervals

- uncertainty about magnitude of effect

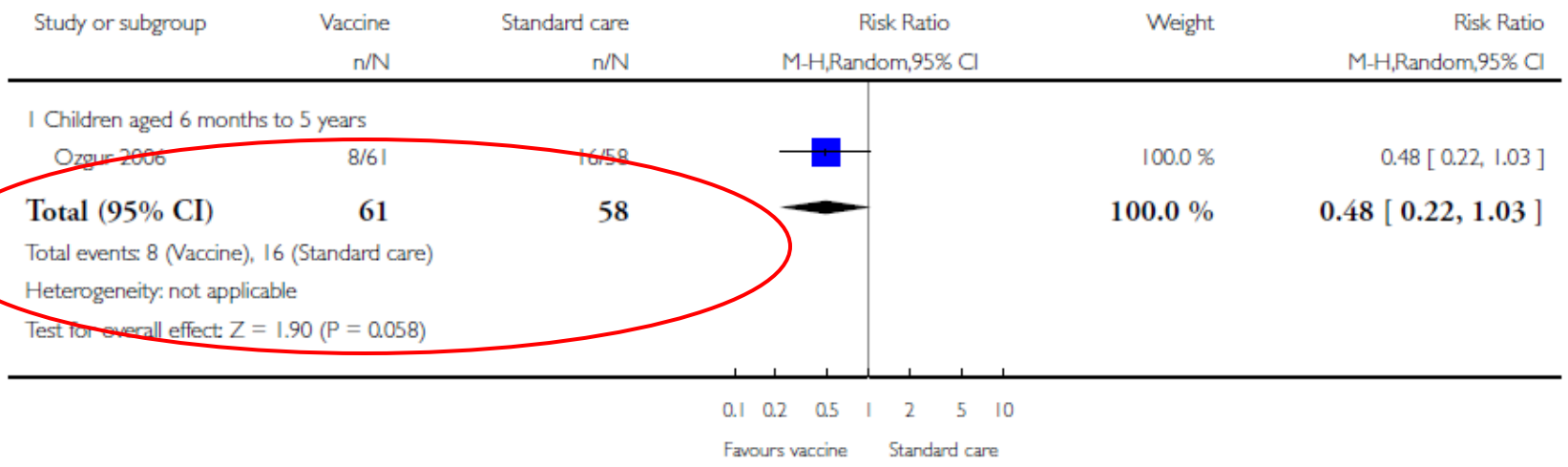
Example: Immunization in children

Analysis 4.3. Comparison 4 Inactivated vaccines - (cohort studies by age group), Outcome 3 Otitis media.

Review: Vaccines for preventing influenza in healthy children

Comparison: 4 Inactivated vaccines - (cohort studies by age group)

Outcome: 3 Otitis media



Citation: Jefferson T, Rivetti A, Harnden A, Di Pietrantonj C, Demicheli V. Vaccines for preventing influenza in healthy children. *Cochrane Database of Systematic Reviews* 2008, Issue 2. Art. No.: CD004879. DOI: 10.1002/14651858.CD004879.pub3.

Is the imprecision…

- Not serious
- Serious
- Very Serious

…that is confidence/certainty in the result is reduced?

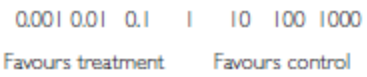
Analysis 6.1. Comparison 6 Inactivated vaccine versus placebo (RCTs), Outcome 1 Influenza.

Review: Vaccines for preventing influenza in healthy children

Comparison: 6 Inactivated vaccine versus placebo (RCTs)

Outcome: 1 Influenza

Study or subgroup	Vaccine n/N	Control n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
1 Inactivated vaccines (one dose)					
Beutner 1979a	28/300	82/275		41.8 %	0.31 [0.21, 0.47]
Clover 1991	9/54	36/82		16.6 %	0.38 [0.20, 0.72]
Gruber 1990	10/54	37/77		18.7 %	0.39 [0.21, 0.71]
Hoberman 2003a	15/273	22/138		17.7 %	0.34 [0.18, 0.64]
Hoberman 2003b	9/252	4/123		5.2 %	1.10 [0.35, 3.50]
Subtotal (95% CI)	**933**	**695**		**100.0 %**	**0.36 [0.28, 0.48]**
Total events: 71 (Vaccine), 181 (Control)					
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 4.13$, df = 4 (P = 0.39); $I^2 = 3\%$					
Test for overall effect: Z = 7.42 (P < 0.00001)					
2 Inactivated vaccines (two doses)					



Is the imprecision…

- Not serious
- Serious
- Very Serious

…that is confidence/certainty in the result is reduced?

Imprecision

Consider

Sample sizes and number of events

- assess according to effect size, control event rates, Optimal information size (OIS)

Width of confidence intervals

- Wide confidence intervals indicate uncertainty about the effect
- Includes null effect and appreciable benefit or harm (rule of thumb: RR<0.75 or >1.25)

Optimal information size (OIS)

if the total number of patients included in a systematic review is **less than** the number of patients generated by a conventional sample size calculation for a single adequately powered trial, consider rating down for imprecision

http://www.stat.ubc.ca/~rollin/stats/ssize/

Optimal information size implications: Consider the total number of events

Reasonable threshold for rating down for imprecision = 300 events

Total Number of Events	Relative Risk Reduction	Implications for meeting OIS threshold
100 or less	≤ 30%	Will almost never meet threshold whatever control event rate
200	30%	Will meet threshold for control event rates for ~ 25% or greater
200	25%	Will meet threshold for control event rates for ~ 50% or greater
200	20%	Will meet threshold only for control event rates for ~ 80% or greater
300	≥ 30%	Will meet threshold
300	25%	Will meet threshold for control event rates ~ 25% or greater
300	20%	Will meet threshold for control event rates ~ 60% or greater
400 or more	≥ 25%	Will meet threshold for any control event rate
400 or more	20%	Will meet threshold for control event rates of ~ 40% or greater

Rules of thumb

Dichotomous outcomes

300 events

Continuous outcomes

400 people providing outcome measures

3. Inconsistency of results (Heterogeneity)

if inconsistency, look for explanation

- patients, intervention, comparator, outcome

if unexplained inconsistency lower quality

Reminders for immunization uptake

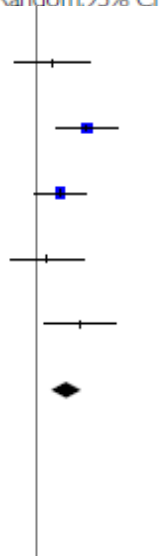
Analysis 2.1. Comparison 2 letter reminders vs. control, Outcome 1 Immunized.

Review: Patient reminder and recall systems to improve immunization rates

Comparison: 2 letter reminders vs. control

Outcome: 1 Immunized

Study or subgroup	Letter reminders n/N	Control n/N	Odds Ratio M-H,Random,95% CI	Odds Ratio M-H,Random,95% CI
2 Preschool-child				
Campbell1994T87	54/87	59/105		1.28 [0.71, 2.28]
Lieu1997T69	82/153	47/136		2.19 [1.36, 3.52]
Lieu1998T82	72/162	78/219		1.45 [0.95, 2.19]
Oeffinger1992T27	33/116	31/122		1.17 [0.66, 2.07]
Young1980T63	51/106	34/105		1.94 [1.11, 3.39]
Subtotal (95% CI)	**624**	**687**		**1.58 [1.26, 1.99]**



Total events: 292 (Letter reminders), 249 (Control)

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 4.08$, $df = 4$ ($P = 0.40$); $I^2 = 2\%$

Test for overall effect: $Z = 3.92$ ($P = 0.000088$)

Citation: Jacobson Vann JC, Szilagyi P. Patient reminder and recall systems to improve immunization rates. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art. No.: CD003941. DOI: 10.1002/14651858.CD003941.pub2.

Would you downgrade for inconsistency?

- [ ] No, there **is no serious inconsistency**
- [ ] Yes, there is serious inconsistency
- [ ] Yes, there is very serious inconsistency

Analysis 6.1. Comparison 6 patient & provider reminder vs. control, Outcome 1 Immunized.

Review: Patient reminder and recall systems to improve immunization rates

Comparison: 6 patient % provider reminder vs. control

Outcome: 1 Immunized

Study or subgroup	Patient % Provider R n/N	Control n/N	Odds Ratio M-H,Random,95% CI	Weight	Odds Ratio M-H,Random,95% CI
2 Preschool-child					
Rodewald1999T95	616/648	532/719		30.0 %	6.77 [4.57, 10.02]
Soljak1987T35	539/709	382/613		31.2 %	1.92 [1.51, 2.43]
Subtotal (95% CI)	**1357**	**1332**		**61.1 %**	**3.57 [1.03, 12.41]**
Total events: 1155 (Patient % Provider R), 914 (Control)					
Heterogeneity: $Tau^2 = 0.78$; $Chi^2 = 29.55$, df = 1 ($P<0.00001$); $I^2 = 97\%$					
Test for overall effect: Z = 2.00 (P = 0.046)					
3 Influenza-adult					

Citation: Jacobson Vann JC, Szilagyi P. Patient reminder and recall systems to improve immunization rates. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art. No.: CD003941. DOI: 10.1002/14651858.CD003941.pub2.

Would you downgrade for inconsistency?

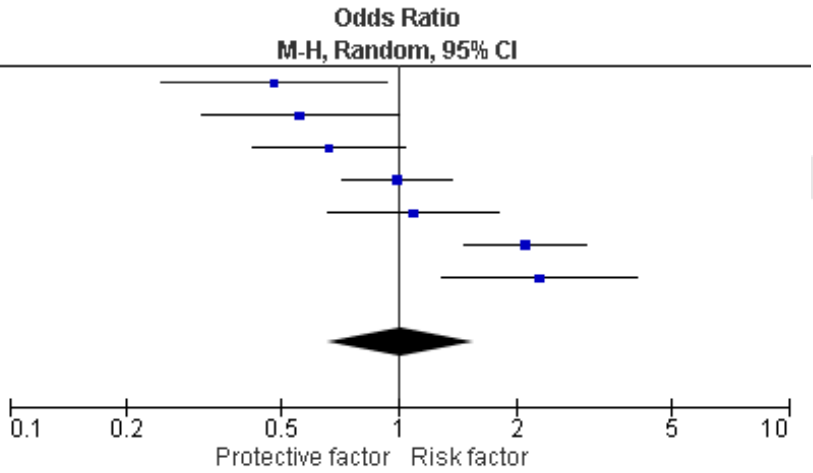
- [ ] No, there **is no serious inconsistency**
- [ ] Yes, there is serious inconsistency
- [ ] Yes, there is very serious inconsistency

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: $Tau^2 = 0.28$; $Chi^2 = 35.73$, df = 6 ($P < 0.00001$); $I^2 = 83\%$

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.

Would you downgrade for inconsistency?

- [ ] No, there **is no serious inconsistency**
- [ ] Yes, there is serious inconsistency
- [ ] Yes, there is very serious inconsistency

Inconsistency

I^2

P-value

Overlap in CI

Difference in point estimates

4. Directness of Evidence generalizability, transferability, applicability

differences in

- populations/patients (HIC – L/MIC)
- interventions (new anticoagulants - warfarin)
- comparator appropriate (newer antibx – old)
- outcomes (important – surrogate; signs and symptoms – mortality)

indirect comparisons

- interested in A versus B
- have A versus C and B versus C

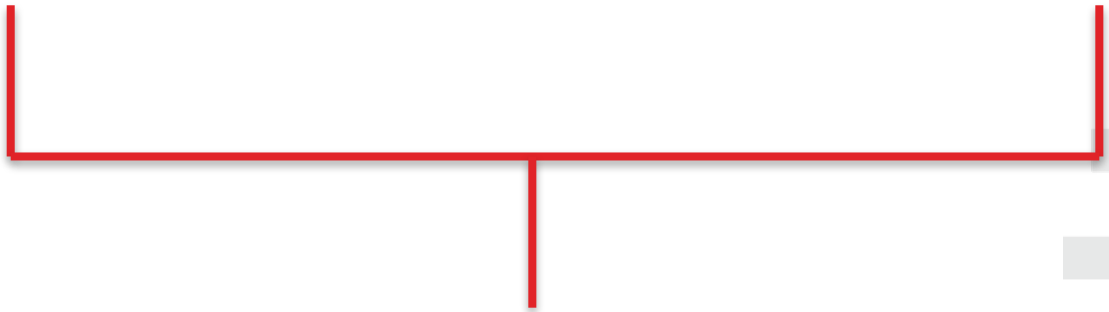
Indirect evidence from RCTs

A versus C

OR 0.70 (95% CI 0.23 to 2.11

C versus B

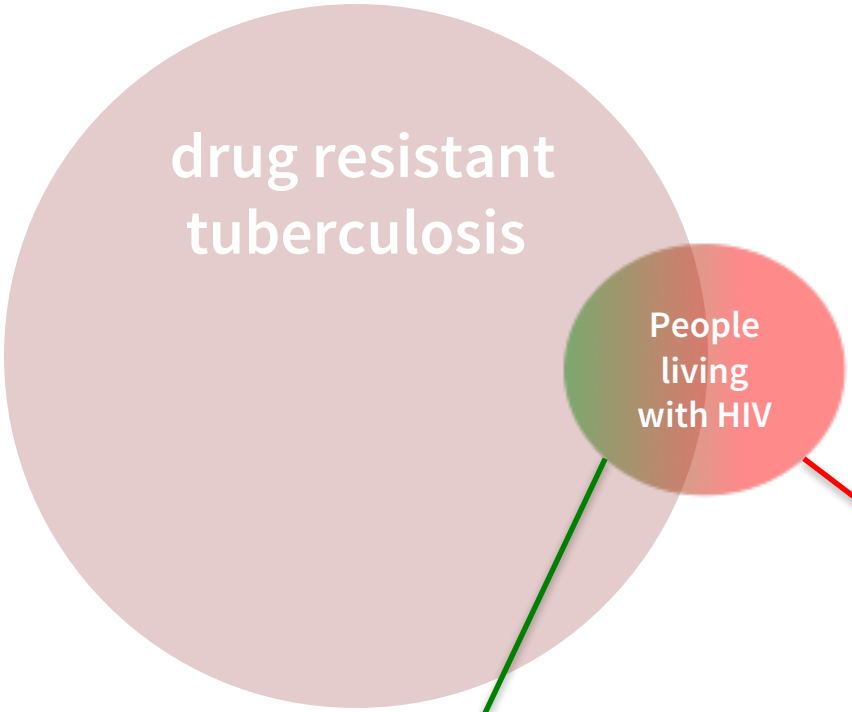
OR 1.32 (95%CI 0.37 to 4.79)



OR 0.93 (95% CI, 0.17 to 5.03)

A versus B

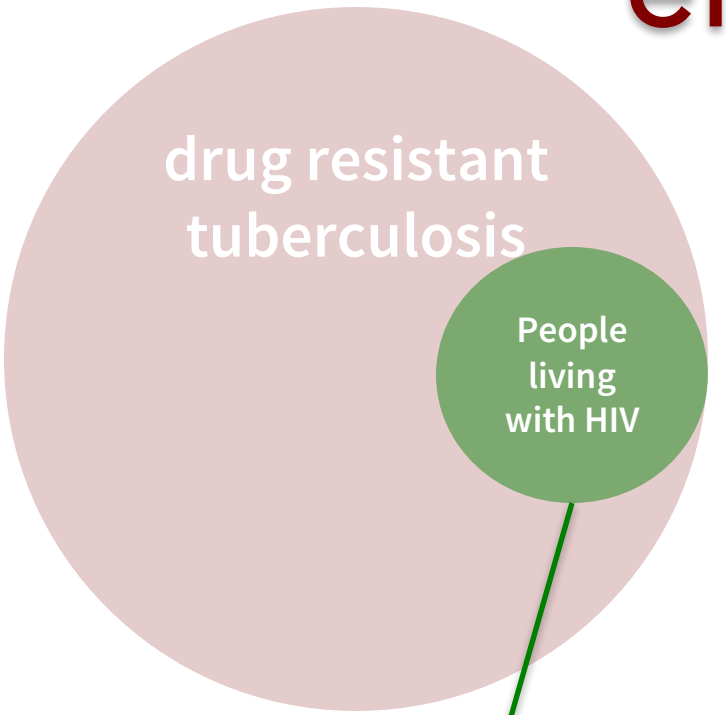
Indirectness - population



No concerns about directness (applicability)
No downgrading
Same recommendation

Concerns about directness
Downgrade ↓
Separate recommendation

Indirectness – different effects?



No concerns about directness (applicability)
No downgrading
Same recommendation

Concerns about directness
Downgrade ↓
Separate recommendation

Example: When is evidence indirect for population?

A systematic review asks

'what are the effects of self management programmes in people with asthma?'

Consider these separate scenarios for 'quality of life'…

1. 5/6 studies include people with severe asthma
2.
3.

Do the results apply directly to people with mild asthma?

Example: When is evidence indirect for population?

A systematic review asks

'what are the effects of self management programmes in people with asthma?'

Consider these separate scenarios for 'quality of life'…

1.
2. 4/6 studies are in children ages 1-15
3.

Do the results apply directly adults?

Example: When is evidence indirect for population?

A systematic review asks

'what are the effects of self management programmes in people with COPD (chronic obstructive pulmonary disease)?'

Consider these separate scenarios for 'quality of life'…

1.
2.
3. All studies are in men

Do the results apply directly to women as well?

Surrogate outcomes

Examples of surrogate outcomes

Symptomatic pulmonary PE versus severe PE

2 or 3 months smear for TB – cure

Cholesterol – cardiovascular disease

Bone mineral density - Risk of fracture

Calcium phosphate levels – coronary artery disease

HIV viral load – morbidity

Tumour size – survival

Reduction in air pollution – morbidity or respiratory disease

Making judgments when faced with surrogate outcomes

Is there a strong association between the surrogate and the patient important outcome?

Can you present the patient important outcomes instead of the surrogates?

Note: it is still important to indicate when there is no data for patient important outcomes