NIPH
Norwegian Institute of Public Health

GRADE

Assessing the confidence in the evidence

Eva Denison, senior researcher

1

Agenda

- 09.00-10.00
  - Presentation of GRADE
  - We use data from Kaufman 2018 as examples
- 10.15-11.00
  - GRADE exercise in break-out groups
  - You use data from another SR
- 11.15-12.00
  - GRADE evidence tables and Summary of findings tables
  - We complete the groups' assessments

2

Steps in conducting a SR

1.Formulate the question

2.Define criteria for inclusion- and exclusion

3.Identify (locate) studies

4.Select studies

5.Assess methodological quality of studies (bias)

6.Extract data

7.Analyse data

8.GRADE

9.Present and interpret results

NIPH -

3

What is GRADE?

https://gdt.gradepro.org/app/handbook/handbook.html#h.svwngs6pm0f2

- The GRADE approach is a system for rating the quality of a body of evidence in systematic reviews and other evidence syntheses, such as health technology assessments, and guidelines and grading recommendations in health care.
- GRADE offers a transparent and structured process for developing and presenting evidence summaries and for carrying out the steps involved in developing recommendations.
- It can be used to develop clinical practice guidelines (CPG) and other health care recommendations (e.g. in public health, health policy and systems and coverage decisions).

NIPH -

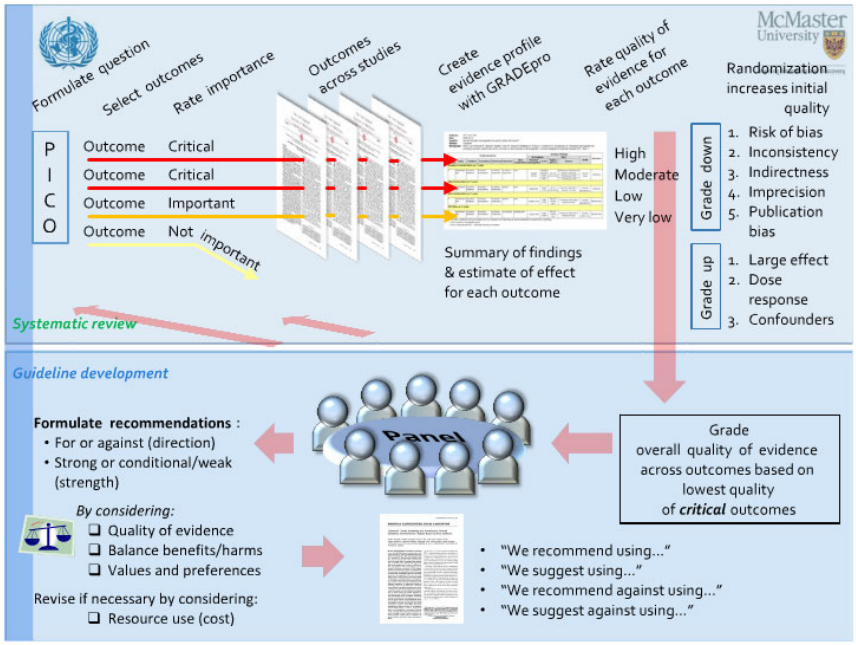
4

GRADE in systematic reviews

https://gdt.gradepro.org/app/handbook/handbook.html#h.svwngs6pm0f2

- Systematic reviews should provide a comprehensive summary of the evidence but they should typically not include health care recommendations.
- Use of the GRADE approach by systematic review authors terminates after rating the quality of evidence for outcomes and clearly presenting the results in an evidence table, i.e. an GRADE Evidence Profile or a Summary of Findings table.

5



6

Grading of Recommendations
Assessment,
Development and Evaluation

http://www.gradeworkinggroup.org/

Table: GRADE's approach to rating quality of evidence (aka confidence in effect estimates)

For each outcome based on a systematic review and across outcomes (lowest quality across the outcomes critical for decision making)

1. Establish initial level of confidence		2. Consider lowering or raising level of confidence		3. Final level of confidence rating
Study design	*Initial confidence in an estimate of effect*	*Reasons for considering lowering or raising confidence*		*Confidence in an estimate of effect across those considerations*
		↓ Lower if	↑ Higher if*	
Randomized trials →	High confidence	Risk of Bias Inconsistency Indirectness Imprecision Publication bias	Large effect Dose response All plausible confounding & bias • would reduce a demonstrated effect or • would suggest a spurious effect if no effect was observed	High ⊕⊕⊕⊕
				Moderate ⊕⊕⊕◯
Observational studies →	Low confidence			Low ⊕⊕◯◯
				Very low ⊕◯◯◯

*upgrading criteria are usually applicable to observational studies only.

7

Reasons for lowering the level of confidence

One level = serious: -1 Two levels = very serious: -2

- Study limitations/risk of bias
  - In relation to preferred study design for your question, e.g. RCT, cohort, cross-sectional
- Indirectness
  - Eligibility criteria: PIO/PICO/PEO/PECO
- Inconsistency/heterogeneity
  - Widely differing estimates: direction of effect, non-overlap of confidence intervals (I-square, Chi-square, Tau)
- Imprecision
  - Few participants, few events, wide confidence intervals
- Publication bias
  - Due to decisions by authors, peer reviewers, editors

8

GRADE assesment of **study limitations** across studies

Risk of bias	Across studies	Interpretation	Considerations	GRADE assessment of study limitations
Low risk of bias.	Most information is from studies at low risk of bias.	Plausible bias unlikely to seriously alter the results.	No apparent limitations.	No serious limitations, do not downgrade.
Unclear risk of bias.	Most information is from studies at low or unclear risk of bias.	Plausible bias that raises some doubt about the results.	Potential limitations are unlikely to lower confidence in the estimate of effect.	No serious limitations, do not downgrade.
			Potential limitations are likely to lower confidence in the estimate of effect.	Serious limitations, downgrade one level.
High risk of bias.	The proportion of information from studies at high risk of bias is sufficient to affect the interpretation of results.	Plausible bias that seriously weakens confidence in the results.	Crucial limitation for one criterion, or some limitations for multiple criteria, sufficient to lower confidence in the estimate of effect.	Serious limitations, downgrade one level.
			Crucial limitation for one or more criteria sufficient to substantially lower confidence in the estimate of effect.	Very serious limitations, downgrade two levels.

9

RoB Kaufman 2018

Figure 2. Risk of bias summary: review authors' judgements about each risk of bias item for each included study

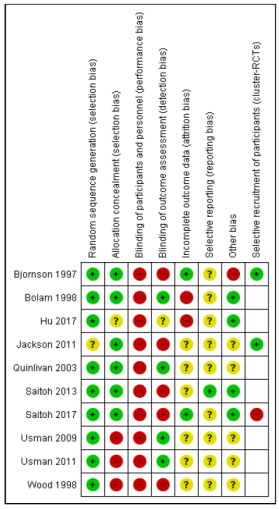
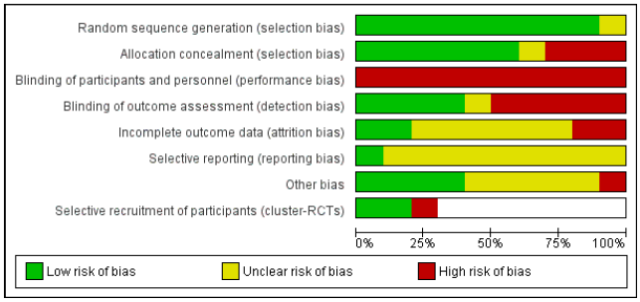


Figure 3. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies



10

Indirectness Kaufman 2018

- Inclusion criteria
- Types of studies
  - RCT; cluster RCT
- Types of participants
  - Children
  - Parents
  - Vaccine program organizers
- Types of interventions
  - Face-to-face communication interventions directed to parents to inform or educate them about routine childhood vaccinations.
- Types of outcomes
  - Vaccination status

- Included studies
  - 7 RTC, 3 cluster RCT
- Participants
  - Mothers and parents, or expectant mothers or parents
- Interventions
  - All studies assessed face-to-face interventions
  - Variations in intensity, content, length, control group intervention
- Outcomes
  - Vaccination status measures in 9/10 studies
  - Final/only measurement time point for most studies was 3 months

11

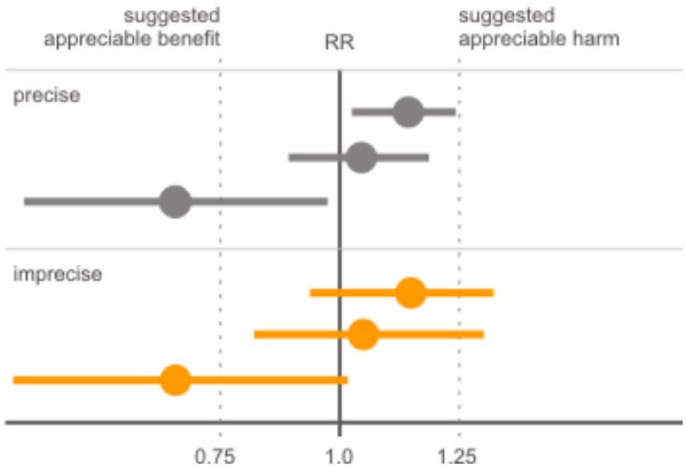
Inconsistency Kaufman 2018

Analysis 1.1. Comparison 1 Face-to-face education versus control or non-face-to-face education (all), Outcome 1 Vaccination status (stratified by length).

Study or subgroup	face-to-face n/N	control (all) n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
1 Short (1 to 10 min)					
Saitoh 2013 (1)	24/70	3/36		1.4 %	4.11 [1.33, 12.75]
Saitoh 2017 (2)	22/51	21/45		6.8 %	0.92 [0.59, 1.44]
Usman 2009 (3)	242/375	205/375		19.2 %	1.18 [1.05, 1.33]
Usman 2011 (4)	228/376	149/378		17.8 %	1.54 [1.33, 1.79]
Subtotal (95% CI)	**872**	**834**		**45.2 %**	**1.32 [1.01, 1.73]**
Total events: 516 (face-to-face), 378 (control (all))					
Heterogeneity: $Tau^2 = 0.05$; $Chi^2 = 14.05$, df = 3 (P = 0.003); $I^2 = 79\%$					
Test for overall effect: Z = 2.03 (P = 0.043)					
2 Long (11+ min)					
Bolam 1998 (5)	179/205	169/198		20.9 %	1.02 [0.95, 1.11]
Hu 2017 (6)	376/418	359/433		21.6 %	1.08 [1.03, 1.14]
Jackson 2011 (7)	18/19	18/25		12.3 %	1.32 [1.01, 1.72]
Subtotal (95% CI)	**642**	**656**		**54.8 %**	**1.07 [1.00, 1.16]**
Total events: 573 (face-to-face), 546 (control (all))					
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 3.85$, df = 2 (P = 0.15); $I^2 = 48\%$					
Test for overall effect: Z = 1.89 (P = 0.059)					
Total (95% CI)	**1514**	**1490**		**100.0 %**	**1.20 [1.04, 1.37]**
Total events: 1089 (face-to-face), 924 (control (all))					
Heterogeneity: $Tau^2 = 0.02$; $Chi^2 = 41.31$, df = 6 (P<0.00001); $I^2 = 85\%$					
Test for overall effect: Z = 2.53 (P = 0.011)					
Test for subgroup differences: $Chi^2 = 2.11$, df = 1 (P = 0.15), $I^2 = 53\%$					

12

Imprecision



13

Imprecision Kaufman 2018

Analysis 1.1. Comparison 1 Face-to-face education versus control or non-face-to-face education (all), Outcome 1 Vaccination status (stratified by length).

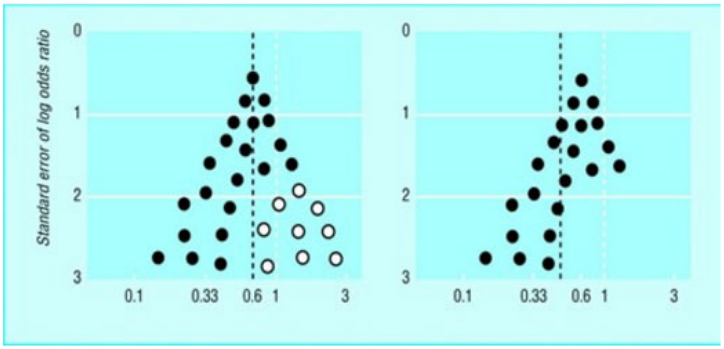
Study or subgroup	face-to-face n/N	control (all) n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
1 Short (1 to 10 min)					
Saitoh 2013 (1)	24/70	3/36		1.4 %	4.11 [1.33, 12.75]
Saitoh 2017 (2)	22/51	21/45		6.8 %	0.92 [0.59, 1.44]
Usman 2009 (3)	242/375	205/375		19.2 %	1.18 [1.05, 1.33]
Usman 2011 (4)	228/376	149/378		17.8 %	1.54 [1.33, 1.79]
Subtotal (95% CI)	**872**	**834**		**45.2 %**	**1.32 [1.01, 1.73]**
Total events: 516 (face-to-face), 378 (control (all))					
Heterogeneity: $Tau^2 = 0.05$; $Chi^2 = 14.05$, df = 3 (P = 0.003); $I^2 = 79\%$					
Test for overall effect: Z = 2.03 (P = 0.043)					
2 Long (11+ min)					
Bolam 1998 (5)	179/205	169/198		20.9 %	1.02 [0.95, 1.11]
Hu 2017 (6)	376/418	359/433		21.6 %	1.08 [1.03, 1.14]
Jackson 2011 (7)	18/19	18/25		12.3 %	1.32 [1.01, 1.72]
Subtotal (95% CI)	**642**	**656**		**54.8 %**	**1.07 [1.00, 1.16]**
Total events: 573 (face-to-face), 546 (control (all))					
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 3.85$, df = 2 (P = 0.15); $I^2 = 48\%$					
Test for overall effect: Z = 1.89 (P = 0.059)					
Total (95% CI)	**1514**	**1490**		**100.0 %**	**1.20 [1.04, 1.37]**
Total events: 1089 (face-to-face), 924 (control (all))					
Heterogeneity: $Tau^2 = 0.02$; $Chi^2 = 41.31$, df = 6 (P<0.00001); $I^2 = 85\%$					
Test for overall effect: Z = 2.53 (P = 0.011)					
Test for subgroup differences: $Chi^2 = 2.11$, df = 1 (P = 0.15), $I^2 = 53\%$					

0.5 0.7 1 1.5 2
Favours control (all) Favours face-to-face

14

Publication bias

- Funnel plot



No Publication Bias Publication Bias

15

Publication bias

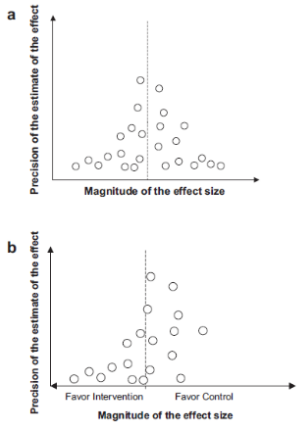
Table 5.8: Possible sources of publication bias throughout the publication process

Phases of research publication	Actions contributing to or resulting in bias.
Preliminary and pilot studies	Small studies more likely to be "negative" (e.g. those with discarded or failed hypotheses) remain unpublished; companies classify some as proprietary information.
Report completion	Authors decide that reporting a "negative" study is uninteresting; and do not invest the time and effort required for submission.
Journal selection	Authors decide to submit the "negative" report to a nonindexed, non-English, or limited-circulation journal.
Editorial consideration	Editor decides that the "negative" study does not warrant peer review and rejects manuscript.
Peer review	Peer reviewers conclude that the "negative" study does not contribute to the field and recommend rejecting the manuscript. Author gives up or moves to lower impact journal. Publication delayed.
Author revision and resubmission	Author of rejected manuscript decides to forgo the submission of the "negative" study or to submit it again at a later time to another journal (see "journal selection" above).
Report publication	Journal delays the publication of the "negative" study. Proprietary interests lead to report getting submitted to, and accepted by, different journals.

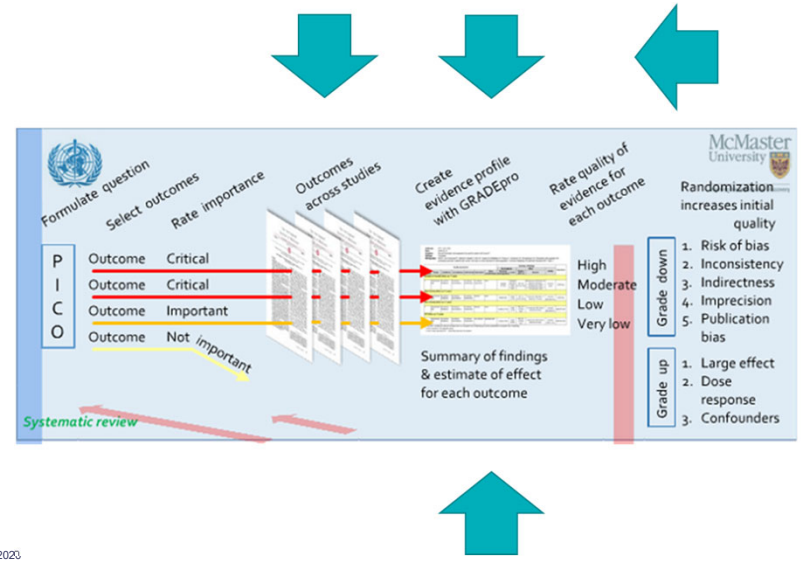
16

Publication bias Kaufman 2018

- We did not downgrade any outcomes for issues of publication bias.



Summary of findings (SoF) table



Summary of Findings table Kaufman 2018

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON *[Explanation]*

Face-to-face interventions directed to parents for informing or educating parents about early childhood vaccination, as compared with control

Patient or population: parents of preschool-aged children or expectant parents
Settings: clinics, antenatal classes, or the mother's home
Intervention: face-to-face information or educational interventions
Comparison: control (no education, other education, or control not described)

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed (baseline) risk	Corresponding (intervention) risk				
	Control (no face-to-face information or education)	Face-to-face information or education				
Vaccination status Final time point (3, 6, or 12 months post-intervention)	**55 per 100**[1]	**66 per 100** (57 to 75)	**RR 1.20** (1.04 to 1.37)	3004 (7 studies)	⊕⊕◯◯ **low**[2]	The results for this outcome were variable, so the true result may be substantially higher or lower than this estimate

2 We downgraded the evidence for this outcome for **risk of bias (-1)**. One trial was at unclear risk for **sequence generation**, two trials were at high, and one at unclear risk of bias for **allocation concealment**. We also downgraded for **inconsistency** (-1) because, while the nature of the interventions and participants were relatively similar across studies, there was **considerable statistical heterogeneity that was not easily explained (I^2 = 85%, Chi^2 P < 0.00001)**.

Online tool

GRADEpro GDT

https://gradepro.org/

GRADE exercise for break-out groups

21