Systematic Reviews and Meta-Analysis

Mohamed Eltorki MBChB, MSc, FRCPC
Associate Professor of Pediatrics, Section of Pediatric Emergency Medicine
University of Calgary
Alberta Children's Hospital

Objectives

1) Understand different types of reviews
2) Brief overview of the steps in conducting a systematic review and meta-analysis
   - Focus and examples are for interventional SR/MA

Reasons to do a systematic review:

Answer a specific research question

Base your answer on the totality of available evidence

What is a systematic review?

- A systematic review is a structured and rigorous approach to summarizing and analyzing existing literature on a specific research question.
- It follows a predefined protocol and employs systematic search strategies to identify, select, and critically appraise relevant studies.

Is IN Ketamine as effective as IN fentanyl for severe pain in children? Is it safe?

RESULTS OF SINGLE RCT Graudins 2014: Total sample size 68

PAIN SCORES

Variable	Intranasal Ketamine			Intranasal Fentanyl			Mean Difference (95% CI)
Pain scores on the 100 mm visual analogue scale (VAS)	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	0.00 [-8.77, 8.77]
	30	19.2	36	30	18.5	35	

ADVERSE EVENTS

Variable	Intranasal Ketamine	Intranasal Fentanyl	Difference (95% CI)
Patients with any adverse events	28/36 (78%)	15/37 (40%)	38 (-58 to 16)

Problems with single RCT?

- Can be only single centre
- Different geographic region than mine
- Small sample size
- Not powered for any other outcomes of interest, e.g., adverse events

Decided to do a comprehensive systematic review

- Comprehensive database search based on pre-defined search criteria and a PICO question
- Extracted aggregate data out of studies
- Can test for rare outcomes (adverse events)
- Pooled the data in a meta-analysis

4 studies, total sample size 263.
IN ketamine vs. IN fentanyl

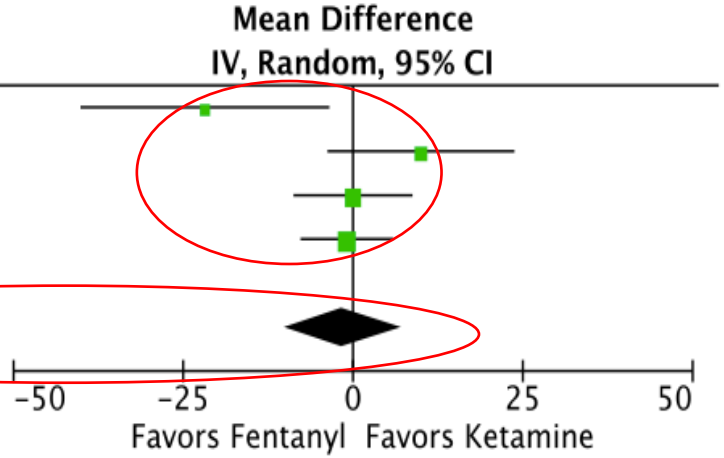
Mean difference in pain 0-100 mm on a pain scale. WMD 1.42 mm

A

Study or Subgroup	Intranasal Ketamine Mean	SD	Total	Intranasal Fentanyl Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Quinn 2018	28.2	25.5	11	50	17.7	11	14.5%	-21.80 [-40.14, -3.46]
Reynolds 2017	36	34	42	26	31	44	20.6%	10.00 [-3.77, 23.77]
Graudins 2014	30	19.2	36	30	18.5	35	30.2%	0.00 [-8.77, 8.77]
Frey 2018	24.4	15.9	42	25.3	16	42	34.6%	-0.90 [-7.72, 5.92]
Total (95% CI)			**131**			**132**	**100.0%**	**-1.42 [-9.95, 7.10]**

Heterogeneity: $Tau^2 = 42.52$; $Chi^2 = 7.43$, df = 3 (P = 0.06); $I^2 = 60\%$

Test for overall effect: Z = 0.33 (P = 0.74)



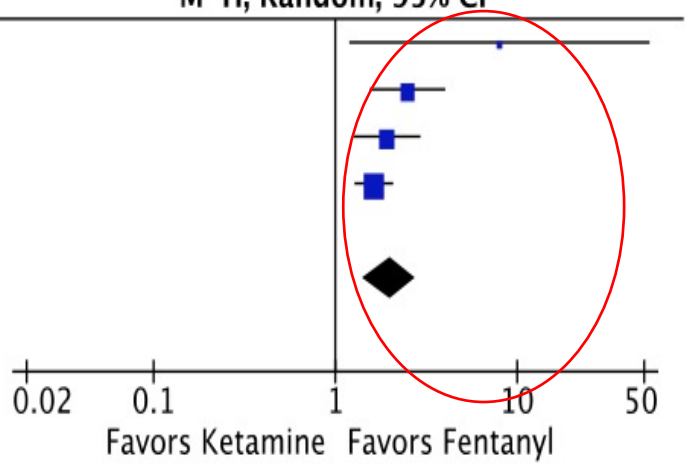
How about Adverse events now?

A

Study or Subgroup	Intranasal Ketamine Events	Total	Intranasal Fentanyl Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Quinn 2018	8	11	1	11	2.9%	8.00 [1.19, 53.67]
Frey 2018	34	44	13	42	25.8%	2.50 [1.55, 4.03]
Graudins 2014	28	36	15	37	28.9%	1.92 [1.25, 2.94]
Reynolds 2017	41	41	25	41	42.4%	1.63 [1.27, 2.08]
Total (95% CI)		**132**		**131**	**100.0%**	**2.00 [1.43, 2.79]**
Total events	111		54			

Heterogeneity: $Tau^2 = 0.05$; $Chi^2 = 5.94$, df = 3 (P = 0.11); $I^2 = 49\%$

Test for overall effect: Z = 4.06 (P < 0.0001)

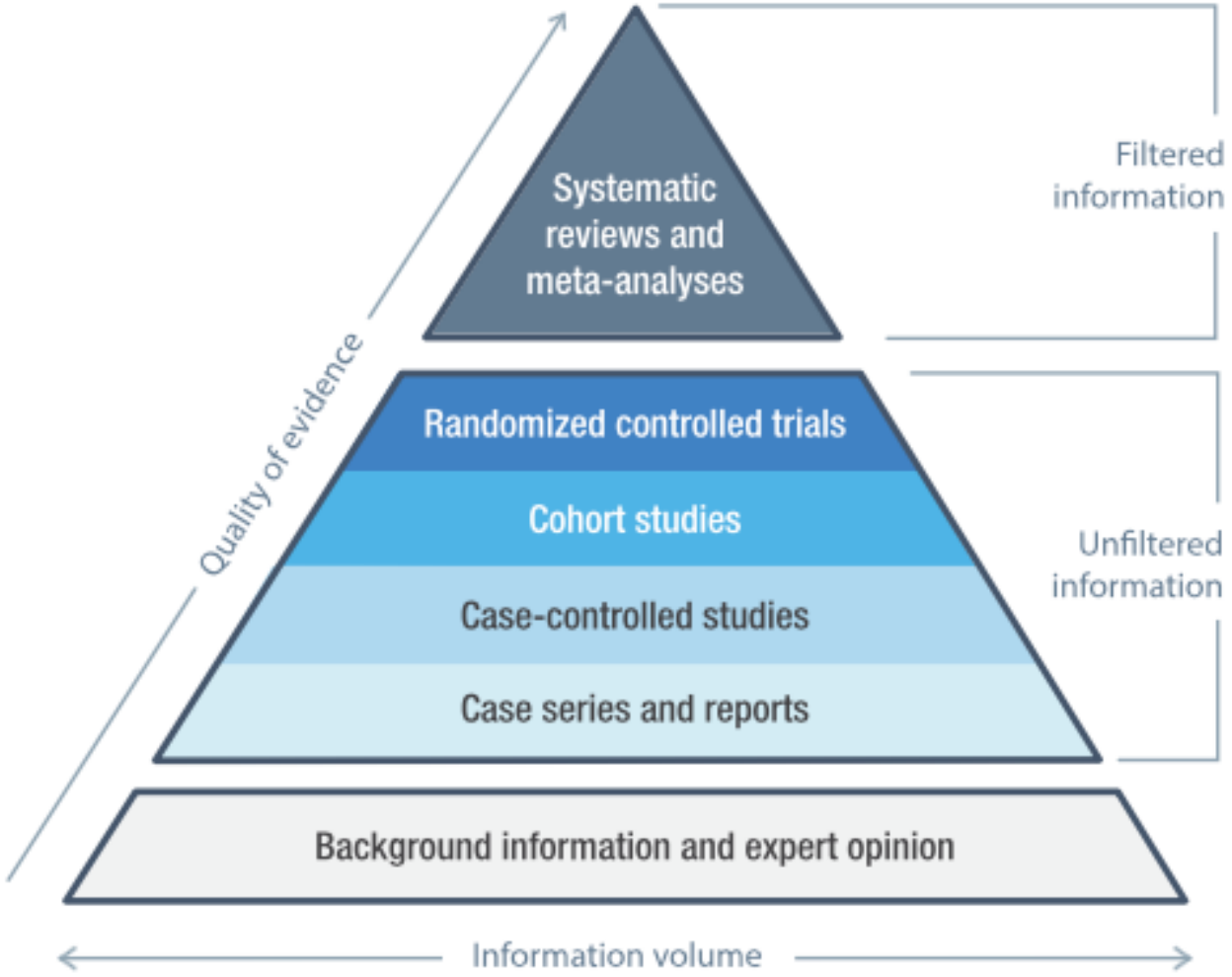


So what can our review tell us now?

IN Ketamine is equivalent to IN Fentanyl in pain relief for children with severe pain from fractures

IN ketamine results in more adverse events

PYRAMID OF EVIDENCE BASED MEDICINE



What this review DOES NOT tell you

- **SR DO NOT GIVE RECOMMENDATIONS…**
- All children with acute limb fractures should NEVER use IN ketamine
  - This is a clinical decision
  - **Guidelines** bring in some missing pieces from SR/MA
    - Clinician preference
    - System issues
    - Cost
    - Severity of those adverse events
    - Parent preference, opioid fears

What are the different types of reviews

- Systematic reviews
  - Intervention
  - Diagnostic
  - Prognostic
  - Prevalence
- Narrative reviews
- Scoping reviews

Scoping Review

> CJEM. 2025 Sep 15. doi: 10.1007/s43678-025-00982-7. Online ahead of print.

Topical non-steroidal anti-inflammatory drug use for pediatric acute musculoskeletal pain: a scoping review

Domenic F Alaimo [1], Marah Al Masri [2], Mohamed Eltorki [3]

Narrative Review (expert review)

REVIEW ARTICLE

Shiga Toxin–Producing *Escherichia coli* and the Hemolytic–Uremic Syndrome

Authors: Stephen B. Freedman, M.D.C.M. , Nicole C.A.J. van de Kar, M.D. , and Phillip I. Tarr, M.D. Author Info & Affiliations

Published October 11, 2023 | N Engl J Med 2023;389:1402-1414 | DOI: 10.1056/NEJMra2108739

VOL. 389 NO. 15 |

Types of reviews

	Systematic Review	Narrative Review	Scoping Review
Purpose	Answer a specific research question for existing literature	Provide comprehensive overview and interpretation of existing literature on a topic	Map existing literature on a topic “lay of the land”
Methodology	Protocolized, follows a specific a-priori structure, explicit sampling strategy, critical appraisal and analysis plan	Less structure, non protocolized, flexible, narrative presentation “expert telling a story”	Systematic approach for search/selection but less rigorous with critical appraisal and data extratction
Objective vs. Subjective	Objective	Subjective	Objective
Resources	Intensive and time consuming	Less resource-intensive and quick	Resource efficient
Value for EBM	Highly valuable	Nice insights from experts in the field	Valuable for identifying gaps in research

How do we do a SR?

Let’s dive in…

Steps for a Systematic Review

1. Define your PICO question
2. Plan your sampling strategy
3. Plan your search strategy
4. Search for studies
5. Screen results of the search against your eligibility criteria
6. Extract data from included studies
7. Assess risk of bias from included studies
8. Analyze results (can combine and synthesis or summarize)
9. Assess certainty of the evidence you found: GRADE
10. Conclude

Developing your question and eligibility criteria

- Identify any previous reviews in your topic
- Could be a new review or an update of a previous one
- Updates are frequently done every 5 years or if there was a large study published
- Need at least 2 studies to pool data together and at least 1 outcome

PICO

- P- Population
- I-Intervention(s)
- C-Comparison(s)
- O-Outcome(s)
- Can be single or several interventions, comparators or outcomes.

Example

- What is the effect of ibuprofen on pain?

Is this a good PICO question for SR/MA?

What is missing?

Example continued

- In children aged 5 to 17 years who present to the Emergency Department with extremity fractures, what is the efficacy of oral ibuprofen compared to oral morphine in pain relief as measured on a visual analogue scale?
- Population
- Intervention
- Comparator
- Outcome
- Study types?

Broad or narrow

- Can decide if you want your search to be broad, but still be specific
- Example of broad:

Comparing any form or dose of ibuprofen with any other form or dose of non-ibuprofen comparator for pain relief

- Example of narrow:

Comparing oral ibuprofen given at 10mg/kg to oral morphine given at 0.5 mg/kg for pain relief.

Population: Questions that help you define further

- How is the disease/condition defined?
- Any relevant demographic factors? I only want to include females, or only ≥80 years old or only minority groups
- Setting? Hospital (inpatient, outpatient, ED, ICU) or community
- Diagnosis is based on what? Clinician, set of diagnostic criteria?
- How will you handle studies that only includes a subset of your population (e.g., interested in children only and a study enrolled age 12-65)

If you decide to be broad

- Consider having subgroups- examples:
  - Does the intervention work differently in specific subgroups of people?
  - E.g., limb fractures vs. all other
  - Burns vs. all other
  - Headache vs. all other
  - Adults vs. children
  - High risk of bias studies vs low risk of bias studies
  - Community studies vs. hospital studies

What is the rationale for subgroups

- Does the intervention work differently in specific groups of people?
- Hypothesis generating..

Intervention: Refine

- Does the intervention have variations in form, dosing, components, frequency, duration, timing?
- Do you need to have a threshold at which you will not accept an intervention, e.g., will not include ibuprofen if not given at 10mg/kg Q6-8H for 2 days scheduled
- Will you allow co-interventions? e.g., will you include trials that allowed acetaminophen to be given with ibuprofen?

Intervention details

- Formulation
- Dose
- Route
- Timing and frequency and duration
- Equipment (e.g., MDI)
- Personnel administering (e.g., parent vs RN)
- Location of intervention
- Monotherapy or combined with other interventions

Comparison

- What are you interested in comparing the intervention to?
  - Active intervention?
  - Non-active intervention?
  - No intervention?
- What is the usual alternative? What is the usual care?

Can have multiple comparators

Intervention	Comparator
Oral ibuprofen	Oral placebo
Oral ibuprofen	Oral acetaminophen
Oral ibuprofen	Oral morphine
Oral ibuprofen	Oral Cannabidiol oil
Oral ibuprofen	Oral ANYTHING together

Conduct a separate analysis for each comparison or POOL all the comparators together but BE READY TO JUSTIFY IT

Outcomes: questions to refine

- What are the important outcomes? are they surrogate or patient important? E.g., ejection fraction vs. how many blocks I can walk?
- Are those outcomes part of the search criteria?
- **Moved away from primary and secondary outcomes in SR/MA**
- How are those outcomes measured? Is it validated?
- Do they use the same measurement tools?
- Have you included adverse events?
- What time point will you use for measuring outcome?
- Did you consult stakeholders for the outcomes of interest?
  - E.g., clinicians, guideline developers, decision makers, patients, caregivers, etc.

Why do we need to chose outcomes before we even do the search?

- Selective reporting….

This occurs when the reporting of outcomes is based on the results!

Authors can then pick and chose the outcomes that provide more robust results and conclusions!

Develop your search strategy

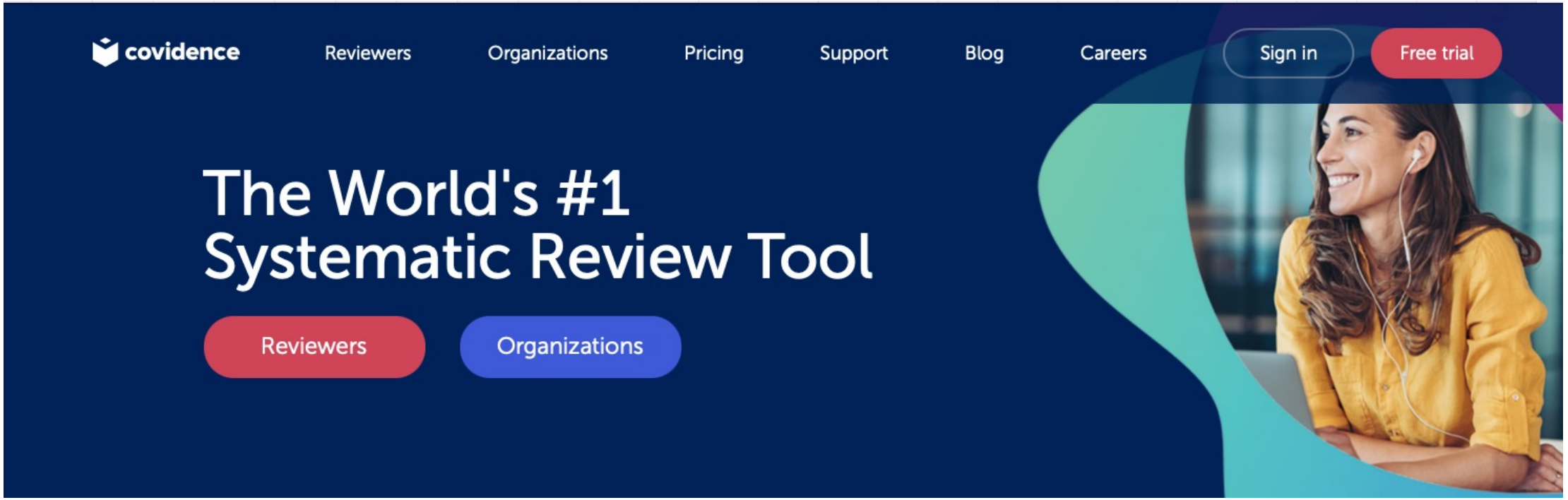
- Using your PICO- Consult with a librarian to formulate a search strategy
- Use Large Databases (CENTRAL, Medline, Embase)
- If you are limiting to a time frame; must justify
- Use OVID platform to run the search

Search and Screen for studies

- Where to search:

1. Database of publications
2. Database of trials in progress
3. Google scholar
4. Abstracts from conferences
5. Reference lists of studies or previous systematic reviews

Import search results into reference manager software



Selecting studies

1. Software removes duplicates
2. Create a screening form based on PICO and pilot whether the team applies it correctly
3. Selection should be performed in duplicate and independently from each other and have a third person to adjudicate
4. Screen titles and abstracts to remove irrelevant studies
5. Retrieve full texts for potentially relevant studies
6. Screen full texts and remove irrelevant studies
7. Adjudicate all disagreements between the two screeners
8. Make final decisions

Example of a screening form:

1. Is this article about children with pain from a limb fracture seen in the ED?
2. Does the article compare ibuprofen to any other comparator
3. Is the duration of treatment at least for 2 days?
4. Does the article report on adverse events?
5. Is the study a randomized controlled trial?

For any Q:

If answer is no exclude, if yes or unclear go to next STEP (FULL TEXT REVIEW)

Title and abstract screening

- In duplicate and independently
- Exclude if BOTH people excluded
- Include to full text review if “include or unsure” by either person

Full text screening

- Do a pilot first (10-20 studies)
- Re-convene with the team and discuss any issues with the piloted form
- Two people assess a study to be included or not
- There must be agreement on the studies to include
  - If no agreement, use a third adjudicator

Full text screening

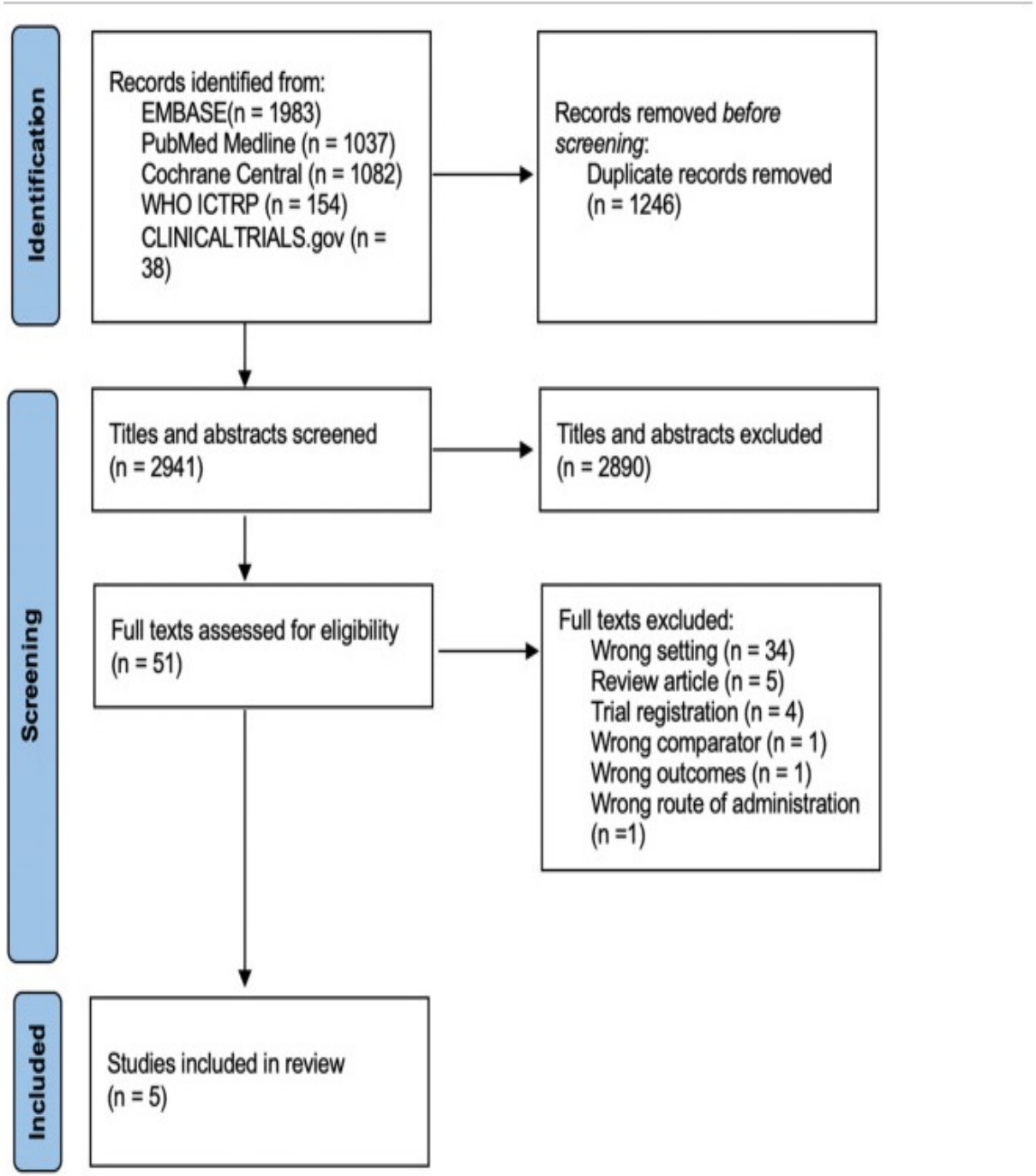
- Must indicate REASON for exclusion
  - Covidence allows you to chose from a list of exclusion criteria
  - E.g., excluded as outcome of interest not reported, or wrong comparator, or incorrect study design….

POPULATE PRISMA FLOW CHART

www.prisma-statement.org

Established guidelines on reporting of SR/MA

Has a PRISMA flow chart and a checklist of items that needs to be included in your SR/MA



Summary page

- Selecting studies

1. Examine titles and abstracts to remove obviously irrelevant reports
2. Retrieve full texts of the potentially relevant results
3. Examine full-text reports
4. Compare decisions between the two investigators
5. Make final decisions

Process

- Decided on systematic review question (PICO)
- Outlined eligibility criteria
- Conducted search
- Screened by title and abstract and then looked at full text
- Confirmed studies to include
- **Abstract data**
- **Assess risk of bias of studies**
- **Abstract data from studies**

Data abstraction and Risk of Bias

- **What do we need from each study?**

Population

Intervention

Comparison

Outcomes of interest

Raw data

- **What do we not care about?**

Their calculated relative risk or odds ratio or p values

Their conclusions

How are we abstracting data?

- Pilot the form; use an easy trial and a hard trial for pilot
- Establish standard operating procedures for data conversions/using tools to extract data from graphs
- Preferably; again in duplicate and independent

Build consensus or have an adjudicator at the end for disagreements

Table 1: Summarize your included studies

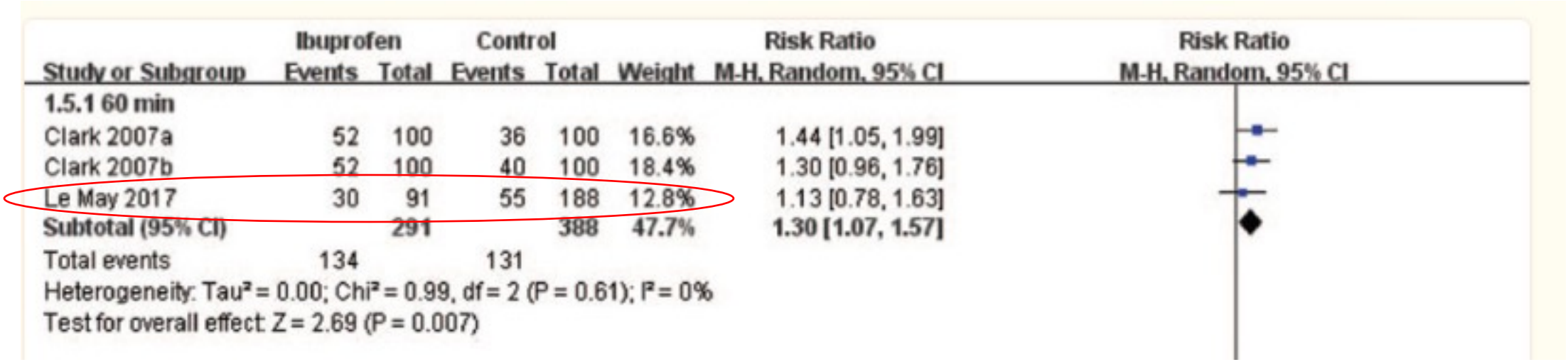
Study Author and Year	Study Design	Type of pain/condition	Inclusion Criteria	Exclusion Criteria	Ketorolac Groups/Dosages and Route	Longest Follow Up (h)	Rescue Analgesia or PRN	No. of Patients Low/Medium /High Dose Ketorolac	Mean or Median Age in Years (SD or 95% CI)	Mean or Median (SD or 95% CI) Baseline Pain Score
Oosterlink 1990	Multi center RCT	Flank pain due to renal colic	1. Suffering pain due to renal colic, described as at least moderate according to a 4-point verbal rating scale 2. Diagnosis of renal colic required radiological evidence of a renal stone or acute renal obstruction 3. Age 18-75 years 4. Weight between 45kg and 100kg	1. Known history of allergy or previous adverse reaction to salicylates or NSAIDs 2. Patients known to abuse alcohol, narcotics, or other drugs 3. Patients with a temperature above 37.5 degrees Celsius	High dose: 90mg IM x 1 Low dose: 10mg IM x 1	12h	Unspecified, at discretion of treating clinician	High dose: 37 Low dose: 45	High dose: 41 (21-69) Low dose: 40 (21-71)	80.95 (15.74)* *VAS 100mm
Motov 2017	Single center RCT	Acute flank, abdominal, musculo-skeletal, or headache pain with an intensity of 5 or greater on a standard 0 to 10 numeric scale	1. Adults aged 18 to 65 years who presented to the ED primarily for management of acute flank, abdominal, musculoskeletal, or headache pain with an intensity of 5 or greater on a standard 0 to 10 numeric rating scale 2. Would routinely be treated with intravenous ketorolac in ED as determined by the treating attending emergency physician	1. Age >65 years 2. Pregnancy or breastfeeding 3. Active peptic ulcer disease 4. Acute gastrointestinal hemorrhage 5. Known history of renal or hepatic insufficiency 6. Allergy to NSAIDs 7. Unstable vital signs (systolic blood pressure <90 or >180 mm Hg; pulse rate <50 or >150 beats/min) 8. Patients having already received analgesic medication	High dose: 30mg IV x 1 Medium dose: 15mg IV x 1 Low dose: 10mg IV x 1	2h	Morphine 0.1mg/kg IV x 1	High dose: 80 Medium dose: 80 Low dose: 80	40.1 (11.93)	7.69 (1.60) on 10 point numeric rating scale

Types of outcomes and measuring a treatment effect

Dichotomous

Continuous

Dichotomous outcomes: proportion who had adequate pain relief at 60 minutes



Study or Subgroup	Ibuprofen Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
1.5.1 60 min						
Clark 2007a	52	100	36	100	16.6%	1.44 [1.05, 1.99]
Clark 2007b	52	100	40	100	18.4%	1.30 [0.96, 1.76]
Le May 2017	30	91	55	188	12.8%	1.13 [0.78, 1.63]
Subtotal (95% CI)		**291**		**388**	**47.7%**	**1.30 [1.07, 1.57]**
Total events	134		131			

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.99$, $df = 2$ ($P = 0.61$); $I^2 = 0\%$

Test for overall effect: $Z = 2.69$ ($P = 0.007$)

Risk Ratio M-H, Random, 95% CI

Dichotomous Outcomes

We are taking raw numbers out of study

	Pain relief (event YES)	No pain relief (event NO)	Total
Ibuprofen	30	61	91
Control	55	133	188
Total	85	194	714

Treatment effect
Risk ratio

	Pain relief (event yes)	No pain relief (event no)	Total
Ibuprofen	30	61	91
Control	55	133	188
Total	85	194	714

- Risk of pain relief with ibuprofen?

30/91=0.32=32%

- Risk of pain relief with control?

55/188=0.29=29%

- Risk ratio= intervention risk/control risk

=0.32/0.29=1.1

Where RR of 1=no difference between group…

Treatment effect
Odds ratio

- Odds of pain relief (event) with ibuprofen= 30/61=0.49
- Odds of pain relief (event) event with control= 55/133=0.41

	Pain relief (event yes)	No pain relief (event no)	Total
Ibuprofen	30	61	91
Control	55	133	188
Total	85	194	714

Odds ratio= odds of event in ibuprofen/odds of event in control

0.49/0.41=1.19

Odds ratio of 1 = no difference between groups

Treatment effect
Risk difference

- Risk of pain relief with ibuprofen?

30/91=0.32=32%

- Risk of pain relief with control?

55/188=0.29=29%

	Pain relief (event yes)	No pain relief (event no)	Total
Ibuprofen	30	61	91
Control	55	133	188
Total	85	194	714

Risk difference (Absolute risk reduction)= intervention risk – control risk

0.32-0.29=0.03 or 3% increase in risk of pain improvement at 60 minutes

This allows you to calculate NNT= 1/ARR=1/0.03=33 patients need to be treated with ibuprofen to achieve 1 pain relief over the control.

In systematic reviews

Treatment effect is usually measured using

Risk Ratio is what is often calculated and an **Absolute risk difference** that then allows you to calculate **Number Needed to Treat**

Sometimes Odds Ratio (more common in meta-regression)

Collect data for 2x2 table from studies.

Ideally, you want to extract

Number of events in each arm

AND

Number of participants in each arm

However……

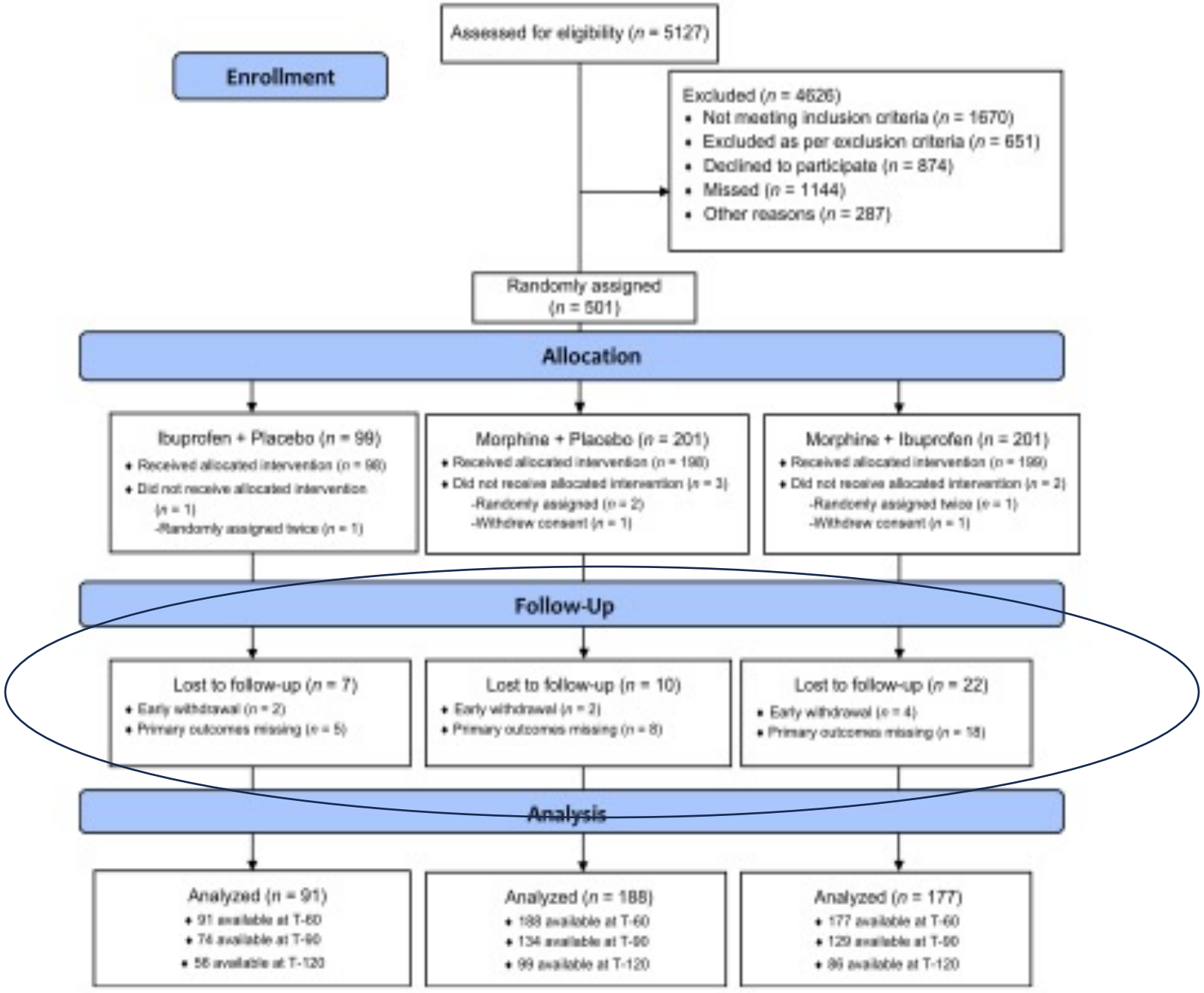


FIGURE 1
Consolidated Standards of Reporting Trials flowchart of subjects' enrollment and/or allocation and study proceedings.

Decide what you will do with missing data in your protocol

- Complete case analysis or available case analysis
- Include data from people that are followed
- Remove data for people not followed up (i.e subtract them from the denominator)
- Include adherents and non adherents; keep them in the groups they were randomized to

Other data issues: different forms

- E.g., study only reported percentages
- One study reported pain on a 100 mm visual analogue scale and another reported on a facial pain scale (look up ways to convert different scales to a common scale or use standardized mean difference)
- Follow up time (differing intervals, which one will you chose?)
- Only relative risks are provided (can still use the data)

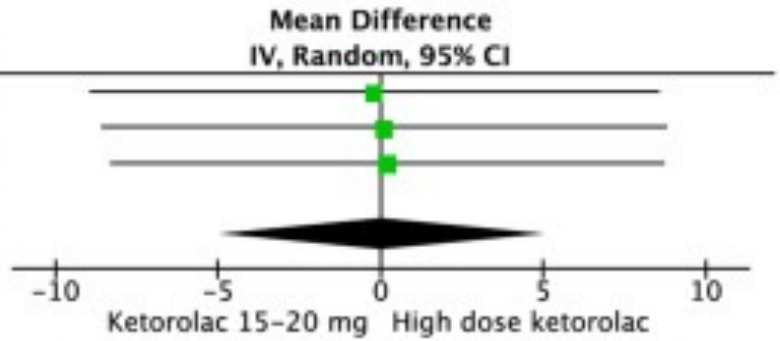
Continuous outcomes

- Can be any value within a specific range
- Intervals are equally spaced
  - E.g., PRAM Score for asthma; weight, temperature, depression scales, pain scores, creatinine level, quality of life score

Study or Subgroup	Ketorolac 15–20 mg Mean [mm]	SD [mm]	Total	High dose ketorolac Mean [mm]	SD [mm]	Total	Weight	Mean Difference IV, Random, 95% CI
Motov 2017	40.9	26.1	77	41.1	29.3	77	32.5%	−0.20 [−8.96, 8.56]
Eidenejad 2020	7.4	19.5	36	7.3	18.8	39	33.1%	0.10 [−8.58, 8.78]
Turner 2021	29.9	23.1	55	29.7	22.5	55	34.4%	0.20 [−8.32, 8.72]
Total (95% CI)			**168**			**171**	**100.0%**	**0.04 [−4.96, 5.03]**

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.00$, df = 2 (P = 1.00); $I^2 = 0\%$

Test for overall effect: Z = 0.01 (P = 0.99)



Data needed for continuous outcome

1) Point estimate (e.g., mean, median)
2) Variance measure (e.g., standard deviation)

Mean difference= intervention score – control score

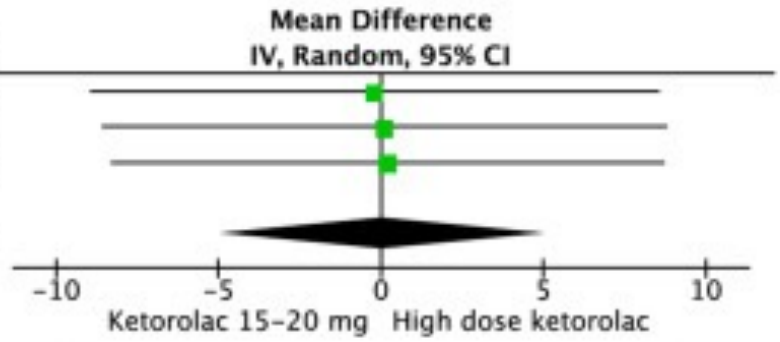
- E.g., Turner 2021

Mean difference between ketorolac 15-20 mg and high dose ketorolac = 29.9-29.7=0.2

What does that mean? What is the scale? What is the direction of difference?

	Ketorolac 15–20 mg			High dose ketorolac				Mean Difference
Study or Subgroup	Mean [mm]	SD [mm]	Total	Mean [mm]	SD [mm]	Total	Weight	IV, Random, 95% CI
Motov 2017	40.9	26.1	77	41.1	29.3	77	32.5%	-0.20 [-8.96, 8.56]
Eidenejad 2020	7.4	19.5	36	7.3	18.8	39	33.1%	0.10 [-8.58, 8.78]
Turner 2021	29.9	23.1	55	29.7	22.5	55	34.4%	0.20 [-8.32, 8.72]
Total (95% CI)			**168**			**171**	**100.0%**	**0.04 [-4.96, 5.03]**

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.00$, df = 2 (P = 1.00); $I^2 = 0\%$
Test for overall effect: Z = 0.01 (P = 0.99)



Risk of Bias Assessments

- What is bias?

Bias is deviation from the truth due to systematic error

It may over or under-estimate the results of a study

Risk of bias assessment tools

- RCT: Cochrane risk of bias tool - RoB 2.0 for individual RCT

https://www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2

Extensions exist for 1) Cluster trials 2) cross over trials

- Observational trials: Cochrane ROBINS-I tools OR Newcastle-Ottawa scale

RoB 2.0

- Provides you with a working excel file and embedded macros to assist in risk of bias assessment and populate tables/button graphs at the end
- 5 domains of assessment leading to
  - Low risk of bias
  - Some concern of bias
  - High risk assessment of bias

Domain 1: Randomization and Allocation Concealment

- Bias with randomization process
  - Randomization not done well
  - Simple unpredictable methods used (flipping coin) vs. computer generated
- Bias with allocation concealment
  - The person allocating the participant to a group is able to predict which group patient will go to (e.g., envelops that are not opaque, person allocations is also the person who created the randomization table)

Domain 2: Bias due to deviation of intended interventions

Participants should be analyzed in the group they were randomized to

Intention to Treat analysis is appropriate

Per protocol analysis not appropriate unless good justification

Domain 3: Bias due to missing outcome data

- Will it lead to bias?
  - Reasons for missing data?
  - Is missing data balanced between groups?
  - Is the amount of missing data have an impact? E.g.,>20% missing data

Domaine 4; Bias in measuring the outcome

- Outcome measurement tool: is it valid (does it measure what it should) is it reliable (does it measure it the same way each time)
- Who is measuring the outcome? Are they blinded if the outcome is subjective?

Domain 5: Bias in the selection of the reported results

- We only care about the outcome we want to extract…
- Did they say they will measure it in methods?
- Did they report it?

Other sources of bias that were included in the past….

- Power/sample size: who cares if we are pooling the studies anyways!
- Industry funded vs. not: if the trial had problems because of industry funding, it will come out in other domains! Remember, university or government funded investigators also have their own biases. Assess the study objectively without assumptions about the investigator

Risk of Bias assessments

- Should be done for each study
- Should be done for each outcome

This is not a hard rule, but this is what Cochrane suggests.

Table 1. Reaching an overall risk-of-bias judgement for a specific outcome.

Overall risk-of-bias judgement	Criteria
Low risk of bias	The study is judged to be at **low risk of bias for all domains** for this result.
Some concerns	The study is judged to raise **some concerns** in at least one domain for this result, but not to be at high risk of bias for any domain.
High risk of bias	The study is judged to be at **high risk of bias** in at least one domain for this result. Or The study is judged to have **some concerns** for **multiple domains** in a way that substantially lowers confidence in the result.

Table. Risk of bias of included studies.

Author, y	Bias Arising from the Randomization Process	Bias Due to Deviations from Intended Interventions	Bias Due to Missing Outcome Data	Bias in the Measurement of the Outcome	Bias in the Selection of the Reported Result	Overall RoB
Oosterlink,[36] 1990	Some concerns	Some concerns	Low	Low	Low	**Some concerns**
Motov,[16] 2017	Low	Low	Low	Low	Low	**Low**
Chao,[34] 2020	Low	Low	Low	Low	Low	**Low**
Eidenejad,[35] 2020	Low	High	Low	Low	Low	**High**
Turner,[17] 2021	Low	Low	Low	Low	Low	**Low**

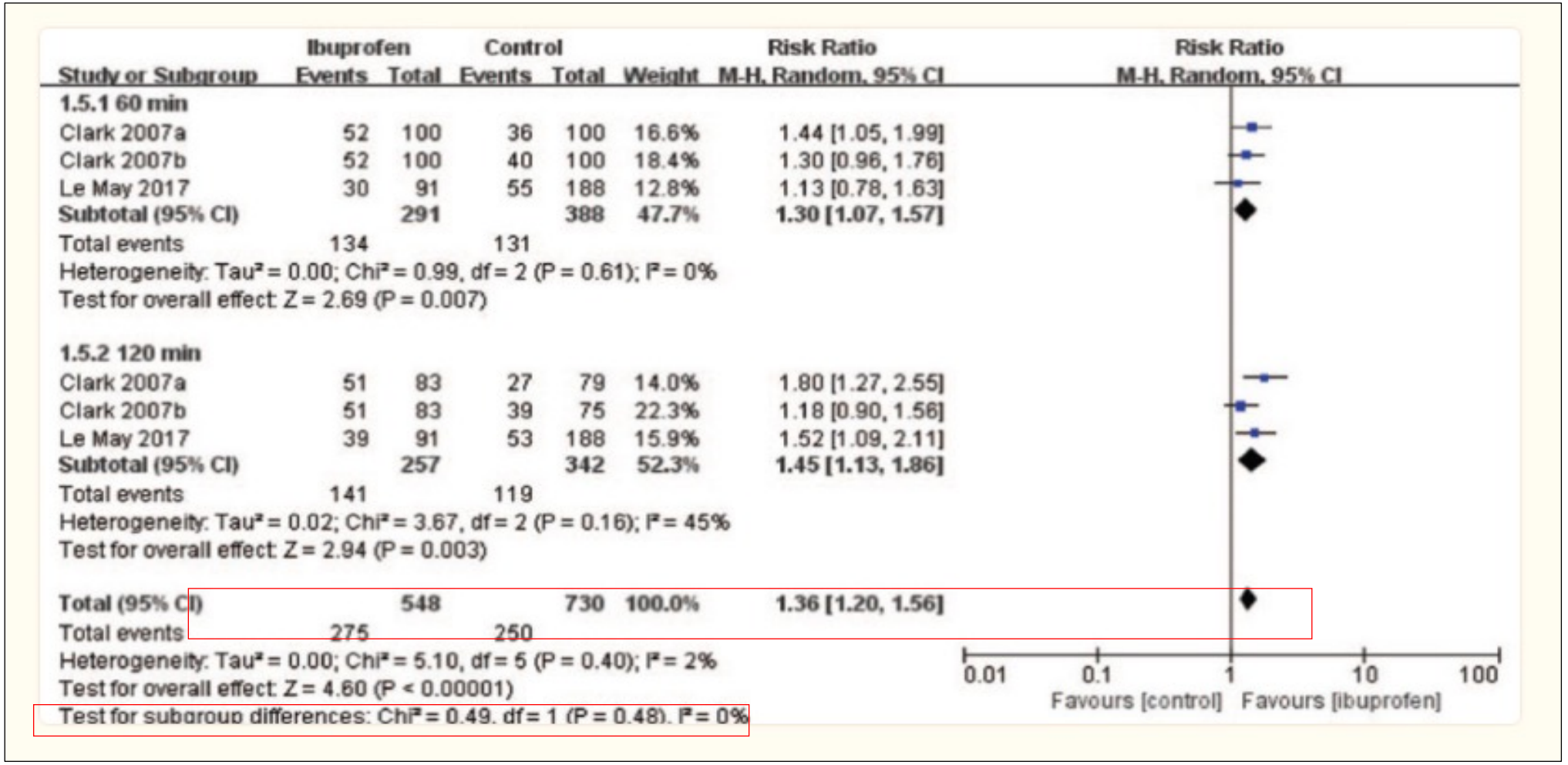
Other Consideration in SR/MA To Think About

Meta-analysis; synthesizing evidence

- Not every systematic review needs a meta-analysis!

- Ask yourself this Question; Should I be pooling these studies together? Are those studies alike? Is the interventions and comparators somewhat similar? How much data is missing? Do they use the same scale to measure outcome?

The Power of a Meta-analysis-Subgroups?

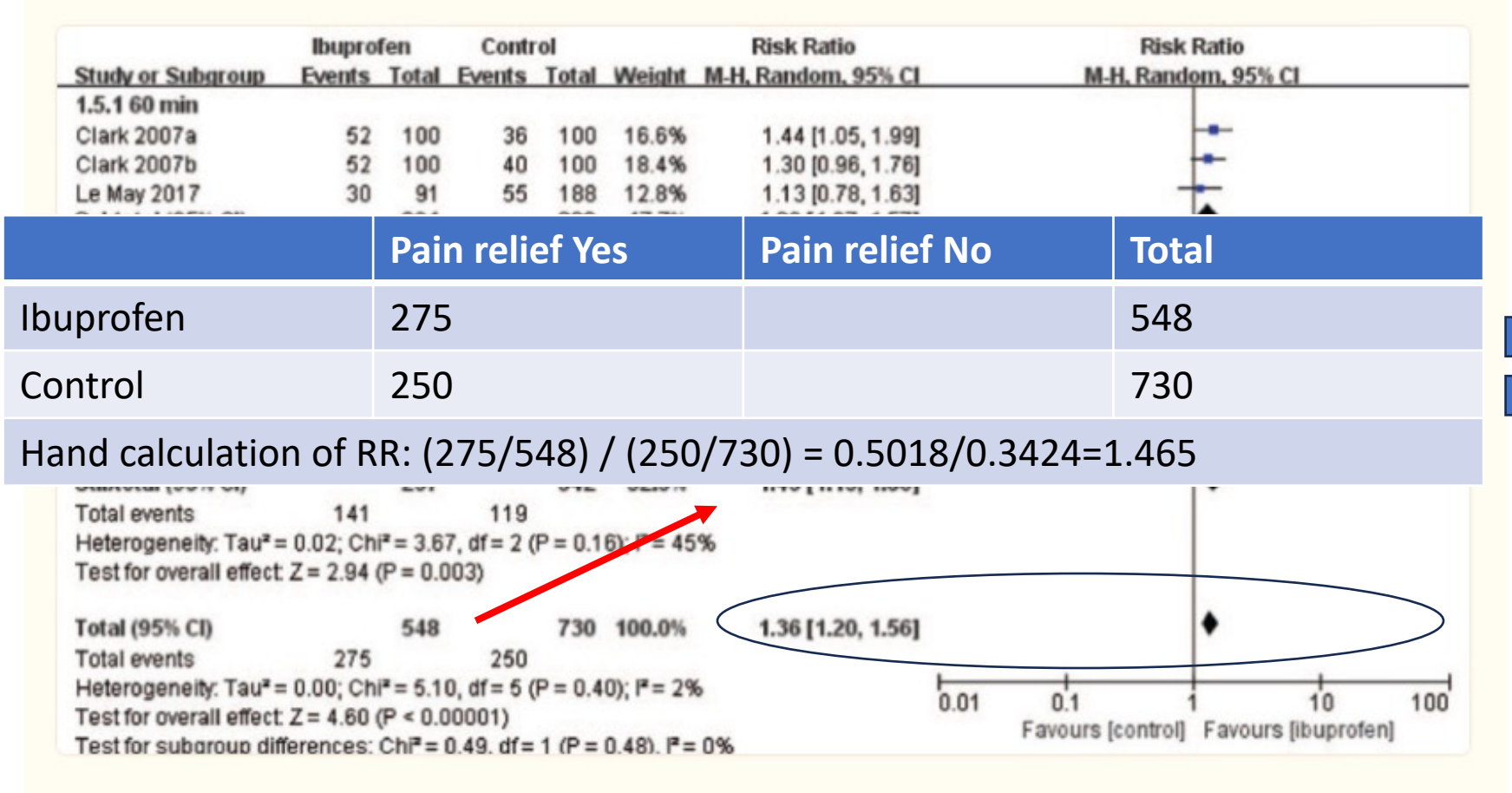


Pain improved 1.36 times with ibuprofen compared to control (from 1.2x more to 1.56x more) or 36% more pain relief at all time points compared to control

What to consider when doing a Meta-analysis

- Organize your data
- Which studies are you going to pool? Need at least 2 for MA
- Need to match intervention/comparator/outcome in a logical way

Is our hand calculation similar to the total RR calculated in the Meta-analysis? Why not?



	Pain relief Yes	Pain relief No	Total
Ibuprofen	275		548
Control	250		730
Hand calculation of RR: (275/548) / (250/730) = 0.5018/0.3424=1.465			

≠ 1.36

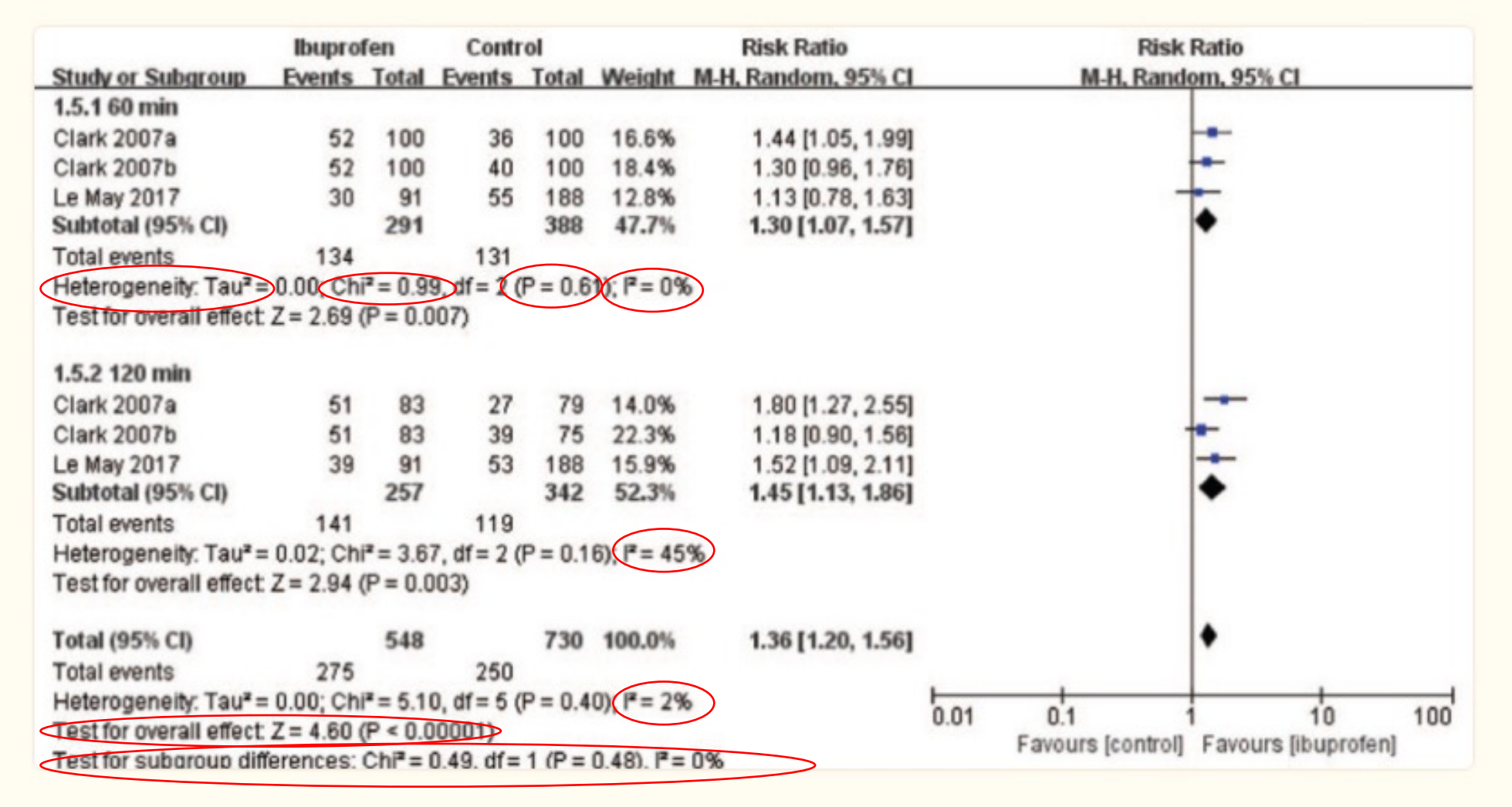
Hand calculation of RR: (275/548) / (250/730) = 0.5018/0.3424=1.465

Why is it different?!

- In a meta-analysis, the calculation of risk ratio uses Weighted averages to combine individual study results. This is a more comprehensive way and statistically rigorous that takes into account:

1) Sample size of each trial
2) Variance within each trial
3) Heterogeneity among the different studies

De-mystify all the gibberish in a forest plot



Common mis-understandings of Confidence interval

“The CI tells us where 95% of the data lie”

“If two CIs overlap, there’s no significant difference.”

“A wide CI means bad results.”

“If the CI excludes the null (like 1.0 or 0), the result is definitely clinically important”

“If the CI includes the null (like 1.0 or 0), the result is definitely NOT clinically important”

What is more important p value or CI?

Feature	P-Value	Confidence Interval (CI)
Definition	Probability of observing the data (or more extreme) if the null hypothesis were true	Range of plausible values for the true effect size
Focus	Tests “is there an effect?”	Describes “how big and how precise is the effect?”
Interpretation	Tells whether result is statistically significant (below chosen α)	Shows both magnitude and uncertainty of effect
Significance threshold	Arbitrary cutoff (e.g., $p < 0.05$)	No arbitrary cutoff; interpretation depends on whether range crosses null (1.0 or 0)
Clinical insight	None — only yes/no	Rich — conveys direction, strength, and precision
Common misuse	Interpreted as “probability the null is true”	Misread as “95% chance the true value is in this interval”

Use REVMAN software for meta-analysis

- Free of charge
- Easy to use
- Available for Mac and PC

Steps for the Meta-analysis

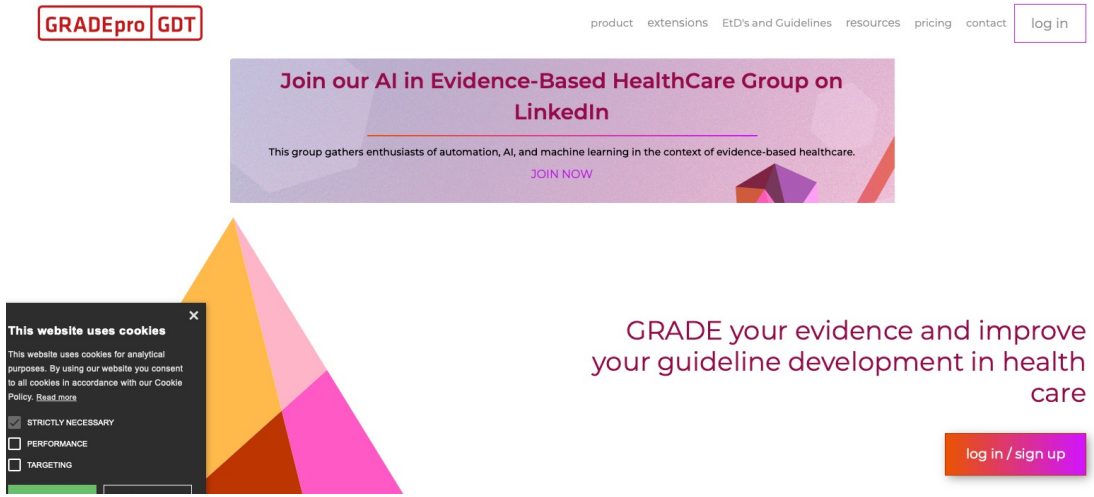
Enter the studies in REVMAN

1. Pick an analysis model
   - Fixed-effect vs. random-effects mode
   - Depends on sources of variability
     - If within study
     - If within study and between studies
2. Pick a treatment effect measure (RR, OR, AR, MD, SMD)
3. Pick a weighting method
   - Mantel-Haenszal; dichotomous outcomes
   - Peto; pooling odds ratios
   - Inverse variance; MD or SMD use inverse

Final Step: Assess certainty of the evidence!

Instead of significant or not significant; let's tell the reader about the certainty or uncertainty of the results!

Multiple things to consider here….



Certainty levels

High certainty: We are **very confident** that the **true effect lies close to that of the estimate of the effect**

Moderate certainty: We are **moderately confident** in the effect estimate: The **true effect is likely to be close** to the estimate of the effect, but there is a possibility that it is substantially different

Low certainty: **Our confidence** in the effect estimate **is limited**: The true effect may be substantially different from the estimate of the effect

Very low certainty: We have **very little confidence** in the effect estimate: The **true effect is likely to be substantially different** from the estimate of effect

Grading of Recommendations Assessment, Development, and Evaluation (GRADE)

GRADE domains	**Rating** (circle one)	**Footnotes** (explain judgements)	**Certainty of evidence** (Circle one)
Risk of Bias	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) No Plausible confounding (+1)		

Incorporating GRADE in reporting results

Recall RR of 1.30 (1.2, 1.56) for pain relief at 60 minutes favoring ibuprofen. It was a SIGNIFICANT result if we stopped there!

However, suppose **high degree of bias**, **indirectness**, **publication bias**, this would lead to an overall assessment of **low-certainity.**

Conclusion;

The relative risk of persistent pain at 60 minutes was found to be 0.76 (0.64, 0.83) with ibuprofen compared to controls. The overall certainty of evidence is low. Ibuprofen **may or may not** improve pain at 60 minutes compared to control

Steps for a Systematic Review

1. Define your PICO question
2. Plan your sampling strategy
3. Plan your search strategy
4. Search for studies
5. Screen results of the search against your eligibility criteria
6. Extract data from included studies
7. Assess risk of bias from included studies
8. Analyze results (can combine and synthesis or summarize)
9. Interpret results and draw conclusions: GRADE
10. Conclude

In Summary

- **Systematic reviews are at the top of the research pyramid since they answer important questions using the totality of the evidence**
- **A good SR/MA requires:**
  - Asking the RIGHT question
  - Pre-defined protocol
  - Registration (Prospero, OSF)
  - Robust search criteria
  - Piloted and accurate screening and data extraction
  - Thoughtful data synthesis
  - Risk of Bias Assessment
  - Assessment of the certainty of evidence

Questions?

Email: mmeltork@ucalgary.ca