An Introduction to Critical Appraisal of Systematic Reviews

Dr. Zacharias Dimitriadis
Associate Professor

Learning outcomes

Until the end of the lectures the students are expected to be able to:

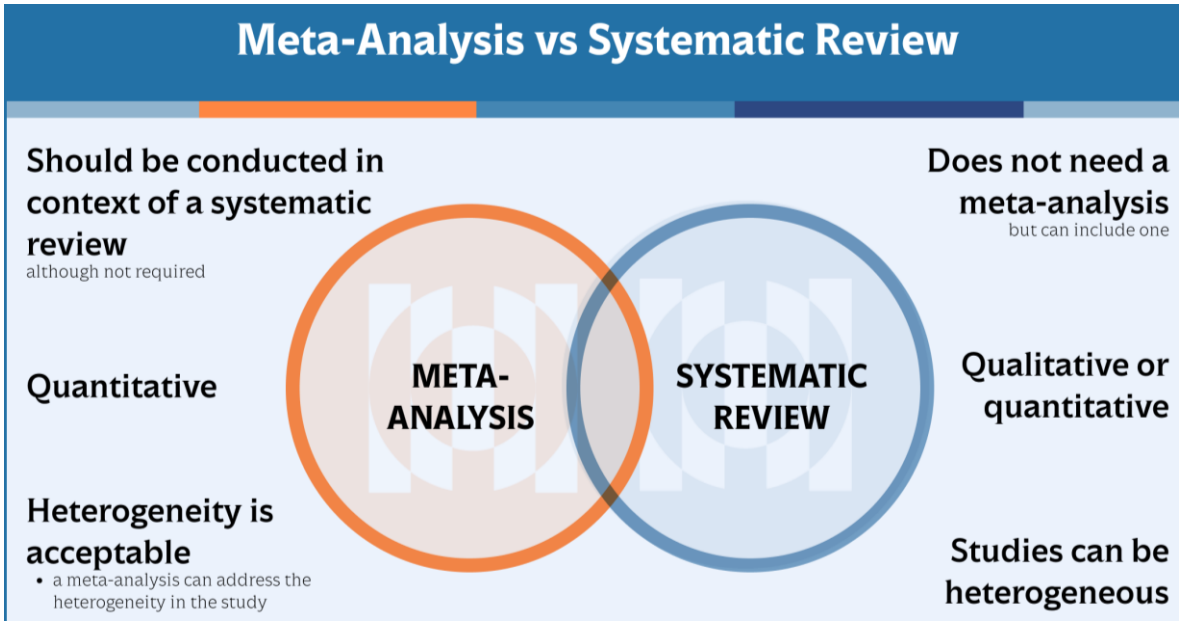
- *understand what is a systematic review*
- *understand the level of evidence it can provide*
- *understand the phases of conducting a systematic review*
- *know how to critically appraise a systematic review*

Review Methodologies

- Systematic review
- Scoping review
- Narrative review
- Umbrella review
- Mapping review
- Critical review
- etc.

Level of Research Evidence

What is the difference between a systematic review and a meta-analysis?



How we can critically appraise a systematic review?

(based on PRISMA guidelines, 2020)

TITLE

TITLE

Q1) Does the title **include all the relative information of the research question** (usually the PICO acronym for interventions) and **recognize the study as a systematic review**?

- It can be written as a) question, b) purpose or c) conclusion.

Example

"The effectiveness of neurodevelopmental approach on gait of patients with cerebrovascular accident: a systematic review"

INTRODUCTION

INTRODUCTION

Q2) Does the introduction of the systematic review include information about:

a) **rationale** - the rationale of the systematic review

b) **novelty** - what it adds to these which are already known, what is new in relation to the existent studies

c) **Importance** – why it is important to perform this systematic review how it can contribute to clinical practice

Q3) Have the objectives of the study been:

a) **clearly stated including all the relative information** (usually using thee PICO acronym)?

b) supported by an appropriate argumentation?

Example

The aims of this study were a) to conduct ***a systematic review of the literature regarding the effectiveness of neurodevelopmental approach on gait of patients with cerebrovascular accident*** *and b) to conduct meta-analyses in order to investigate the effectiveness of neurodevelopmental approach on gait of patients with cerebrovascular accident*

METHODS

ELIGIBILITY CRITERIA

Q4) Have the inclusion and exclusion criteria for the review been clearly defined?

Are they appropriate?

The PICO acronym is frequently used for helping owards this direction

e.g.

-Type of studies
-Participants characteristics
-Type of intervention
-Outcome measures

Example

Eligibility Criteria

(P)opulation*: a) diagnosis of stroke, b) age > 18 y.o., c) recent stroke onset (< 1 month)*

(I)ntervention: *physiotherapy program based on Bobath principles, b) no application of technology, c) therapy duration of at least 2 weeks*

(C)ontrol*: -*

(O)utcome*: a) inclusion of at least one outcome measure of gait, b) at least two assessments of gait (one att baseline and one post-treatment)*

Design*: Only RCTs*

Other: *a) written in English language, b) published in international peer-reviewed journal*

INFORMATION SOURCES

Q5) Is there a clear description of:

a) the databases used
b) the date of search/access
c) Other sources used (e.g. contact with authors, references included in papers etc)

Are these appropriate (e.g. is there any important source that has been omitted)?

Example

The study was performed according to PRISMA guidelines.[21] The sources that were included in the search were international databases including PubMED, MEDLINE, EBSCOhost, Google scholar, SportDiscus, COCHRANE and EMBASE. Search was performed at April 26th, 2019. Studies were also searched in the information provided in the manuscript of the detected studies as well as after consultation with experts.

SEARCH STRATEGY

Q6) Have the full search strategies for all databases been appropriately selected and described?

Is the **search strategy based on an acronym** (e.g. PICO)?

Is the acronym appropriate?

Are the **key-words used** for search appropriate?

Are there important key-words omitted increasing the danger of missing relative articles?

Is the connection of the keywords appropriate?

Example

(P)opulation	stroke OR “cerebrovascular accident”
	AND
(I)ntervention	Bobath OR “neurodevelopmental treatment” OR “neurodevelopmental approach”
	AND
(C)ontrol	-
	AND
(O)utcome	gait OR walking

SELECTION PROCESS

Q7) Have the methods used to decide whether a study will be included in the qualitative synthesis of study been described and are these methods appropriate?

- Is the **methodology of the study selection** complete, clear and accurate enough in order to be reproducible?

The phases of the review:

- **Identification:** Records indentified through the initial search on databases or other sources

- **Screening:** Initial screening of the detected articles for findings potentially eligible articles (e.g. duplicates are removed, articles that are irrelative based on title and abstract are removed)

- **Eligibility:** Assess of the full text of he potentially eligible articles in order to apply the eligibility criteria and decide which are really eligible for inclusion in the synthesis

How many reviewers participated in these phases? **Did they work independently** of each other?

Did the researchers made any effort possible in order to retrieve inaccessible articles?

Example

During the "Identification phase" the studies were identified by the database searching and search in other sources (content of studies, experts). During the "Screening phase" duplicates and non-related studies, according to their title and abstract, were removed. During the "Eligibility phase" the full texts of the remaining studies were assessed for eligibility. The Identification phase was completed by the main author. For the other two phases (Screening and Eligibility phase) an additional reviewer examined the titles, abstracts and full texts of the articles for eligibility. Disagreements were resolved by discussion between the two reviewers. If no agreement was reached, the decision was taken by a third reviewer.

DATA COLLECTION PROCCESS

Q8) Have the methods used to collect data from reports been described and are they appropriate?

- Have the methods for collecting the data from the reports (e.g. data extraction sheets) been described?
- How many reviewers collected the data?
- Did they work independently?
- How disagreements were resolved?
- Have processes or obtaining or confirming relevant data been described?

"We designed a data extraction form based on that used by Lumley 2009, which two review authors (RC and TC) used to extract data from eligible studies. Extracted data were compared, with any discrepancies being resolved through discussion. RC entered data into Review Manager 5 software (Review Manager 2014), double checking this for accuracy. When information regarding any of the above was unclear, we contacted authors of the reports to provide further details."[175]

DATA ITEMS

Q9) Have all variables for which data were sought been listed and defined? Are the data selected appropriate?

Data may be relevant to

a) Outcome measures

b) Other variables

Example regarding the outcome measures

"Any measure of gait was eligible for inclusion. It was expected that individual would report data for multiple gait outcomes. Specifically, a single study may report results for:

a) kinematic gait characteristics based on motion analysis,

b) kinematic gait characteristics based on clinical tests

c) kinematic gait characteristics based on specialized equipment

d) functional gait based on scales. Any time-point measurement of gate was included in the review."

Example regarding other variables

Data was collected on the:

report: author, year, country

Study: research design

Population: type of stroke, stroke sequence, time from stroke onset, age, gender, functional level

Intervention: type of NDT, duration of each session, number of sessions, frequency

EFFECT MEASURES

Q10) Has the effect measure(s) used in the presentation of results been specified for each outcome? Is it appropriate?

- Have the **effect measure(s) of each outcome** (e.g. risk ratio, mean difference, d, Pearson) been specified?

- Have **thresholds or ranges to interpret the size of effect** (e.g. small, moderate, large effect) been described?

Note: Presentation of the results of the studies by simply reporting the probability level (p value) is not ideal.

Example

The results of each study were presented by using Cohen's d. Cohen's d can be interpreted as small (d=0.2-0.5), moderate (d=0.5-0.8) or large (d>0.8).

STUDY RISK OF BIAS ASSESSMENT

Q11) Have the methods used to assess risk of bias in the included studies been specified? Are they appropriate?

- **How the bias of the included studies was assessed?**
- **Were the tools used appropriate (valid and reliable)?**
- **How many researchers assessed the studies?**
- **Did they work independently?**

- Simple scores for each study are not enough (each internal validity threat should be assessed)

- ROB II and PEDro scale are the most frequent tools for experimental studies

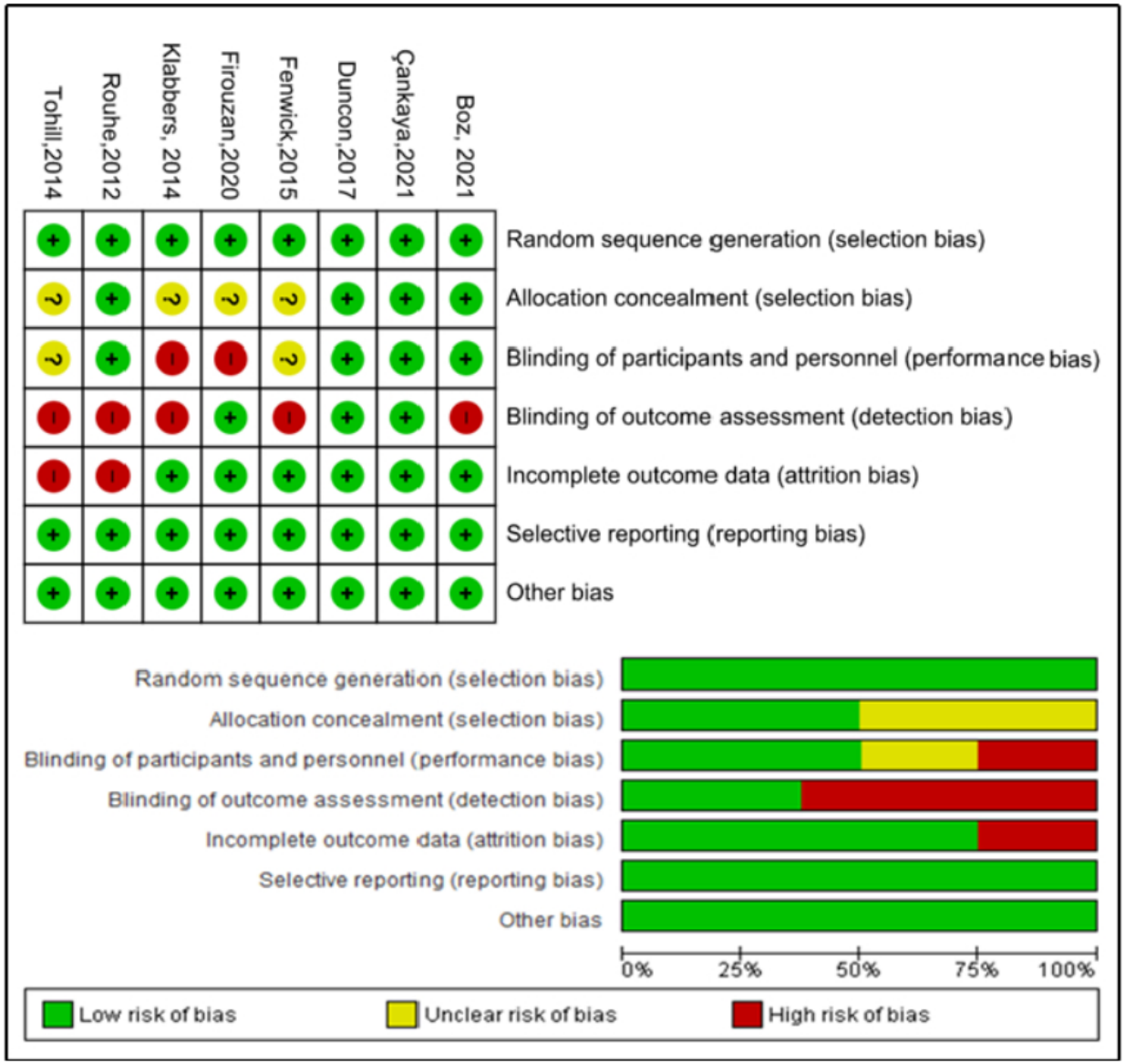
- Each methodology has its own tool

PEDro scale

1. **Eligibility** criteria were specified	no ❑ yes ❑ where:
2. **Subjects** were randomly allocated to groups (in a crossover study, subjects were randomly allocated an order in which treatments were received)	no ❑ yes ❑ where:
3. **Allocation** was concealed	no ❑ yes ❑ where:
4. **The** groups were similar at baseline regarding the most important prognostic indicators	no ❑ yes ❑ where:
5. **There** was blinding of all subjects	no ❑ yes ❑ where:
6. **There** was blinding of all therapists who administered the therapy	no ❑ yes ❑ where:
7. **There** was blinding of all assessors who measured at least one key outcome	no ❑ yes ❑ where:
8. Measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups	no ❑ yes ❑ where:
9. All subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome was analysed by "intention to treat"	no ❑ yes ❑ where:
10. The results of between-group statistical comparisons are reported for at least one key outcome	no ❑ yes ❑ where:
11. The study provides both point measures and measures of variability for at least one key outcome	no ❑ yes ❑ where:

PEDro Scale. Modified from PEDro Physiotherapy Evidence Database:http://www.pedro.org.au/25

ROB II


Tohill,2014
Rouhe,2012
Klabbers, 2014
Firouzan,2020
Fenwick,2015
Duncon,2017
Çankaya,2021
Boz, 2021
Random sequence generation (selection bias)
Allocation concealment (selection bias)
Blinding of participants and personnel (performance bias)
Blinding of outcome assessment (detection bias)
Incomplete outcome data (attrition bias)
Selective reporting (reporting bias)
Other bias
Random sequence generation (selection bias)
Allocation concealment (selection bias)
Blinding of participants and personnel (performance bias)
Blinding of outcome assessment (detection bias)
Incomplete outcome data (attrition bias)
Selective reporting (reporting bias)
Other bias
0%
25%
50%
75%
100%
Low risk of bias
Unclear risk of bias
High risk of bias

Example

Quality assessment of the trials

The quality of the included trials was assessed with the PEDro scale.[22] PEDro scale is an 11-item scale for assessing the validity of randomized controlled trials. The first item is about external validity and it is not included in the total score. Each other item represents an internal validity criterion which takes 1 point at its fulfillment. The final score may range from 0 (low validity) to 10 (high validity). These items are concerned with the random allocation, allocation concealment, baseline comparability, blinding of therapists, patients and raters, experimental mortality, intention-to-treat analysis, statistical comparisons and point measures and measures of variability. The validity of each study can be considered as poor (PEDro score ≤ 3), fair (PEDro score 4-5), high (PEDro score 6-8) or excellent (PEDro score 9-10). Selective reporting was assessed with the fifth domain of the Revised Cochrane risk of bias tool for randomized trials (RoB 2) which classify the studies as low risk, some concerns and high risk of bias.[23] Each study was assessed by two reviewers. Disagreements were resolved by discussion between them. If no agreement was reached, the decision was taken by a third reviewer.

REPORTING BIAS ASSESSMENT

Q12) Have any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases) been described?

"Bias due to missing results" may arise from "reporting biases" such as:

a) selecting non-publication (publication bias)
b) selective non-reporting of the results

- Funnel plots, Egger's test (for publication bias)
- Comparison with the registered protocols (for reporting bias)

- ROB II (6th dimension)

"To assess small-study effects, we planned to generate funnel plots for meta-analyses including at least 10 trials of varying size. If asymmetry in the funnel plot was detected, we planned to review the characteristics of the trials to assess whether the asymmetry was likely due to publication bias or other factors such as methodological or clinical heterogeneity of the trials. To assess outcome reporting bias, we compared the outcomes specified in trial protocols with the outcomes reported in the corresponding trial publications; if trial protocols were unavailable, we compared the outcomes reported in the methods and results sections of the trial publications."[187]

CERTAINTY ASSESSMENT

Q13) Have any methods used to assess certainty (or confidence) in the body of evidence for an outcome been described? Are they appropriate?

- How many researchers performed the assessment?
- How they reach to an agreement?

✓ The **GRADE approach** is potentially the most popular tool to assess certainty

GRADE rating of Quality of evidence

Table 1. GRADE certainty ratings

Certainty	What it means
Very low	The true effect is probably markedly different from the estimated effect
Low	The true effect might be markedly different from the estimated effect
Moderate	The authors believe that the true effect is probably close to the estimated effect
High	The authors have a lot of confidence that the true effect is similar to the estimated effect

- Risk of bias
- Imprecision
- Inconsistency
- Indirectness
- Publication bias

- Large magnitude of effect
- Dose-response gradient
- All residual confounding would decrease magnitude of effect (in situations with an effect)

Example

Quality of the evidence

Quality of evidence for pain and disability was assessed by using the GRADE approach. According to this approach, the evidence can be downgraded based on limitations in study design or execution (1-2 levels), inconsistency of results (1-2 levels), indirectness of evidence (1-2 levels), imprecision (1-2 levels), publication bias (1-2 levels) and can be upgraded based on a large magnitude of effect size (1-2 levels), a dose-response gradient (1 level) and an effect of plausible residual confounding (1 level). Quality of evidence may be rated as very low, low, moderate and high.

RESULTS

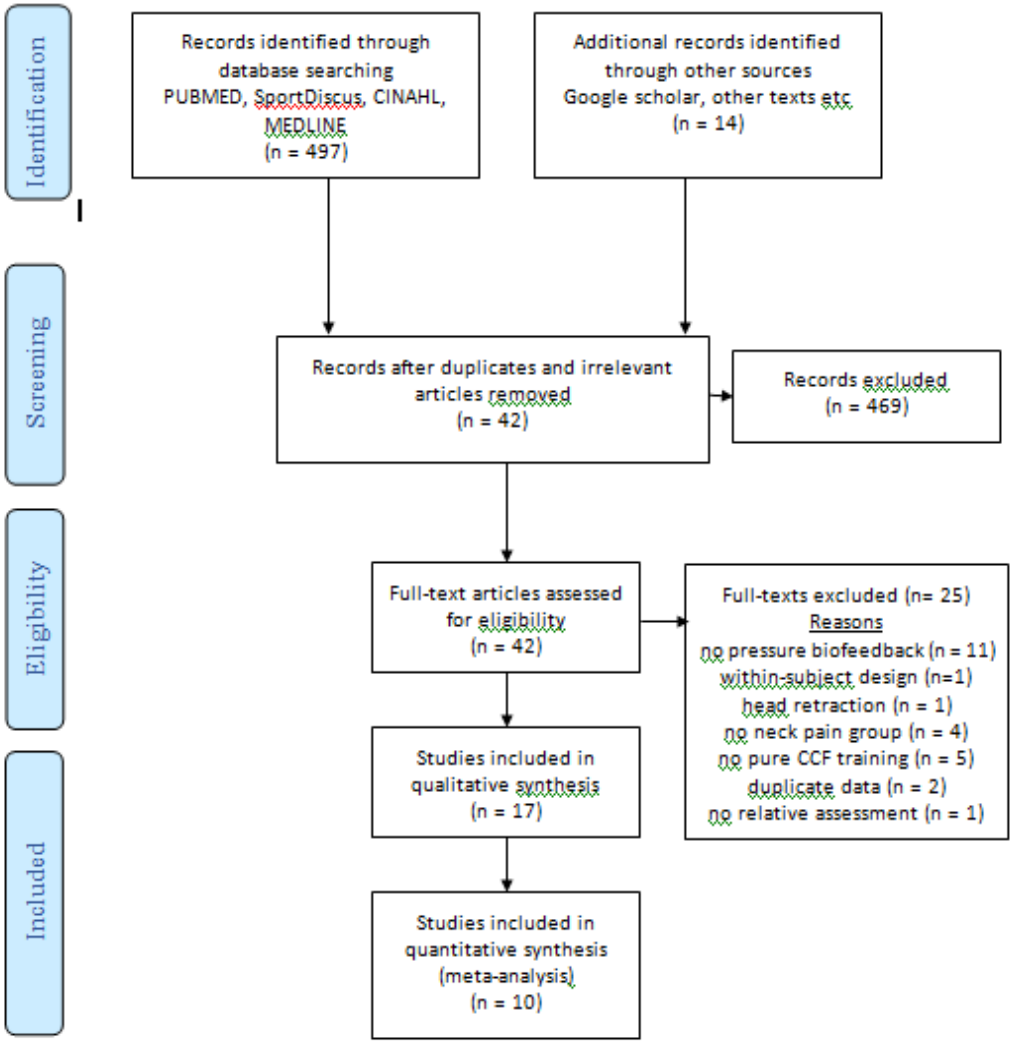
STUDY SELECTION

Q14) Have the results of the search and selection process been clearly described and presented?

- **Number of records identified** in each sage should be stated
- **Excluded studies should be cited** and reasons or exclusion should be stated (during the eligibility phase)
- **A flow diagram** is strongly advised to be used.

Example

Figure 5.1: Study selection flow chart



STUDY CHARACTERISTICS

Q15) Has each included study be cited and its characteristics be presented?

- Usual characteristics which are described are:

  - Authors, date
  - Country of origin
  - Study design features
  - Characteristics of Participants
  - Characteristics of the Interventions
  - Data Collection / Measurements / Outcome Measures

- The characteristics are usually presented with the use **Tables**.

- For studies with interventions, an additional able with the characteristics of the intervention is suggested

Example

Table 1: Description of the included studies

STUDY	DESIGN	SAMPLE	INTERVENTIONS	ASSESSMENT	RESULTS
Borisut et al., 2013	RCT	100 female patients with intermittent work-related chronic neck pain > 6 months, 20-35 y, VAS > 30 mm Experimental group n=25, age 30.4±3.54 y, VAS pain 56.04±22.66, NDI 29.96±4.51 Control group A n=25, age 32.72±3.11 y, VAS pain 55±10.98, NDI 28.2±5.56 Control group B n=25, age 30.16±2.96 y, VAS pain 61.48±16.68, NDI 29.23±5.27 Control group C n=25, age 29.32±3.11 y, VAS pain 59.04±10.49, NDI 31.56±5.14	Experimental group Craniocervical flexion exercise group. Pain-free low load DNF training with a pressure biofeedback device from supine (15 repetitions each trial, 10 sec trial duration for each target level, 10 sec relaxation intervals between the trials) (daily, 12 weeks) Control group A Strength-endurance exercise group. Progressive resistance exercise program. Pain-free progressive resistance exercises for neck flexors and extensors. Neck flexion and extension from a supine and prone position. During the first stage (4 weeks) the patients performed 12-15 repetitions with an individualized weight. During the second stage (8 weeks), the patients performed 3 sets of 15 repetitions with 1-minute intervals between the sets (daily, 12 weeks) Control group B Combined exercise group. Both craniocervical flexion exercises and strength endurance training. For each session, the patients initially performed craniocervical flexion training and after a 5-minute rest period they performed strength-endurance exercises (daily, 12 weeks)	Disability (NDI) and neck pain intensity (VAS). 2 assessments (baseline, 12 weeks)	Pain and disability were significantly improved for the experimental (MDpain: 13 mm, MDdisability: 15.55), control A (MDpain: 16.32 mm, MDdisability: 13.51) and control B (MDpain: 44.6 mm, MDdisability: 13.52), but not for the control C group (MDpain: -2.28 mm, MDdisability: -2.3). The control B group presented the significantly highest improvement in pain and control group C the lowest improvement. Experimental and control group A presented statistically similar improvement. Disability was improved statistically similarly for all groups, except for the control C group.

			Control group C Control group. No intervention is described.		
Cho et al., 2018	RCT	31 patients with mechanical neck pain and forward head posture Experimental group n=16, age: 24±2.5 y, f/m 11/5, NPRS pain: 4.4±1.2, pain chronicity: 31.7±24.9 months Control group n=15, age: 24.1±3 y, f/m 10/5, NPRS pain: 5±1.4, pain chronicity: 18.1±15.4 months	Experimental group Low load DNF training with a pressure biofeedback device from supine (3 sets, 10 sec contractions, 5-sec intervals between the trials, 30-sec intervals between sets, total exercise < 10 minutes) (4 weeks, 10 sessions). Control group Mobilization treatment. C1-2 bilateral segmental mobilizations from a sitting position for improving flexion. T1-2 segmental mobilisations from a prone position for improving extension of the upper thoracic spine. Mobilizations were based on the Kaltenborn's approach (convex-concave rule) (totally < 10 minutes with 10-sec intervals per joint mobilization, 30 sec for intersegmental movements) (4 weeks, 10 sessions).	Pain intensity (NPRS). 3 assessments (baseline, 4 weeks, 6 weeks follow-up).	Both groups improved in pain intensity after the completion of interventions [experimental group: MD 1.2(95% CI 0.7, 1.7), control group: MD 2.5(95% CI 2.0, 3.0)]. This improvement remained after the 2-week follow-up period (experimental group: MD 0.1 change, control group: MD 0.1 change). The pain intensity improvement was significantly higher for the mobilisation (control) group [(4 wk: MD 1.3(95% CI -0.7, 2.0), 6wk: MD 1.4(95% CI 0.5, 2.2)].
Chung et al., 2012	Non-RCT	35 patients with chronic neck pain > 3 months Experimental group n=17, age: 35 y, NDI 19.71±6.48, pain chronicity: 12 months Control group n=18, age: 37 y, NDI 20.17±5.68,	The exercises were initially performed in the hospital and progressively they were performed more and more frequently at home over the period of the study. Both groups performed stretching exercises for each cervical direction as warm-up and cool down exercises. Experimental group Low-load DNF training with a pressure biofeedback device from supine (from 20mmHg to 40 mmHg at 4 mmHg increements, week 1: 12 10-sec holds, week	Disability (Korean-NDI). 2 assessments (baseline, 8 weeks).	Both groups significantly improved in disability (experimental: MD 9.47±5.02, control: MD 6.0±1.53). The improvement in disability for the experimental group was significantly 3.47 units higher than the improvement for the control group.

RISK OF BIAS IN STUDIES

Q16) Have assessments of risk of bias for each included study been appropriately presented?

- **Assessment of each dimension** of internal validity / risk of bias or each study separately
- The most frequent tools for RCTs/CCTs → **PEDRO, ROB II**
- Other tools for other methodologies → Newcastle-Otawa Scale, etc.
- **Use of tables or figures** indicating for each study the risk of bias in each domain/component/item assessed and overall study-level risk of bias

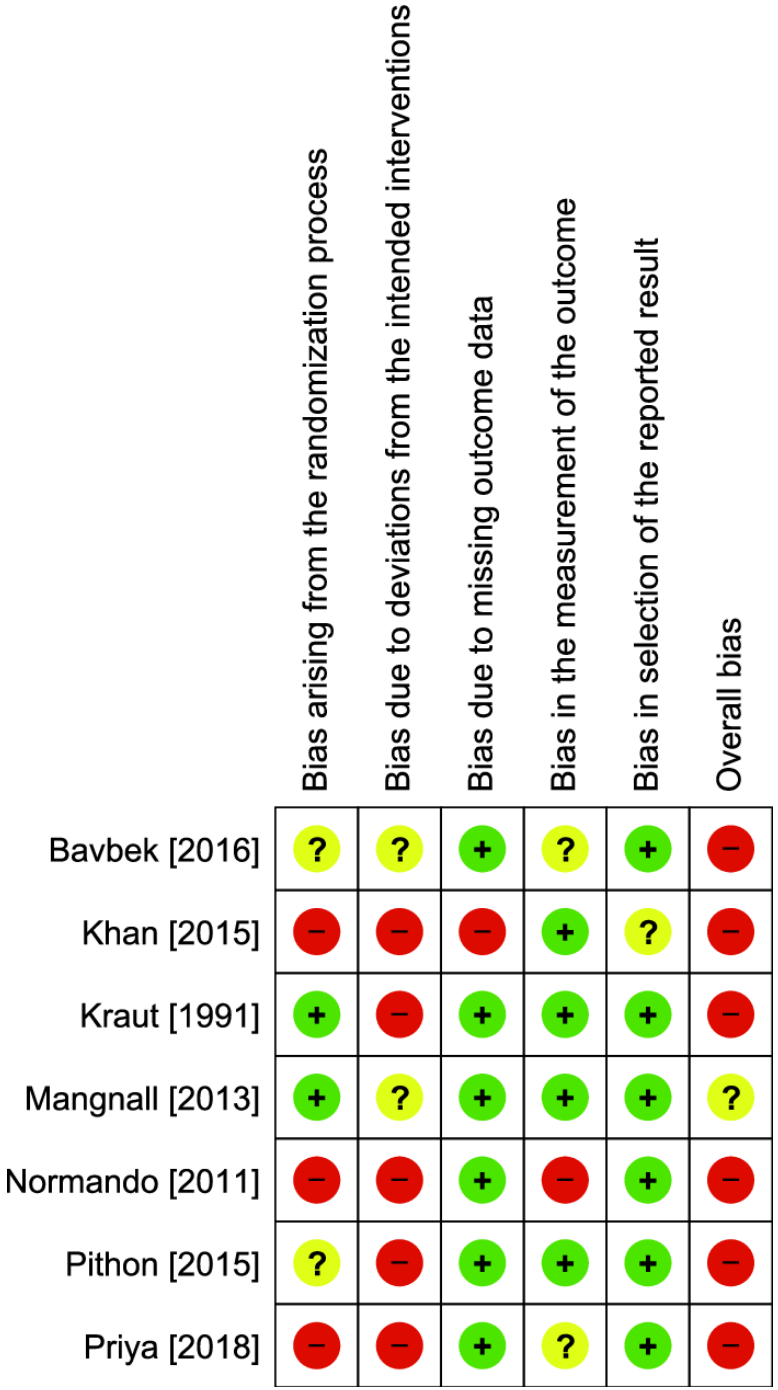
Example

PEDro

Study	1	2	3	4	5	6	7	8	9	10	11	Tota
Akbari et al. (2006) [33]	-	Yes	No	Yes	No	No	Yes	No	No	Yes	Yes	5
Coroian et al. (2018) [29]	-	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes	7
Dehno et al. (2021) [30]	-	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes	7
Fernandes et al. (2015) [32]	-	No	No	No	No	No	No	Yes	No	No	Yes	2
Fernandez et al. (2016) [34]	-	Yes	No	Yes	No	No	Yes	Yes	No	Yes	Yes	6
Flansbjer et al. (2008) [35]	-	Yes	No	Yes	No	No	No	Yes	Yes	Yes	No	5
Mun et al. (2019) [36]	-	Yes	No	Yes	No	No	No	No	No	Yes	Yes	4
Lattouf et al. (2021) [37]	-	Yes	No	Yes	No	No	No	No	No	Yes	Yes	4
Patten et al. (2013) [31]	-	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes	7

Example

ROBII



RESULTS OF INDIVIDUAL STUDIES

Q17) For all outcomes, for each study should be presented:

(a) summary statistics for each group (where appropriate) and

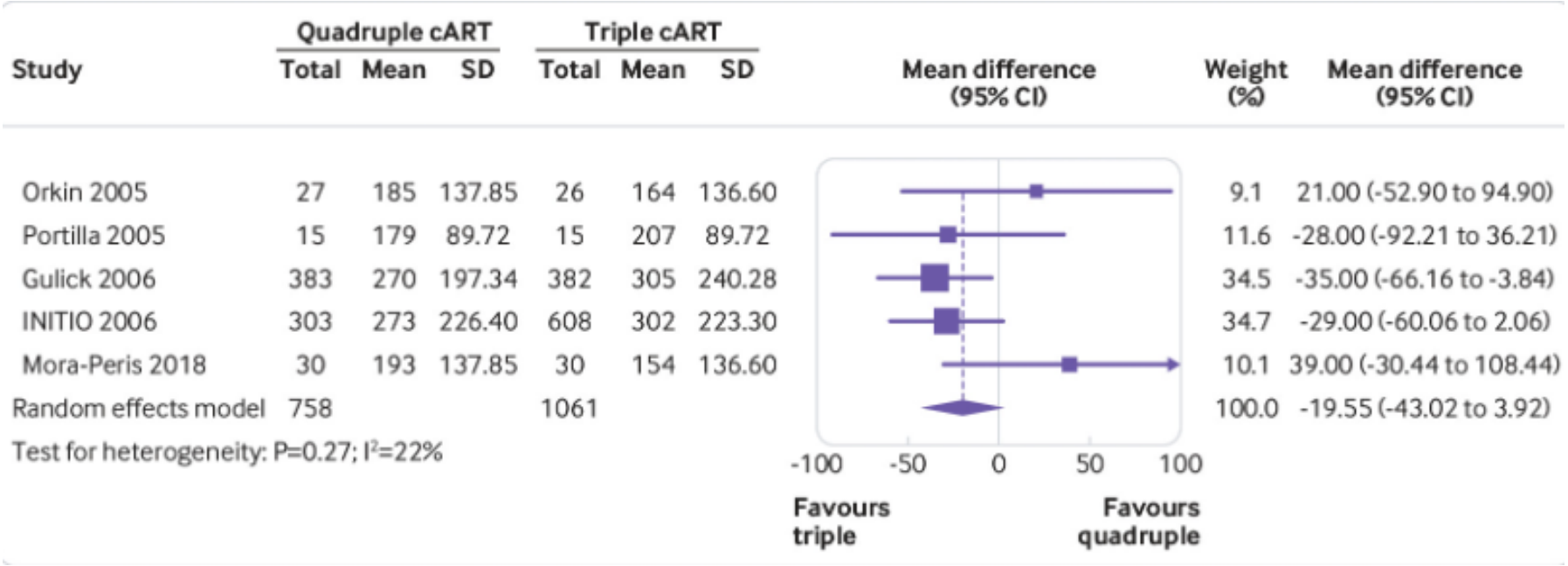
(b) an effect estimate and its precision (such as confidence/credible interval), ideally using structured tables or plots

Example

PAIN INTENSITY DATA												
STUDY	TIME	PRESSURE BIOFEEDBACK			STRENGTH-ENDURANCE			MEAN DIFF BEF-AFT (PB)	MEAN DIFF BEF-AFT (S-E)	Intervention period	Intention to treat analysis	Pooled baseline SD
		M	SD	n	M	SD	n					
Borisut et al., 2013	BEFORE	56.04	22.66	25	55	10.98	25	13	16.32	12 weeks	No need	
	AFTER	43.04	18.56	25	38.68	9.49	25					
Chung and Jeong, 2018	BEFORE	4.85	1.56	25	5.26	0.99	25	2.13	1.29	8 weeks	YES	
	AFTER	2.72	1.28	22	3.97	0.87	19					
Falla et al., 2006	BEFORE	3.6	2	29	4.7	2	29	0.9	1.1	6 weeks	No need	
	AFTER	2.7	N / A	29	3.6	N / A	29					
Ghaderi et al., 2017	BEFORE	61.35	27.9	20	59.73	22.6	20	39.62	39	10 weeks	No need	
	AFTER	21.73	15.9	20	20.73	11.3	20					
Gupta et al., 2013	BEFORE	5.27	0.704	15	5.33	0.724	15	1.47	0.6	4 weeks	No need	
	AFTER	3.80	0.676	15	4.73	0.704	15					
Javanshir et al., 2015	BEFORE	4.97	2.39	30	5.07	2.15	30	2.7	2.32	10 weeks	No need	
	AFTER	2.27	1.51	30	2.75	1.41	30					
Jull et al., 2009	BEFORE	4.5	1.6	23	4.2	2.1	23	1.7	1	6 weeks	No need	
	AFTER	2.8	N / A	23	3.2	N / A	23					
Kim and Kwag, 2016	BEFORE	5.2	2.1	15	5.1	2.7	15	1.7	1.3	4 weeks	NO	
	AFTER	3.5	2	14	3.8	2	14					
O'Leary et al., 2012	BEFORE	33.2	13	20	29.9	14.5	20	19.2	9	10 weeks	YES	
	AFTER	14	10.2	19	20.9	18	20					

Example

Forest-plot



RESULTS OF SYNTHESES

Q18) Is there an appropriate informative qualitative synthesis of data?

- In systematic reviews, **qualitative synthesis** is performed
- It is a narrative, textual approach, analyzing and assessing he body of knowledge included in the review

In this synthesis:

- A general summary of the characteristics and findings of the included studies is provided
- The relationships between studies, exploring patterns and investigating heterogeneity are analyzed
- PICO is a nice guidance for such an analysis

Example

Studies characteristics

Quality of studies

Fifteen out of the 17 detected studies were RCTs[19,48-58,60-62], whereas only two were non-RCTs.[20,59]

Based on PEDro scale, 2 studies were of poor quality[20,59], 7 studies were of fair quality[19,52-54,56,58,62], and 8 studies were of high quality[48-51,55,57,60-61](Table 2).

Participants

Fourteen studies included patients with chronic NP[20,48-58,61-62], whereas only 3 included mixed samples of patients with acute or chronic NP.[19,59-60]

Four studies included only female patients with NP[48-50,52], whereas the remaining studies seem to have included mixed samples of male and female patients.[19,20,51,53-62]

Intervention

The motor control training of DNF with a pressure stabilizer was compared with different interventions across different studies. Some of these studies included multiple groups and therefore the intervention of interest was compared with more than one intervention.[51-52,57,62] More specifically, the motor control training of DNF with a PB device was compared in 10 studies with an endurance-strength training of cervical muscles[20,48,50-54,56,58,61], in 1 study with motor control training of DNF without PB[19] in 1 study with motor control training of deep neck extensors[62], in 1 study with stabilization training with a Swiss ball[59], in 1 study with mobilization treatment[60], in 2 studies with proprioceptive training of cervical muscles[49,55], in 1 study with no intervention at all[52], in 2 studies with active cervical movements[51,57] and in 1 study with a conventional physiotherapy program.[62] All the studies with the exception of one[20] used the classic approach of the motor control training of DNF with PB.

Outcome measures

All the studies included assessment of both NP intensity and disability, with the exception of one study[20] which included only assessment of NP-induced disability and another study[60] which included only assessment of NP intensity.

Disability was assessed with Neck Disability Index in all included studies. NP intensity was assessed with visual analog scales or numeric rating scales in all included studies.

Example

Effectiveness of the intervention

After classifying the studies according to their results in the two dependent variables (NP intensity, disability), the following results derive:

Neck pain intensity

All the studies showed that motor control training of DNF with PB is effective on reducing NP intensity. The duration of the intervention ranged from 4 to 12 weeks across studies. In terms of comparisons with other interventions, in 6 studies the motor control training of DNF with PB was more effective on pain reduction in comparison with the control condition[19,53,56-57,59,61], in 7 studies both interventions had similar effectiveness[48-51,54-55,58] and in 1 study the results favored the control condition.[60] In another study[52], PB training was found to be more effective on pain reduction than 'no treatment', equal effective to strength-endurance training and less effective than a combined treatment (PB training + strength/endurance training) of cervical muscles. In another study[62] training with PB had better effectiveness in comparison with the one control intervention (conventional physiotherapy), but equal effectiveness to motor control training of deep neck extensors. Only 4 studies[51,56,60,62] used a follow-up period (2-16 weeks) with variable results.

Disability

All the studies showed that motor control training of DNF with PB is effective on reducing NP-induced disability. The duration of the intervention ranged from 4 to 12 weeks across studies. In terms of comparisons with other interventions, in 6 studies the motor control training of DNF with PB was more effective on reduction of NP-induced disability in comparison with the control condition[19,20,53,56-57,59] and in 8 studies both interventions had similar effectiveness.[48-51,54-55,58,61] In another study[52], PB training was found to be more effective on disability improvement than 'no treatment', but equal effective to strength-endurance training or a combined treatment (PB training + strength/endurance training) of cervical muscles. In another study[62] training with PB had better effectiveness in comparison with the one control intervention (conventional physiotherapy), but equal effectiveness to motor control training of deep neck extensors. Only 3 studies[51,56,62] used a follow-up period (4-16 weeks) with rather promising effects for the PB training.

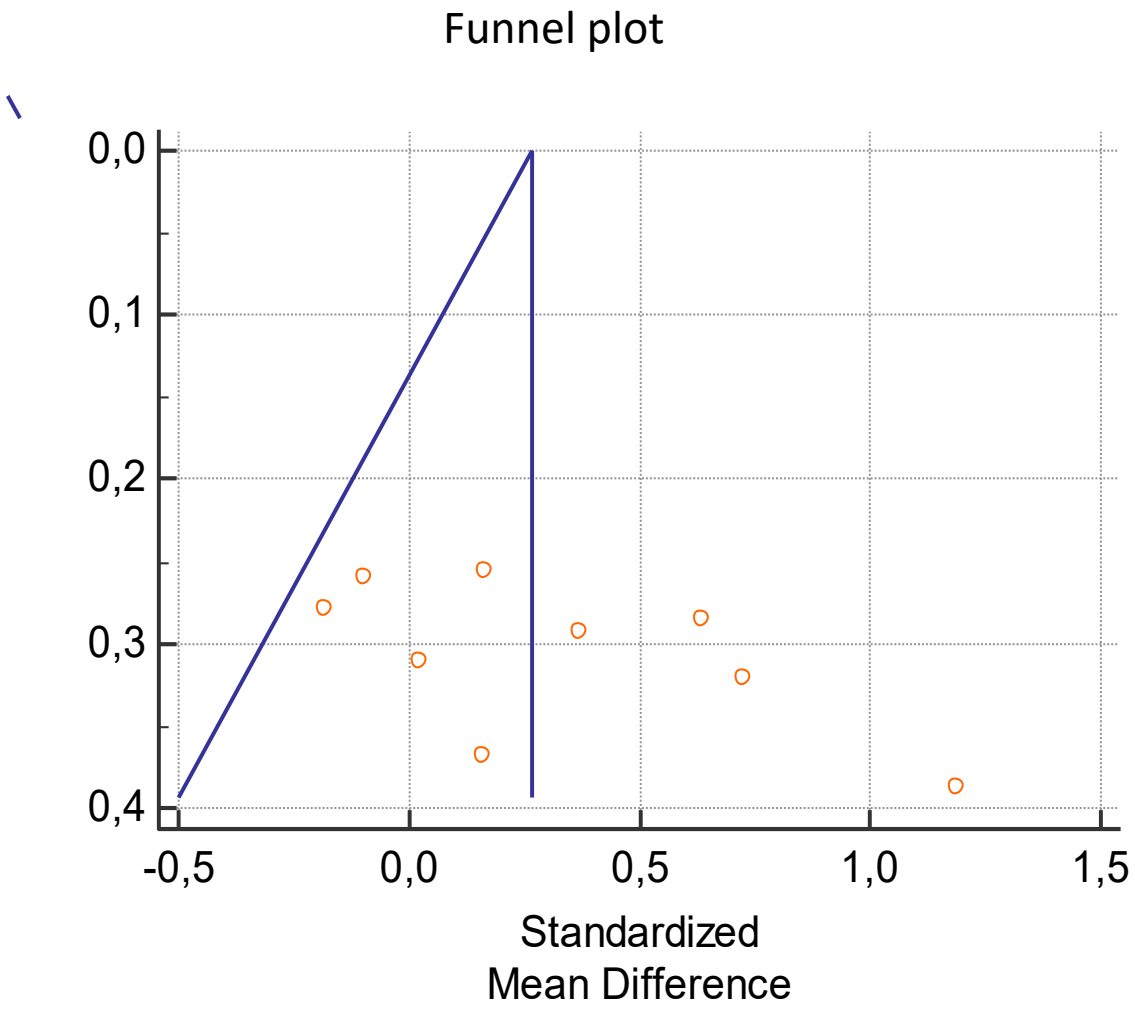
RISK OF REPORTING BIASES IN SYNTHESIS

Q19) Have the assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed been presented?

- Have the results of reporting bias been presented?

- Selective non-reporting, funnel plots

Example


Funnel plot
0,0
0,1
0,2
0,3
0,4
-0,5
0,0
0,5
1,0
1,5
Standardized
Mean Difference

CERTAINTY OF EVIDENCE

Q20) Have the assessments of certainty (or confidence) in the body of evidence for each outcome assessed been presented?

- Have the results of the selected tool (e.g. **GRADE approach**) have been reported (overall level of certainty)?

- Have the reasons for upgrading or downgrading the level of evidence been provided?

Example

Based on the GRADE approach the quality of evidence was graded as 'low' for both pain and disability. There were no considerable problems with inconsistency, indirectness, imprecision and publication bias to downgrade the quality of evidence. The effect sizes were not of large magnitude to upgrade the quality of evidence. However, the study design and execution were not satisfactory. The nature of the studies did not permit the blinding of subjects, therapists or assessors. Allocation concealment was usually not applied. Furthermore, an adequate follow-up and an intention-to-treat analysis were not always appropriately applied or reported. These methodological problems led to downgrade the quality of evidence by 2 levels for both pain and disability.

DISCUSSION

DISCUSSION

Q21) Is there an appropriate discussion which includes:

a) a general interpretation of the results in the context of other evidence?

b) limitations of the studies included in the systematic review?

c) limitations of the current systematic review?

d) clinical or policy implications?

e) implications and suggestions or future research

QUESTIONS?