Supplementary Table 1. Criteria evidence for the GRADE

Downgrade	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias
-1	If one or more of the three criteria (randomization, masking, drop-out rate ≤30%) is not met in **10-30%** of trials included	**I^2 50-74%**	The question being addressed by the guideline panel is **different** from the available evidence regarding the PICO or regarding the characteristics of those who will deliver the intervention	(a)The **overall number** of individuals included in trials is low (less than **200 individuals,** both treatment arms) **OR** (b) the 95% confidence interval includes both 1) no effect and 2) appreciable benefit (RR: ≤0.75) or appreciable harm (RR: ≥1.25) [a]	-
-2	If one or more of the three criteria (randomization, masking, drop-out rate≥30%) is not met in **>30%** of trials included	**$I^2 \geq 75\%$**	The question being addressed by the guideline panel is **markedly different** from the available evidence regarding the PICO or regarding the characteristics of those who will deliver the intervention	(a) the **overall number** of individuals included in trials is **very low** (**less than 200 individuals,** both treatment arms) **AND** (b) the 95% confidence interval includes both 1) no effect and 2) appreciable benefit (RR: ≤0.75) or appreciable harm (RR: ≥1.25) [a]	**Egger's test** (p-value) <0.05

[a] For **continuous outcomes** "***no effect***" means a SMD with a confidence interval that ***crosses zero***; **appreciable** benefit or appreciable harm means that the **upper or lower confidence limit crosses an effect size of 0.5** in either direction.
For **dichotomous outcomes** "***no effect***" means an estimate with a confidence interval that ***crosses one***; **appreciable** benefit or appreciable harm means that the upper or lower confidence limit **crosses a risk of 1.25 or 0.75**.

Abbreviations: OR: Odds ratio**; PICO:** Population, Intervention, Comparison and Outcomes; **RR**: Risk Ratio

Supplementary Table 2. Additional analyses for the meta-analyzable outcomes

Population	Outcome	Type of ES	Number of RCTs	Effect size	Low 95%CI	High 95%CI	P-value	I^2	Egger's test	p-value Egger	Largest study significant	Low 95% prediction intervals	High 95%prediction intervals
OA	Lesquene index	SMD	3	0.363	0.202	0.524	0.00001	0.0	7	0.126	yes	-0.68	1.41
KOA	JSW 3 years	SMD	2	0.432	0.235	0.628	0.00002	0.0	29.79198	NA	yes	NP	NP
OA	Pain	SMD	21	-0.646	-0.910	-0.382	0.00002	88.0	-4.46366	0.003	no	-1.84	0.55
OA	JS narrowing	SMD	2	0.410	0.210	0.600	0.00004	0.0	-1.96	NA	yes	NP	NP
KOA	OA progression	OR	2	0.382	0.216	0.677	0.00100	0.0	-1.53391	NA	yes	NP	NP
OA	JSW	MD	2	0.250	0.120	0.380	0.00242	0.0	1.367442	NA	yes	NP	NP
KOA	pain VAS	WMD	5	-9.507	-17.218	-1.797	0.01567	71.5	-5.0462	0.298	no	-35.11	17
OA	Mobility	SMD	2	0.501	0.091	0.912	0.01664	0.0	-0.39596	NA	yes	NP	NP
KOA	WOMACi	WMD	6	-3.903	-7.148	-0.658	0.01841	52.4	-0.68138	0.622	yes	-12.84	5.04
OA	JSW	SMD	2	-0.192	-0.385	0.001	0.05128	0.0	15.37	NA	yes	NP	NP
KOA	WOMAC pain stiffness	WMD	4	-0.525	-1.055	0.005	0.05210	78.0	2.498699	0.134	no	-2.74	1.69
KOA	WOMAC physical function	WMD	7	-2.947	-6.068	0.173	0.06414	80.4	-0.98144	0.667	no	-13.03	7.14
OA	Being a responder	RR	4	1.548	0.940	2.548	0.08598	67.5	1.682166	0.133	yes	0.2	11.95
OA	VAS pain	SMD	4	0.326	-0.053	0.705	0.09183	61.3	1.429706	0.658	yes	-1.18	1.83
KOA	WOMAC pain	WMD	7	-1.160	-2.839	0.520	0.17595	92.7	2.045143	0.676	yes	-7.07	4.75
KOA	JSW 1 year	SMD	2	0.079	-0.129	0.286	0.45892	12.6	-277.01	NA	no	NP	NP
OA	AdE	OR	5	1.236	0.623	2.454	0.54454	46.0	0.593484	0.337	no	0.17	8.91

The outcomes are ranked by p-value under random effect model

Abbreviations: ES: effect size; RCT: randomized controlled trial; CI: confidence interval; SMD: standardized mean difference; OR: odds ratio; MD: mean difference; WMD: weighted mean difference; OA: osteoarthritis; KOA: knee osteoarthritis; JSW: joint space width; JS: joint space; VAS: visual analogue scale; WOMAC: Western Ontario and McMaster Universities Arthritis Index; AdE: adverse events.

Supplementary Table 3. AMSTAR 2 quality assessment of meta-analyses.

AMSTAR 2 items [a, c]																
Author, Year [Reference]	**1**	**2** [b]	**3**	**4** [b]	**5**	**6**	**7** [b]	**8**	**9** [b]	**10**	**11** [b]	**12**	**13** [b]	**14**	**15** [b]	**16**	**Overall rating** (based on critical domains)[d]
Dostrowski 2011	Y	N	Y	Y	Y	Y	Y	Y	Y	N	No MA conducted	No MA conducted	N	N	No MA conducted	Y	Critically Low
Eriksen 2014	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	High
Gallagher 2014	Y	N	Y	Y	N	N	N	N	Y	N	Y	Y	Y	Y	Y	Y	Critically low
Gregori 2018	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	High
Honvo 2019	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	High
Knapik 2018	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	High
Lee 2009	Y	N	Y	Y	N	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Critically Low
Melo 2018	Y	Y	Y	Y	Y	Y	N	PY	Y	N	No MA conducted	No MA conducted	Y	Y	No MA conducted	Y	Critically Low
Richy 2003	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Low
Simental-Mendìa 2018	Y	N	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y	N	N	Y	Critically Low
Sodha 2011	Y	N	Y	PY	Y	Y	N	Y	Y	N	No MA conducted	No MA conducted	Y	N	No MA conducted	Y	Critically Low

[a] Yes, No, Other
[b] Critical Domains
[c] AMSTAR 2 items:

1. **Did the research questions and inclusion criteria for the review include the components of PICO (Population, Intervention, Comparator group, Outcome)?** YES/NO. For yes, must have all four.
2. **Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?** YES, PARTIAL YES, NO. For Partial YES: the authors state that they had a written protocol or guide that included ALL the following (review question(s), a search strategy, inclusion/exclusion criteria, a risk of bias assessment). For YES: as for partial yes, plus the protocol should be registered and should also have specified: a meta-analysis/synthesis plan, if appropriate, and a plan for investigating causes of heterogeneity, justification for any deviations from the protocol.
3. **Did the review authors explain their selection of the study designs for inclusion in the review?** YES/NO. For YES, the review should satisfy one of the following: explanation for including only RCTs, or explanation for including only NRSI, or explanation for including both RCTs and NRSI.
4. **Did the review authors use a comprehensive literature search strategy?** YES, PARTIAL YES, NO. for PARTIAL YES must have all of the following: searched at least 2 databases (relevant to research question), provided key word and/or search strategy, justified publication restrictions (eg. Language). For YES should also have all of the following: searched the reference lists/biographies of included studies, searched trial/study registries, included/consulted content experts in the field, searched for grey literature where relevant, conducted search within 24 months of completion of the review.

5. **Did the review authors perform study selection in duplicate?** YES/NO. for YES, either ONE of the following: at least two reviewers independently agreed on selection of eligible studies and achieved consensus on which studies to include OR two reviewers selected a sample of eligible studies and achieved good agreement (at least 80 per cent) with the remainder selected by one reviewer.
6. **Did the review authors perform data extraction in duplicate?** YES/NO. For YES, either one of the following: at least two reviewers achieved consensus on which data to extract from included studies OR two reviewers extracted data from a sample of eligible studies and achieved good agreement (at least 80 per cent) with the remainder extracted by one reviewer.
7. **Did the review authors provide a list of excluded studies to justify the exclusions?** YES, PARTIAL YES, NO. FOR partial yes must provide a list of all potentially relevant studies that were read in full text form but excluded from the review. For YES must also have justified the exclusion from the review of each potentially relevant study.
8. **Did the review authors describe the included studies in adequate detail?** YES, PARTIAL YES, NO. For PARTIAL YES, must describe all of the following: populations, interventions, comparators, outcomes, research designs. For YES should also have all of the following: described populations in detail, described intervention and comparator in detail (including doses where relevant), described study setting, timeframe or follow-up.
9. **Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review? For RCTs**: YES, PARTIAL YES, NO, INCLUDES ONLY NRSI. For PARTIAL YES must have assessed RoB from unconcealed allocation and lack of blinding of patients and assessors when assessing outcomes (unnecessary for objective outcomes such as all cause mortality); for YES must also have assessed RoB from allocation sequence that was not truly random and selection of the reported result from among multiple measurements or analyses of a specified outcome. **For NRSI** (Non Randomized Studies of Intervention)**:** YES, PARTIAL YES, NO, INCLUDES ONLY RCTs. For PARTIAL YES must have assessed RoB from confounding and from selection bias. For YES, must also have assessed methods used to ascertain exposures and outcomes, and selection of the reported results from among multiple measurements or analyses of a specified outcome.
10. **Did the review authors report on the sources of funding for the studies included in the review?** YES/NO. For YES: must have reported on the sources of funding for individual studies included in the review. Note: reporting that the reviewers looked for this information but it was not reported by study authors also qualifies
11. **If meta-analysis was performed, did the review authors use appropriate methods for statistical combination of results? For RCTs:** YES, NO, NO META-ANALYSIS. For YES: the authors justified combining the data in a meta-analysis and they used an appropriate weighted technique to combine study results and adjusted for heterogeneity if present and investigated the causes of heterogeneity. **For NRSI:** YES, NO, NO META-ANALYSIS CONDUCTED. For YES: the authors justified combining the data in a meta-analysis and they used an appropriate weighted technique to combine study results, adjusting for heterogeneity if present, and they statistically combined effects estimates from NRSI that were adjusted for confounding, rather than combining raw data, or justified combining raw data when adjusted effect estimates were not available, and they reported separate summary estimates for RCTs and NRSI separately when both were included in the review.
12. **If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?** YES, NO, NO META-ANALYSIS INCLUDED. For YES: included only low risk of bias RCTs or, if the pooled estimate was based on RCTs and/or NRSI at variable RoB, the authors performed analysis ton investigate possible impact of RoB on summary estimates of effect.
13. **Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?** YES/NO. for YES: included only low risk of bias RCTs or, if RCTs with moderate or high RoB, or NRSI were included, the review provided a discussion of the key impact of RoB on the results
14. **Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?** YES/NO. For Yes: there was no significant heterogeneity in the results OR if heterogeneity was present the authors performed an investigation of sources of any heterogeneity in the results and discussed the impact of this on the results of the review
15. **If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?** YES, NO, NO META-ANALYSIS CONDUCTED. For YES: performed graphical statistical tests for publication bias and discussed the likelihood and magnitude of impact of publication bias
16. **Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?** YES/NO. For Yes: the authors reported no competing interests OR the authors described their funding sources and how they managed potential conflicts of interest.

[d] Rating overall confidence in the results of the review:

HIGH: *no or one non-critical weakness*: the systematic review provides an accurate and comprehensive summary of the results of the available studies that address the question of interest
MODERATE: *more than one non critical weakness* (multiple non-critical weaknesses may diminish confidence in the review and it may be appropriate to move the overall appraisal down from moderate to low confidence): the systematic review has more than one weakness but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review
LOW: *one critical flaw with or without non-critical weaknesses*: the review has a critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest
CRITICALLY LOW: *more than one critical flaw with or without non-critical weaknesses*: the review has more than one critical flaw and should not be relied on to provide an accurate and comprehensive summary of the available studies

---

[c] AMSTAR items:
1. Was an 'a priori' design provided? The research question and inclusion criteria should be established before the conduct of the review. *Note: Need to refer to a protocol, ethics approval, or pre-determined/a priori published research objectives to score a "yes."*
2. Was there duplicate study selection and data extraction? There should be at least two independent data extractors and a consensus procedure for disagreements should be in place. *Note: 2 people do study selection, 2 people do data extraction, consensus process or one person checks the other's work.*
3. Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include years and databases used (e.g., Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found. *Note: If at least 2 sources + one **Appendix** strategy used, select "yes" (Cochrane register/Central counts as 2 sources; a grey literature search counts as **Appendix**).*
4. Was the status of publication (i.e., grey literature) used as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc. *Note: If review indicates that there was a search for "grey literature" or "unpublished literature," indicate "yes." SIGLE database, dissertations, conference proceedings, and trial registries are all considered grey for this purpose. If searching a source that contains both grey and non-grey, must specify that they were searching for grey/unpublished lit.*
5. Was a list of studies (included and excluded) provided? A list of included and excluded studies should be provided. *Note: Acceptable if the excluded studies are referenced. If there is an electronic link to the list but the link is dead, select "no."*
6. Were the characteristics of the included studies provided? In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported. *Note: Acceptable if not in table format as long as they are described as above.*
7. Was the scientific quality of the included studies assessed and documented? 'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant. *Note: Can include use of a quality scoring tool or checklist, e.g., Jadad scale, risk of bias, sensitivity analysis, etc., or a description of quality items, with some kind of result for EACH study ("low" or "high" is fine, as long as it is clear which studies scored "low" and which scored "high"; a summary score/range for all studies is not acceptable).*
8. Was the scientific quality of the included studies used appropriately in formulating conclusions? The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations. *Note: Might say something such as "the results should be interpreted with caution due to poor quality of included studies." Cannot score "yes" for this question if scored "no" for question 7.*
9. Were the methods used to combine the findings of studies appropriate? For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e., , Chi-squared test for homogeneity, I2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should

be taken into consideration (i.e., , is it sensible to combine?). *Note: Indicate "yes" if they mention or describe heterogeneity, i.e., , if they explain that they cannot pool because of heterogeneity/variability between interventions.*

10. Was the likelihood of publication bias assessed? An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test, Hedges-Olken). *Note: If no test values or funnel plot included, score "no". Score "yes" if mentions that publication bias could not be assessed because there were fewer than 10 included studies.*

11. Was the conflict of interest included? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies. *Note: To get a "yes," must indicate source of funding or support for the systematic review AND for each of the included studies.*