Appendix 1. Search strategy in the Cochrane Database of Systematic Reviews

1. Meta-analysis/ (166)
2. Exp meta-analysis as topic/ (598)
3. Meta analy*.mp. (39440)
4. Metaanaly*.mp. (915)
5. 1 or 2 or 3 or 4 (1867)
   Online publication date from Jan 1, 2015 to April 1, 2017

Appendix 2. Data Extraction Form

Screener initials:	**Study ID:**	**Author, year: _____, _____**

1.	**Review includes appropriate meta-analysis (MA)?**	☐ No ☐ Network MA ☐ No MA ☐ Yes ☐ Standard MA ☐ Individual participant data-MA	☐ **Exclude**, stop here
2.	**Eligible review with statistically significant PRO(s) reported in the abstract?**	☐ No ☐ No PRO ☐ No statistically significant PRO ☐ Yes	☐ **Exclude**, stop here
3.	**Review described as comparing two or more interventions in humans?**	☐ No, *check all that apply* ☐ Animal studies included ☐ Review of diagnostic test accuracy ☐ Review of prognosis ☐ Yes	☐ **Exclude**, stop here
4.	**Does the review include only randomized controlled trials (RCTs)?**	☐ No ☐ Non-randomized studies (NRS) ☐ Mixed NRS and RCTs ☐ Yes	☐ **Exclude**, stop here

Please fill out this box for each study

☐ **Include**

☐ **Exclude**

☐ **3rd reviewer needed** *(no consensus between 2 reviewers)*

BACKGROUND INFORMATION		
5.	Clinical area *Check only one*	**Medical** ☐ (1)Dermatology ☐ (2)Cardiology ☐ (3)Endocrinology ☐ (4)Gastro Intestinal ☐ (5)Hematology ☐ (6)Intensive Care ☐ (7)Infectious Diseases ☐ (8)Neurology ☐ (9)Oncology ☐ (10)Psychiatric ☐ (11)Renal ☐ (12)Respiratory ☐ (13)Rheumatology ☐ (14)Other *(specify)*: **Surgical** ☐ (1)Cardiac surgery ☐ (2)General surgery ☐ (3)Obstetrics/ Gynecology ☐ (4)Ophthalmology ☐ (5)Orthopedic surgery ☐ (6)Otorhinolaryngology (ENT: Ear Nose Throat) ☐ (7)Neurosurgery ☐ (8)Plastic surgery ☐ (9)Thoracic surgery ☐ (10)Urologic surgery ☐ (11)Vascular surgery ☐ (12)Other *(specify)*:
6.	Intervention *Check only one*	☐ (1)Pharmacological ☐ (2)Surgery/ Invasive procedure ☐ (3)Non-surgical invasive procedure ☐ (4)Rehabilitation ☐ (5)Behavioral intervention ☐ (6)Complementary and alternative medicine *(specify)*: ☐ (7)Lifestyle modification ☐ (8)Other *(specify)*:
7.	Control *Check only one*	☐ (1)Standard care ☐ (2)Placebo ☐ (3)Pharmacological ☐ (4)Surgery/ Invasive procedure ☐ (5)Non-surgical invasive procedure ☐ (6)Rehabilitation ☐ (7)Behavioral intervention ☐ (8)Complementary and alternative medicine *(specify)*: ☐ (9)Lifestyle modification ☐ (10)Waiting list ☐ (11)Other *(specify)*:
8.	Use of GRADE	☐ Yes ☐ No
9.	List all of the non-PROs (i.e. outcomes of morbidity, mortality, surrogate outcomes) assessed in the review	*[text]*
THE REMAINING SECTIONS OF THE FORM WILL BE FILLED OUT FOR EACH PRO THAT IS REPORTED AS STATISTICALLY SIGNIFICANT IN THE REPORT		
CHARACTERISTICS OF PROMs		
10.	What is the construct that the PRO instrument is measuring?	*[text]*
11.	List the full name of the PROM(s) included in the MA of the PRO	*[text]*
12.	Do the authors cite or describe measurement properties associated with any of the PROMs?	☐ Yes, *specify, for each PROM, whether a description, citation, or both was provided:* _____ ☐ No
13.	Approach to PROM selection when more than one PROM capturing the same construct(s) is reported	*[text]*

	within a single trial. If not reported, specify "NR". If only one PROM reported in trial, specify "NA".	
14.	Did the meta-analysis pool same or different PROMs?	☐ Same ☐ Different
CALCULATION AND PRESENTATION OF POOLED PROM EFFECT ESTIMATES		
15.	Summary effect measure – primary analysis	☐ Mean difference (MD) ☐ Standardized mean difference (SMD) ☐ Ratio of means (RoM) ☐ Minimal important difference (MID) units ☐ Risk difference (RD) ☐ Number needed-to-treat (NNT) ☐ Relative risk (RR) ☐ Odds ratio (OR) ☐ Other *(specify)*:
16.	How was the effect measure calculated?	*[text]*
17.	Number of trials	###
18.	# participants in intervention group	###
19.	# participants in control group	###
20.	Effect estimate	###
21.	95% confidence interval	###
22.	Use of control group response rate?	☐ Yes, single value ☐ Yes, range of plausible values ☐ No
23.	If 'yes, a range of possible values' to question 22 above, how many corresponding effect estimates are reported?	###
24.	If 'yes' to question 22 above, provide a source (citation and/or description) for the control group response rate(s). If not reported, specify "NR".	*[text]*
25.	Use of MID(s)?	☐ Yes, single value ☐ Yes, range of plausible values ☐ No
26.	If 'yes, a range of possible values' to question 22 above, how many corresponding effect estimates are reported?	###
27.	If 'yes' to question 22 above, provide a source (citation and/or description) for the MID(s). If not reported, specify "NR".	*[text]*
28.	Did authors pool differences in post-test scores or change (from baseline) scores?	☐ Post-test scores ☐ Change scores ☐ Both
Questions 15-28 will also be completed for sensitivity analysis(es)		
INTERPRETATION OF PROM EFFECT ESTIMATES		
29.	Authors inferences regarding magnitude of effect	*[text]*
30.	Basis of these inferences	*[text]*
31.	Criteria for judging authors' interpretation of pooled estimates from PROMs *Select all that apply*	☐ Dichotomous presentation of PROM data presented (relative or absolute effects) informed by MID or some other meaningful threshold ☐ MID units – presentation of summary effect estimate ☐ RoM – presentation of summary effect estimate ☐ MID – characterization of magnitude of effect ☐ Other meaningful threshold – characterization of magnitude of effect ☐ Cohen's interpretation of effect sizes – characterization of magnitude of

		effect ☐ PROM's total score range – characterization of magnitude of effect ☐ Patient's intuition – characterization of magnitude of effect ☐ Clinician's intuition – characterization of magnitude of effect ☐ Descriptive terms (e.g. trivial, small but important, moderate, large) – characterization of magnitude of effect ☐ Comment on the magnitude of effect outweighing or not outweighing the burden or harms associated with the intervention ☐ Authors indicate a need for empirically determined thresholds to quantify the importance of apparent effects ☐ Other *(specify)*:
32.	Authors' discussion of limitations or uncertainties regarding characterization of effect	*[text]*
33.	Discussion of congruency in presentation approaches? If not reported, specify "NR". If only one presentation approach reported in trial, specify "NA".	*[text]*
34.	Authors' conclusions about the treatment?	☐ Explicit – for ☐ Explicit – against ☐ Explicit – no recommendation ☐ Implicit – for ☐ Implicit – against ☐ Implicit – no recommendation
35.	Basis of the conclusion	*[text]*

Appendix 3. Summary and rationale for small and large treatment effects

Approach	Small effect	Large effect
Standardized mean difference	0.20	0.80
Mean difference in minimal important difference units	0.60	2.00
Mean difference in natural units	0.60 point reduction	2.00 point reduction
Ratio of means	0.92, 8% less	0.63, 37% less
Relative risk	0.80 (80%), 20% RRR	0.50 (50%), 50% RRR
Risk difference	0.04 (4%), 4 per 100 fewer, NNT=25	0.20 (20%), 20 per 100 fewer, NNT=5

Reproduced from Johnston BC, Alonso-Coello P, Freidrich JO, et al. Do clinicians understand the size of treatment effects? A randomized survey across 8 countries. *Canadian Medical Association Journal* 2016;188(1):25-32.

Answers are based on consensus among the group of investigators. The following empirical data and guidelines were used to justify small and large treatment effects.

Standardized Mean Difference (SMD): as a rule of thumb, 0.2 SD represents a small difference, and 0.8 represents a large effect (1). Cohen proposed these values on theoretical grounds, which have subsequently became accepted.

Minimal Importance Difference (MID) units: a consensus statement suggests that a 1 point change on a 10 point pain scale constitutes an MID (2). We have previously published guidelines indicating that half (0.5) the MID is a small treatment effect (3). Reporting results in MID units risks naïve misinterpretation. For example, some may be under the impression that for estimates above 1 MID, the treatment has important benefits for all patients, and estimates below 1 for none. Even if the pooled estimate lies between 0 and 1 (or 0 and -1), treatment may have an important impact on many patients (4). We suggest the following guide for interpretation: if the pooled estimate is greater than 1 MID unit, many patients are likely to gain important benefits from treatment. If the estimate of effect lies between 0.5 and 1 MID unit, the treatment may benefit an appreciable number of patients. As the pooled estimate falls below 0.5 MID units it becomes progressively less likely that an appreciable numbers of patients will achieve important benefits from treatment (3,5).

Mean Difference (MD) in natural units: Please see above.

Ratio of Means (RoM): our estimates were based on Cohen's proposed values, with SMDs of 0.2, and 0.8 corresponding to increases in RoM of approximately 8% and 37%, respectively. This is based on work correlating the SMD to the RoM using 232 meta-analyses that included at least 5 trials and that reported continuous outcomes (6).

Relative Risk (RR) estimates are based on published guidelines on the interpretation of these measures of treatment effect (7).

Risk Difference (RD) estimates are based on published guidelines on the interpretation of these measures of treatment effect (8,9).

References

1. Cohen J. Statistical power analysis for the behavioral sciences. 2nd edition. Hillside, NJ: Lawrence Erlbaum Associates; 1988. pp. 24-7.
2. Dworkin RH, Turk DC, Wyrwich KH, et al. Interpreting the clinical importance of treatment outcomes

in chronic pain clinical trials:
IMMPACT recommendations. Journal of Pain 2009;9:105-21.
3. Johnston BC, et al. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important
difference units. Health Qual Life Outcomes 2010 Oct 11;8:116.
4. Guyatt GH, et al. Interpreting treatment effects in randomised trials. BMJ 1998, 316:690-3.
5. Johnston BC, et al. Patient-reported outcomes in meta-analyses-part 2: methods for improving interpretability for decision-makers.
Health Qual Life Outcomes 2013 Dec 21;11:211.
6. Friedrich JO, et al. Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods. J
Clin Epidemiol. 2011;64(5):556-64.
7. Guyatt GH, et al. GRADE Working Group. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol. 2011
Dec;64(12):1311-6.
8. Guyatt GH, et al. Part B, Therapy; Chapter 8. Confidence Intervals. In: Guyatt G, Rennie D, Meade M, Cook D. Users' Guides to the
Medical Literature: A Manual for Evidence-Based Clinical Practice, Second Edition. McGraw-Hill Professional, 2008.
9. Jaeschke R, et al. Part B, Therapy; Chapter 7. Does Treatment Lower Risk? Understanding the Results. In: Guyatt G, Rennie D, Meade
M, Cook D. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, Second Edition. McGraw-Hill
Professional, 2008.