Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Physical Therapy for Knee Pain Secondary to Osteoarthritis*

Draft review available for public comment from November 2011 to December 2011.

Research Review Citation: Shamliyan TA, Wang S-Y, Olson-Kellogg B, Kane RL. Physical Therapy Interventions for Knee Pain Secondary to Osteoarthritis. Comparative Effectiveness Review No. 77. (Prepared by the Minnesota Evidence-based Practice Center under Contract No. 290-2007-10064 I) AHRQ Publication No. 12(13)-EHC115-EF. Rockville, MD: Agency for Healthcare Research and Quality; November 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Source: http://effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1285
Published Online: November 6, 2012

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	Finally, please incorporate any of these recommendations you implement into the Executive Summary as well so it is consistent with the main report. One discrepancy was already noted above.	We revised the report to avoid discrepancies in data presentation.
John M. Eisenberg Center	Executive Summary	REPORT ERROR: Consistency of SOE ratings per finding across the report. Page ES-20, Table B, for Aerobic Exercises, SOE for Transfers function is listed as Insufficient. However, on page ES-22 Table C and page 26 Table 3, it is listed as Low.	We revised the report to avoid discrepancies in strength of evidence presentation.
John M. Eisenberg Center	Executive Summary	REPORT ERROR: Unrated findings of stated conclusion/outcomes. a. Page ES-20, Table B, for Proprioception Exercises, finding for the outcome of Strength has been listed as Improvement. However, the SOE for this outcome has not been listed. b. Page ES-20, Table B, for Aquatic Exercises, finding for the outcome of Strength has been listed as Inconsistent. However, the SOE for this outcome has not been listed. c. Page ES-20, Table B, for E-stim, finding for the outcome of Strength has been listed as Improvement. However, the SOE for this outcome has not been listed.	We revised the report to avoid discrepancies in strength of evidence presentation.

Commentator & Affiliation	Section	Comment	Response
John M. Eisenberg Center	Executive Summary	TRANSLATION CHALLENGE: Applicability of findings is unclear. Page ES-14 of the report states:"Our review was complicated by the discrepancy between the actual practice of physical therapy and the design of research evaluating these interventions. Published studies have focused on individual physical therapy interventions. In practice, however, physical therapists or physical therapist assistants most often treat their patients with a combination of modalities." Additionally, page 52 of the report states:"Several factors affected the applicability of the research base. This lessened the degree to which our synthesis can fully and accurately address the efficacy and comparative effectiveness of physical therapy interventions for knee OA. Most important, current standards of physical therapy practice involve combined interventions, whereas published studies have examined individual physical therapy interventions." Thus, the report questions the applicability of the findings to the intended audience (clinicians/physical therapists, patients, policymakers). However, the CER does not provide answers to these questions.	We clarified that all RCTs of health care intervention share the same problem of low applicability to real clinical practice. We revised the discussion of applicability of RCTs that examined PT interventions as follows: "Our review was complicated by the fact that published studies rarely have focused on combined physical therapy interventions commonly used in PT settings In addition, clinical care for adults with knee OA includes pharmacological interventions, while our review was limited to nonpharmacological treatments, further complicating our effort. To address such complexity, our review focused on randomized trials since these equally distribute concomitant treatments among treatment groups an thus provided valid estimates of effects of the examined interventions. When trials identified noneffective individual modalities, a combination of such would not result in significant benefits neither i future research not in clinical practice. Randomized trials are gold standard in establishing benefits with health care interventions. However, applicability of findings is limited to the similar settings, treatment, and patient populations. In our review, while randomization might equally distribute the effect of pain relievers, for example, it would not prevent the dampening of potential effects from physical therapy interventions. The trials we examined rarely provided information about all other treatments patients may have received. Nor did the trials analyze outcomes separately in patient subgroups by concomitant treatments. We tried to examine how pain relievers (a common concomitant treatment) may influence physical therapy outcomes for pain, but rare and inconsistent reporting of drug treatments impeded synthesis of evidence. Finally, heterogeneity in populations, treatments, and definitions of the outcomes downgraded strength of evidence to low or moderate in most cases."

Commentator & Affiliation	Section	Comment	Response
John M. Eisenberg Center	Executive Summary (Results)	REPORT ERROR: Insufficient findings of stated conclusion/outcome On pages 26-30, Table 3, (and summarized on page ES-22, Table C), several conclusions/findings statements on effectiveness have been reported with the SOE listed as "Insufficient". According to AHRQ methods of SOE ratings, a declared finding statement rated insufficient should include or be limited to a statement that "conclusions could not be drawn." We would appreciate your clarification	We revised the report to clarify the ranking of the SOE. We defined the evidence from individual high risk of bias study as insufficient. We revised our definitions of precision, following the recently published GRADE recommendations.
John M. Eisenberg Center	The results	One of our Medical Writers, Goethe Ashanti, has reviewed the most recent final draft version of the CER on Physical Therapy for Knee Pain Secondary to Osteoarthritis. There continue to be several inconsistencies and disclarity in SOE rated findings in the report, including some we have identified before and some new ones that have come up since our initial review. Although it is unlikely that we will be producing summary translation products for this report, we want to call these to your attention to ensure that the published final report is accurate. Our initial review identified that there were only insufficient rated findings, which have now been changed to low or moderate. However, the text of the report apparently was not changed to reflect or justify these new ratings.	Following your recommendations we provide rating of strength of evidence at hypotheses level rather than overall body of evidence. We downgraded strength of evidence if risk of bias was moderate or high, if heterogeneity was statistically significant, or if estimates were inconsistent or imprecise. We defined evidence from single studies with high risk of bias as insufficient.
John M. Eisenberg Center	The results	As a result, there are several instances in the report where the SOE indicated in the text is different from the SOE indicated in the corresponding table. I have listed two (but not all) examples of such occurrences below: *For key question 1,* *On page 17, under Comparative Effectiveness of Physical Therapy Interventions, lines 8-10, the report states: Evidence was insufficient for the majority of comparisons (Tables E-F in the Executive Summary and Table 4).* *However, all findings in Table F in the Executive Summary and in Table 4 are rated Low.*	We meant that comparative effectiveness evidence was not available for majority of the possible comparisons the examined PT intervention. We revised the report to avoid confusion as follows:" Limited direct evidence of comparative effectiveness of physical therapy interventions from single studies showed low SOE for the majority of comparisons (Tables C and D)." We also revised all places in the report consistently providing the same ranks of evidence for examined hypotheses in the tables and in the text. We eliminated references to rating SOE for combined areas and instead rated SOE only for specific hypotheses.

Commentator & Affiliation	Section	Comment	Response
John M. Eisenberg Center	The results	*For key question 2,* *On page 33, lines9-10, the report states: Evidence for the association between intermediate and clinical outcomes was limited to individual studies and thus was insufficient to draw valid conclusions (Table 5). And again on page 37, lines 6-10, the report states: In summary, disability measures were associated with gait, mobility restrictions, muscle strength, and range of motion measures, but the magnitude and clinical importance of the associations remain unclear. Individual studies offered insufficient evidence for determining which intermediate outcomes strongly and consistently predict patient-centered outcomes. And yet, Table 5 (page 40) indicates that for findings related to association between intermediate and clinical outcomes; strength of evidence is low from individual observational studies.*	We revised the report consistently reporting the same strength of evidence in the tables and in the text:"Evidence for the association between intermediate and clinical outcomes was limited to individual observational studies and did not show a strong consistent association between changes in intermediate and patient centered outcomes (Table 5)." We removed the overarching SOE rating for this generalization.
John M. Eisenberg Center	The results	We are unclear of the specific rated findings of Key Question 2. *On page 37, the statement is made that Adults with good self-efficacy had an 11 percent relative decrease in risk of poor WOMAC function with reference to Table 8 on page 47. Yet, we are having difficulty determining from this table how this conclusion was drawn.* *Likewise, we do not see a table representing the specific SOE rated findings (with clear conclusion statements) for this Key Question. We understand this data is from observational study, but are having difficulty determining which statements are the actual rated findings for this question.*	We corrected the table to restore decimals in the estimate" 0.89/2.5points. We clarified in the Methods section that we assessed strength of evidence from diagnostic studies according to the United States Preventive Services Task Force criteria.[29] We clarified the title of Table 5 as follows:" Association between intermediate and clinical outcomes; low strength of evidence from individual observational studies" with foot not explanation that "studies not provide strong consistent and unbiased estimates" We corrected the header of this table.
Peer Reviewer #1	Introduction	The authors state "Meanwhile, many physical therapies for knee OA have yet to be evaluated." The cited clinical practice guidelines from the American Academy of Orthopedic Surgeons (http://www.aaos.org/research/guidelines/GuidelineOAKnee.asp) include the evidence tables on which the guidelines are based and among them are systematic reviews (not all were included in their review) on braces/orthoses, patellar taping/bracing, lateral wedge orthoses, and aerobic/strengthening exercises. So either I misunderstood the author's sentence or the breadth of available systematically reviewed interventions is more extensive than claimed by the authors. However I do agree that updating the evidence is warranted, and the current review is timely.	We deleted this sentence.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Introduction	2/5-11 The first paragraph on this page cites non-rehabilitation specific references to claim that patient-centered rather than instrumental tests (intermediate outcomes) should be measured and considered when making treatment and reimbursement decisions. However, later on the page li 35-38 the authors acknowledge that prior systematic reviews have not examined the relationship between these two sets of measures in knee OA. The International Classification of Functioning, Disability and Health framework (http://www.who.int/classifications/icf/en/) provides a context to understand this relationship and why both sets of measurements could be important (instead of only one type). If a person has quadriceps weakness and is unable to rise from a seated position which in turn limits their ability to live independently, this can be treated several ways. One option is to modify their chairs at home (elevate the seat, assure chairs have arms that can be used to assist) to reduce the demands of the task so the person can rise from a chair with their existing level of quadriceps strength. This may allow him/her to continue living alone independently. The second strategy is to implement quadriceps strengthening exercises so the person gains sufficient strength to rise from the greatest variety of chairs possible and can function independently not only in the home but also in the community. A patient-centered outcome measure will identify when this goal has been achieved, but interim measures of muscle strength will document 1) that the quadriceps weakness could be the reason for the disability and help select the appropriate treatment (strengthening exercises) then 2) document that progress is being made in the strengthening exercise program before the critical threshold of strength is achieved that is needed to rise from a chair. An intermediate outcome measure that documents this progress also may help with patient adherence/compliance to the exercise program and ultimately promote success in achieving the patient-centered outcome all desire. With only a measure of disability, the physical therapist may not have sufficient information to develop a treatment plan. But as the authors note, more evidence is needed to support or refute the relationship among these outcome measures. This conclusion could be emphasized much more strongly in the future research section of the report.	We revised this paragraph as follows: “Measurement of physical therapy benefits should address patient-centered outcomes rather than the results of instrumental tests. Additionally, clinicians and policymakers should consider patient-centered outcomes when making treatment and reimbursement decisions. We need to recognize the importance and the relationship between patient-centered and intermediate outcomes for elderly patients with knee OA. With only a measure of disability, the physical therapist may not have sufficient information to develop a treatment plan to prevent disability and improve quality of life. While intermediate outcome measures are helpful to document gradual progress in function and also may help with patient adherence/compliance to the exercise program and ultimately promote success in achieving the patient-centered outcome all desire, research based on patient-centered outcomes provides patients and clinicians the necessary information for effective and informed decisions about physical therapy and other health care services.”
Peer Reviewer #1	Introduction	In the interim, more caution is advised with some of the declarative statements in this section of the report that, as the authors have already noted, are not supported by the evidence.	We revised the report to avoid declarative statements that are not supported by strong evidence.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	Introduction	good intro given with appropriate definitions, etc	Thank you.
Peer Reviewer #4	Introduction	Again, meaningful and well defined.	Thank you.
Reviewer # 5	Introduction	The report is clinically meaningful, the target population and audience are well defined. The key questions are appropriate and clearly stated	Thank you.
Peer Reviewer # 6	Introduction	The introduction is clear and makes a good case for updating existing guidelines. However the authors did not make clear what "physical therapy" interventions are. As noted above first the authors need to explain their rationale for choosing interventions. Their original list of interventions seemed to include every non-pharmocologic intervention that was ever used whether the interventions were actually plausible PT interventions. Balneotherapy is spa therapy found in Europe, massage is not used by PTs for knee OA. Diathermy is rarely or never used anymore by PTs. Conversely the authors include "electrical stimulation" which is not a single intervention. E-stim can be used for pain relief (TENS) or for muscle strengthening (neuromuscular electrical stimulation.) Examining the effectiveness of e-stim as a whole makes no sense.	We clarified that we defined and selected PT interventions according to the Guide to Physical Therapist Practice. We provided a comprehensive list of eligible interventions in Table 1. Physical therapy interventions eligible for review. We synthesized the evidence according to exact intervention and measures definitions including types, regimes, and doses of electrical stimulation. We conducted pooled analyses then definitions of interventions and outcomes were deemed to be similar.
Peer Reviewer #1	Methods	The authors should be commended for their decision to use a best available evidence approach to guide the study inclusion criteria for the report. Unfortunately few well designed RCTs with low risk of bias have been published on the interventions within the scope of this review. However, as highlighted in the introduction, knee OA is a prevalent disorder and patients and providers need to make the best informed decisions they can now about which treatments to select. Study exclusion criteria on pg 7 and in Appendix D do not match exactly. Appendix D states “Adults with physical therapy treatment combined with drug treatment” were excluded but this criterion does not appear on pg 7, and 12/56-57 notes “Adults in 114 studies (44 percent) were taking anti-inflammatory drugs or pain relievers.” I assume that studies were only excluded if medication was studied as an independent variable, but included if existing medications were maintained as constant as possible during the physical therapy intervention study.	We clarified our exclusion criteria as follows: “We excluded studies with drugs examined as an independent variable but included studies where existing medications were maintained as constant as possible during the physical therapy intervention study.”

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Methods	Regardless, more details on how medication use was used as an exclusion criterion is needed. It then needs to clarified throughout the report.	We clarified in the data extraction section that: 'We abstracted reported proportions of the patients taking anti-inflammatory and analgesic medications, types, and doses of the drugs." We clarified in our methods section that we "compared the effects of the examined physical therapy interventions across the studies according to reported proportion of the patients taking concomitant drug treatments, and types of the drugs."
Peer Reviewer #1	Methods	The second exclusion criterion of concern that appears both on pg 7 and in Appendix D is "Adults with knee OA who had knee arthroplasty on the "study limb" within 6 months before the study." Presumably the review could include studies of subjects with arthroplasy greater than 6 months ago. Please provide more details on subjects from the included studies that met this criteria (appears to be a total of 7 based on pg F-28 of Table F9). How was a diagnosis of knee OA defined after surgery and the joint surfaces have been replaced? Did this apply to those with a pre-op diagnosis of knee OA, and any late post-op PT treatment studies could be included?	Most of these studies used the Kellgren/Lawrence scale as one of the inclusion criteria. Three studies included cases of knee arthroplasty performed more than 6 months before enrollment. The rest of the studies did not specify the type of prior surgery. Unfortunately, the authors did not discuss how they made the OA diagnosis. OA knee diagnosis might be based on pre-op x-ray, as one study specified the x-ray should be performed within 3 years.
Peer Reviewer #1	Methods	This is such a different subgroup, more discussion is needed about these subjects if detail is available from the primary studies.	We did not include studies examining post-surgical PT treatments.
Peer Reviewer #1	Methods	This also should be discussed in the sections on applicability and study limitations later in the report.	We acknowledge our study limitations in the report.
Peer Reviewer #1	Methods	It is not clear whether studies with high risk of bias or poor reporting were included or excluded from the review. In the methods 8/18-19 the authors state "We did not downgrade methodological quality of poorly reported studies; however, we synthesized evidence from these studies separately." While in the discussion 52/27-28 the authors state, "We excluded those with poorly reported results or high risk of bias."	We revised the report and synthesized study characteristics from all studies we abstracted. In meta-analysis; we had to exclude poorly reported studies, as they could not provide adequate input for pooled analysis. For individual studies with high risk of bias, which we could not pool for analyses, we estimated the effect size, judged the evidence as insufficient, and did not report such results in the Executive Summary to avoid giving misleading conclusions. We clarified in the methods section that "We defined strength of evidence as low when evidence is limited to an individual study that is low or medium risk of bias, and evidence from single studies with high risk of bias as insufficient."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Methods	In the results, they report 82 or 85 studies had low risk of bias, 121 medium, 42 or 45 high risk of bias and 9 were poorly reported (numbers differed between pgs ES-7 and 12). That totals 257 compared to the 256 studies they say contributed to synthesis of evidence and conclusions (ES-7/13). The flow diagram in Figure B does not list the number of studies excluded for poor reporting or high risk of bias. Whether these studies were included or excluded needs to be clarified and reported consistently throughout the report.	When the numbers of the studies were sufficient to conduct meta-analyses, we included studies with high risk of bias. We revised the report clarifying that we did analyze the effect modification by risk of bias and we did exclude from the analyses poorly reported studies: "Most often, strength of evidence was low due to exclusion of patients from the analyses, inadequate allocation concealment, or unmasked outcome assessment. Few studies reported that the researchers who assessed outcomes were unaware of the treatment status of the patients. The majority of trials had moderate risk of bias. We explored with meta-regression and subgroup analyses how risk of bias could modify treatment effect and did not find consistently significant changes. We excluded those with poorly reported results. We also excluded trials that enrolled patients with knee or hip OA without separately reporting the outcomes. Moreover, many trials failed to provide sufficient details about the interventions themselves, their intensity, or the involvement of a physical therapist."
Peer Reviewer #1	Methods	0/9-11 "For continuous variables we calculated mean differences with 95 percent CI. For pain, we calculated mean differences at 99 percent CI." The decision to treat pain differently needs greater explanation. As described later in the report, the risk of harms from treatment is small, especially compared to long term use of medications typically prescribed for OA. So there does not seem to be a need to have wider confidence limits with greater possibility of including a null effect for pain as compared to other outcomes. Stakeholder input is described to guide other methodological decisions during the review, but it is not clear that patient stakeholders agreed that the balance of benefits and harms warranted this more conservative choice.	We revised the report calculating all results at 95 percent CI.
Peer Reviewer #1	Methods	There appear to be reporting errors in the document. All pain related CIs in all tables and figures, including Appendix F are noted as 95%. The only exception was the title for Figure 2. The CI for mean differences in pain is cited as 99% in the Table of Contents, but 95% on pg 56. The actual Cis need to be verified, and if not 95%, corrected in the report and a more thorough justification provided	We deleted all results at 99 percent CI.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Methods	There is also inconsistent use of the terms labeling the patient-oriented outcomes. Sometimes function is included in the list, sometimes quality of life, sometimes disability but all terms are not used together consistently (example p. ES-15, line 45). The low level of evidence rating for the exercise interventions is in contradiction to other reviews. “Low” is defined as “further research is likely to change the confidence in the estimate of effect and is likely to change the estimate”. I disagree with this, the evidence for exercise has been seen for decades and has been relatively consistent. I believe these findings are weakened because of the inclusion of non-PT delivered interventions in this review.	We focused on patient-centered outcomes including pain, disability, and quality of life. We categorized intermediate outcomes as measurements of gait, strength, balance, transfers, endurance, joint function, or composite measure of functional performance. We revised the text to say, “Our report analyzes patient-centered outcomes including pain, disability, and quality of life with physical therapy interventions available in the United States for adults with knee OA.” Following the guidelines from the Methods Guide for Effectiveness and Comparative Effectiveness Reviews from AHRQ, we judged the strength of evidence for each major outcome according to risk of bias, consistency, directness, and precision. Whether we included non-PT delivered interventions did not substantially change our conclusions.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Methods	Low versus high intensity (ES-12 line 12) • Taping – for patellafemoral OA? (p. ES-9, line 49) • "reasonably narrow confidence intervals" (p. 8) • "physical agents/modalities used infrequently" (p. 13) – need a reference for this; I doubt this is true	We summarized in the report that "Although definitions of intensity differed among studies, prior research indicated similar benefits for low- and high-intensity exercise, defined by one study as 40 percent and 70 percent heart rate reserve, respectively. In one study, low- and high-intensity exercises similarly improved function, gait, and pain. Another study found that the effects of high-resistance strength training (>60 percent of one repetition maximum) appear larger than those of low-resistance strength training (10 percent of one repetition maximum), but the differences were statistically insignificant." We clarified that taping was examined for generalized knee OA. We integrate GRADE 6 concept for precision. Specifically, because side effects of PT are rare and not serious, we define the effect size to be precise if the 95 percent CI of effect size does not include 0. For 95 percent CI of estimated standardized effect size does include 0, the 95 percent CI should be within ±0.5 to be precise so that we do not miss potential benefits/harms. Rather than stating that "physical agents/ modalities used infrequently," we argued that physical agents/modalities are infrequently used in isolation. We revised as: "Second, many of the interventions studied were physical agents/modalities (i.e., orthotics, ultrasound, taping, etc.). Again, this contradicts the practice of physical therapy, in which physical agents/modalities are infrequently used in isolation, but rather in combination with other more 'active' interventions (i.e., exercises)."
Peer Reviewer #2	Methods	The major flaw in this review is the criteria used to include and exclude studies. Eligible interventions "within the scope of PT practice but not necessarily administered by PTs or PTAs" combines actual clinical therapeutic exercise with community-based exercise delivered by trained exercise leaders. In my opinion this is mixing "apples and oranges".	We re-analyzed the data and revised the report reporting the results by PT involvement in delivered interventions. We analyzed separately actual clinical therapeutic exercise and community-based exercise delivered by trained exercise leaders with subgroup analyses.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Methods	First of all, patients using PT delivered interventions in a health care setting are different than individuals using self-directed or community-based exercise programs. PT requires people to have access to the health care system, a referral from their physician (some states and health plans have direct access), have health insurance, can afford co-pays, and can fit PT into their schedule. These patients have worse disease (pain, lower function, work disability, etc.) and have been referred to physical therapy because of this. Community-based exercise programs tend to attract people with higher degrees of functioning and less disability, does not require health insurance or access to the health care system, can be self-directed or administered by trained lay-leaders or exercise professionals and include people with many different types of arthritis including osteoarthritis, rheumatoid arthritis, lupus and fibromyalgia. Community-based exercise is a public health approach to improving pain, function, quality of life and reducing risk of disability. Physical therapy is a clinical approach. Both approaches are needed, and compliment each other. While there is overlap between clinical and public health interventions for OA, because of the differences in training level of the implementers, patient characteristics and health system access these two approaches are “apples and oranges”. Kelley et al discussed this point in their meta-analysis of community-delivered exercise programs for people with arthritis. (Kelley GA, Kelley KS, Hootman JM, Jones DL. Effects of community-deliverable exercise on pain and physical function in adults with arthritis and other rheumatic diseases: a meta-analysis. Arthritis Care Res (Hoboken). 2011,Jan;63(1):79-93.)	We reanalyzed the data and revised the report to explore the role of physical therapist involvement on effects with exercises. We performed subgroup analyses by PT involvement for all outcomes with aerobic or strengthening exercises. We emphasized that our intention is not to evaluate the value of PT involvement as the OA in the PT involvement group would likely be more severe. While in few outcome comparisons the no PT involvement groups had a larger effect size, in general we found effect sizes with PT involvement are larger than that without PT involvement by 0.1 standardized effect size (0.1 standard deviations). Furthermore, the results in the PT involvement group tended to be consistent without heterogeneity. However, the sample size of most pooled analyses decreased, and whether or not we restricted to only the PT involvement group did not substantially change our conclusions (Appendix Table F24).
Peer Reviewer #2	Methods	A good example of this is the FAST study which is included in this review. Although a well-designed and executed study that has shown good efficacy for exercise for knee OA, the intervention was not a physical therapy intervention. It was an intensive fitness facility based exercise program delivered by an exercise physiologist followed by home-based exercise. It was conducted in a university setting at the university exercise center not a physical therapy clinic or hospital. It is more like a community-based public health approach study, with the exception of having used a relatively highly trained person (at least masters level exercise physiologist) to deliver it, which is not feasible for mass implementation of exercise in community settings.	Same as above.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Methods	This inclusion of studies not using PT/PTAs as the implementer applies differentially to the different types of interventions. It is not legal for non-licensed persons to use modalities such as diathermy, ultrasound, electrical stimulation etc. So, as figure 4 shows, there is a tremendous amount of disparity between the modality or orthotic interventions and the exercise interventions regarding the involvement of the PT. To have the most generalizability to PT clinical practice, the exercise interventions not administered by PTs should be removed from the review.	Same as above. Following the public discussion, and the recommendations from the stakeholder and the TEP, we included exercise that may require various PT involvement.
Peer Reviewer #2	Methods	It is unclear how braces, specifically unloading braces, were included. At some points they are vaguely mentioned under orthotics but their inclusion is not consistent and not at all represented in the results.	Thank you for your comment. We stated that "We were unable to perform a pooled analysis to draw valid conclusions. In one study, unloader brace improved disability, composite, joint, and gait functions."
Peer Reviewer #3	Methods	Yes, explicitly discussed inclusion/exclusion criteria. many definitions explained, Statistics appear appropriate	Thank you.
Peer Reviewer # 5	Methods	Inclusion/exclusion criteria are justifiable, search strategy logical and defined, outcome measures are appropriate and statistical methods are sound.	Thank you.
Peer Reviewer #4	Methods	Inclusion and exclusion criteria justifiable and logical. I'm not sure the outcome measures are appropriate, given that so much of what is observed is subject to bias or lack of quantifiable measures (esp. with regard to pain). Later criteria (strength of quadracep measured in nm/x are better determiners.	We revised the report to provide a distribution of pain measurements. We analyzed clinical importance of pain reduction using the same measure (VAS).
Peer Reviewer # 6	Methods	I also have concern about how strength of the evidence is assessed. I was surprised to see that evidence for the benefits of exercise in knee osteoarthritis (KOA) was assessed as "low" when virtually all other guidelines have determined that the evidence for exercise is moderate to strong. It appears that the authors have lumped studies of high quality with those of low quality, effectively reducing the strength of evidence.	Following the guidelines from the Methods Guide for Effectiveness and Comparative Effectiveness Reviews from AHRQ, we judged the strength of evidence for each major outcome according to risk of bias, consistency, directness, and precision. We incorporated risk of bias in synthesis of evidence and explored differences in treatment benefits among low, moderate, and high risk of bias studies. We defined as insufficient the evidence from individual RCTs with high risk of bias. We revised the report to clarify risk of bias assessment, ranking strength of evidence, and reasons for exclusion from the analyses of individual studies with high risk of bias.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer # 6	Methods	It also seemed that the list of references used in the final analysis was incomplete. For example Wang, et al 2009 an RCT examining the effectiveness of Tai Chi is not included. And systematic reviews were not included without explanation.	We did include the article "Tai Chi Is Effective In Treating Knee Osteoarthritis: A Randomized Controlled Trial. by Wang C, Schmid CH, Hibberd PL, et al. Arthritis & Rheumatism. 2009;61(11):1545–553." We were unable to pool-analyze this paper because their results were reported in "change," not mean and its standard deviation. In addition, the comparison in this trial was Tai Chi vs. stretching exercise. We had reported the results in the section of Comparative Effectiveness of Physical Therapy Interventions (reference 200 in the prior version).
Reviewer #1	Results	16/35-40 The authors were particularly insightful in relating KQ1 to KQ2 in their results by comparing the magnitude of the effect of treatment on pain and disability observed in the clinical trials to the information they tabulated in the studies on minimal clinically important differences.	Thank you.
Reviewer #1	Results	17/10 Adherence is used as the subheading for this section, but compliance tends to be the favored term in most of the report. This may reflect usage in the primary studies. Some providers consider these synonyms, others do not. Those who perceive a distinction consider adherence to imply a more active role in treatment which is consistent with participation in an exercise program. The authors should clarify their use of these terms within the context of the report given the debate about their usage	Thank you for your comments. We changed adherence to compliance. We identified five studies examining this issue: Two used the term "adherence" and three used "compliance." They were interchangeable, as either compliance or adherence was measured by the percentage of classes attended.
Reviewer #1	Results	18/25 Please give examples of what is included under comorbidity	The authors defined comorbidity as the presence of knee OA plus other two or more clinical conditions. The results indicated that aerobic exercise improved function and reduced pain independently of the presence of comorbidity.
Reviewer #1	Results	What is composite function? The term is used liberally in the results of the Executive Summary and main report, but not clearly defined. It appears to be first described as "composite measure of function[al performance]" ES-6/8, 10/31 but that is inadequate to know if it represents several different constructs from the primary studies, or a grouping of several tools defined by the authors of this report. More detail is needed.	Appendix Table F11 provided the detailed information on composite function. We prioritized WOMAC physical function than Lequesne's index. Others included KOOS: Symptoms subscale or functional performance tests involving multiple tasks.

Commentator & Affiliation	Section	Comment	Response
Reviewer #1	Results	40/Table 5 “Function” (particularly walking speed/time) is listed in both the clinical outcome and intermediate outcome columns of Table 5. The authors acknowledge that the designation of outcome categories was somewhat artificial 33/13. I realize the decision for this review has already been made, but a slightly more complex framework such as the ICF may have been needed to look at the relationships among impairments in body structures and functions (e.g. strength, range of motion), physical activities (e.g. balance, walking) and participation including the patient-centered outcomes of dis/ability. This approach would have easily allowed functional activities such as walking to be related both to impairments and disabilities (Table 5), or between pain and disability, two patient centered outcomes as noted 33/22.	We revised the report pointing out that: “Integrated approach in evaluating the relationships among impairments in body structures and functions (e.g. strength, range of motion), physical activities (e.g. balance, walking) and participation in activities of daily living would allow better prediction of patient-centered outcomes of disability and quality of life.”
Reviewer #1	Results	33/45-51 The authors have noted that some studies used logistic or Cox regression to examine the association between intermediate and patient-centered outcomes. The definitions of intermediate outcomes in Table 7 show the cut points used in the studies to create categories for these outcome measures (with more detail provided in Tables F32-40). Many studies using this approach found a significant relationship between impairments (such as muscle strength or balance) and disability (as measured by WOMAC or SF-36) at the higher levels of impairment but not at the lower levels (e.g. O’Reilly 1998, CAS(K)/Thomas 2008). While obvious, this observation also deserves mention in this section.	We included all results from the studies by O’Reilly, 1998 and Thomas, 2008 in Tables 6 and 7. We added in the results section that: “The studies demonstrated the importance of the appropriate cut points to categorize muscle strength reporting a significant relationship between impairments in muscle strength and disability at the higher levels of impairment but not at the lower levels.”
Reviewer #1	Results	38/42 Authors note that 30 studies reported minimal clinically important differences (MCIDs). One concern is that the MCID Tables (F54-57) included more than just the MCID to document important change over time. One example is the Minimal Detectable Change (MDC), the smallest change that exceeds the measurement error of the instrument, which may not necessarily be clinically important. The text of the report should clarify and define all the differences that were reviewed within the scope of the report and included in the tables.	We revised appendix tables to focus on clinically important changes in the outcomes.

Commentator & Affiliation	Section	Comment	Response
Reviewer #1	Results	48/42-49 The authors review the methods used in the studies to determine minimal clinically important differences, including relative and absolute differences and the distribution method. They note that the relative difference method incorporates baseline severity and the distribution method accounts for the distribution of changes in outcome. Another approach that was used in at least one cited study divided baseline severity into categories, which is complementary to the Cox and logistic regression approaches described under the main KQ2. Salaffi 2004 examined both absolute and relative changes in pain scores in each of three categories based on cut points using a numeric rating scale and found that clinically significant changes in pain were non-uniform across the scale. This approach and finding is also worthy of comment in the results. If this approach is included with the category of the distribution method, then a better description is required in the text.	We revised the report emphasizing diagnostic value of the changes in the Pain Numeric Rating Scale as follows: "In one prospective cohort study, a Pain Numeric Rating Scale score of <4 had a conclusive diagnostic value for patient perception of clinically significant improvement (Appendix Table F31). The study examined both absolute and relative changes in pain scores in each of three categories based on cut points using a numeric rating scale and found that clinically significant changes in pain were non-uniform across the scale. A reduction of 15 percent in the Pain Numeric Rating Scale score represented minimal clinically important changes while a reduction of 33 percent represented "much better" improvement in the patient's global impression of change."
Peer Reviewer #1	Results	Paragraph 13/16-25 notes the numbers of articles on various interventions (most for aerobic exercise then strengthening). The summary on 13/37-43 highlights the interventions with some benefits (aerobic and strengthening exercise decrease pain and improve function and aquatic exercise improves disability). Descriptions of individual interventions (pg 13-16) therefore should start with exercise, specifically aerobic, strengthening, and aquatic, followed by other forms of exercise, the other treatments as listed, and education last since it is the least physical therapy specific (includes dietary).	We reorganized the order as the reviewer suggested.
Peer Reviewer #1	Results	Also recommend changing the presentation of results of KQ1 in the summary (pg 50) so that the authors lead with the positive results first. Patients and providers want to know what works, then follow with findings on lack of effectiveness or insufficient evidence. Don't bury your lead. Much of the evidence presented in the report is based on insufficient strength of evidence, many times because it is derived from only one study. I appreciate the authors including the findings in the summary text and tables, but would prefer that the results with greater strength of evidence (above insufficient) be easier to locate in the tables. Options might include bolding the stronger evidence (e.g. Table 5), or placing it first in a long list (e.g. Table 10).	We revised the report to present the results according to the changes in pain and disability by strength of evidence and then in functional outcomes. We revised strength of evidence grading and focused on sufficient evidence from pooled analyses. We added a table demonstrating positive results with examined PT interventions on pain and disability.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Results	The results section is relatively comprehensive. As stated above, exercise studies not using a PT should be excluded from the analysis. Sensitivity analyses are appropriate but Appendix tables 15-18 s how that the studies with known involvement of the PT are clearly different in outcomes. This supports the factor that PT involvement should be considered an effect modifier and therefore at minimum the results should be stratified by this variable. But these tables are buried in the appendix. Ideally these non-PT studies should be excluded per the reasons stated above	We revised our analysis and the report to explore PT involvement as an effect modifier and therefore report the results among subgroup of the studies with different level of PT involvement. We added this information in the text of the report.
Peer Reviewer #2	Results	I think the finding reported on p. 16 regarding the clinically important improvement in pain, disability and joint function in the majority of individual RCTs for aerobic exercise is under-promoted. It is not mentioned in the key findings. I think the key finding that no one intervention improved all outcomes should be minimized. The key finding that few interventions showed efficacy is a rather negative and uninformative leading finding. There were interventions that did show efficacy, suggest leading with that.	We revised the report clarifying ranking the strength of evidence. We present the results following the analytical framework and focusing on pain, disability, and quality of life and then intermediate outcomes.
Peer Reviewer #3	Results	good discussion of the results. Multiple studies are discussed	Thank you.
Peer Reviewer #4	Results	Amount of detail very good in results section, as are key messages. Figures and tables seem adequate and descriptive. It seems that very little could have been overlooked with the list of thousands of publications researched.	Thank you.
Peer Reviewer # 5	Results	Results are presented clearly in a set of tables and discussed in the manuscript. I am not aware of any major studies overlooked	Thank you.
John M. Eisenberg Center	Results	REPORT ERROR: Insufficient findings of stated conclusion/outcome On pages 30-31, Table 4, (and summarized on page ES-25, Table F), several conclusions/findings statements have been reported with the SOE listed as “Insufficient”. According to AHRQ methods of SOE ratings, a declared finding statement rated insufficient should include or be limited to a statement that “conclusions could not be drawn.” We would appreciate your clarification.	We revised the report to avoid discrepancies in SOE presentation.
John M. Eisenberg Center	Results	REPORT ERROR: Insufficient findings of stated conclusion/outcome On page 40, Table 5, several conclusions/findings statements have been reported with the SOE listed as “Insufficient”. According to AHRQ methods of SOE ratings, a declared finding statement rated insufficient should include or be limited to a statement that “conclusions could not be drawn.” We would appreciate your clarification.	We revised the report to clarify our use of the AHRQ ranking system to judge strength of evidence in therapeutic studies. We ranked level of evidence form diagnostic studies using the USPSTF criteria.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer # 6	Results	As suggested above, I believe the methods lack sufficient explanation for how quality of studies was assessed and the authors provide insufficient data for the individual assessment of each study. This makes interpreting their conclusions regarding strength of evidence very difficult. A table explaining the strength of each study would be extremely helpful. It took me a very long time to even find any information about each study. But what was provided was insufficient to understand their quality assessment.	Appendix Table F8 provided quality of studies we assessed. We updated this Table providing information with respect to whether the particular study was included in the pooled analyses, and the reason of not including in the pooled analyses.
Peer Reviewer # 6	Results	I believe the "key messages" are misleading and will be a disservice to patients The first message is that few PT interventions provide benefit. Actually exercise has been shown to be effective and that seems to be the most important message. Another message is that ultrasound is effective in decreasing pain. Yet ultrasound is rarely if ever provided in isolation. But it is a lucrative intervention for the provider, faster and cheaper to provide and well reimbursed. These conclusions would serve to justify a PT's use of ultrasound (a passive intervention) instead of exercise (an active intervention). I believe such a message is contrary to all existing knowledge of non-pharmocologic treatment of OA.	We revised key messages to focus on effective interventions according to pooled analyses. We emphasized the benefits from exercises.
Peer Reviewer # 6	Results	Another key message is that "no single physical therapy intervention improved all interventions." That is not surprising with outcomes ranging from pain, function, gait, disability, health perception, quality of life and psychological disability. Does any intervention address *all* outcomes? This conclusion again minimizes the effectiveness of some treatments, especially exercise and will lead providers to abandon all physical therapy interventions.	We revised the report focusing on the primary outcome of the review, which is pain. We clarified that the ultimate goal of all interventions for older patients with knee OA is reducing disability and improving quality of life. We revised key messages as follows: “No single PT intervention was shown to improve all examined outcomes.” In addition, we analyzed all eligible outcomes that the principal investigators aimed to improve.
Peer Reviewer # 6	Results	The authors also conclude that orthotic devices and taping are ineffective. Yet these conclusions are based on limited data from low quality studies. Why did the authors come to this conclusion rather than reporting that there was insufficient data to draw a conclusion?	We ranked strength of evidence with all interventions following the same criteria. We revised the report to clarify the definition of insufficient evidence from individual RCTs with high risk of bias.
Peer Reviewer # 6	Results	The authors examined if patient characteristics influenced effectiveness of any treatments. However the authors examined the associations across all interventions, which seems illogical. Malalignment may have not effect on the effectiveness of ultrasound but a significant effect on strengthening exercises.	We explored heterogeneity in treatment effects with *a priori* defined patient characteristics. We did synthesize the evidence according to the feasible and expected PT intervention effect modification by patient characteristics.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer # 6	Results	I found the tables difficult to interpret. First, is there a rationale for the order of interventions listed in the tables? I can find no logic in the order. Passive therapies are not listed first followed by active ones. Please explain the rationale or list the interventions alphabetically.	We clarified that we conduct the review following definitions and classification of PT interventions from the Guide to Physical Therapist Practice.
Peer Reviewer # 6	Results	The authors need more detailed legends. Specifically what is the "category " in Table 12? Table F12 (appendix) does not include units. Intensity of intervention ranges from 1-5. Is this the number of visits /week? If so this is a poor measure of intensity. Treatment frequency is *not* indicative of intensity of ultrasound or intensity of exercise, etc. Please provide more detail about figures as well. Figures 4 and 5 lack an explanation of the x-axis and do not include all the interventions, particularly ultrasound.	We revised Table 12 clarifying that we categorized the studies according to the reporting intervention components. We provided explanation of horizontal axis for figures 4 and 5. We clarified definition of intensity in Table F12.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Discussion/ Conclusion	The authors concluded that variability in studies resulted in insufficient evidence to draw conclusions about the association between intermediate and patient centered outcomes and that the magnitude and clinical importance of these associations remain unclear. The discussion of these points and the recommendation for future research that lead from them can be made much stronger. This is important because many physical therapy treatments are impairments directed with the goal of ultimately reducing disability. Some possible questions that the authors might consider incorporating into the future research section include: Does there need to be a tested and accepted theoretical framework that describes the relationship between impairments and disability to guide future studies, or is there a candidate framework that investigators are just not using to design their studies? The review authors even acknowledged that their distinction between intermediate and patient centered outcomes was somewhat artificial, implying that an alternative may be more appropriate. Another area of future research would be to try to develop models with stronger correlations between intermediate and patient-centered outcomes. Based on the primary studies reviewed, should a linear model between impairment and disability be tested or an asymptotic one? (Does the slope of the relationship change after a critical threshold is reached? e.g. Does an increase in strength from unable to move against gravity to "normal/healthy/average" produce the same benefit to ADL as an increase in strength from "normal/healthy/average" to that of a body builder?) Future research might also evaluate how the ability to compensate might modify the relationship between impairments and disability. Does the relationship between impairment and disability get stronger with increasing numbers of impairments that fall below a clinically important threshold? (e.g. Are number of weak muscles or joints with restricted motion below threshold a better predictor of disability than the magnitude of weakness in any one particular muscle?) The variety of methods and results reviewed under KQ2 that prevented firm conclusions should lead the authors to make much stronger and more specific recommendations for future research in Table 13.	We revised the report to point out that "Many physical therapy treatments are directed at impairments with the goal of ultimately reducing disability. Future research should address an accepted theoretical framework that describes the relationship between impairments and disability to guide future studies." We proposed that "benefits from physical therapy interventions should be defined as clinically important improvement in pain, independence in ADL, and quality of life. Treatment success should be estimated using rates of the patient-centered outcomes." We propose categorizing changes in intermediate outcomes according to thresholds of clinical importance rather than linear regression that results in statistically significant coefficients of questionable clinical importance. Researchers could categorize patients according to the clinical importance of the changes they experience and analyze rates of patient-centered outcomes. Assuming investigators were willing to share their data, meta-analyses of individual patient data could also provide good estimates of treatment effects in patient subpopulations by age, comorbidity, severity of knee OA, and concomitant treatments." We revised the tables with future research needs addressing non linear relationship between correlated impairments and importance of clinically important threshold that can predict disability and quality of life.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Discussion/ Conclusion	The strength of this section is the future research recommendations which are neatly laid out in a tabular format. There is a critical need for studies that evaluate PT interventions as they are delivered clinically (combined modalities versus individually) and for dose response studies. The second area of critical research is to not exclude participants with multimorbidities as this adds to the generalizability of the findings to real clinical practice (e.g., pragmatic trials).	We revised the report to recommend pragmatic trials to improve applicability of findings to real clinical practice.
Peer Reviewer #2	Discussion/ Conclusion	I feel there should be more discussion of the limitations. For instance, if there was insufficient evidence (e.g., for quality of life outcomes) was it because quality of life was not measured in the study or because it was not well-reported in the manuscript. This is two different issues, the former can be addressed by identifying the minimal patient-oriented outcomes that should be used in clinical trials and the latter can be addressed by better implementation of reporting guidelines.	We revised limitations of our review clarifying that: "Our report has several limitations. We relied on published information and did not contact the principal investigators of poorly reported or unpublished studies. We evaluated selective outcome reporting according to the outcomes described in the methods sections. Very few trials examined quality of life as a clinical outcome. We did not contact the authors clarifying whether the trials measured quality of life but did not report it or whether the investigators did not aim to examine this important outcome. Future research should identify minimal patient-oriented outcomes that should be used in clinical trials. All clinical trials examining PT interventions should be registered in Clinicaltrials.gov with detailed protocol and links to the publications."
Peer Reviewer #2	Discussion/ Conclusion	The same issue is true for one of the characteristics used in level of evidence. In many studies, it is impossible to blind the patient to the intervention. For example, a patient is going to know they are exercising or wearing a brace versus not. This is probably not as important an issue as it is in drug trials. But the critical piece is, is the outcome assessor blinded to group assignment. It is not clear if these were assessed separately or not and how much they influenced the finding of "low level of evidence".	We revised the report to clarify the ranking of evidence according to the AHRQ guidelines. We did not downgrade masking of the treatment status for PT interventions that are not feasible to mask.
Peer Reviewer #3	Discussion/ Conclusion	not easily tranlated into future research due to the limitations cited (in the clinical world, more than one intervention is usually utilized).	Thank you.
Peer Reviewer #4	Discussion/ Conclusion	Implications are clearly stated, not sure about the limitations. The future research section is clear but could probably be more detailed, unless you are looking for a springboard for more thought in the matter.	We revised a table with future research needs and provided additional details about study design and objectives.
Peer Reviewer # 5	Discussion/ Conclusion	Implications of findings and limitations of review are clearly stated. Future research needed is discussed clearly	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer # 6	Discussion/ Conclusion	As noted earlier, the first conclusions presented are that few PT interventions work and that none improves all outcomes. The 5th bullet notes that exercise (aerobic, strengthening and aquatic) improve pain, function and/or disability. The order of these conclusions diminishes the positive findings. The authors findings are in conflict with existing clinical guidelines. but the authors fail to discuss this difference and to put their findings in the context of existing guidelines.	We revised the order of key messages following your recommendations.
Peer Reviewer # 6	Discussion/ Conclusion	However the authors do a service by listing the needs of future research. They rightly note the need for common outcome measures and additional research to identify characteristics of patients that benefit from each intervention. They also correctly identify the need for more data on the dose-response for these interventions. While they outline appropriate needs for future research I believe this could have been more detailed.	We revised a table with future research needs providing more details about methods and objectives.
Peer Reviewer #1	Discussion	Under the subquestion related to time of follow up, a related area of future research to explore is whether changes in intermediate outcomes precede changes in disability measures and if so by how much.	We revised the table with future research needs addressing this point.
Peer Reviewer #1	Discussion	54/27 The authors refer to “physical therapists and their assistants.” Physical Therapist Assistants (PTAs) complete an accredited education program leading to an associate degree, they sit for a national examination, and most states require a license, certificate, or registration to practice. PTAs work as part of a team to provide physical therapy services under the direction and supervision of a physical therapist. The term is not synonymous with a PT aide or technician, who are on-the-job trained and not eligible to provide physical therapy by many payers. PTAs should be referred to by name, not as “their assistants.”	We revised the report using the appropriate term as “physical therapist assistants.”
Associate Editor	Figures	Figure A in ES: check the KQ), something is missing here Why are there gray boxes in Figure 3 (in Discussion)?	We corrected KQ in Figure A and revised Figure 3.
Associate Editor	Tables	Table F11 in appendices: check word "yes" in the 4th column header, should it be there?	Corrected.
Associate Editor	Tables	Table F26: title: Interventions spelled wrong; also see the column Risk difference for a typo to correct.	Corrected.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Appendix	A spot check identified several possible reporting errors in the tables. The review deadline prevented an extensive review of all tables cells for every error, but a few examples include: Table F48 Multiple rows refer to Barker 2004 on pg F-155, and 3 cells in column 1 compare walking speed to leg extensor power. The results in column 4 are only correct for the last row (0.519). In the 1st of the 3 rows, 0.357 is the correlation between walking speed and WOMAC and in the 2nd row 0.394 is the correlation between walking speed and pain-VAS. So the values are taken from the primary study, but the outcomes compared are not reported correctly in the table.	We rechecked all tables to assure correct abstraction.
Peer Reviewer #1	Appendix	A similar problem appears again for the Barker reference in Table F51 on pg F-231. 2 rows site the correlation between the WOMAC pain subscale and leg extensor power. The non-bolded entry is correct (0.388). The bolded entry is the correlation of two pain measures – WOMAC pain with pain-VAS (0.654). Table F54 Tubach 2005 pg F-302. A row entry appears to be missing for satisfactory scores for the highest tertile. In general, the detected problems were found when the same study was entered into multiple table rows with only small changes required for some of the table cells. It appears that some of the needed edits were missed during a cut and paste. A thorough search for duplicate entries in particular, and a review of accuracy in general is required for Appendix F.	We rechecked all tables to assure correct abstraction and deleted duplicated information from the tables.
Peer Reviewer #1	General	Quality of the Report: Superior Number of Hours Spent to Review the Report: 20 a. General Comments: The authors are congratulated for their analysis of a very large retrieved and synthesized literature in the scope of this review. This review makes a unique contribution to the evidence because it not only includes physical therapy treatments for knee osteoarthritis, but also the breadth of measurement tools used to document treatment outcomes. The description of the measurement tools in Table F28 and the definitions of minimal clinically important differences in Tables F54-57 are very useful resources in the report and should be highlighted when the report is disseminated to clinicians and researchers.	Thank you.
Peer Reviewer #2	General	Quality of the Report: Fair	

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	General	The key questions are clear and appropriate explicit. The target population and audience is clearly defined but the clinically meaningfulness of the report is limited by the inclusion/exclusion criteria used for study selection. By including studies that evaluated interventions that were in the scope of physical therapy practice but not necessarily delivered by a physical therapist (PT) or PT assistants, the findings cannot be truly applied to PT practice. The authors state they used the Guide to Physical Therapist Practice from the APTA, which I assume clearly defines PT clinical practice and does not include non-PT delivered interventions. This is particularly true for the exercise-based interventions as therapeutic exercise done in the context of PT clinical practice is "apples" compared to community-delivered exercise delivered by trained but non-PT providers are "oranges". The fact that only 28% of the aerobic exercise studies reported that PTs delivered the intervention (p 52) also supports this issue of apples versus oranges.	We performed subgroup analyses by PT involvement for all outcomes with aerobic or strengthening exercises (Appendix Table F24). We emphasize that our intention is not to evaluate the value of PT involvement as the population in PT involvement group should have more severe OA. While the no PT involvement groups had a larger effect size in few outcome comparisons, in general we found effect sizes with PT involvement are larger than that without PT involvement by 0.1 standardized effect size. Furthermore, the results in the PT involvement group tended to be consistent without heterogeneity. However, the sample size of most pooled analyses decreased, and whether or not we restricted to only the PT involvement group did not substantially change our conclusions.
Peer Reviewer #2	General	There needs to be more explanation of the exact instruments and constructs that make up the outcomes of interest. For example, what measures were used to assess quality of life? Just the SF-36? How was disability defined and what instruments were used to measure disability in these studies?	We reported the exact instruments used for pain, disability, quality of life, and composite function in the Appendix Table F11.
Peer Reviewer #2	General	The conclusion stating that no intervention improved all outcomes is not expected. Suggesting that one PT intervention could improve all outcomes is not really feasible nor reasonable and should not be an assumption of this review. The finding that many of these PT intervention evaluations were done in isolation and not combined as is usually done in practice supports this. It is possible clinicians combine interventions specifically that affect different outcomes. There is some inconsistent use of efficacy and effectiveness, which are not interchangeable terms. This needs to be addressed because this review seems to be targeting effectiveness not efficacy or maybe both	We revised the report to emphasize that pain was a primary outcome of the review and all interventions for adults with knee OA target reductions in disability and improvement in quality of life. We synthesized the evidence of treatment effects on all eligible outcomes that were reported in the original studies (primary or secondary) assuming that investigators expected to improve the outcomes they measured. All studies with active comparators as defined as usual care or another PT intervention was considered as effectiveness studies.
Peer Reviewer #3	General	Quality of the Report: Superior Yes, this report was meaning. Target populations and audience identified well. multiple key questions, but covered the breadth of information available from these studies	Thank you.
Peer Reviewer #4	General	Quality of the Report: Superior Yes, clinically meaningful, but is 4000 cases enough when considering 15 or so modalities?	We revised the report to point out lack of power in individual studies and low precision of pooled estimates.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	General	Target population is very broad (no children, no people in nursing home) with a myriad of cases involving work life, early traumatic injury, etc., not considered. Key questions appropriate and stated explicitly.	We clarified that we explore whether patient characteristics can modify treatment effects. We present the results of meta-regression by patient, study, and treatment characteristics. We pointed out that vast variability in patient populations, PT interventions, and outcome measurements hampered synthesis of evidence.
Associate Editor	General	Suggest checking "Hodges" throughout and replace with "Hedges"	We corrected this typo in the report and appendices.
APTA	General	Recommendation 1 (Use of the Term Physical Therapy): APTA does not feel that the title of this report accurately reflects the methodology of the review. Physical therapy is a field of professional practice, and physical therapy treatment regimens for this patient population will typically incorporate a variety of specific interventions and modalities. In this report, the interventions were individually assessed for their effect on functional and clinical outcome metrics. The authors address this limitation in the discussion and use the terminology “physical therapy interventions” in the results and discussion. Although physical therapists may be the professionals most commonly utilizing these interventions in patients with knee OA it is important to make the distinction clear that this report has assessed the effect of specific interventions, which may be used by physical therapists and other qualified health care professionals, to impact outcomes measures in this patient population.	We revised the report clarifying that we defined, selected, and classified interventions according to the Guide to Physical Therapist Practice. We finalized a list of eligible interventions after the public discussion and consultation with stakeholders and the TEP. We re-analyzed the data following your recommendation to explore the role of PT involvement in exercise benefits. We acknowledged that our review was complicated by the discrepancy between the actual practice of physical therapy and the design of research evaluating these interventions. Published studies have focused on individual physical therapy interventions. In practice, physical therapists or physical therapist assistants most often treat their patients with a combination of modalities. However, when trials identified ineffective individual modalities that would not result in significant benefits in clinical practice. As physical therapists are an essential member of the health care team who provide evaluation and treatment for individuals who have knee OA, this report expects to contribute toward advances in the field of physical therapy.

Commentator & Affiliation	Section	Comment	Response
APTA	General	Recommendation 2 (Patient Population) The patient population as defined in this report is extremely heterogeneous with respect to clinical presentation and individual impairments. There is a weak or absent relationship between radiographic disease severity and symptomatic-severity of this disease. These differences may arise from individual perceptions of pain, joint laxity, the number of compartments involved, the number of limbs/joints involved and measurement technique used to diagnose or quantify OA. Therefore, the treatment of these individuals will be dependent on their measured impairments and reported symptoms. Individual modalities and interventions may be appropriate for one patient, but not another with an identical radiographic diagnosis. Concluding that few specific interventions are effective may be a consequence of a heterogeneous patient population and reviews/trials that only evaluate a single component of physical therapy treatment, and not that physical therapy treatment on an individual basis is ineffective. The individual interventions assessed in this report are typically targeted towards a specific impairment, and the outcome from these interventions should be evaluated using a metric of that impairment. It is not expected that a treatment to improve strength would necessarily improve pain, so determining such a treatment to be unsuccessful in that respect is not appropriate. APTA believes that this was not addressed or discussed in the report and the authors make the point several times that "no single intervention improved all outcomes."	We revised the report and key messages focusing on benefits with exercise demonstrated in pooled analyses. We revised the report clarifying our primary outcome as pain and focusing on patient disability and quality of life. We clarified that an ultimate goal of all interventions for patients with knee OA is preventing disability and improving quality of life. We did, however, synthesize treatment effects on all reported outcomes the principal investigators expected to improve. While we stated that, "no single intervention improved all outcomes," we provided information of benefits of individual intervention, such as aerobic and aquatic exercise improved disability; aerobic, strengthening exercise, and ultrasound reduced pain and improved function. Except e-stimulation, we were unable to perform meta-analyses with respect to muscle strength because of variation in muscle strength measures. Further studies are warranted.

Commentator & Affiliation	Section	Comment	Response
APTA	General	Recommendation 3 (Patient-Centered Outcomes) APTA believes that there is an important distinction between patient-centered outcomes and patient-reported, or patient-perception outcomes. We would like to note that many of the clinical metrics described as intermediate outcomes in this report are integral to patient-centered outcomes and patient-centered goals in physical therapy. Patients will commonly report they would like to walk further, walk faster, get out of a chair easier, not require a handrail during stairs, improve balance, etc. Metrics that evaluate these factors are not considered patient-centered outcomes in the current report; the authors only consider patient-centered outcomes to be patient self-reports that evaluate the patient's perception of their functional ability or their self-reported quality of life or pain. As physical therapists and experts on improving patient function and patient mobility, it is our job to define clinically meaningful outcomes that directly address functional impairments. The ability to walk longer, faster, and on uneven surface is certainly a patient-centered goal and outcome, even if it is not quantified through a patient self-report questionnaire. These performance-based tests (identified in this report as intermediate outcomes) are essential to fully evaluate a patient's functional ability. Stratford and Kennedy[5] argue that in patients with knee OA, "self-report measures comment on the experience associated with the physical activities and that time or distance provides information concerning the ability to perform the activity. Furthermore, we suggest that the two assessment methods provide complementary information that is essential to both clinical research and clinical practice." In fact in the post-arthroplasty population, there is a marked discrepancy between the patient's perception of their abilities and actual ability to perform the task. Interventions that reduce pain, such as arthroplasty, or physical therapy treatments, may affect perception and performance differently. Mizner et al. make the push that clinical and research outcomes should be evaluated by patient-centered outcomes. However, APTA does not believe that patient-centered outcomes are limited to patient self-report questionnaires.	We revised the report emphasizing the role of patient self reported outcomes and clinical importance of the measures of function. We revised future research needs emphasizing necessity to define clinically meaningful outcomes and association between functional impairments and patient disability and quality of life. We acknowledge that gait speed and function, transfers, and other performance-based tests are essential to fully evaluate a patient's functional ability. We also synthesized the evidence about patient perception of clinical importance in performance-based tests. It should be noted that studies by Stratford and Kennedy and Mizner et al. were post-op studies, whereas we only focused on the non-surgery participants. Given physical function could drop profoundly after surgery, the findings in the references may not apply to our population. Furthermore, our patient-centered outcomes were selected by the TEP members in the beginning of the project. Finally, we also provided results on performance-based measures.

Commentator & Affiliation	Section	Comment	Response
APTA	General	Recommendation 4 (Successful Interventions) One of the primary conclusions of the report is that "few physical therapy interventions were shown to be effective". APTA would conclude that out of the interventions included in the analysis, the interventions most commonly employed by physical therapists to treat patients with knee OA (exercise and strengthening) were effective at reducing pain and disabilities, improving composite function and gait even in the long-term (12 months). It is possible that the authors conclude that "few" interventions are effective because they gave equal weight to modalities that are less frequently used (diathermy, PEMF, Tai Chi) or not used as a primary treatment intervention, but rather as an adjunct to more traditional care (heat, cryotherapy, massage). The pooled analysis revealed that the core physical therapy interventions were effective.	We revised the report providing separate conclusions for core PT interventions and for less frequently used passive interventions.
APTA	General	Recommendation 5 (Use of PASS) APTA supports the use of patient-centered outcomes such as the Patient Acceptable Symptom State (PASS). However, we believe that there are clinically important differences in PASS versus the use of a traditional visual analog scale (VAS). PASS is based on the concept of a pain threshold for which a patient could live with and not seek additional care. Although this conceptually makes sense, some patients may have preferences for different health states. In addition there is the potential for clinicians and or patients to game the system if they knew they would not receive care for reporting a certain level on a VAS.	We revised the report and conducted sensitivity analysis of treatment effects on pain measured with VAS. We then judged clinical importance of pain reduction according to minimum clinically important differences in VAS that have been determined in observational studies comparing VAS scores with patient perception of treatment success.
APTA	General	Recommendation 6 (Future Considerations for Physical Therapists) APTA agrees with the authors that physical therapists, as research professionals, should incorporate or evaluate the number of individuals who meet clinically important differences or demonstrate clinical improvements in addition to simply reporting average scores for continuous variables.	Thank you.
Peer Reviewer # 5	General	Quality of the Report: Good Number of Hours Spent to Review the Report: 5 I have assumed that the structure of the report has been determined by AHRQ. It is unfortunate that this structure makes the report so long and requires so much repetition of the same information.	We revised the report focusing on sufficient evidence from pooled analyses. We improved readability of the report by condensing key messages and tabulating synthesized evidence.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer # 6	General	Quality of the Report: Poor Number of Hours Spent to Review the Report: 6 I have tried to very carefully and thoroughly review this report. I know it represents an enormous amount of work. However I have several concerns that I believe will limit its usefulness and indeed may be counter-productive to quality health. First is the question: What is physical therapy? In many states it is the treatment provided by a physical therapist. That is, it is defined by the deliverer rather than the specific modality. The authors appear to have defined physical therapy as various modalities. Consequently they have included many interventions that are not typical physical therapy interventions including balneotheraoy (spa therapy), virtually unheard of in the US, and massage which is rarely if ever used by a PT for osteoarthritis. These inclusions and others like them limit the utility of the recommendations. The key questions listed by the group are very good. However we on the expert panel warned the authors that it was unlikely they would find useful data to answer these questions as they learned eventually.	We revised the report to clarify that our analytical framework was built according to the definitions and classifications of interventions and measures from the Guide to Physical Therapist Practice. We defined physical therapy and selected the interventions and methods to assess the outcomes in accordance with the Practice Pattern 4E: Impaired Joint Mobility, Motor Function, Muscle Performance, and Range of Motion Associated with Localized Inflammation from the Guide to Physical Therapist Practice. We also re-analyzed the data to explore the role of PT involvement in treatment benefits with examined interventions. We refer to balneotherapy, therapeutic touch, and massage following public discussion and recommendations from the stakeholders. We did synthesize evidence for PT interventions the stakeholders found applicable to PT practice in the U.S.
Peer Reviewer #1	General (Clarity and Usability)	Main points of the review are clearly and succinctly presented. Even the results section presents a summary of the findings from the reviewed studies rather than extensive descriptions of each included study. This approach results in a highly readable report. I have just a few suggestions that I hope might enhance accessibility to the main findings even further. I recommend the authors consider reorganizing the order in which the results for each treatment are presented for KQ1 so positive results are presented earlier. The current scheme is not intuitive from the introductory text.	Thank you for your suggestion. We summarized pain and disability outcomes associated with each PT intervention by strength of evidence. We added a table showing strength of evidence for positive results.
Peer Reviewer #2	General (Clarity and Usability)	I am not sure this report is going to have significant influence on PT practice. Reimbursement is still dependent on billing for procedures such as US, diathermy, e-stim, etc. Because studies are scarce that evaluate the combined interventions used typically in PT practice this isn't that compelling for the usual practitioner. I also feel the data are too weak for health plans to make policy decisions regarding covering certain interventions or not. In some cases where there is insufficient data or data supporting low efficacy, it is due to lack of large studies (low power),lack of measuring patient-oriented outcomes and exclusion of patients with cormobidities and not true non-effectiveness.	Our goal was a comprehensive synthesis of evidence with scientifically sound and transparent methods of analyses. Our findings can help policymakers in informed decisions regarding reimbursement for the PT interventions that showed little benefit in clinical trials.

Commentator & Affiliation	Section	Comment	Response
Reviewer #3	General (Clarity and Usability)	Report discusses many studies, results are not inherently usable in the clinical setting due to constraints discussed by the authors.	Thank you.
Reviewer #4	General (Clarity and Usability)	Yes, well structured and organized, clearly presented. I would be very careful with the conclusions being used to inform policy. The whole idea of physical therapy and the dozen approaches to treating OA are very individual in their perceived effectiveness in the patient. And where one day electro-stim seemed to reduce joint pain, another day it would not seem to have done anything at all. Depression, mental well-being, and all the other subjective areas are very difficult to put into quantitative terms, even though these may have a direct impact on the perceived effectiveness of weight training or heat therapy or anything else at any given time.	Thank you. We revised the report to provide additional information that can help in informed decisionmaking. We explored whether exercise effects differ in the trials with different PT involvement. We added thin information in the report.
Peer Reviewer # 5	General (Clarity and Usability)	clearest part of the report is the Executive Summary and this will be useful to inform both policy and practice	Thank you.
Peer Reviewer # 6	General (Clarity and Usability)	While the overall manuscript is well organized with the limitations identified above, it is cumbersome and difficult to use. As noted above the tables and figures need more detail to be helpful. Yes this report could be used for policy and practice decisions. However I believe this would be an unfortunate outcome. I believe the authors have drawn inaccurate conclusions regarding the strength of evidence for exercise, failed to put their conclusions regarding ultrasound in context, and have presented findings that not only will undermine the utilization of established treatments of exercise, but will also justify the use of other passive interventions based on limited data. Similarly, they seem to rule out interventions whose effectiveness has been studied in a limited way rather than to state that the data are too limited to allow a recommendation.	We revised the report with consistent criteria to rank strength of evidence for all PT interventions. We revised precision definitions following the recently published GRADE recommendations. We clarified that individual RCTs with high risk of bias provided insufficient evidence for valid effectiveness conclusions.

Commentator & Affiliation	Section	Comment	Response
Associate Editor	General (Question 1)	The document states in several places that a major conclusion is that no clinically important changes were found, yet there is no explicit analysis or presentation of the data supporting these statements for Key Question 1. It appears that the individual results of RCTs were checked for clinical significance, but not the pooled results, which I find to be very odd. In a systematic review, pooled results are of far more importance than individual study results. The lack of application of minimally clinically important differences to the pooled results/conclusion from Key Question 1 is a big flaw in the interpretation of the results. An effect may be statistically significant but be too small to be clinically significant. The confidence intervals around a non-significant effect may be small enough to exclude the possibility of a clinically significant effect, in which case a conclusion of "it doesn't work" or "the two treatments appear to be equally effective" would be a more valid conclusion than "it's not statistically significant". The major outcomes for Key Question 1- pain on VAS, SF-36- have well-accepted levels of clinical significance that should be explicitly listed in the document and should be checked against the summary effect sizes. See, for example: http://erj.ersjournals.com/content/37/4/755.full.pdf Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004 Aug;8(4):283-91. PMID: 15207508 Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001 Nov;94(2):149-58. PMID: 11690728 Hanley MA, Jensen MP, Ehde DM, Robinson LR, Cardenas DD, Turner JA, Smith DG. Clinically significant change in pain intensity ratings in persons with spinal cord injury or amputation. Clin J Pain 2006 Jan;22(1):25-31. PMID: 16340590 Braitman LE. Confidence intervals assess both clinical significance and statistical significance [editorial]. Ann Intern Med 1991 Mar 15;114(6):515-7. PMID: 1994799 Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. Pharmacoeconomics 1999 Feb;15(2):141-55. PMID: 10351188 Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988. 567 p.	We revised the report to clarify that we used standardization in all pooled analyses because the studies used different tools to measure the same outcomes. We clarified the units of standard mean differences as differences by standard deviations. We also conducted a new analyses calculating clinical important of non standard mean difference in pain measured with the same scale (VAS). We did a comprehensive literature review about minimally clinically important differences in WOMAC, VAS, and others scales.