Systematic Review

Arthroscopic Partial Meniscectomy Versus Physical Therapy for Degenerative Meniscus Lesions: How Robust Is the Current Evidence? A Critical Systematic Review and Qualitative Synthesis

Erik Hohmann, M.D., Ph.D., Vaida Glatt, Ph.D., Kevin Tetsworth, M.D., and Mark Cote, P.T., D.P.T., M.S.C.T.R.

Purpose: The purpose of this systematic review was to investigate study quality and risk of bias for randomized trials comparing partial meniscectomy with physical therapy in middle-aged patients with degenerative meniscus tears. **Methods:** A systematic review of Medline, Embase, Scopus, and Google Scholar was performed from 1990 through 2017. The inclusion criteria were at least 1 validated outcome score, and middle-aged patients (40 years and older) with a degenerative meniscus tear. Studies with a sham arm, and acute and concomitant injuries were excluded. Risk of bias was assessed with the Cochrane Risk of Bias Tool. The quality of studies was assessed with the Cochrane GRADE tool and quality assessment tool (Effective Public Health Practice Project). Publication bias was assessed by funnel plot and Egger's test. The I^2 statistics was calculated a measure of statistical heterogeneity. **Results:** Six studies were included, and all were assessed as having a high risk of bias. There was no publication bias ($P = .23$). All studies were downgraded (low, n = 5; very low, n = 1). The Effective Public Health Practice Project assessed 1 study as strong, 2 as moderate, and 3 as weak. The overall results demonstrated moderate to low quality of the included studies. The I^2 statistic was 96.2%, demonstrating substantial heterogeneity between studies. **Conclusions:** The results of this systematic review strongly suggest that there is currently no compelling evidence to support arthroscopic partial meniscectomy versus physical therapy. The studies evaluated here exhibited a high risk of bias, and the weak to moderate quality of the available studies, the small sample sizes, and the diverse study characteristics do not allow any meaningful conclusions to be drawn. Therefore, the validity of the results and conclusions of prior systematic reviews and meta-analyses must be viewed with extreme caution. The quality of the available published literature is not robust enough at this time to support claims of superiority for either alternative, and both arthroscopic partial meniscectomy or physical therapy could be considered reasonable treatment options for this condition. **Level of evidence:** Level II, systematic review of Level I and II studies.

Degenerative meniscal lesions are a common source of knee pain and are frequently diagnosed in middle-aged and elderly patients using magnetic resonance imaging.[1,2] These lesions are commonly associated with aging and osteoarthritis.[3,4] Englund et al.[3] demonstrated that the prevalence of meniscal damage in the 50- to 59-year age group was 32%, and 56% in persons aged between 70 and 90 years. However, it is still debated whether the associated symptoms occur as a consequence of the osteoarthritis, the meniscal tear, or the combination of these factors.[5-7]

Treatment of these lesions, if they are or become symptomatic, is currently a matter of considerable controversy.[8-11] The 2013 guidelines from the American Academy of Orthopedic Surgeons do not recommend arthroscopy with lavage and/or debridement in patients with primary symptomatic osteoarthritis of the knee.[12] However, the guidelines make no

From the Department of Orthopaedic Surgery and Sports Medicine, Valiant Clinic/Houston Methodist Group (E.H.), Dubai, United Arab Emirates; School of Medicine, University of Pretoria (E.H.), Pretoria, South Africa; School of Medicine, University of Queensland (E.H.), Queensland, Australia; University of Texas Health Science Center (V.G.), San Antonio, Texas, U.S.A.; Orthopaedic Research Center of Australia (V.G., K.T.), Kogoarah, Department of Orthopaedic Surgery, Royal Brisbane Hospital (K.T.), Herston, Department of Surgery, School of Medicine, University of Queensland (K.T.), Queensland, Australia; and Department of Orthopaedic Surgery, University of Connecticut (M.C.), Farmington, Connecticut, U.S.A.

The authors report the following potential conflicts of interest or sources of funding: E.H. is an associate editor for Arthroscopy Journal. *K.T. is a consultant for the AAOS. M.C. is an associate editor for* Arthroscopy Journal. *Full ICMJE author disclosure forms are available for this article online, as supplementary material.*

Received September 5, 2017; accepted April 24, 2018.

Address correspondence to Erik Hohmann, Valiant Clinic/Houston Methodist Group, PO Box 414296, Dubai 414296, United Arab Emirates. E-mail: ehohmann@hotmail.com

0749-8063/171070/$36.00

https://doi.org/10.1016/j.arthro.2018.04.018

recommendation either for or against partial meniscectomy in osteoarthritic patients with a torn meniscus.[12] Currently accepted indications for surgery are a clear history of mechanical symptoms, such as locking and catching, with joint line pain and/or acute onset of symptoms that have failed nonsurgical treatment.[8,13] Moreover, partial arthroscopic meniscectomy and debridement is also performed with the belief that partial resection treats the underlying cause rather than producing a placebo effect.[14]

Recently, the indication for arthroscopic surgery has been challenged by several randomized studies and meta-analyses, which were unable to demonstrate any clinical benefit from surgical treatment.[14-21] For example, Thorlund et al.[20] reported that knee arthroscopy is associated with harm and is not recommended for middle-aged or older patients with or without signs of osteoarthritis. The results of this study have been criticized for the inclusion of nonrelevant studies and other related biases.[11] Other studies demonstrated a superior outcome of arthroscopic partial meniscectomy in patients with symptomatic meniscal tears.[2,22] Recently, Ha et al.[23] critically reviewed the published literature and determined that valid conclusions cannot be drawn with regard to the optimal treatment for meniscal tears. Buchbinder[5] noted that, despite purportedly high-quality, randomized trials suggesting that arthroscopy is no more effective than placebo or nonoperative treatment, convincing evidence in support of nonoperative treatment for degenerative meniscal tears is also lacking.

The purpose of this systematic review was to investigate the study quality and risk of bias of randomized trials comparing partial meniscectomy with physical therapy in middle-aged patients with degenerative meniscus tears. We hypothesized that the quality of the currently available published literature would not be robust enough to allow valid conclusions of superiority of arthroscopic treatment versus physical therapy.

Methods

The research was conducted according to the methods described in the Cochrane Handbook.[24] The results of this study are reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines statement.[25]

Eligibility Criteria

All randomized Level I and Level II studies comparing arthroscopic partial meniscectomy with physical therapy published between 1990 and 2017 were identified. Studies were included if at least 1 validated outcome score was used, treated patients were of middle-age (defined here as 40 years and older), and had a degenerative meniscus tear. Patients with osteoarthritis were not specifically excluded. The studies were excluded if the protocol included a sham treatment; had Level III, IV, or V evidence; or was a case report, review, or letter to the editor. Concomitant injuries, such as acute and chronic cruciate or collateral ligament injuries, were also excluded. No specific restrictions were used for age to capture all published literature, and a final decision for inclusion or exclusion was based on a full text review.

Literature Research

A systematic review of the literature was performed to identify all relevant publications in the English and German literature on June 14, 2017. Medline, Embase, Scopus, and Google Scholar were searched using the terms and Boolean operators: “meniscus tear” AND “degenerative” AND/OR “knee arthroscopy”; “partial meniscectomy” AND “physical therapy” AND/OR “physiotherapy.” Two reviewers conducted independent title and abstract screening. All eligible articles were manually cross-referenced to ensure that all potential studies were included. Disagreements regarding the included studies were resolved by consensus; if no consensus was reached, studies were subjected to a full text review.

Data Extraction and Quality Assessment

For studies that met the inclusion criteria, an electronic data extraction form was used to obtain the following information from each article: author, journal, year of publication, any conflicts of interest, surgical technique and type, duration and number of physical therapy interventions (if available), sample size, study duration, duration of follow-up, and demographic data of the study population. Two authors independently completed data extraction, and the third reviewer and senior author verified the data.

Risk of bias was assessed adapting the Cochrane Collaboration’s Risk of Bias Tool.[24] The use of this tool allowed for the assessment of the adequacy of patient allocation, allocation concealment, blinding, clarity of outcome data, and the potential for selective reporting. A low risk of bias assigned allocated to studies that had low risk of bias assessments for all key domains; an unclear risk of bias if 1 or more key domains were found to have an unclear risk of bias, and a high risk of bias if 1 or more of the domains were assessed as high risk (Table 8.7a in the Cochrane Handbook).[24]

The GRADE system was used by 2 reviewers to assess the quality of the evidence for each outcome measure; the third reviewer verified these assessments. The recommendations from the Cochrane Handbook were followed, and studies were downgraded if there were limitations in the design, indirectness of evidence, unexplained heterogeneity, imprecision of results, and a high probability of publication bias. All institutional and author information was concealed to the third reviewer, who independently reviewed the included studies. Any disagreement between reviewers was resolved by consensus and/or arbitration of the 2 senior authors. The GRADE assessment served as the main outcome measure to determine whether the conclusions of the included studies were valid or inconclusive. Inconclusive was defined as any double or triple downgrade resulting in a low or very low quality rating. Possible inconclusiveness was defined as a single downgrade resulting in moderate quality. Only a high-quality rating was defined as conclusive. The factors resulting in a downgrade are outlined in Table 12.2.b in the Cochrane Handbook[24] and include limitations in the design, indirectness of evidence, unexplained study heterogeneity or inconsistency of results, imprecision of results, and high probability of publication bias. Factors that may increase the quality of the body of evidence are outlined in Table 12.2.c in the Cochrane Handbook[24] and include a large magnitude of effect. All plausible confounding would reduce a demonstrated effect or suggest a spurious effect when the results show no effect, dose−response gradient.

To include risk of bias assessment into the GRADE quality assessment, the following procedure was followed: studies with an unclear risk of bias were downgraded 1 level, whereas studies with a high risk of bias were downgraded 2 levels. Although this approach is somewhat arbitrary, the authors felt that it was important to incorporate risk of bias for the quality assessment. The results reported herein therefore are presented both with and without the risk of bias assessment.

In addition to GRADE, the Effective Public Health Practice Project (EPHPP) was used as a quality assessment tool. The EPHPP was initially developed for the assessment of public health studies and allows a comprehensive assessment of the overall quality of a quantitative study.[26,27] The tool rates selection bias, study design, confounders, blinding, data collection methods, withdrawals and dropouts, intervention strategy, and analysis. Three different quality ratings can be allocated: strong, moderate, or weak.[28]

Statistical Analysis

Interobserver differences for risk of bias and study quality were measured using Cohen's kappa coefficient. Meta-analysis was not performed because the included studies used different outcome measures, used methods of calculation, or did not consistently report measures of variability (standard deviation), thereby prohibiting pooling of the data to determine a common treatment effect. For the purposes of creating funnel plots to assess publication bias, the Knee Injury and Osteoarthritis Outcome Score,[2,29,30] Lysholm scores,[18,31] and Western Ontario and McMaster Universities Osteoarthritis Index[16] were used to establish treatment effects, Hedge's g, difference of means, and standard difference of means. For these pooled outcomes, a random effects model was selected. If the authors did not report standard deviations, the standard deviation was calculated using the following formula: standard deviation = max−min/4. Hozo et al.[32] demonstrated that this formula provides a good estimate for standard deviation. If publication bias was present based on visual inspection of the funnel plot, Egger's test of intercept was conducted to test for asymmetry. The I^2 statistic, as a measure of statistical heterogeneity, was calculated as a further measure of clinical and methodologic diversity.[33] The degree of heterogeneity was defined as suggested by Higgins et al.[34]: 0% to 25% low, 26% to 50% moderate, 51% to 75% moderately large, and greater than 75% high. Funnel plots, as well as all statistical analyses, were performed using STATA SE (Version 12.0; StataCorp, College Station, Texas) for Windows and the comprehensive meta-analysis software package (CMA), version 3 (Biostat, Englewood, NJ).

Results

Study Selection and Characteristics

The literature search initially identified 166 studies. After removal of duplicates, the abstracts of 76 studies were screened. After reviewing these 76 articles, 37 publications were excluded, and an examination of the remaining 39 full-text manuscripts was conducted. The eligibility criteria were met in only 6 of the 39 articles, and these 6 studies were then included in the analysis (Fig 1).[2,16,18,29-31] Overall, agreement between the 2 reviewers regarding final eligibility was excellent (kappa value of 0.93; 95% confidence interval, 0.91-0.95). All 6 studies were published in English between 2013-2018 and included a total of 905 patients. The study characteristics and results are summarized in Table 1.

Risk of Bias

The findings of the risk of bias assessment are summarized in Table 2. Using the criteria from the Cochrane Handbook, all 6 studies were assessed as having a high risk of bias. The domain blinding of participants and personnel was assessed as a high risk for all studies. The best performing study design was that of Kise et al.,[30] yet even in this publication, in addition to the blinding domain concerns, the domain other bias was assessed as high risk because 38% of those eligible refused participation, and reporting bias was assessed as unclear. Visual inspection of the funnel plot did not imply asymmetry, but 3 studies were outside of the projected triangle, suggesting the possibility of publication bias (Fig 2). However, Egger's intercept value was not significant (P = .23 2-tailed)

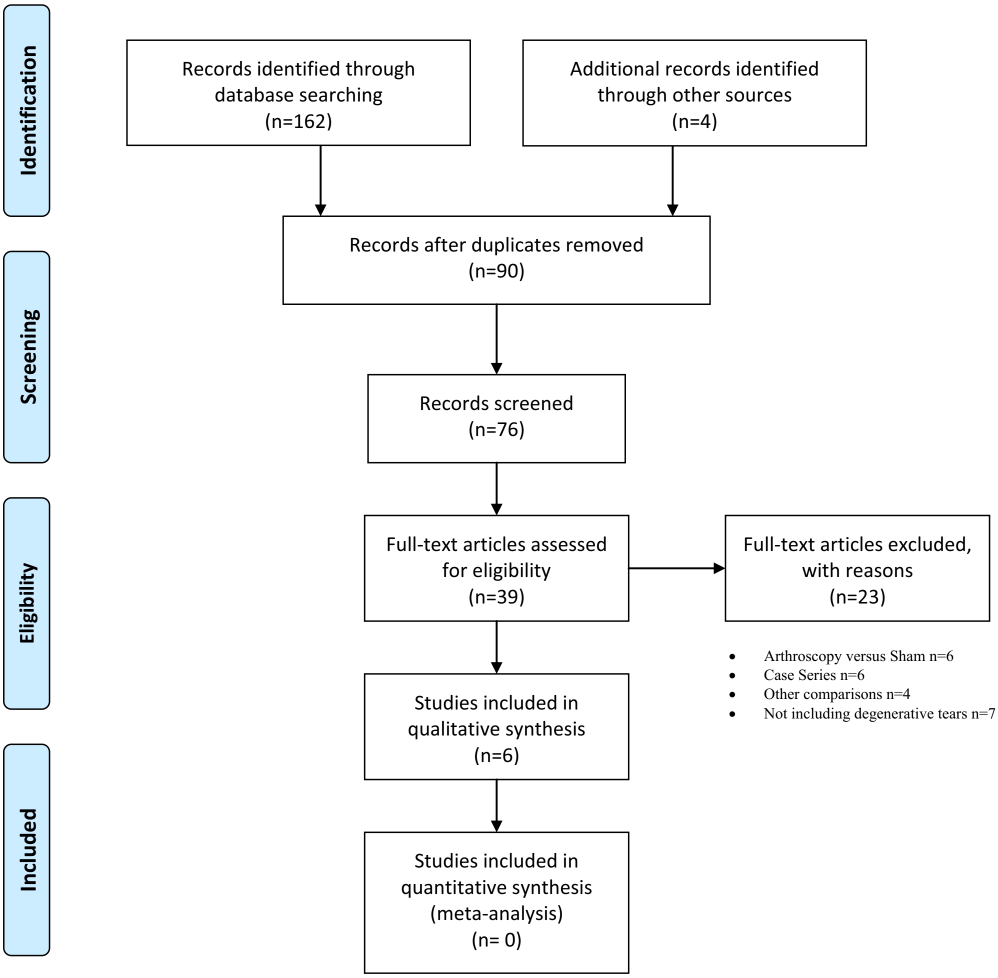

Fig 1. Of the initial 166, only 6 studies were then included in the analysis.

and was calculated to be −4.75 (95% confidence interval, −14.28 to 4.77; t = 1.38), refuting the possibility of publication bias.

Study Quality Assessment and Heterogeneity

Using the GRADE quality assessment criteria, the quality of evidence was double downgraded for 5 studies[2,16,18,29,30] because of limitations in the design and imprecision of results with wide confidence intervals. The study by Herrlin et al.[31] was triple downgraded based on the indirectness of evidence. These authors based inclusion on positive findings on magnetic resonance imaging and clinical history. Patients were then contacted by telephone and after giving written consent were randomized. However, it seemed that the clinical examination was only performed after randomization, and this resulted in at least 1 exclusion (Table 3).[17,31] When incorporating both the risk of bias assessment and the GRADE quality assessment, all 6 studies were downgraded 2 levels for their high risk of bias. Furthermore, when evaluating the studies as a whole, the differences in eligibility criteria, interventions, indirectness of evidence, and imprecision of results, the cumulative body of evidence was triple downgraded to a very low quality.

The results of the quality assessment tool for quantitative studies (EPHPP) are shown in Tables 3 and 4. The study by Kise et al.[30] scored the highest on all of the items and was assessed as a strong study. The studies by Gauffin et al.[2] and Yim et al.[18] had only 1 weak score and were assessed as moderate quality studies, whereas the articles by Herrlin et al.,[31] Østeras et al.,[29] and Katz et al.[16] were all assessed as weak and had 2 or more weak item scores. The I^2 statistic was 96.2%, demonstrating substantial heterogeneity between studies.

Discussion

The results of this systematic review demonstrated high and varying risk of bias, moderate to low methodologic quality, and substantial statistical heterogeneity among the 6 eligible randomized trials comparing arthroscopic partial meniscectomy with physical therapy treatment in middle-aged patients with a degenerative meniscus tear. This study found that differences in eligibility criteria, outcome measures, and the nature

Table 1. Studies Included in the Analysis: Demographic and Treatment Details

Study [LOE]	Mean Age PT/ASC	Male Sex (%) PT/ASC	Physical Therapy	Surgical	Outcome Measures	OA Inclusion PT/ASC	Length of Follow-up (mo)	Loss to Follow-Up PT/ASC	Outcome PT/ASC
Østeras et al., 2012[29] [LOE: 2]	47.0-52.7	88.9-62.5	n = 9 PT 3 times weekly for 3 months	n = 8 "Standard" partial meniscectomy	Pilot study VAS, KOOS	Kellgren grade 0-2 Not reported	3	No loss	VAS 2-2.6 KOOS 39.7-40.9 Compliance cons group: 84%
Herrlin et al., 2013[17] [LOE: 1]	56-54	63.3%-67.8%	n = 49 Exercise program for 8 weeks twice weekly	n = 47 "Standard protocol," details not mentioned	KOOS, Lysholm, Tegner	Ahlbäck, baselines not reported	60	4.1%-4.2%	Lysholm 95-99 Tegner 3-3 VAS 0-0
Yim et al., 2013[18] [LOE: 1]	57.6-54.9	80-82	n = 52 Physical exercise 60 min 3× weekly under PT guidance ×3 weeks then 8 weeks home exercise program ×8 weeks	n = 50 Limited debridement articular lesions, partial meniscectomy	VAS, Lysholm, Tegner	Kellgren grade 0-3 0: 67.7%-78% 1: 32.7%-22%	24	Not reported	VAS 1.7-1.8 Lysholm 84.3-83.2 Pain relief Complete 67%-68% Improved 23%-26% Persistent 10%-6%
Katz et al., 2013[16] [LOE: 1]	57.8-59	42.6-44.1	n = 169 Home-based exercise	n = 161 Partial meniscectomy, excision of loose cartilage	WOMAC, KOOS, SF36	Kellgren Grade 0-3 0: 21.3%-21.1% 1: 20.7%-16.1% 2: 23.1%-23% 3: 23.1%-28%	12	Not reported	WOMAC physical function improvement 18.5-20.9 at 6 mo
Gauffin et al., 2014[2] [LOE:1]		74.7-70.1	n = 75 3 months twice weekly, gym and home-based exercise program	n = 75 Partial meniscectomy	KOOS, EQ5D, VAS	Kellgren Grade 0-2 0: 43%-49% 1: 48%-45% 2: 9%-5%	12	13%	KOOS Pain: 13.9-16.6 KOOS Symp:9.8-15 KOOS ADL: 10.8-11.7 KOOS Sport: 9.2-21.1 KOOS QOL: 10.5-21.9 EQ5D: 0.06-0.16 VAS: 7-9.1
Kise et al., 2016[30] [LOE: 1]	50.2-48.9	61%-61%	n = 70 Neuromuscular and strength exercise over 12 weeks, 2-3 weekly session	n = 70 Partial meniscectomy	KOOS, SF36, 1-leg-hop, 6-m timed-hop	Kellgren 0-3 0: 70%-73% 1: 26%-23% 2: 3%-4% 3: 1%-%	24	14.3%-8.6%	KOOS: 25.3-24.4 improvement @ 24 mo

NOTE. Left column figures for PT, right column data for ASC.

ADL, activities of daily living; EQ5D, EuroQOL-5D; KOOS, Knee Injury and Osteoarthritis Outcome Score; LOE, level of evidence; PT/ASC, physical therapy/arthroscopy; SF36, Short Form 36; VAS, visual analog scale; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

Table 2. Risk of Bias Assessment

	Random Sequence Generation (Selection Bias)	Allocation Concealment (Selection Bias)	Blinding of Participants and Personnel (Performance Bias)	Blinding of Outcome Assessment (Detection Bias)	Incomplete Outcome Data (Attrition Bias)	Selective Reporting (Reporting Bias)	(Other Bias)
Østeras et al., 2012[29]	?	?	−	−	+	+	−
Herrlin et al., 2013[17]	+	+	−	−	+	?	−
Yim et al., 2013[18]	?	?	−	−	?	+	−
Katz et al., 2013[16]	+	?	−	?	+	?	−
Gauffin et al., 2014[2]	+	+	−	−	+	?	−
Kise et al., 2016[30]	+	+	−	+	+	?	−

+, low risk of bias; ?, unclear risk of bias; −, high risk of bias.

of nonoperative interventions, coupled with generally small samples contributed to a diverse but not generalizable group of studies that lacked precision. Consequently, it is clear that the evidence from these studies remains insufficient to reach meaningful conclusions regarding the superiority of 1 treatment over the other.

Recent systematic reviews and meta-analyses have questioned the benefit of surgical interventions in middle-aged patients with a symptomatic degenerative meniscus lesion.[19,20,23] Thorlund et al.[20] performed a systematic review comparing arthroscopic partial meniscectomy and/or debridement with various control treatment ranging from placebo surgery to exercise and concluded that the benefit of surgery was small and absent only 1 to 2 years after surgery. They also suggested that arthroscopy might be associated with harm, such as deep vein thrombosis, pulmonary embolism, infection, or death.[20] However, 8 of the 9 included studies were considered to have a high risk of bias, which considerably weakened the confidence of the results.[24] Although Thorlund et al.[20] did not assess methodologic quality, 4 of the 9 studies in their analysis were included in the present review and were deemed to be of only moderate to low quality.[2,16,18,29] Given the high risk of bias and moderate study quality,

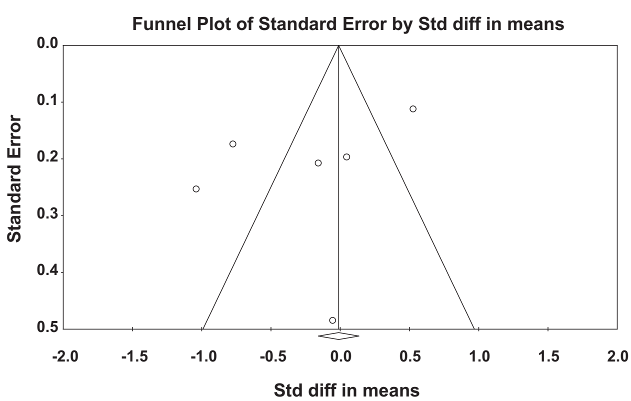

Fig 2. On visual inspection, the funnel plot seemed to be symmetric, but 3 studies were outside of the projected triangle, suggesting the possibility of publication bias. However, Egger's intercept value was not significant. Std dff, standard difference.

conclusions drawn from the synthesis of these studies would be considered dubious, at best.

The authors have also concluded that knee arthroscopy was associated with harm. They have included separate studies for this analysis and the scientific validity of this approach must be questioned. Hetsroni et al.[35] investigated the incidence of symptomatic pulmonary embolism in 413,323 outpatient procedures and reported 117 adverse events (0.03%), identifying age and operating time as the only variables that significantly increased the risk. Maletis et al.[36] investigated 20,770 cases and documented a 0.25% incidence of symptomatic venous thromboembolism and 0.17% for pulmonary embolism after elective knee arthroscopy, with only 1 postoperative death attributable to an embolism. A study by Hame et al.[37] identified the incidence of pulmonary embolus and deep vein thrombosis in 314,578 patients undergoing arthroscopic meniscectomy over the age of 65 from Medicare data. In their study cohort, 982 patients developed pulmonary embolus (0.3%) and 2,507 patients developed deep vein thrombosis (0.8%). In contrast, Katz et al.[16] directly compared adverse events between arthroscopic partial meniscectomy and physical therapy. They did not observe any differences between the 2 groups, suggesting that these adverse events might not be related to the surgical intervention, and instead reflect the normal incidence and prevalence of these phenomena. Katz et al.[16] performed an a priori sample size calculation that was based on a 10-point between-group difference. The difference is representative of the

Table 3. Quality Assessment

Study	Grade*	EPHPP
Østeras et al., 2012[29]	Low (1,4)	Weak
Herrlin et al., 2013[17]	Very low (1, 2, 4)	Weak
Yim et al., 2013[18]	Low (1 ,4)	Moderate
Katz et al., 2013[16]	Low (1, 4)	Weak
Gauffin et al., 2014[2]	Low (1, 4)	Moderate
Kise et al., 2016[30]	Low (1, 4)	Strong

EPHPP, Effective Public Health Practice Project.

*1, limitations in study design or execution; 2, indirectness of evidence; 3, unexplained heterogeneity; 4, imprecision of results; 5, publication bias.

Table 4. Results of the Quality Assessment using the Effective Public Health Practice Project

	Selection Bias	Study Design	Confounders	Blinding	Data Collection Methods	Withdrawals Dropouts	Total
Østeras et al., 2012[29]	3	1	2	3	1	1	Weak
Herrlin et al., 2013 [17]	3	1	1	3	1	1	Weak
Yim et al., 2013 [18]	2	1	1	3	1	1	Moderate
Katz et al., 2016[16]	3	1	1	3	1	1	Weak
Gauffin et al., 2014[2]	1	1	1	3	1	1	Moderate
Kise et al., 2016[30]	2	1	1	2	1	1	Strong

Component ratings: 1, strong; 2, moderate; 3, weak.

minimum clinically important difference of the Western Ontario and McMaster Universities Osteoarthritis Index functional scale and was also observed in their preliminary observational pilot data. The target sample size was a total of 340 patients, but only 334 were included in the final analysis and 48 patients crossed from physical therapy to the operative arm of the study. Applying strict scientific criteria, a type II error cannot, therefore, be entirely excluded.

In a systematic review by Khan et al.,[19] 7 studies were included. Five studies compared physical exercise with arthroscopic meniscectomy, 1 study sham with meniscectomy, and 1 study intra-articular steroid injection with meniscectomy. Similar to Thorlund et al.,[20] the risk of bias of the included studies was unclear owing to the small and nonsignificant effect size; therefore, study quality was not assessed. The conclusions by Khan et al.[19] are cautiously worded, suggesting no benefit from arthroscopic surgery in patients with degenerative meniscal tears. However, it is possible that unidentified limitations of the included studies did not even allow them to reach valid conclusions.

For the practicing physician, it is challenging to remain current with the huge volume of contemporary medical literature and the rapidly changing state of knowledge.[38] It has been suggested that systematic reviews and meta-analyses are the highest forms of evidence and should be considered a guideline to stay up to date on new clinical advances.[38-40] However, it is essential to understand that the quality of these reviews depend greatly on the quality of the primary studies, and their limitations may not allow any valid conclusions to be reached.[41] For example, Ioannidis[42] demonstrated that approximately 20% of the currently produced meta-analyses were inherently flawed, 13% were misleading, 17% were acceptable but not useful, 27% were redundant and unnecessary, and only 3% were of good quality and useful. Therefore, the application of these evidence-based recommendations may not be beneficial, and may even be considered harmful.[43]

Examining potential differences in the outcomes between operative or nonoperative treatment can be challenging. Blinding of the subject and the provider is generally not practical. Although this reality creates an inevitable assessment of high in the blinding domain on the Risk of Bias tool, additional measures can be included to ensure that all other aspects of the study are conducted in a manner to reduce the threat of bias. In the present review, 4 of the 6 studies did not blind the study groups to the personnel making outcome assessments. Furthermore, knowledge of the group assignment introduces the potential for detection bias, and there were various domains where the bias assessment was rated as unclear. Three of 6 studies did not adequately describe methods of allocation concealment. This factor may reflect a reporting problem, because the allocation of group assignments may have been concealed but not reported in the manuscripts. However, having the knowledge of group assignment would have introduced the potential for selection bias.

In addition to the defined domains of the Risk of Bias tool, all studies included in this review had further concerns of additional biases. Four studies reported a lack of participation by eligible subjects, creating concern for overselection bias.[16-18,30] Both cross-over and intention-to-treat analyses potentially introduce additional risk of bias. Cross-over refers to patients who were allocated to 1 group and switched to another against protocol. It likely contributed to the heterogeneity and certainly contributed to the studies being lower quality. One-way cross-over could introduce bias owing to nonadherence to the randomization protocol.[44] With 2-way cross-over, the likelihood of bias is decreased; however, if more patients switch from the failed nonsurgical group into the surgical treatment group, similar conclusions would apply. In some cases, the biases introduced may result in conclusions that are wrong.[44] Four studies reported cross-over ranging from 10% to 30%,[2,16,17,30] 1 study did not specifically report cross-over,[18] and the pilot study by Østeras et al.[29] had no cross-over between the 2 small groups. With the addition of cross-over bias to the other methodologic flaws, the validity of the conclusions made by the individual authors of systematic reviews and meta-analyses, including this review, must be interpreted with caution.

In an effort to fully evaluate the existing evidence, this systematic review used 2 different quality assessment tools in addition to the Cochrane Risk of Bias Tool. Using the GRADE quality assessment, 5 of the

6 studies[2,16,18,29,30] were of low quality, and 1 study was of very low quality.[31] When applying GRADE across the 6 included studies, the overall quality had to be further downgraded owing to the differences in eligibility criteria, interventions, and outcomes from study to study. In addition, the high risk of bias and heterogeneity further weakened the quality and confidence of the results.[24] In contrast with GRADE, EPHPP assigned a strong quality to the study by Kise et al.[30] Despite the strong rating, the components or selection bias and blinding were assessed as only moderate, introducing a potential bias.

Bias and methodologic quality aside, the included studies varied substantially in terms of eligibility criteria, outcomes, and interventions. Five of the 6 studies identified knee pain or daily symptoms as the main factors for inclusion and used the Kellgren Lawrence scale as an inclusion criteria; however, the distribution of the grades varied substantially across studies.[2,16,18,29,31] In addition, only Yim et al.[18] specifically mentioned mechanical symptoms affecting daily living as a mandatory requirement to determine the treatment outcomes, yet they failed to elaborate on what constituted mechanical symptoms. The nature of nonoperative intervention was generally described and varied widely across studies. The lack of a well-described and structured rehabilitation programs makes it difficult to determine exactly what was being compared in and amongst the studies.

Sample size was also consistently problematic among the eligible studies reviewed here. All of the included studies performed an a priori sample size calculation, but only the study by Kise et al.[30] recruited the required number of patients. Furthermore, when performing a post hoc calculation, the calculated power was only 32%, and in the studies by Østeras et al.[29] and Yim et al.[18] the calculated post hoc power was only 11%. Post hoc power analysis can be considered a futile exercise, confirming studies not reaching significance are not adequately powered. However, the wide confidence intervals among the studies included here indicate a lack of precision in estimating differences in treatment effects between arthroscopic partial meniscectomy and physical therapy treatment. Studies with larger sample sizes are absolutely necessary to definitively determine if meaningful differences in outcomes exist between these 2 groups. Randomized clinical trials in orthopaedic surgery have several additional weaknesses such as validity limited to a specific study population reducing external validity, outcome measures not correlating with outcomes of interest, resource intensity, and that completion may not occur until after the introduction of new treatment methods.[45,46] Consensus statements based on pooling expert opinions may be a very reasonable alternative to the current evidence-based approach.[46]

Limitations

Quality assessment of published research depends on the subjective assessment of the investigators and may reflect their own biases. Even though GRADE and EPHPP allow for a certain amount of subjectivity, the high kappa value of 0.93 indicated excellent agreement and strongly suggests that the subjectivity in this review was low and almost certainly did not influence the outcome. However, this only mitigates but does not eliminate the risk of bias, because the 2 authors can theoretically have bias in their assessment but agree with each other. The principle limitations of systematic reviews and meta-analyses are always directly related to the limitations of the included studies. However, these limitations are not applicable here because the main purpose of this systematic review was to investigate these limitations and to assess the rigor and strength of the currently available evidence.

Conclusions

The results of this systematic review strongly suggest there is currently no compelling evidence to support arthroscopic partial meniscectomy versus physical therapy. The studies evaluated here exhibited a high risk of bias, and the weak to moderate quality of the available studies, the small sample sizes, and the diverse study characteristics do not allow any meaningful conclusions to be drawn. Therefore, the validity of the results and conclusions of prior systematic reviews and meta-analyses must be viewed with extreme caution. The quality of the available published literature is not robust enough at this time to support allegations of superiority for either alternative, and both arthroscopic partial meniscectomy or physical therapy could be considered reasonable treatment options for this condition.

References

1. Englund M, Guermazi A, Gale D, et al. Incidental meniscal findings on knee MRI in middle-aged and elderly persons. *N Engl J Med* 2008;359:1108-1115.
2. Gauffin H, Tagesson S, Meunier A, Magnusson H, Kvist J. Knee arthroscopic surgery is beneficial to middle-aged patients with meniscal symptoms: A prospective randomised single blinded study. *Osteoarthritis Cartilage* 2014;22: 1808-1816.
3. Englund M, Haugen IK, Guermazi A, et al. *Osteoarthritis Cartilage* 2016;24:270-273.
4. Bhattacharyya T, Gale D, Dewire P, et al. The clinical importance of meniscal tears demonstrated by magnetic resonance imaging in osteoarthritis of the knee. *J Bone Joint Surg Am* 2003;85:4-9.
5. Buchbinder R. Meniscectomy in patients with knee osteoarthritis and a meniscal tear? *N Engl J Med* 2013;368: 1740-1741.
6. Marx RG. Arthroscopic surgery for osteoarthritis of the knee? *N Engl J Med* 2008;359:1169-1170.

7. Mezhov V, Teichtahl AJ, Strasser R, Wluka AE, Cicuttini FM. Meniscal pathology: The evidence for treatment. *Arthritis Res Ther* 2014;16:206.
8. Krych AJ, Carey JL, Marx RG, et al. Does arthroscopic knee surgery work? *Arthroscopy* 2014;30:544-545.
9. Howell JR, Handoll HH. Surgical treatment for meniscal injuries of the knee in adults. *Cochrane Database Syst Rev* 2000:CD001353.
10. Elattrache N, Lattermann C, Hannon M, Cole B. New England Journal of Medicine evaluating the usefulness of meniscectomy is flawed. *Arthroscopy* 2014;30:542-543.
11. Bollen SR. Is arthroscopy of the knee completely useless? *Bone Joint J Br* 2015;97:1591-1592.
12. Jevsevar DS. Treatment of osteoarthritis of the knee: Evidence-based guideline, 2nd edition. *J Am Acad Orthop Surg* 2013;21:571-576.
13. Conaghan PG, Dickson J, Grant RL, Guideline development group. Care and management of osteoarthritis in adults: Summary of NICE guidance. *BMJ* 2008;336:502-503.
14. Sihvonen R, Englund M, Turkiewicz A, Jarvinnen TLN. Mechanical symptoms and arthroscopic partial meniscectomy in patients with degenerative meniscus tears. *Ann Intern Med* 2016;164:449-455.
15. Sihvonen R, Englund M, Turkiewicz A, Jarvinnen TLN. Mechanical symptoms as an indication for knee arthroscopy in patients with degenerative meniscus tear: A prospective cohort study. *Osteoarthritis Cartilage* 2016;24:1367-1375.
16. Katz JN, Brophy RH, Chaisson CE, et al. Surgery versus physical therapy for a meniscal tear and osteoarthritis. *N Engl J Med* 2013;368:1675-1684.
17. Herrlin SV, Wange PO, Lapidus G, Hallander M, Werner S, Weidenhielm L. Is arthroscopic surgery beneficial in treating non-traumatic degenerative medial meniscus tears? A five year follow-up. *Knee Surg Sports Traumatol Arthrosc* 2013;21:358-364.
18. Yim JH, Seon JK, Soong EK, et al. A comparative study of meniscectomy and nonoperative treatment for degenerative horizontal tears of the meniscus. *Am J Sports Med* 2013;41:1565-1570.
19. Khan M, Evaniew N, Bedi A, Ayeni OR. Arthroscopic surgery for degenerative tears of the meniscus: A systematic review and meta-analysis. *CMAJ* 2014;186:1057-1064.
20. Thorlund JB, Juhl CB, Roos EM, Lohmander LS. Arthroscopic surgery for degenerative knee: Systematic review and meta-analysis of benefits and harms. *BMJ* 2015;350:h2747.
21. Marsh JD, Birmimgham TB, Giffin JR, et al. Cost-effectiveness analysis of arthroscopic surgery compared with non-operative management for osteoarthritis of the knee. *BMJ Open* 2016;5:e009949.
22. El Ghazaly S, Rahman AAA, Yusry AH, Fathalla MM. Arthroscopic partial meniscectomy is superior to physical rehabilitation in the management of symptomatic unstable meniscal tears. *Int Orthop (SICOT)* 2015;39:769-775.
23. Ha AY, Shalvoy RM, Voisinet A, Racine J, Aaron RK. Controversial role of arthroscopic meniscectomy of the knee: A review. *World J Orthop* 2016;7:287-292.
24. Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions. Version 5.1.9 [updated March 2011]*. London, UK: The Cochrane Collaboration, 2011.
25. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analysis: The PRISMA statement. *Int J Surg* 2010;8:336-341.
26. Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews Evid Based Nurs* 2004;1:176-184.
27. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: A comparison of the Cochrane Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract* 2012;18:12-18.
28. National Collaborating Centre for Methods and Tools (2008). Quality assessment tool for quantitative studies method. Hamilton, ON: McMaster University. http://www.nccmt.ca/resources/search/15. Published 13 April 2010. Accessed June 17, 2017.
29. Østeras H, Østeras B, Torstensen TA. Medical exercise therapy, and not arthroscopic surgery, resulted in decreased depression and anxiety in patients with degenerative meniscus injury. *J Bodywork Movement Ther* 2012;16:456-463.
30. Kise NJ, Risberg MA, Stensrud S, Ranstam J, Engebretsen L, Roos EW. Exercise therapy versus arthroscopic partial meniscectomy for degenerative meniscal tear in middle-aged patients: Randomised controlled trial with two year follow-up. *BMJ* 2016;354:i3740.
31. Herrlin S, Hallander M, Wange P, Weidenhielm L, Werner S. Arthroscopic or conservative treatment of degenerative medial meniscus tears: A prospective randomised trial. *Knee Surg Sports Traumatol Arthrosc* 2007;15:393-401.
32. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005;5:13.
33. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;1:1539-1558.
34. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analysis. *BMJ* 2003;327:557-560.
35. Hetsroni I, Lyman S, Do H, Mann G, Marx RG. Symptomatic pulmonary embolism after outpatient procedures of the knee: The incidence and risk factors in 418,323 arthroscopies. *Bone Joint Surg Br* 2011;93:47-51.
36. Maletis GB, Inacio MC, Reynolds S, Funahashi TT. Incidence of symptomatic venous thromboembolism after elective knee arthroscopy. *J Bone Joint Surg Am* 2012;94:714-720..
37. Hame SL, Nguyen V, Ellerman J, Ngo SS, Wang JC, Gamradt SC. Complications of arthroscopic meniscectomy in the older population. *Am J Sports Med* 2012;40:1402-1405.
38. Delaney A, Bagschaw SM, Ferland A, Manns B, Laupland KB, Doig CJ. A systematic evaluation of the quality of meta-analyses in the critical care literature. *Crit Care* 2005;9:R575-R582.
39. Cook DJ, Meade MO, Fink MP. How to keep up with the critical care literature and avoid being buried alive. *Crit Care Med* 1996;24:1757-1768.

40. Bartels EM. How to perform a systematic search. *Best Pract Res Clin Rheumatol* 2013;27:295-306.
41. Bhandari M, Morrow F, Kulkarni AV, Tornetta P III. Meta-analysis in orthopaedic surgery. *J Bone Joint Am* 2001;83:15-24.
42. Ioannidis JPA. The mass production of redundant, misleading, and conflicting systematic reviews and meta-analyses. *Milbank Q* 2016;94:485-514.
43. Stahel PF, Mauffrey C. Evidence-based medicine. A hidden threat for patient safety and surgical innovation? *Bone Joint J* 2014;96B:997-999.
44. Lubowitz JH, D'Agostino RB Jr, Provencher MT, Rossi MJ, Brand JC. Can we trust knee meniscus studies? One-way confounds intent-to-treat statistical methods. *Arthroscopy* 2016;32:2187-2190.
45. Frieden TR. Evidence for health decision making: Beyond randomized controlled trials. *N Engl J Med* 2017;377:465-475.
46. Hohmann E, Brand JC, Rossi MJ, Lubowitz JH. Expert opinion is necessary: Delphi panel methodology facilitates a scientific approach to consensus. *Arthroscopy* 2018;34:349-351.