European Radiology (2023) 33:5269–5281
https://doi.org/10.1007/s00330-023-09548-6

European Radiology
ESR EUROPEAN SOCIETY OF RADIOLOGY

REVIEW

Does paraspinal muscle morphometry predict functional status and re-operation after lumbar spinal surgery? A systematic review and meta-analysis

Gengyu Han[1,2,3] · Haotian Wu[1,2,3] · Jinyue Dai[1,2,3] · Xinhang Li[1,2,3] · Lihao Yue[1,2,3] · Zheyu Fan[1,2,3] · Qiaoyu Li[1] · Qirui Shao[1] · Yu Jiang[1,2,3] · Weishi Li[1,2,3]

Received: 31 October 2022 / Revised: 12 February 2023 / Accepted: 24 February 2023 / Published online: 28 March 2023
© The Author(s) 2023

Abstract

Objectives Whether paraspinal muscle degeneration is related to poor clinical outcomes after lumbar surgery is still indistinct, which limits its clinical application. This study aimed to evaluate the predictive value of paraspinal muscle morphology on functional status and re-operation after lumbar spinal surgery.

Methods A review of the literature was conducted using a total of 6917 articles identified from a search of PubMed, EMBASE, and Web of Science databases through September 2022. A full-text review of 140 studies was conducted based on criteria including an objective assessment of preoperative paraspinal muscle morphology including multifidus (MF), erector spinae (ES), and psoas major (PS) in addition to measuring its relationship to clinical outcomes including Oswestry disability index (ODI), pain and revision surgery. Meta-analysis was performed when required metrics could be calculated in ≥ three studies, otherwise vote counting model was a good alternative to show the effect direction of evidence. The standardized mean difference (SMD) and 95% confidence interval (CI) were calculated.

Results A total of 10 studies were included in this review. Of them, five studies with required metrics were included in the meta-analysis. The meta-analysis suggested that higher preoperative fat infiltration (FI) of MF could predict higher postoperative ODI scores (SMD = 0.33, 95% CI 0.16–0.50, $p = 0.0001$). For postoperative pain, MF FI could also be an effective predictor for persistent low back pain after surgery (SMD = 0.17, 95% CI 0.02–0.31, $p = 0.03$). However, in the vote count model, limited evidence was presented for the prognostic effects of ES and PS on postoperative functional status and symptoms. In terms of revision surgery, there was conflicting evidence that FI of MF and ES could predict the incidence of revision surgery in the vote count model.

Conclusion The assessment of MF FI could be a viable method to stratify patients with lumbar surgery by the risk of severe functional disability and low back pain.

Key Points

- *The fat infiltration of multifidus can predict postoperative functional status and low back pain after lumbar spinal surgery.*
- *The preoperative evaluation of paraspinal muscle morphology is conducive for surgeons.*

Keywords Treatment outcome · Paraspinal muscles · Surgical procedures, operative · Meta-analysis

Gengyu Han and Haotian Wu contributed equally to this article

✉ Weishi Li
puh3liweishi@163.com

[1] Department of Orthopaedics, Peking University Third Hospital, No. 49 NorthGarden Road, Haidian District, Beijing 100191, China

[2] Engineering Research Center of Bone and Joint Precision Medicine, Ministry of Education, Beijing, China

[3] Beijing Key Laboratory of Spinal Disease Research, Beijing, China

Abbreviations

CI	Confidence interval
CSA	Cross-sectional areas
DDD	Degenerative disc disease
ES	Erector spinae
FI	Fat infiltration
JBI	Joanna Briggs Institute
LBP	Low back pain
LDH	Lumbar disc herniation
LLIF	Lateral lumbar interbody fusion
LSS	Lumbar spinal stenosis

MF	Multifidus
NRS	Numerical rating scale
ODI	Oswestry disability index
OR	Odd ratio
PEM	Paraspinal extensor muscle
PLIF	Posterior lumbar interbody fusion
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PS	Psoas major
SMD	Standardized mean difference
SWiM	Synthesis without meta-analysis
tCSA	Total cross-sectional area
TLIF	Transforaminal lumbar interbody fusion
VAS	Visual analogue scale

Introduction

Spine surgery is a common adjunct treatment for degenerative spinal diseases, which has increased significantly in recent years [1]. However, this is associated with a growth in the frequency of inferior postoperative outcomes [2, 3]. Studies have revealed that the degeneration of paraspinal muscles, which can be generally found among elderly patients, is implicated in multiple degenerative lumbar pathologies [4–6].

Currently, the value of paraspinal muscle morphometry on image examination serving as a prognostic factor for several surgical disciplines including metastatic disease, trauma, and fracture is being unearthed [7–9]. Multifidus (MF), erector spinae (ES), and psoas major (PS), serving as primary extensor and flexor muscles, are generally quantified by the cross-sectional areas (CSA) and fat infiltration (FI) on magnetic resonance image (MRI). Although abundant work has been carried out to identify potential factors for the prognosis of spine surgery [10, 11], few unequivocal predictive factors related to paraspinal muscle have come to light.

Preoperative assessment of paraspinal muscle morphology may be conducive to identifying patients who tend to have unsatisfactory clinical outcomes and thus making precautionary measures in advance. Two systematic reviews have concluded the degree of preoperative paraspinal muscle degeneration in relation to several complications after spinal surgery [12, 13]. However, both of them could not conduct a meta-analysis because of high heterogeneity. Besides, they did not focus on the patients' functional status and re-operation after surgery. Moreover, there exist conflicting results on whether paraspinal muscle morphology was associated with clinical outcomes in patients with lumbar surgery [14, 15]. In consequence, our systematic review and meta-analysis primarily aim to elucidate the predictive value of preoperative paraspinal muscle morphology on functional status, symptoms, and re-operation in patients with surgery for degenerative lumbar diseases.

Materials and methods

Search strategy and eligibility criteria

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement was used to structure this systematic review and meta-analysis. To retrieve interrelated articles, we conducted a search in the following three databases: PubMed, EMBASE, and Web of Science databases through September 2022. All fields were searched for these terms: "paraspinal muscle," "paravertebral muscle," "multifidus," "erector spinae," or "psoas major"; and "surgery," "operative," "clinical outcome," or "functional status"; and "lumbar" or "lumbosacral." Two authors assessed all abstracts and titles to rate adherence to review criteria. Inclusion criteria consisted of the following: (1) articles including adults with degenerative lumbar diseases; (2) assessment of any lumbar paraspinal muscle characteristic on MRI or computed tomography preoperatively; (3) assessment of any clinical outcomes after lumbar surgery; (4) analyzed the relationship between preoperative imaging data and postoperative outcomes; (5) articles were published in English. Studies were excluded if they included subjects < 18 years of age; assessed lumbar muscle through nonconventional MRI (such as functional MRI, MRS, and chemical-shift MRI) and ultrasonography; included only postsurgical subjects. Studies included randomized controlled trials, cohort studies, case–control studies, and case series.

Assessment of risk of bias

We used Quality in Prognosis Studies (QUIPS), a widely accepted tool for evaluating the risk of bias in prognostic studies [16]. All articles meeting review criteria were evaluated independently for risk of bias by two authors, with any differences in assessment resolved by discussions until consensus was reached. The QUIPS contained 6 separate domains: study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis and reporting. Each domain was judged as low, moderate, or high risk of bias, according to published criteria [17]. The overall risk of bias for an included study was defined as low risk with ≥ 4 low- and no high-risk domains, moderate risk with < 4 low- and no high-risk domains, and high risk with ≥ 1 high-risk domain [18].

Data extraction

Two authors independently extracted the following information from included studies: study design, participant characteristics, details of assessments of preoperative lumbar muscle characteristics, and study results that were relevant

to our research question. Any disagreement would be adjudicated by a third author.

Measures and outcomes

The FI and atrophy of paraspinal muscles were evaluated by MRI or computed tomography. The parameters of high FI covered the increased percentage of the fat area or signal intensity of muscles, and the parameters of muscle atrophy included the declined area of total or lean paraspinal muscles. We defined the paraspinal extensor muscle (PEM) group as the integrity of MF and ES [19]. Oswestry disability index (ODI) scores were adopted to evaluate the postoperative functional disability [20]. Visual analogue scale (VAS) and numerical rating scale (NRS) were used to determine the postoperative pain at low back or legs [21]. The rate of revision surgery was also recorded.

For the applicability of synthesis, we have conducted subgroup analyses with a consistently measured paraspinal muscle and its method of morphology measure to reduce the possible heterogeneity among studies. Similar subgroup analyses were conducted in previous studies [19, 22].

Data analysis and levels of evidence

The literature presented data in various forms, including the means with/without standard deviations (SDs) for continuous outcomes, odds ratios (ORs) for dichotomous outcomes, and correlation coefficient (r) in correlation analysis. In view of that, a standardized mean difference (SMD) was calculated from the reported means and SDs [23]. The SDs could also be estimated from *t* test value or *p* value with the degree of freedom, if not directly reported [24]. Other different metrics (ORs and correlation coefficients) were transformed into SMDs, for the feasibility of overall comparison among various metrics [25]. The effect size was defined as small, medium, large, or very large with a SMD at 0.10–0.34, 0.35–0.64, 0.65–1.19, or > 1.20 respectively [26].

Meta-analyses were performed to show the weighted effect size for homogenous comparisons when required metrics could be calculated in ≥ 3 studies according to previous studies [27, 28]. The analysis was completed by Revman 5.4.1 and Rstudio software [29]. The forest plots were used to report SMDs and 95% CIs for separate studies and weighted effects. The heterogeneity was evaluated by Cochran's Q and I^2 statistics tests. The fixed-effects model was selected when the heterogeneity was acceptable with $p > 0.10$ and $I^2 < 50\%$; otherwise, a random-effects model was fitted [27]. Publication bias was displayed by funnel plots and Egger's test [23, 27].

When the meta-analysis was inappropriate due to the lack of metrics reported in < 3 studies, a vote-counting model was conducted according to previous studies, with the direction of effect [30–32]. The model followed the Cochrane and Synthesis without meta-analysis (SWiM) reporting guidelines [33, 34]. We performed a qualitative summary of evidence for lumbar muscle characteristics as predictors of postoperative outcomes. The model followed the definition for levels of evidence in previous reviews [35–37]: "strong" evidence was defined as consistent findings (≥ 75%), among which ≥ 2 studies were at low risk; "moderate" evidence was defined as consistent findings (≥ 75%), among which one study was at low-risk and ≥ 1 studies were at moderate- or high-risk; "limited" evidence was defined as consistent findings (≥ 75%), among which one study was at low- or moderate-risk or ≥ 2 studies were at high-risk; "conflicting" evidence was defined as inconsistent findings irrespective of study quality. One high-risk study was not considered in the vote-counting model. The effect direction plot displayed the direction of effect in the model [28, 34].

Results

For the review of preoperative paraspinal muscle characteristics and their predictive value in clinical outcomes after spinal surgery, 6917 studies were identified through database searching. After the removal of duplicate records, 3482 studies were screened in our study. Of these articles, 140 were deemed to be eligible for full-text review. Finally, a total of ten articles were included. Of them, five studies with required metrics were included in the meta-analysis. The search flow diagram is shown in Fig. 1. The meta-analyses for the relationships between preoperative MF FI, postoperative ODI, and postoperative back pain were performed in a fixed-effects model since the heterogeneities were both acceptable ($p = 0.13$, $I^2 = 47\%$ for ODI; $p = 0.43$, $I^2 = 0\%$ for back pain).

Study characteristics and risk of bias

Of ten included studies, six studies investigated participants with lumbar disc herniation (LDH, $n = 1$) [38] or lumbar spinal stenosis (LSS, $n = 5$) [14, 15, 21, 39, 40], one reported spondylolisthesis [41], and three reported multiple lumbar degenerative diseases [42–44]. All articles examined the relationship between preoperative paraspinal muscle and postoperative symptoms and functional status and two articles examined the re-operation. All studies assessed the muscle morphology by MRI. There were five cohort studies, four case series, and one secondary analysis in the randomized controlled trial. After the assessments by QUIPS, there were six, one, and three studies considered to have a low, moderate, and high risk of bias, respectively (Fig. 2 and Supplementary table 1). The characteristics of the included studies are summarized

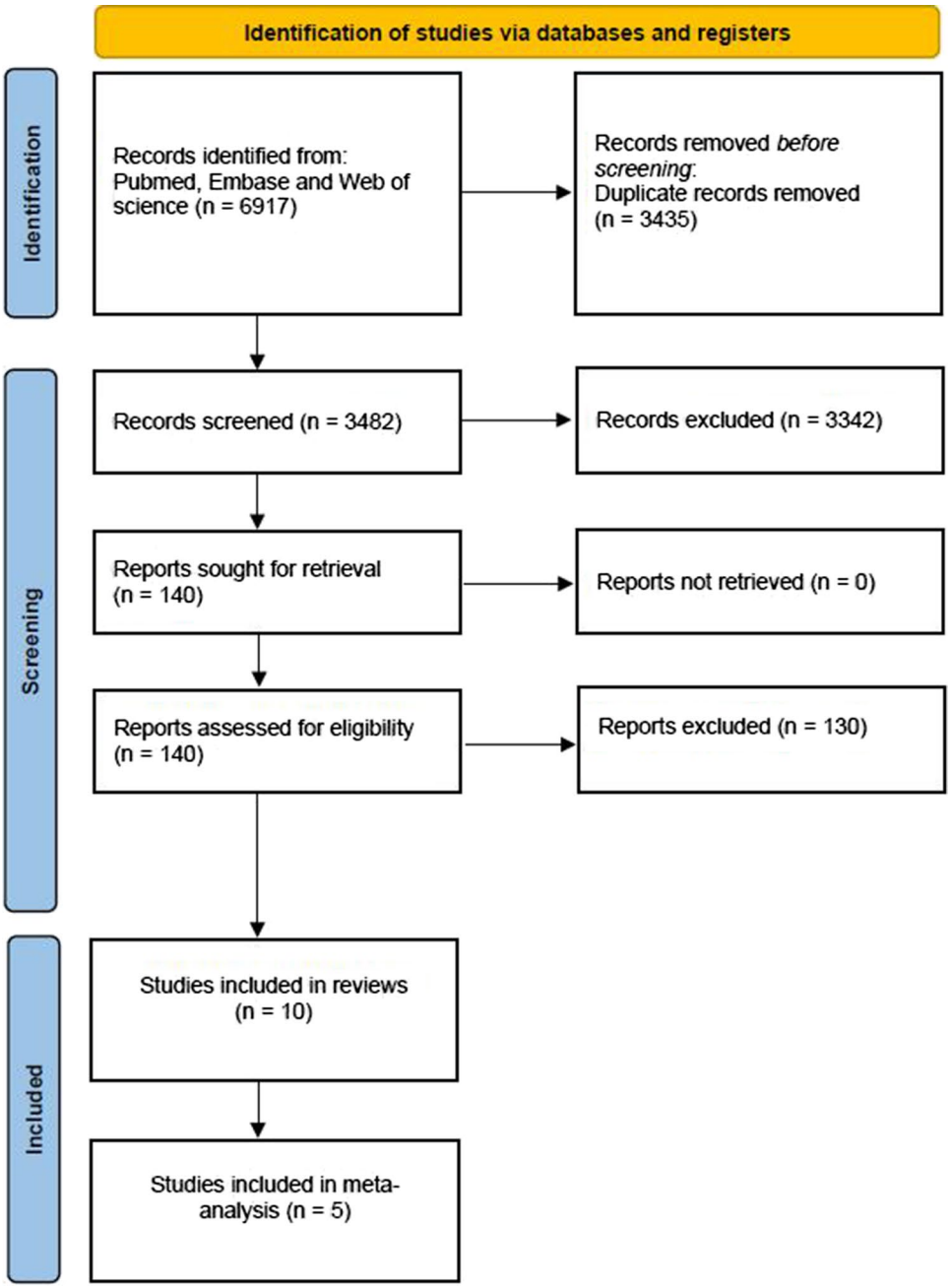


Fig. 1 Flowchart of the study selection and inclusion process. A total of 6917 articles were identified from databases through September 2022. After the removal of duplicate records, 3482 studies were screened. Of these articles, 140 were eligible for full-text review. A total of 10 articles were included in the review and 5 articles were included in the meta-analysis

in Table 1. In addition, publication bias was not observed in meta-analyses based on funnel plots and Egger's test (Supplementary Figs. 1 and 2).

Postoperative functional status and symptoms

Postoperative ODI

Eight studies investigated the relationship between preoperative paraspinal muscle morphology and postoperative ODI [14, 15, 38–43]. A meta-analysis including four studies with sufficient data in the fixed-effects model was performed for comparison in postoperative ODI scores between groups with high or low FI in MF [39, 40, 42, 43]. The meta-analysis in the fixed-effects model revealed that patients with a high grade of preoperative MF FI had higher postoperative ODI scores, compared to those with a low grade of MF FI (SMD = 0.33, 95% CI 0.16–0.50, $p = 0.0001$; Fig. 3).

In the vote-counting model, whether MF atrophy was related to postoperative ODI remained indistinct (Table 2). Zotti et al [14] found that greater tCSA (< 8.5 cm^2) of MF at L4-5 predicted a larger improvement in ODI (SMD = 0.85, $p = 0.006$). Chen et al [41] reported that a higher muscle/vertebra CSA ratio (> 0.60) of MF at L4-5 was associated

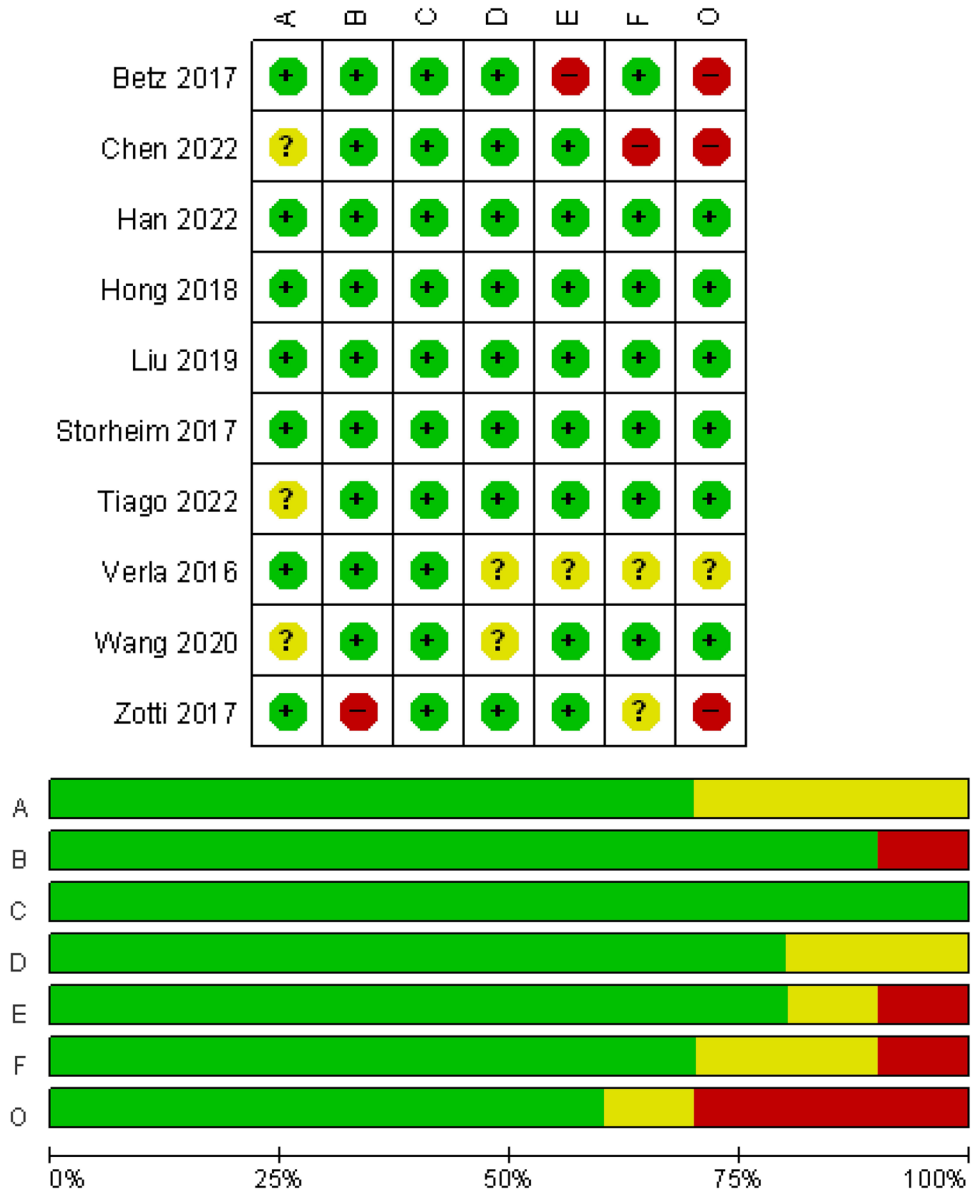


Fig. 2 Risk of bias for the included studies. A = study participation, B = study attrition, C = prognostic factor measurement, D = outcome measurement, E = study confounding, F = statistical analysis and reporting, O = overall risk of bias. The overall risk of bias for an included study was defined as low risk with ≥ 4 low- and no high-risk domains, moderate risk with < 4 low- and no high-risk domains, and high risk with ≥ 1 high-risk domains

with more improvement in ODI (SMD = 0.55, $p = 0.010$). On the contrary, Wang et al [15] found that the lower scores and better improvement in postoperative ODI were not correlated to a higher tCSA of MF at L4 in patients with LSS after PLIF. As for ES, only one single study by Wang et al assessed the prognostic effect of CSA and FI on ODI [15]. Limited evidence revealed that FI of ES could predict the postoperative ODI after lumbar surgery, while atrophy of ES could not (Table 2 and Fig. 4). Moreover, PS thickness could be a possible indicator for change in ODI with limited evidence (Table 2). One study by Verla et al [44] measured the thickness of PS at each segment and revealed that more improvement (> 50%) in ODI was associated with greater psoas muscle thickness at L3-5 ($p = 0.017$) in patients with multiple degenerative lumbar diseases after PLIF.

Postoperative pain

Eight studies evaluated the association between postoperative pain and FI of paraspinal muscles [21, 38–44]. A meta-analysis including five studies with sufficient data in the fixed-effects model was conducted for comparison in low back pain (LBP) between groups with high or low FI in MF [21, 39, 40, 42, 43]. The meta-analysis in the fixed-effects model reported that the patients with higher preoperative FI of MF had greater postoperative LBP compared to those with low FI of MF (SMD = 0.17, 95%CI 0.02–0.31, $p = 0.03$; Fig. 4).

In the vote counting model, whether FI of MF could predict persistent leg pain remained conflicting (Table 3). Although some studies by Hong et al [38], Liu et al [40], and Chen et al [41] showed that MF FI failed to predict the improvement in

Table 1 Study characteristic of the included studies

Authors	Year	Design	Participants, n (male), age	Disease	Intervention	Radiology assessment	Outcome assessment	Metrics	Follow-up
Verla	2016	Retrospective case-series	257 (45.3%), 58.15 ± 14.96 y	Multiple degenerative lumbar diseases	LIF	Thickness of PS, by axial T1-MRI at level L1-5	ODI, VAS	Difference without SD and *p* value in subgroups	1.5 m
Storheim	2017	Secondary analysis in RCT	83 (47.6%), 41.0 ± 7.2 y	Chronic LBP and degenerative disc	total disc replacement	MF FI by axial T2- MRI (T1 if T2 was lacking) at level L3-4, L4-5 and L5-S1 FI grading 0 = < 20% of total cross-section (both sides) contains fat, 1 = 20–50% of cross-section contains fat, 2 = ≥ 50% of cross-section contains fat	ODI, VAS	Mean, SD, and P value	2 y
Zotti	2017	Prospective cohort	66 (37.9%), 67 (29–86) y	LSS	Spinal decompression	tCSA and FI of MF by axial and sagittal T1 and T2-MRI at level L4-5 tCSA: low = < 8.5 cm^2, high = ≥ 8.5 cm^2 Kader FI grade 0 = no fat streaks, 1 = < 10% of fat streaks, 2 = < 50% of fat streaks, 3 = > 50% of fat streaks	ODI	Patient number in subgroups and *p* value	2 y
Betz	2017	Prospective cohort	165 (44.8%), 73.8 ± 8.5 y	LSS	Decompression	PEM FI by axial T2-MRI (1.5 Tesla and 3 Tesla) at level L3 Goutallier/Fuchs FI grade 0 = no fatty streaks, 1 = some fatty streaks, 2 = important fatty streaks, but more muscle than fat, 3 = as much fat as muscle, 4 = more fat than muscle good paravertebral muscle quality: grade ≤ 1 bad paravertebral muscle quality: grade > 1	NRS, revision surgery	Patient number in subgroups, mean, SD, and *p* value	1 y
Hong	2018	Retrospective case-series	664 (63.4%), 41.2 ± 12.3 y	Single-level LDH	Microendoscopic discectomy surgery	MF FI by axial T2-MRI at level L3-4, L4-5 and L5-S1 Three-scale FI grading system: normal/mild = FI < 10%, moderate = 10–50%, severe = > 50%, the most severe atrophy = left–right asymmetry in affected level	ODI, VAS, revision surgery	OR and *p* value	63.8 ± 20.0 (24–96) m
Liu	2019	Retrospective cohort	118 (52.5%), 59.8 ± 8.6 y	L4-5 single-segment DLSS	Decompression and fusion	MF FI by axial T2-MRI (3 Tesla) at the superior endplate level of S1 insignificant FI = < 25% of the tCSA significant FI = ≥ 25% of the tCSA	ODI, VAS	Mean, SD, and *p* value	18 m

Table 1 (continued)

Authors	Year	Design	Participants, n (male), age	Disease	Intervention	Radiology assessment	Outcome assessment	Metrics	Follow-up
Wang	2020	Retrospective case-series	69 (37.7%), 60.6 ± 9.4 y	Symptomatic LSS	PLIF	FI of MF and ES by axial T2-MRI (3 Tesla) at the inferior vertebral endplate of L4 FI (the ratio of tCSA minus fat-free CSA to tCSA)	ODI	Correlation coefficient and *p* value	6 m
Tiago	2022	Retrospective case-series	75 (32.0%), 52.3 y	Degenerative lumbar diseases	TLIF or PLIF	MF FI by axial T2-MRI at the interbody fusion vertebral level and adjacent levels Tamai FI grade: 0 = no intramuscular fat, 1 = some fatty streaks present, 2 = fat evident, but less than muscle tissue, 3 = amounts of fat equal to amount of muscle, 4 = more fat than muscle tissue	ODI, VAS, self-reported leg pain	Number of patients in subgroups mean without SD, and *p* value	2 y
Han	2022	Retrospective cohort	320 (65.0%), group 1: 60.7 ± 7.0 y; group 2: 59.9 ± 6.7 y	LSS	PLIF	MF FI by axial T2-MRI (3 Tesla) at L4 group 1 = FI ≥ 25%, group 2 = FI < 25%	ODI, VAS	Mean, SD, and *p* value	1 y
Chen	2022	Retrospective cohort	58 (48.3%), 51.4 ± 7.3 y	L4-5 single-segment spondylolisthesis	LLIF	MF CAS and FI by axial T1 and T2-MRI (1.5 Tesla) at level L4/5 muscle/vertebra CSA ratio grade: 1 = > 0.80, 2 = 0.60–0.80 and 3 = < 0.60 Kalichman FI grade: 1 = < 10% of tCSA, 2 = 11–50% of tCSA, 3 = > 50% of tCSA	ODI, VAS	Number of subgroup, mean, SD, and *p* value	20.6 ± 4.5 m

RCT, randomized controlled trial; *LBP*, low back pain; *LSS*, lumbar spinal stenosis; *LDH*, lumbar disc herniation; *LIF*, lumbar interbody fusion; *TLIF*, transforaminal lumbar interbody fusion; PLIF, posterior lumbar interbody fusion; *LLIF*, lateral lumbar interbody fusion; *MRI*, magnetic Resonance Imaging; *CT*, computed tomography; *MF*, multifidus; *ES*, erector spinae; *PS*, psoas major; *PEM*, paraspinal extensor muscle; *CSA*, cross-section area; *tCSA*, total cross-section area; *FI*, fatty infiltration; *ODI*, Oswestry disability index; *VAS*, visual analogue scale; *NRS*, numeric rating scale; *OR*, odds ratio; *CI*, confidence interval; *SD*, standard deviation

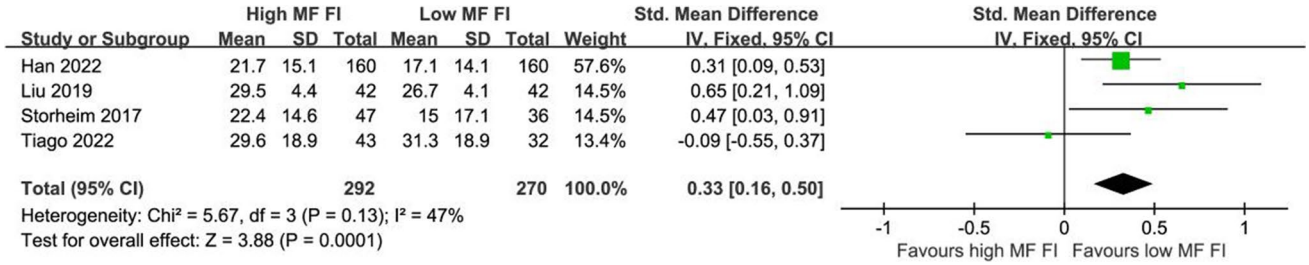


Fig. 3 Forest plot of postoperative ODI between patients with high and low MF FI for lumbar surgery. Patients with high-grade preoperative MF FI had higher postoperative ODI scores, compared to those with low-grade MF FI (SMD = 0.33, 95% CI 0.16–0.50, $p = 0.0001$)

Table 2 Levels of evidence for paraspinal muscle characteristics as predictors of postoperative ODI after lumbar surgery in vote counting model

Study	Support	Outcomes with metrics	Risk of bias	Level of evidence
Atrophy of MF can predict the postoperative ODI				Conflicting
Zotti 2017	Support	Patients with greater (> 8.5 cm^2) MF CSA were more likely to obtain greater improvement (> 40%) in ODI (> 8.5 cm^2: 13/19 vs. < 8.5 cm^2: 15/47, $p = 0.006$) Standardized metrics: OR = 4.622, SMD = 0.85	High	
Chen 2022	Support	Greater MF CSA predicted higher ODI scores ($p < 0.001$) and more improvement (> 40%: $n = 40$, FI mean = 0.57, SD = 0.18 vs. < 40%: $n = 18$, FI mean = 0.48, SD = 0.11, $p = 0.045$) in patients with spondylolisthesis after LLIF Standardized metrics: improvement of ODI: SMD = 0.55 (95%CI − 0.02 to 1.11)	High	
Wang 2020	Not support	Higher MF tCSA failed to predict the lower level ($r = -0.005$, $p > 0.05$) or better improvement ($r = -0.218$, $p > 0.05$) of ODI Standardized metrics: level: SMD = − 0.01; improvement: SMD = − 0.45	Low	
Atrophy of ES cannot predict the postoperative ODI				Limited
Wang 2020	Support	ES tCSA failed to predict the level ($r = -0.010$, $p > 0.05$) or improvement ($r = -0.135$, $p > 0.05$) of ODI Standardized metrics: level: SMD = − 0.02, improvement: SMD = − 0.28	Low	
FI of ES can predict the postoperative ODI				Limited
Wang 2020	Support	Lower ES FI was correlated with a high level ($r = 0.438$, $p < 0.01$) although not correlated with better improvement ($r = -0.142$, $p > 0.05$) of ODI Standardized metrics: level: SMD = 0.97, improvement: SMD = − 0.29	Low	
Atrophy of PS can predict thepostoperative ODI				Limited
Verla 2016	Support	Patients with greater improvement (> 50%) in ODI had greater PS thickness at level L3-4 (difference = 3.541 mm, $p = 0.017$) Standardized metrics: SMD cannot be calculated with limited information	Moderate	

MF, multifidus; *ES*, erector spinae; *PS*, psoas major; *CSA*, cross-section area; *tCSA*, total cross-section area; *FI*, fatty infiltration; *ODI*, Oswestry disability index; *OR*, odds ratio; *SMD*, standardized mean difference (SMD < 0 represented that the metric in patients with greater FI or CSA was lower than those with lower FI or CSA); *SD*, standard deviation; *95%CI*, 95% confidence interval; *r*, correlation coefficient; *p*, *p* value

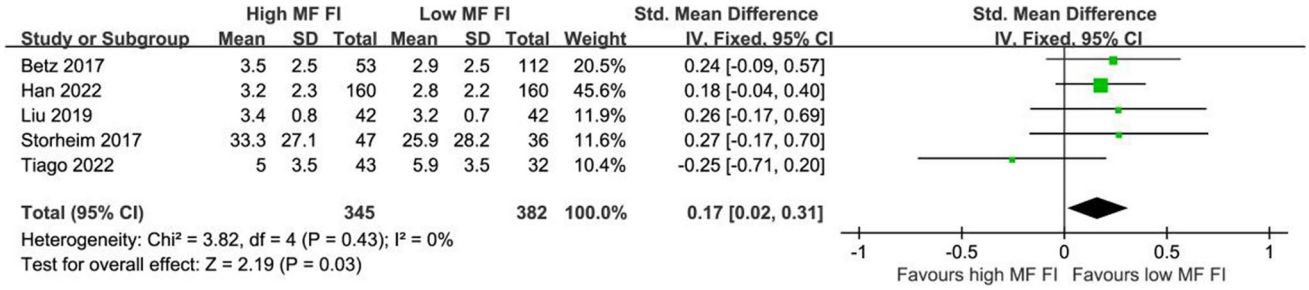


Fig. 4 Forest plot of postoperative back pain between patients with high and low MF FI for lumbar surgery. Patients with higher preoperative FI of MF had greater postoperative LBP compared to those with low FI of MF (SMD = 0.17, 95%CI 0.02–0.31, $p = 0.03$)

VAS for leg pain, Tiago et al [42] uncovered the predictive evidence of lower MF FI for greater relief of self-reported leg pain post surgeries (grades 3–4: 35/43 vs. grades 1–2: 31/32, SMD = − 1.09, $p = 0.04$). Han et al [39] also discovered a higher VAS for leg pain in LSS patients with higher MF FI (≥ 25% tCSA: n = 160, mean = 2.8, SD = 2.4) than lower MF FI (< 25% tCSA: $n = 160$, mean = 2.3, SD = 2.5) after PLIF (SMD = 0.20, $p = 0.039$). Besides, limited evidence showed that the atrophy of PS could predict persistent LBP (Table 3). Verla et al [44] found that more improvement (> 50%) in VAS was associated with a greater PS thickness at L2-3 ($p = 0.032$), L3-4 ($p = 0.043$), and L4-5 ($p = 0.022$) in patients with multiple degenerative lumbar diseases after PLIF.

Re-operation

In the vote counting model, conflicting evidence existed in the prognostic value of MF and ES on the re-operation rate (Table 4). Betz et al [21] quantified the preoperative quality of the PEM group by the degree of fatty degeneration (according

Table 3 Levels of evidence for paraspinal muscle characteristics as predictors of postoperative pain after lumbar surgery in vote counting model

Study	Support	Outcomes with metrics	Risk of bias	Level of evidence
FI of MF can predict the persistent leg pain				Conflicting
Tiago 2022	Support	Regarding the 'worst' level (level with the highest grade of MF FI), higher MF FI was associated with a lower rate of self-reported relief in leg pain (grades 3–4: 35/43 vs. grades 1–2: 31/32, $p = 0.04$) Standardized metrics: OR = 0.14, SMD = − 1.09	Low	
Han 2022	Support	Patients with smaller (< 25%) MF FI had lower VAS ($n = 160$, mean = 2.3, SD = 2.5) for leg than those with MF FI ≥ 25% ($n = 160$, mean = 2.8, SD = 2.4) ($p = 0.039$) Standardized metrics: SMD = 0.20 (95%CI − 0.02 to 0.42)	Low	
Liu 2019	Not support	Patients with insignificant (< 25%: $n = 41$, mean = 3.1, SD = 0.6) and significant (≥ 25%: $n = 41$, mean = 3.2, SD = 0.5) MF FI had similar VAS-leg at 6 months after a propensity score matching including 82 patients ($p = 0.423$) Patients with insignificant (< 25%: $n = 42$, mean = 3.2, SD = 0.7) and significant (≥ 25%: $n = 42$, mean = 3.2, SD = 0.6) MF FI had similar VAS-leg at 18 months after a propensity score matching including 84 patients ($p = 0.950$) Standardized metrics: at 6 months: SMD = 0.18 (95%CI − 0.25 to 0.61), at 18 months: SMD = 0 (95%CI − 0.43 to 0.43)	Low	
Chen 2022	Not support	MF FI failed to predict VAS-leg ($p = 0.213$) and improvement ($p > 0.05$) in patients with spondylolisthesis after LLIF Standardized metrics: SMD cannot be calculated with limited information	High	
Hong 2018	Not support	MF FI failed to predict the rate of > 30% improvement of VAS-leg ($p > 0.05$) Standardized metrics: SMD cannot be calculated with limited information	Low	
Atrophy of PS can predict the persistent LBP				Limited
Verla 2016	Support	Patients with greater improvement (> 50%) in VAS had greater PS thickness (difference at level L2-3 = 4.015 mm, $p = 0.032$; at L3-4 = 4.128 mm, $p = 0.043$; at L4-5 = 5.012, $p = 0.022$) Standardized metrics: SMD cannot be calculated with limited information	Moderate	

MF, multifidus; *PS*, psoas major; *FI*, fatty infiltration; *VAS*, visual analogue scale; *LBP*, low back pain; *OR*, odds ratio; *SMD*, standardized mean difference (SMD < 0 represented that the metric in patients with greater FI or CSA was lower than those with lower FI or CSA); *SD*, standard deviation; *95%CI*, 95% confidence interval; *p*, *p* value

Table 4 Levels of evidence for paraspinal muscle characteristics as predictors of re-operation after lumbar surgery in vote counting model

Study	Support	Outcomes with metrics	Risk of bias	Level of evidence
FI of PEM can predict the incident of revision surgery				Conflicting
Hong 2018	Support	Patients with higher MF FI had higher rates of revision surgeries (OR = 2.095, $p = 0.001$) Metrics: SMD = 0.41	Low	
Betz 2017	Not support	Patients with lower (grades 0–1: 11/112) or higher (grades 2–4: 4/53) PEM FI had a similar rate of revision surgery ($p = 0.55$) Standardized metrics: OR = 0.74, SMD = −0.17	High	

PEM, paraspinal extensor muscle; *FI*, fatty infiltration; *OR*, odds ratio; *SMD*, standardized mean difference (SMD < 0 represented that the metric in patients with greater FI or CSA was lower than those with lower FI or CSA); *p*, *p* value

to the Goutallier classification) at L3. They found that there was no difference in revision surgery between symptomatic LSS patients who underwent revision surgery with good or bad muscle quality (SMD = −0.17, $p = 0.55$). Whereas Hong et al [38] confirmed that MF atrophy (OR = 2.095, SMD = 0.41, $p = 0.01$) at the involved level according to a three-scale grading system on MRI was a statistically significant predictor of a need for re-operation in patients with microendoscopic discectomy for single-level LDH.

In our analysis, all of the pieces of evidence in the vote-counting model were shown in an effect direction plot (Fig. 5).

Discussion

Paraspinal muscle degeneration can be visibly characterized by a muscle atrophy and an increased FI. Multiple studies have considered paraspinal muscle degeneration as

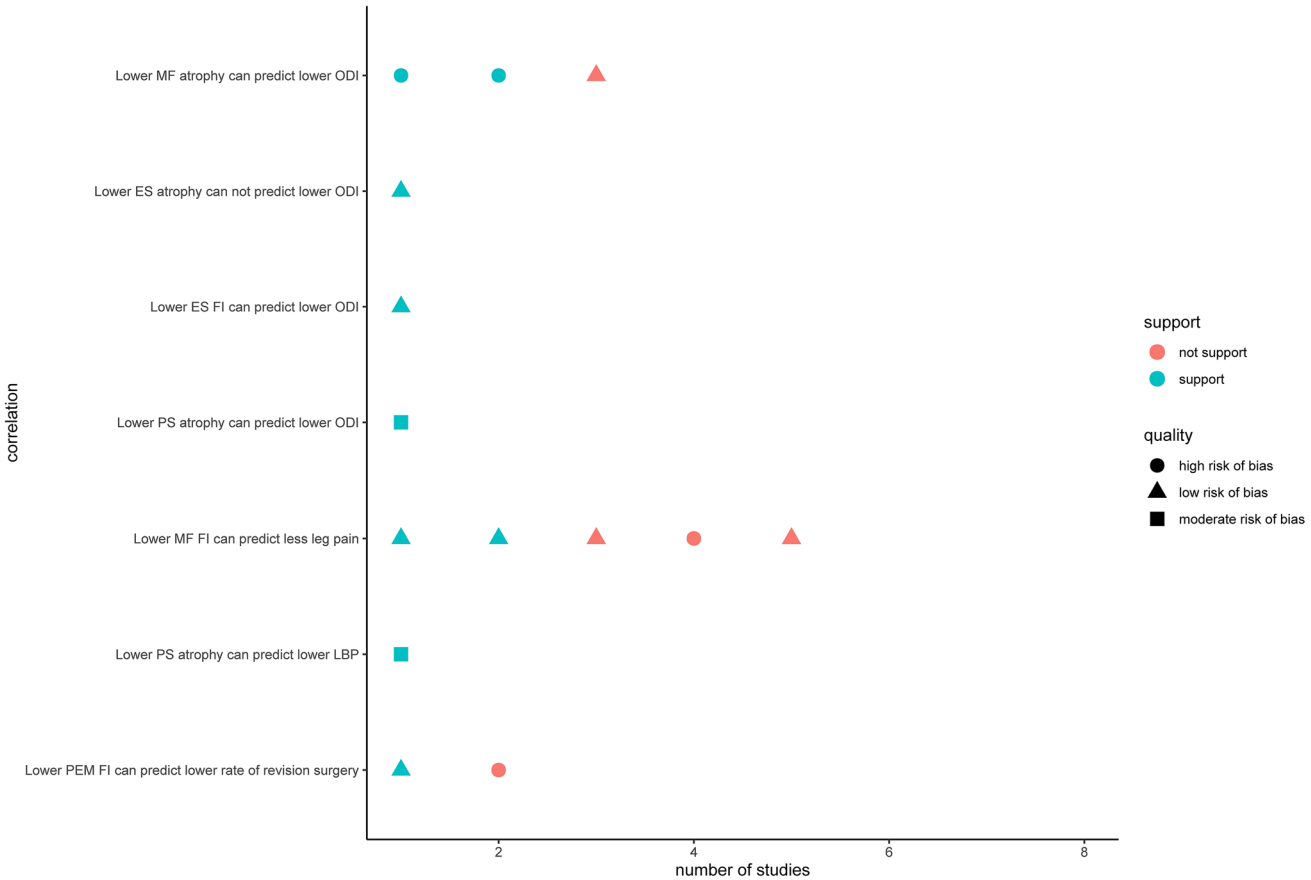


Fig. 5 Effect direction plot for vote counting model. Evidence remained conflicting in terms of whether MF atrophy could predict ODI, whether MF FI could predict leg pain, and whether PEM FI could predict revision surgery. Besides, evidence remained limited in terms of whether ES atrophy, ES FI, or PS atrophy could predict ODI, and whether PS atrophy could predict low back pain

a prognostic factor for surgical disciplines [7–9]. However, whether paraspinal muscle degeneration is related to a poor improvement of functional status after lumbar surgery is still indistinct. This is the first study that investigated the predictive value of three main back muscles on postoperative functional status and re-operation.

This systematic review and meta-analysis included ten studies providing evidence for relationships between lumbar muscle characteristics and postoperative functional status and symptoms. First, the meta-analysis found an association between MF FI and postoperative ODI. Previous studies have also demonstrated the relationship between MF FI and postoperative complications, which was accordant to our findings [12, 13]. The reason why patients with higher MF FI have poorer ODI might be interpreted by the fact that increased muscle FI is correlated to poorer muscle function and low physical activity [45–48]. Besides, only two studies assessed ES and PS. As a result, the evidence for ES and PS as prognostic factors is limited.

In terms of postoperative pain, higher MF FI could predict persistent LBP after lumbar surgery in our meta-analysis. Compared to other paraspinal muscles, MF was more closely related to the vertebral plate and spinous process [49]. Thus, the ability of MF to stabilize and control the movement of the lumbar spine may be a crucial factor in the development of symptoms, since those with severer symptoms have lower activity in MF [50]. In addition, our findings suggested that FI, not atrophy, was a good predictor for postoperative outcomes, which would be helpful for clinical decision-making. The reason might be that the relationship between atrophy and muscle function was not as significant as that of FI. A study demonstrated that FI of paraspinal muscles, rather than CSA, remained a significant predictor of extensor strength in multivariate regression [45]. Moreover, we found limited evidence that showed that the atrophy of PS could predict persistent LBP. As PS is an indicator of sarcopenia and correlates to clinical outcomes in several surgeries [7, 9], it may be a potential factor for predicting LBP.

In our study, two articles examined the relationship between preoperative paraspinal muscle and the re-operation. There was conflicting evidence that MF and ES could predict the re-operation rate in lumbar surgery. The implementation of revision surgery may be due to the failure to achieve osseous fusion, complications resulting from surgical implants, adjacent segment diseases, and persistent pain, part of which were associated with paraspinal muscle degeneration [13, 51]. Consequently, considering the small amount of included studies, the relationship between paraspinal muscles and re-operation needs more high-quality studies.

There are some limitations in this systematic review and meta-analysis. First, the amount of included studies was relatively small, which might produce bias. However, the included studies have been more than previous systematic reviews that investigated factors of clinical outcomes in degenerative lumbar spine diseases [19, 23]. Besides, although the funnel plots and Egger's test indicated no evidence for publication bias in our study, the publication bias assessment based on a small number of included studies might be unreliable. Second, heterogeneity existed in our study, such as diseases and operations. Previous reviews investigating paraspinal muscle degeneration could not perform a meta-analysis due to high heterogeneity [12, 13]. However, in our review, we made subgroup analyses to reduce the heterogeneity ($I^2 < 50\%$) and made it possible for meta-analyses.

In conclusion, FI of MF could be a predictive factor of postoperative ODI and LBP. Whereas, for postoperative leg pain, MF FI could not be a predictor with conflicting evidence. Besides, limited evidence was presented for the prognostic effects of ES and PS on postoperative functional status and symptoms. Additionally, there was conflicting evidence that FI of MF and ES could predict the incidence of revision surgery. The results suggested that the assessment of paraspinal muscle degeneration could be a viable method to stratify patients by risk of postoperative functional status and pain to some extent. A rehabilitation approach focused on training MF before and after surgery might improve clinical outcomes. Considering that limited studies were included and conflicting or limited evidence also existed in this study, more research needs to focus on this field for assisting surgeons to predict the prognosis.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-023-09548-6.

Funding The authors state that this work has not received any funding.

Declarations

Guarantor The scientific guarantor of this publication is Weishi Li.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was not required for this study because the article type is a systematic review and meta-analysis.

Ethical approval Ethical approval was not required for this study because the article type is a systematic review and meta-analysis.

Methodology
- Not applicable
- Not applicable
- Performed at one institution

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Yoshihara H, Yoneoka D (2015) National trends in the surgical treatment for lumbar degenerative disc disease: United States, 2000 to 2009. Spine J 15:265–271. https://doi.org/10.1016/j.spinee.2014.09.026
2. Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG (2010) Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. JAMA 303:1259–1265. https://doi.org/10.1001/jama.2010.338
3. Goz V, Weinreb JH, McCarthy I, Schwab F, Lafage V, Errico TJ (2013) Perioperative complications and mortality after spinal fusions: analysis of trends and risk factors. Spine (Phila Pa 1976) 38:1970–1976. https://doi.org/10.1097/BRS.0b013e3182a62527
4. Beneck GJ, Kulig K (2012) Multifidus atrophy is localized and bilateral in active persons with chronic unilateral low back pain. Arch Phys Med Rehabil 93:300–306. https://doi.org/10.1016/j.apmr.2011.09.017
5. Belavy DL, Armbrecht G, Richardson CA, Felsenberg D, Hides JA (2011) Muscle atrophy and changes in spinal morphology: is the lumbar spine vulnerable after prolonged bed-rest? Spine (Phila Pa 1976) 36:137–145. https://doi.org/10.1097/BRS.0b013e3181cc93e8
6. Kalichman L, Carmeli E, Been E (2017) The association between imaging parameters of the paraspinal muscles, spinal degeneration, and low back pain. Biomed Res Int 2017:2562957. https://doi.org/10.1155/2017/2562957
7. Bourassa-Moreau É, Versteeg A, Moskven E et al (2020) Sarcopenia, but not frailty, predicts early mortality and adverse events after emergent surgery for metastatic disease of the spine. Spine J 20:22–31. https://doi.org/10.1016/j.spinee.2019.08.012
8. Tee YS, Cheng CT, Wu YT et al (2020) The psoas muscle index distribution and influence of outcomes in an Asian adult trauma population: an alternative indicator for sarcopenia of acute diseases. Eur J Trauma Emerg Surg 47:1787–1795. https://doi.org/10.1007/s00068-020-01360-x
9. Byun SE, Kim S, Kim KH, Ha YC (2019) Psoas cross-sectional area as a predictor of mortality and a diagnostic tool for sarcopenia in hip fracture patients. J Bone Miner Metab 37:871–879. https://doi.org/10.1007/s00774-019-00986-1
10. Sigmundsson FG, Jönsson B, Strömqvist B (2013) Impact of pain on function and health related quality of life in lumbar spinal stenosis. A register study of 14,821 patients. Spine (Phila Pa 1976) 38:E937-945. https://doi.org/10.1097/BRS.0b013e3182955989
11. Knutsson B, Michaëlsson K, Sandén B (2013) Obesity is associated with inferior results after surgery for lumbar spinal stenosis: a study of 2633 patients from the Swedish spine register. Spine (Phila Pa 1976) 38:435–441. https://doi.org/10.1097/BRS.0b013e318270b243
12. Khan AB, Weiss EH, Khan AW, Omeis I, Verla T (2017) Back muscle morphometry: effects on outcomes of spine surgery. World Neurosurg 103:174–179. https://doi.org/10.1016/j.wneu.2017.03.097
13. Gengyu H, Jinyue D, Chunjie G et al (2021) The predictive value of preoperative paraspinal muscle morphometry on complications after lumbar surgery: a systematic review. Eur Spine J 31:364–379. https://doi.org/10.1007/s00586-021-07052-3
14. Zotti MGT, Boas FV, Clifton T, Piche M, Yoon WW, Freeman BJC (2017) Does pre-operative magnetic resonance imaging of the lumbar multifidus muscle predict clinical outcomes following lumbar spinal decompression for symptomatic spinal stenosis? Eur Spine J 26:2589–2597. https://doi.org/10.1007/s00586-017-4986-x
15. Wang W, Sun Z, Li W, Chen Z (2020) The effect of paraspinal muscle on functional status and recovery in patients with lumbar spinal stenosis. J Orthop Surg Res 15:235. https://doi.org/10.1186/s13018-020-01751-1
16. He K, Head J, Mouchtouris N et al (2020) The implications of paraspinal muscle atrophy in low back pain, thoracolumbar pathology, and clinical outcomes after spine surgery: a review of the literature. Global Spine J 10:657–666. https://doi.org/10.1177/2192568219879087
17. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C (2013) Assessing bias in studies of prognostic factors. Ann Intern Med 158:280–286. https://doi.org/10.7326/0003-4819-158-4-201302190-00009
18. Jiang C, Wang Y, Fu W et al (2022) Association between sarcopenia and prognosis of hepatocellular carcinoma: a systematic review and meta-analysis. Front Nutr 9:978110. https://doi.org/10.3389/fnut.2022.978110
19. Jermy JE, Copley PC, Poon MTC, Demetriades AK (2020) Does pre-operative multifidus morphology on MRI predict clinical outcomes in adults following surgical treatment for degenerative lumbar spine disease? A systematic review. Eur Spine J 29:1318–1327. https://doi.org/10.1007/s00586-020-06423-6
20. Ballatori AM, Shahrestani S, Nyayapati P et al (2022) Influence of patient-specific factors when comparing multifidus fat infiltration between chronic low back pain patients and asymptomatic controls. JOR Spine 5:e1217. https://doi.org/10.1002/jsp2.1217
21. Betz M, Burgstaller JM, Held U et al (2017) Influence of paravertebral muscle quality on treatment efficacy of epidural steroid infiltration or surgical decompression in Lumbar Spinal Stenosis-Analysis of the Lumbar Spinal Outcome Study (LSOS) data: a Swiss prospective multicenter cohort study. Spine (Phila Pa 1976) 42:1792–1798. https://doi.org/10.1097/brs.0000000000002233
22. Pishnamaz M, Schemmann U, Herren C et al (2018) Muscular changes after minimally invasive versus open spinal stabilization of thoracolumbar fractures: a literature review. J Musculoskelet Neuronal Interact 18:62–70
23. Jones JJ, Oduwole S, Feinn R, Yue JJ (2021) Postoperative bracing on pain, disability, complications, and fusion rate following 1–3+ level lumbar fusion in degenerative conditions: a meta-analysis. Clin Spine Surg 34:56–62. https://doi.org/10.1097/bsd.0000000000001060
24. Higgins JP, Li T, Deeks JJ (2019) Choosing effect measures and computing estimates of effect. In: Cochrane handbook for systematic reviews of interventions. pp 143–176. https://doi.org/10.1002/9781119536604.ch6
25. Polanin JR, Snilstveit B (2016) Converting between effect sizes. 12:1-13. https://doi.org/10.4073/cmpn.2016.3
26. Schober P, Mascha EJ, Vetter TR (2021) Statistics From A (Agreement) to Z (z Score): A guide to interpreting common measures of association, agreement, diagnostic accuracy, effect size, heterogeneity, and reliability in medical research. Anesth Analg 133:1633–1641. https://doi.org/10.1213/ane.0000000000005773

27. Shu L, Yu D, Jin F (2022) Healthy dietary patterns, foods, and risk of glioma: a systematic review and meta-analysis of observational studies. Front Nutr 9:1077452. https://doi.org/10.3389/fnut.2022.1077452
28. Tweed EJ, Thomson RM, Lewer D et al (2021) Health of people experiencing co-occurring homelessness, imprisonment, substance use, sex work and/or severe mental illness in high-income countries: a systematic review and meta-analysis. J Epidemiol Community Health 75:1010–1018. https://doi.org/10.1136/jech-2020-215975
29. Patro S, Meto A, Mohanty A et al (2022) Diagnostic accuracy of pulp vitality tests and pulp sensibility tests for assessing pulpal health in permanent teeth: a systematic review and meta-analysis. Int J Environ Res Public Health 1910.3390/ijerph19159599
30. Boyland E, McGale L, Maden M et al (2022) Association of food and nonalcoholic beverage marketing with children and adolescents' eating behaviors and health: a systematic review and meta-analysis. JAMA Pediatr 176:e221037–e221037. https://doi.org/10.1001/jamapediatrics.2022.1037%JJAMA Pediatrics
31. Kunzler AM, Lindner S, Röthke N et al (2023) Mental health impact of early stages of the COVID-19 pandemic on individuals with pre-existing mental disorders: a systematic review of longitudinal research. Int J Environ Res Public Health 2010.3390/ijerph20020948
32. O'Keefe H, Shenfine R, Brown M, Beyer F, Rankin J (2023) Are non-invasive or minimally invasive autopsy techniques for detecting cause of death in prenates, neonates and infants accurate? A systematic review of diagnostic test accuracy. BMJ Open 13:e064774. https://doi.org/10.1136/bmjopen-2022-064774
33. McKenzie JE, Brennan SE (2019) Synthesizing and presenting findings using other methods. In: Cochrane handbook for systematic reviews of interventions. pp 321–347. https://doi.org/10.1002/9781119536604.ch12
34. Campbell M, McKenzie JE, Sowden A et al (2020) Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. 368: l6890. https://doi.org/10.1136/bmj.l6890%JBMJ
35. Domingues L, Cruz EB, Pimentel-Santos FM et al (2018) Prognostic factors for recovery and non-recovery in patients with non-specific neck pain: a protocol for a systematic literature review. BMJ Open 8:e023356. https://doi.org/10.1136/bmjopen-2018-023356
36. Rustenburg C, Faraj S, Ket J, Emanuel K, Smit T (2019) Prognostic factors in the progression of intervertebral disc degeneration: which patient should be targeted with regenerative therapies? JOR Spine 2(3):e1063. https://doi.org/10.1002/jsp2.1063
37. Varkevisser RDM, van Stralen MM, Kroeze W, Ket JCF, Steenhuis IHM (2019) Determinants of weight loss maintenance: a systematic review. Obes Rev 20:171–211. https://doi.org/10.1111/obr.12772
38. Hong X, Shi R, Wang Y-T, Liu L, Bao J-P, Wu X-T (2018) Lumbar disc herniation treated by microendoscopic discectomy: prognostic predictors of long-term postoperative outcome. Orthopade 47:993–1001. https://doi.org/10.1007/s00132-018-3624-6
39. Han G, Zou D, Li X et al (2022) Can fat infiltration in the multifidus muscle be a predictor of postoperative symptoms and complications in patients undergoing lumbar fusion for degenerative lumbar spinal stenosis? A case-control study. J Orthop Surg Res 17:289. https://doi.org/10.1186/s13018-022-03186-2
40. Yang L, Yuzeng L, Yong H, Tie L, Li G, Chen X (2019) Fat infiltration in the multifidus muscle as a predictor of prognosis after decompression and fusion in patients with single-segment degenerative lumbar spinal stenosis: an ambispective cohort study based on propensity score matching. World Neurosurgery 128:E989–E1001. https://doi.org/10.1016/j.wneu.2019.05.055
41. Chen J, Li J, Sheng B, Li L, Wu S (2022) Does preoperative morphology of multifidus influence the surgical outcomes of stand-alone lateral lumbar interbody fusion for lumbar spondylolisthesis? Clin Neurol Neurosurg 215:107177. https://doi.org/10.1016/j.clineuro.2022.107177
42. Amorim-Barbosa T, Catelas D, Pereira C et al (2022) Is preoperative fat infiltration in lumbar spine muscles associated with worse clinical outcomes after lumbar interbody fusion? Eur J Orthop Surg Traumatol. https://doi.org/10.1007/s00590-022-03311-110.1007/s00590-022-03311-1
43. Storheim K, Berg L, Hellum C et al (2017) Fat in the lumbar multifidus muscles - predictive value and change following disc prosthesis surgery and multidisciplinary rehabilitation in patients with chronic low back pain and degenerative disc: 2-year follow-up of a randomized trial. BMC Musculoskelet Disord 18:145. https://doi.org/10.1186/s12891-017-1505-5
44. Verla T, Adogwa O, Elsamadicy A et al (2016) Effects of psoas muscle thickness on outcomes of lumbar fusion surgery. World Neurosurg 87:283–289. https://doi.org/10.1016/j.wneu.2015.11.022
45. Schlaeger S, Inhuber S, Rohrmeier A et al (2019) Association of paraspinal muscle water-fat MRI-based measurements with isometric strength measurements. Eur Radiol 29:599–608. https://doi.org/10.1007/s00330-018-5631-8
46. Goodpaster BH, Park SW, Harris TB et al (2006) The loss of skeletal muscle strength, mass, and quality in older adults: the health, aging and body composition study. J Gerontol A Biol Sci Med Sci 61:1059–1064. https://doi.org/10.1093/gerona/61.10.1059
47. Han G, Wang W, Yue L et al (2022) Age-dependent differences of paraspinal muscle endurance and morphology in chinese community population without chronic low back pain. Global Spine J. https://doi.org/10.1177/21925682221103507
48. Goodpaster BH, Chomentowski P, Ward BK et al (1985) (2008) Effects of physical activity on strength and skeletal muscle fat infiltration in older adults: a randomized controlled trial. J Appl Physiol 105:1498–1503. https://doi.org/10.1152/japplphysiol.90425.2008
49. Wang Z, Zhao Z, Han S et al (2022) Advances in research on fat infiltration and lumbar intervertebral disc degeneration. Front Endocrinol (Lausanne) 13:1067373. https://doi.org/10.3389/fendo.2022.1067373
50. Danneels LA, Vanderstraeten GG, Cambier DC, Witvrouw EE, De Cuyper HJ (2000) CT imaging of trunk muscles in chronic low back pain patients and healthy control subjects. Eur Spine J 9:266–272. https://doi.org/10.1007/s005860000190
51. Deyo RA, Martin BI, Kreuter W, Jarvik JG, Angier H, Mirza SK (2011) Revision surgery following operations for lumbar stenosis. J Bone Joint Surg Am 93:1979–1986. https://doi.org/10.2106/jbjs.J.01292

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.