

# The Statistical Fragility of Marrow Stimulation for Cartilage Defects of the Knee: A Systematic Review of Randomized Controlled Trials

CARTILAGE  
2024, Vol. 15(4) 389–398  
© The Author(s) 2024  
DOI: 10.1177/19476035241233441  
journals.sagepub.com/home/CAR



Avanish Yendluri<sup>1</sup> , Ara Alexanian<sup>2</sup>, Rohit R. Chari<sup>3</sup>,  
John J. Corvi<sup>1</sup>, Nikan K. Namiri<sup>1</sup>, Junho Song<sup>1</sup>, Michael J. Alaia<sup>4</sup>,  
Xinning Li<sup>5</sup>, and Robert L. Parisien<sup>1</sup>

## Abstract

**Objective.** Marrow stimulation is used to address knee cartilage defects. In this study, we used the fragility index (FI), reverse fragility index (rFI), and fragility quotient (FQ) to evaluate statistical fragility of outcomes reported in randomized controlled trials (RCTs) evaluating marrow stimulation. **Design.** PubMed, Embase, and MEDLINE were queried for recent RCTs (January 1, 2010–September 5, 2023) assessing marrow stimulation for cartilage defects of the knee. The FI and rFI were calculated as the number of outcome event reversals required to alter statistical significance for significant and nonsignificant outcomes, respectively. The FQ was determined by dividing the FI by the study sample size. **Results.** Across 155 total outcomes from 21 RCTs, the median FI was 3 (interquartile range [IQR], 2–5), with an associated median FQ of 0.067 (IQR, 0.033–0.010). Thirty-two outcomes were statistically significant, with a median FI of 2 (IQR, 1–3.25) and FQ of 0.050 (IQR, 0.025–0.069). Ten of the 32 (31.3%) outcomes reported as statistically significant had an FI of 1. In total, 123 outcomes were nonsignificant, with a median rFI of 3 (IQR, 2–5). Studies assessing stem cell augments were the most fragile, with a median FI of 2. In 55.5% of outcomes, the number of patients lost to follow-up was greater than or equal to the FI. **Conclusion.** Statistical findings in RCTs evaluating marrow stimulation for cartilage defects of the knee are statistically fragile. We recommend combined reporting of *P*-values with FI and FQ metrics to aid in the interpretation of clinical findings in comparative trials assessing cartilage restoration.

## Keywords

marrow stimulation, statistical fragility, knee cartilage, systematic review, randomized controlled trial

## Introduction

Articular cartilage has a very limited capacity for self repair.<sup>1,2</sup> As such, marrow stimulation techniques were developed to promote functional restoration of cartilage defects within joints instead of solely alleviating symptoms. Microfracture, first described by Steadman in the early 1990s, consists of debriding the defect and perforating small holes into the subchondral bone to induce invasion of progenitor cells and encourage tissue repair.<sup>1–3</sup> Although temporary improvements in joint function have been reported, microfracture results in fibrocartilaginous repair tissue, degradation of subchondral bone, and functional loss long term.<sup>1,4</sup> Subchondral drilling is an even earlier technique first described by Smilie in 1957, which involves

<sup>1</sup>Department of Orthopedic Surgery, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>2</sup>Harbor-UCLA Medical Center, Torrance, CA, USA

<sup>3</sup>University of Maryland School of Medicine, Baltimore, MD, USA

<sup>4</sup>Department of Orthopaedic Surgery, New York University Langone Health, New York, NY, USA

<sup>5</sup>Department of Orthopaedic Surgery, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, USA

Investigation Performed by the Scientific Collaborative for Orthopaedic Research and Education (SCORE) Group

### Corresponding Author:

Avanish Yendluri, Department of Orthopedic Surgery, Icahn School of Medicine at Mount Sinai, 5 East 98th Street, 9th Floor, New York, NY 10029, USA.

Email: [avanish.yendluri@icahn.mssm.edu](mailto:avanish.yendluri@icahn.mssm.edu)



drilling holes (usually larger) in the subchondral bone plate which leads to blood clot formation and fibrocartilage repair tissue.<sup>5</sup> Similar to microfracture, this fibrocartilaginous tissue is structurally and biomechanically inferior to hyaline articular cartilage, leading to decreasing clinical results in as soon as 18 months.<sup>5,6</sup>

To direct repair toward hyaline-like cartilage, cell-based techniques have been developed, including autologous chondrocyte implantation (ACI), matrix-applied chondrocyte implantation (MACI), and osteochondral autologous transplantation (OATS). ACI/MACI involve harvesting chondrocytes and expanding them *ex vivo* and subsequently implanting them into the damaged articular defect as a patch.<sup>7</sup> The current generation, MACI, involves seeding of chondrocyte cells onto a type I/III collagen matrix, which can be cut to fit the size of the defect and implanted by arthroscopy or mini-arthrotomy.<sup>8</sup> In previous systematic reviews of randomized controlled trials (RCTs), ACI and MACI have demonstrated superior clinical improvement compared with marrow stimulation techniques.<sup>9,10</sup> Prior literature also suggests that OATS may lead to significantly higher return to activity, patient-reported outcome measures (PROMs), and lower failure rates compared with marrow stimulation.<sup>11,12</sup> Many recent RCTs have also assessed the use of augments, including stem cells, collagen membranes, extracellular matrix scaffolds, and so on, for marrow stimulation. However, the clinical significance of marrow stimulation augments remains unclear.<sup>10,13</sup>

RCTs represent the highest level of evidence in guiding management of cartilage defects with *P*-values reported widely in the orthopedic literature to indicate statistical significance.<sup>14</sup> Although the *P*-value is essential, it has received criticism for neglecting study design elements and patients lost to follow-up.<sup>15</sup> To supplement the *P*-value, Feinstein introduced the concept of the fragility index (FI) to address the *P*-value's limitations.<sup>16</sup> The FI represents how "fragile" a statistical outcome is and is calculated as the number of iterative outcome event reversals needed to lose statistical significance.<sup>16</sup> This metric has been used widely to assess statistical fragility of RCT findings in the orthopedic literature.<sup>16-24</sup> The reverse fragility index (rFI) was similarly defined to represent the number of outcome event reversals required to turn nonsignificant outcomes into statistically significant findings.<sup>25-27</sup> To take sample size into consideration, the fragility quotient (FQ) is calculated as the FI divided by the sample size and represents the proportion of patients that need an outcome event reversal for significance to be altered.<sup>28,29</sup> The purpose of this study was to evaluate the statistical fragility of RCTs assessing the efficacy of marrow stimulation techniques for cartilage defects of the knee using the FI, rFI, and FQ metrics. Specifically, we evaluated the fragility of RCTs that compared both marrow stimulation versus other cartilage restoration techniques and RCTs evaluating augments for marrow stimulation. We hypothesized

that study findings would be statistically fragile, especially outcomes initially reported as statistically significant.

## Methods

### Literature Review

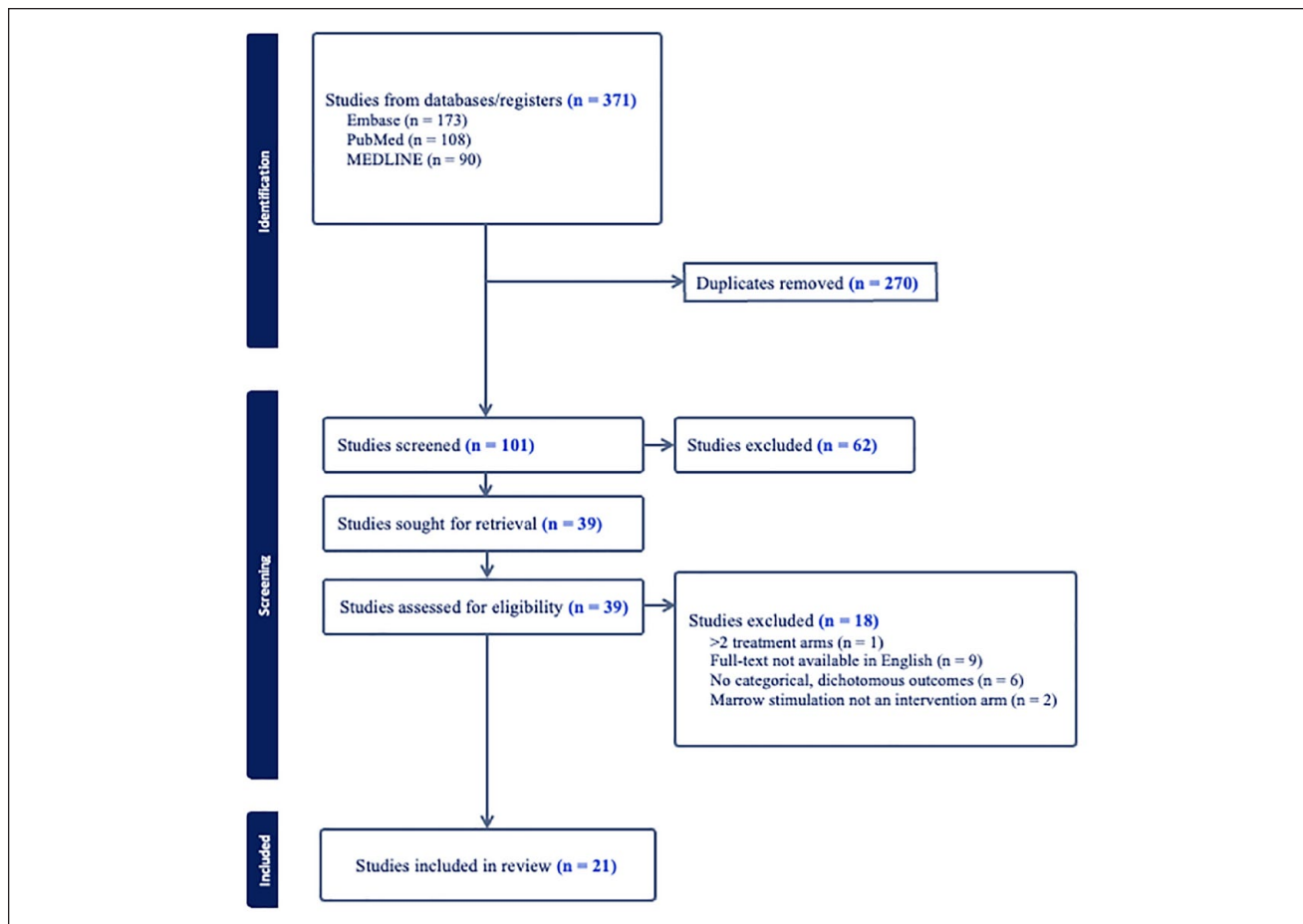
This systematic review was in accordance with the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).<sup>30</sup> The PubMed, Embase, and Medline databases were queried to identify RCTs published from January 1, 2010, to September 5, 2023, related to marrow stimulation for cartilage defects of the knee (**Fig. 1**). The search keywords used across all databases were ((stimulation) OR (microfracture) OR (mfx) OR (drilling)) AND (chondral OR cartilage) AND "knee." Studies met the inclusion criteria if they were RCTs reporting dichotomous, categorical outcomes with an intervention arm related to marrow stimulation techniques (e.g., microfracture, subchondral drilling). Non-English language, cadaveric/biomechanical/animal, *in vitro*, and non-RCT studies were excluded. The same RCT population at multiple follow-up time periods was included if the reported outcomes were distinct. Title/abstract screening and full-text review was performed by 2 independent reviewers and all conflicts were resolved by a third independent reviewer. A risk-of-bias assessment was performed for all included studies. The Cochrane risk of bias tool for assessing bias of randomized trials was used for quality assessment.<sup>31</sup>

### Data Extraction

The first author, year of publication, journal of publication, and treatment intervention of the 2 arms were extracted from each included study. Reporting of clinically meaningful difference metrics was also assessed for each included article (i.e., the minimum clinically important difference [MCID]).<sup>32</sup> Outcome events in each intervention arm, any reported *P*-values, and patients lost to follow-up were recorded for each study outcome. All RCT outcomes were reviewed and outcome categories were established by 2 reviewers for subgroup analysis. Outcome categories included complications/adverse events, volume of cartilage defect filling, failure/reoperation rates, clinical improvement in PROMs, subchondral bone architecture, integration of cartilage repair with adjacent native cartilage, and quality and homogeneity of repair tissue surface and structure.

### Fragility Analysis

A 2-tailed Fisher exact test was used to confirm reported *P*-values for each outcome. Outcomes with *P*-values < 0.05 were considered statistically significant. The FI was calculated by manipulating outcome events until the *P*-value



**Figure 1.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram showing identification, screening, and inclusion of eligible articles from PubMed, Embase, and Medline.

	Defect visible (+)	Defect visible (-)		Defect visible (+)	Defect visible (-)
OAT	2	26		3	25
Microfracture	9	20		9	20
		0.041			0.103

**Figure 2.** Demonstration of statistical significance reversal using a 2 × 2 contingency table with a resulting fragility index = 1. OAT = osteochondral autologous transplantation.

was reversed from  $<0.05$  to  $\geq 0.05$  as demonstrated in **Figure 2**. The rFI was calculated similarly for the  $P$ -value to switch from  $\geq 0.05$  to  $<0.05$ . The FQ was calculated by dividing the FI or rFI for each outcome by the study sample size to represent the proportion of patients that require an

outcome event reversal for significance to be altered for a given outcome. We further performed subgroup analysis based on statistical significance, intervention comparison type, and outcome type. Findings are presented as median FI (interquartile range [IQR]).

## Results

Of 371 RCTs screened, 21 studies were included for analysis. Seven RCTs were from *American Journal of Sports Medicine*, 4 from *Knee Surgery, Sports Traumatology, Arthroscopy*, 3 from *Journal of Bone and Joint Surgery*, 2 from *Orthopaedic Journal of Sports Medicine*, 2 from *Arthroscopy: The Journal of Arthroscopic and Related Surgery*, 2 from *Cartilage*, and 1 from *Regenerative Therapy*. The included RCTs assessed both marrow stimulation versus other cartilage restoration techniques and augmentations for marrow stimulation as indicated in **Table 1**.

There were 155 total outcomes reported across the included RCTs related to marrow stimulation techniques for cartilage defects of the knee. The median FI across all outcomes was 3 (IQR, 2-5), indicating 3 outcome event reversals from the included RCTs alters overall statistical significance. The median FQ across all outcomes was 0.067 (IQR, 0.033-0.010). Thus, an outcome event reversal in 6.7 out of 100 patients alters outcome significance (**Table 2**). In 86 out of 155 (55.5%) of outcomes, the number of patients lost to follow-up was greater than or equal to the FI.

Thirty-two outcomes were statistically significant with a median FI of 2 (IQR 1-3.25), indicating that statistically significant outcomes rely on just 2 outcome events. The associated FQ for significant outcomes was 0.050 (IQR, 0.025-0.069). Thus, outcome event reversals in just 5% of patients reverse statistically significant outcomes. In 10 of the 32 (31.3%) outcomes reported as statistically significant,

the FI was found to be 1. In total, 123 outcomes were reported as statistically nonsignificant; these outcomes were found to have a median rFI of 3 (IQR, 2-5) and FQ of 0.067 (IQR, 0.034-0.010).

Six of the included studies assessed the efficacy of augmentations to marrow stimulation and had a median FI of 3 (IQR, 2-5) across 52 outcomes. Stem cells were the most fragile augment, with a median FI of 2 (IQR, 1-4) across 21 outcomes. Autologous matrix-induced chondrogenesis (AMIC) versus microfracture alone demonstrated a median FI of 3.5 (IQR, 2.75-4) across 8 outcomes.

MACI versus microfracture was the least fragile intervention comparison assessed. Across 5 studies comprising 39 outcomes, we identified a median FI of 4 (IQR, 2-4.5). Four studies assessed ACI versus microfracture and had a median FI of 3.5 (IQR, 2-6) across 16 outcomes. Three studies assessed OATS versus microfracture and had a median FI of 3 (IQR, 2-4) across 20 outcomes.

The 22 outcomes related to cartilage defect filling and 11 outcomes related to the repair tissue structure were the most fragile outcome categories, with a median FI of 2 (IQR, 2-4) and 2 (IQR, 1.5-2.5), respectively (**Table 3**). Complications/adverse events were the most commonly reported outcome category comprising 54 outcomes with a median FI of 4 (IQR, 2-5). Outcomes related to failure/reoperation, clinical improvement, and integration with adjacent cartilage similarly demonstrated fragility, each with a median FI of 3; outcomes relating to the subchondral architecture had a median FI of 4.

**Table 1.** Characteristics of Included Studies Including First Author, Journal/Year of Publication, Title, and Interventions Assessed.

Author	Year	Study Title	Journal	Interventions Assessed
Brittberg	2018	<i>Matrix-Applied Characterized Autologous Cultured Chondrocytes Versus Microfracture: Five-Year Follow-up of a Prospective Randomized Trial</i>	<i>American Journal of Sports Medicine</i>	MACI vs. microfracture
Chung	2014	<i>Cartilage extra-cellular matrix biomembrane for the enhancement of microfractured defects</i>	<i>Knee Surgery, Sports Traumatology, Arthroscopy</i>	Extracellular matrix biomembrane augment for marrow stimulation (AMIC)
Chung	2023	<i>Particulated Costal Allograft Cartilage With Microfracture Versus Microfracture Alone for Knee Cartilage Defects: A Multicenter, Prospective, Randomized, Participant- and Rater-Blinded Study</i>	<i>Orthopaedic Journal of Sports Medicine</i>	Costal allograft cartilage scaffold paste augment for marrow stimulation
Cole	2011	<i>Outcomes after a single-stage procedure for cell-based cartilage repair: a prospective clinical safety trial with 2-year follow-up</i>	<i>American Journal of Sports Medicine</i>	Cartilage autograft implantation system vs. microfracture
Crawford	2012	<i>NeoCart, an autologous cartilage tissue implant, compared with microfracture for treatment of distal femoral cartilage lesions: an FDA phase-II prospective, randomized clinical trial after two years</i>	<i>Journal of Bone and Joint Surgery</i>	Autologous cartilage tissue implant vs. microfracture
Gudas	2012	<i>Ten-year follow-up of a prospective, randomized clinical study of mosaic osteochondral autologous transplantation versus microfracture for the treatment of osteochondral defects in the knee joint of athletes</i>	<i>American Journal of Sports Medicine</i>	OATS vs. microfracture

(continued)

**Table 1.** (continued)

Author	Year	Study Title	Journal	Interventions Assessed
Hashimoto	2019	<i>Transplantation of autologous bone marrow-derived mesenchymal stem cells under arthroscopic surgery with microfracture versus microfracture alone for articular cartilage lesions in the knee: A multicenter prospective randomized control clinical trial</i>	<i>Regenerative Therapy</i>	Bone marrow–derived mesenchymal stem cells augment for marrow stimulation
Hoburg	2021	<i>Matrix-Associated Autologous Chondrocyte Implantation with Spheroid Technology Is Superior to Arthroscopic Microfracture at 36 Months Regarding Activities of Daily Living and Sporting Activities after Treatment</i>	<i>Cartilage</i>	MACI vs. microfracture
Hoburg	2023	<i>Sustained superiority in KOOS subscores after matrix-associated chondrocyte implantation using spheroids compared to microfracture</i>	<i>Knee Surgery, Sports Traumatology, Arthroscopy</i>	MACI vs. microfracture
Ibarra	2021	<i>Arthroscopic Matrix-Assisted Autologous Chondrocyte Transplantation Versus Microfracture: A 6-Year Follow-up of a Prospective Randomized Trial</i>	<i>American Journal of Sports Medicine</i>	MACI vs. microfracture
Kim	2020	<i>Microfractures Versus a Porcine-Derived Collagen-Augmented Chondrogenesis Technique for Treating Knee Cartilage Defects: A Multicenter Randomized Controlled Trial</i>	<i>Arthroscopy: The Journal of Arthroscopic and Related Surgery</i>	Collagen chondrogenesis scaffold gel augment for marrow stimulation
Knutsen	2016	<i>A Randomized Multicenter Trial Comparing Autologous Chondrocyte Implantation with Microfracture: Long-Term Follow-up at 14 to 15 Years</i>	<i>Journal of Bone and Joint Surgery</i>	ACI vs. microfracture
Koh	2016	<i>Adipose-Derived Mesenchymal Stem Cells With Microfracture Versus Microfracture Alone: 2-Year Follow-up of a Prospective Randomized Trial</i>	<i>Arthroscopy: The Journal of Arthroscopic and Related Surgery</i>	Adipose-derived stem cells with fibrin glue augment for marrow stimulation
Lim	2021	<i>Allogeneic Umbilical Cord Blood-Derived Mesenchymal Stem Cell Implantation Versus Microfracture for Large, Full-Thickness Cartilage Defects in Older Patients: A Multicenter Randomized Clinical Trial and Extended 5-Year Clinical Follow-up</i>	<i>Orthopaedic Journal of Sports Medicine</i>	Allogeneic umbilical cord blood–derived mesenchymal stem cells and 4% hyaluronate vs. microfracture
Saris	2014	<i>Matrix-Applied Characterized Autologous Cultured Chondrocytes Versus Microfracture: Two-Year Follow-up of a Prospective Randomized Trial</i>	<i>American Journal of Sports Medicine</i>	MACI vs. microfracture
Solheim	2018	<i>Randomized Study of Long-term (15-17 Years) Outcome After Microfracture Versus Mosaicplasty in Knee Articular Cartilage Defects</i>	<i>American Journal of Sports Medicine</i>	OATS vs. microfracture
Stanish	2013	<i>Novel scaffold-based BST-CarGel treatment results in superior cartilage repair compared with microfracture in a randomized controlled trial</i>	<i>Journal of Bone and Joint Surgery</i>	BST-CarGel scaffold augment for marrow stimulation
Ulstein	2014	<i>Microfracture technique versus osteochondral autologous transplantation mosaicplasty in patients with articular chondral lesions of the knee: a prospective randomized trial with long-term follow-up</i>	<i>Knee Surgery, Sports Traumatology, Arthroscopy</i>	OATS vs. microfracture
VanAssche	2010	<i>Autologous chondrocyte implantation versus microfracture for knee cartilage injury: a prospective randomized trial, with 2-year follow-up</i>	<i>Knee Surgery, Sports Traumatology, Arthroscopy</i>	ACI vs. microfracture
Vanlauwe	2011	<i>Five-year outcome of characterized chondrocyte implantation versus microfracture for symptomatic cartilage defects of the knee: early treatment matters</i>	<i>American Journal of Sports Medicine</i>	ACI vs. microfracture
Yoon	2021	<i>Costal Chondrocyte-Derived Pellet-Type Autologous Chondrocyte Implantation versus Microfracture for Repair of Articular Cartilage Defects: A Prospective Randomized Trial</i>	<i>Cartilage</i>	ACI vs. microfracture

MACI = matrix-associated autologous chondrocyte implantation; AMIC = autologous matrix-induced chondrogenesis; OATS = osteochondral autologous transplantation; ACI = autologous chondrocyte implantation.



**Table 2.** Fragility Data Based on Outcome Significance.

	Number of Outcomes	FI, Median (IQR)	FQ, Median (IQR)
All RCT outcomes	155	3 (2-5)	0.067 (0.033-0.10)
Significant outcomes ( $P < 0.05$ )	32	2 (1-3.25)	0.050 (0.025-0.069)
Nonsignificant outcomes ( $P \geq 0.05$ )	123	3 (2-5)	0.067 (0.034-0.010)

FI = fragility index; IQR = interquartile range; FQ = fragility quotient; RCT = randomized controlled trial.

**Table 3.** Subgroup Analysis Based on Outcome Type.

	Number of Outcomes	FI, Median (IQR)	FQ, Median (IQR)
Complications/adverse events	54	4 (2-5)	0.058 (0.029-0.10)
Cartilage defect filling	22	2 (2-4)	0.059 (0.029-0.071)
Failure/reoperation	20	3 (3-4)	0.063 (0.029-0.09)
Clinical improvement	18	3 (2-5.5)	0.056 (0.035-0.067)
Subchondral architecture	17	4 (2-5)	0.068 (0.051-0.096)
Integration with adjacent cartilage	13	3 (2-5)	0.10 (0.068-0.14)
Repair tissue structure	11	2 (1.5-2.5)	0.071 (0.035-0.090)

FI = fragility index; IQR = interquartile range; FQ = fragility quotient.

**Table 4.** Bias Assessment for Included Studies Evaluated Using Revised Cochrane Risk-of-Bias Tool for Randomized Trials.

First Author	Domain 1: Risk of Bias Arising From Randomization Process	Domain 2: Risk of Bias Due to Deviations From the Intended Interventions	Domain 3: Risk of Bias Due to Missing Outcome Data	Domain 4: Risk of Bias in Measurement of the Outcome	Domain 5: Risk of Bias in Selection of the Reported Result	Overall Risk of Bias
Brittberg	Low risk	Low risk	Low risk	Some concerns	Low risk	Some concerns
Chung (2014)	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Chung (2023)	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Cole	Low risk	Low risk	Some concerns	Low risk	Low risk	Some concerns
Crawford	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Gudas	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Hashimoto	Low risk	Low risk	Low risk	Some concerns	Low risk	Some concerns
Hoburg (2021)	Low risk	Low risk	Low risk	Some concerns	Low risk	Some concerns
Hoburg (2023)	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Ibarra	High risk	Low risk	Some concerns	Low risk	Low risk	High risk
Kim	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Knutsen	Low risk	Low risk	Low risk	Some concerns	Low risk	Some concerns
Koh	Low risk	Low risk	Some concerns	Low risk	Low risk	Some concerns
Lim	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Saris	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Solheim	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Stanish	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Ulstein	Low risk	Low risk	Low risk	Some concerns	Low risk	Some concerns
VanAssche	Low risk	Low risk	Low risk	Some concerns	Low risk	Some concerns
Vanlauwe	Low risk	Low risk	Some concerns	Low risk	Low risk	Some concerns
Yoon	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk

For shading, green shading indicates low risk of bias. Yellow shading indicates some concerns for risk of bias. Red shading indicates high risk of bias. Refer to the online version of the article to view this table in color.

Bias assessment revealed that only one study was found to be at “high risk” of bias (Table 4). Furthermore, only 11 of the 105 domains of bias evaluated across the 21 included studies. Bias was identified most commonly in the domains for missing outcome data (i.e., lost to follow-up) and in measurement of outcomes.

Of the 21 included RCTs, 7 (33.3%) used the MCID metrics to demonstrate clinically significant improvement in reported outcomes.

## Discussion

The purpose of this systematic review was to use the FI, rFI, and FQ metrics to assess the statistical fragility of RCTs evaluating marrow stimulation techniques for cartilage restoration of the knee. Across 155 total outcomes, we demonstrated that just 3 outcome event reversals may alter statistical significance for the 21 included RCTs. In addition, outcome event reversals in just 6.7% of patients may be needed to alter statistical significance. We further demonstrated that the number of patients lost to follow-up was greater than or equal to the FI in over half of all outcomes. Subgroup analysis demonstrated considerable fragility for stem cell augments for marrow stimulation, while findings relating to MACI versus microfracture were most robust. The most fragile outcome categories across the cartilage restoration modalities involved cartilage defect filling and the repair tissue structure.

An important finding in this present study was the median FI of 2 identified for statistically significant outcomes, which indicates that outcome event reversals in just 2 patients may alter statistical significance in our review. In addition, with a median FQ of 0.050, statistical significance of findings may be lost with outcome event reversals in just 5% of patients. Furthermore, given that nearly one-third of the 32 statistically significant outcomes had an FI of 1, statistically significant findings in RCTs assessing marrow stimulation may not be as reliable as previously thought. Given that statistically significant outcomes reported in RCTs evaluating marrow stimulation may hinge on the outcome events of just 2 patients, these findings must be interpreted with caution.

For 86 out of 155 of the outcomes assessed, the number of patients lost to follow-up was greater than or equal to the outcome FI or rFI. This finding raises skepticism over the reliability of outcomes in RCTs assessing marrow stimulation techniques as outcome events lost due to attrition are capable of altering over half of all outcomes. Prior literature has suggested that a substantial portion of RCTs published do not adequately report follow-up data and those that do have high levels of missing outcome data.<sup>33,34</sup> Furthermore, in a fragility analysis of RCTs in the orthopedic sports medicine literature, over 50% of included studies did not report on potential sources of bias.<sup>35</sup> In our bias assessment, only 1 of 21 RCTs demonstrated high risk of bias, which

indicates that the statistical fragility identified likely is not a result of high level of bias among included studies. Ten of the studies demonstrated at least “some concerns” for bias and 4 of the studies were at risk of bias as a result of missing outcome data. Thus, efforts to minimize loss to follow-up may minimize bias and improve the reliability of RCT findings in the orthopedic literature.<sup>34</sup>

In a 2022 systematic review by Wen *et al.*,<sup>36</sup> augmented microfracture was deemed to have superior Lysholm scores and radiographic outcomes compared with microfracture alone. On the contrary, a 2022 systematic review and meta-analysis by Abraamyan *et al.*<sup>10</sup> identified no benefit for augmented microfracture across PROMs assessed. In our subgroup analysis, we identified that the 6 studies assessing augmented marrow stimulation were particularly fragile with just 3 outcome events altering statistical significance. Stem cell augments for marrow stimulation, in particular, were the most fragile with just 2 outcome event reversals required to alter significance. Outcomes relating to AMIC versus microfracture alone were less fragile; however, these outcomes may be reversed with outcome event reversals in just 3.5 patients. AMIC involves a 1-step repair procedure that involves marrow stimulation augmented by application of a collagen I/III matrix to stabilize the blood clot.<sup>37</sup> A systematic review by Kim *et al.*<sup>38</sup> reported significantly greater PROMs and radiographic findings for AMIC compared to microfracture. However, we demonstrate that RCT outcomes involving AMIC versus marrow stimulation alone might be more fragile than previously thought.

Prior literature has also suggested that ACI and MACI produce clinically meaningful improvements in PROMs compared with marrow stimulation especially in younger, more active populations.<sup>10,39</sup> Among the studies comparing cell-based techniques to marrow stimulation, our fragility analysis identified that the 5 studies assessing MACI versus microfracture were the least fragile with a median FI of 4, while the 4 studies assessing ACI versus microfracture had a median FI of 3.5. Although the studies assessing MACI versus microfracture were less fragile in our analysis, an FI of 4 is still of concern and additional comparative trial literature is needed to guide decision-making surrounding the role of cell-based techniques in cartilage restoration.

The *P*-value is a ubiquitous tool for determining statistical significance in the scientific literature.<sup>40</sup> However, when used independently, it has garnered significant criticism as it fails to indicate effect size, clinical significance, or consider patients lost to follow-up.<sup>41,42</sup> Furthermore, Chavalarias *et al.*<sup>43</sup> reported that with increasing use of the *P*-value in the biomedical literature, there has been a considerable bias toward reported significant *P*-values and even reporting data in a way that transforms nonsignificant findings into significant outcomes. Furthermore, Chavalarias *et al.*<sup>43</sup> recommend that the *P*-value not be used in isolation. As Sterne and Poeran noted, the FI and FQ metrics serve as valuable

metrics to clearly convey a result's uncertainty in a way that can be easily interpreted by clinicians.<sup>44</sup> In a 2018 study, Checketts *et al.*<sup>45</sup> identified a median FI of 2 and median FQ of 0.022 across 72 trials regarded as "strong evidence" by the American Academy of Orthopaedic Surgeons Clinical Practice Guidelines. Thus, studies guiding evidence-based medicine are prone to statistical fragility and their findings are not as robust as previously thought.

Statistical fragility in the orthopedic literature has been demonstrated by several studies.<sup>17,18,20,21,23,24,46-56</sup> In a 2019 study, Parisien *et al.*<sup>47</sup> found a median FI of 5 across 102 comparative trials in the sports medicine literature. In a recent study, Lawrence *et al.*<sup>57</sup> identified a median FI of 5 in studies evaluating bone-patellar tendon-bone versus hamstring tendon autografts for anterior cruciate ligament reconstruction. In an analysis of 19 RCTs in the knee cartilage restoration literature from 2000 to 2020, Parisien *et al.*<sup>46</sup> found a median FI of 4 across 60 outcomes. Our fragility analysis focused on RCTs published from 2010 to 2023 evaluating marrow stimulation techniques for knee cartilage restoration. Interestingly, we identified a higher level of statistical fragility (FI of 3) across 155 total outcomes compared to the 2021 study by Parisien *et al.* The findings in this present study thus highlight the continued need for future comparative trials evaluating cartilage restoration approaches such as MACI, OATS, AMIC, and scaffold/extracellular chondral matrix augments in treating chondral lesions.

This present fragility analysis included RCTs from 2010 to present across the PubMed, Embase, and Medline databases in adherence with the PRISMA guidelines. Our 2-directional fragility analysis demonstrated statistical fragility for significant and nonsignificant outcomes and took into account the sample size through the FQ metric. Our subgroup analysis by intervention comparisons, outcome category, and assessment of lost to follow-up among included RCTs further adds credence to the clinical implications of the statistical fragility identified in the marrow stimulation literature.

Given the statistical fragility identified across the orthopedic literature, future research should integrate the FI and FQ metrics in outcome reporting to aid in the interpretation of study findings. Furthermore, future studies should consider using additional statistical tools such as the MCID, substantial clinical benefit, patient-acceptable symptomatic state, and maximal outcome improvement to ensure clinically significant improvement from interventions employed. In the present fragility analysis, just 7 articles used such metrics to indicate clinically significant improvement. Consistent reporting of the MCID or other clinically significant outcome measures may allow for more effective evaluation of the extent of treatment effects.<sup>32</sup> Furthermore, these metrics will aid in standardizing assessment of patient improvement and ensuring evidence-based decision-making for management of cartilage defects of the knee.

## Limitations

This fragility analysis is not without limitations. Our systematic review was limited to the available RCT literature with an intervention arm related to marrow stimulation for cartilage restoration of the knee. In addition, our fragility analysis was limited to dichotomous outcomes, thus leaving out continuous outcomes or trials with greater than 2 intervention arms. Furthermore, while we categorized the extracted outcomes for subgroup analysis, there was heterogeneity across the studies in assessment of the outcomes (e.g., different timepoints of outcome assessment, different PROMs evaluated, different thresholds for volume of defect filling). This review also did not assess studies prior to 2010 which may have limited how comprehensive our fragility analysis of marrow stimulation RCTs was. Finally, there are currently no FI and FQ thresholds set in the literature. However, Baer *et al.*<sup>58</sup> argue against setting a uniform threshold and instead recommends considering the clinical question being addressed and study design characteristics when interpreting fragility. Thus, a comprehensive analysis of outcome robustness should include the *P*-value, FI, and FQ metrics in conjunction with evaluation of study design quality and evidence of bias.

## Conclusion

RCTs assessing marrow stimulation techniques in cartilage restoration of the knee are statistically fragile. Only 3 outcome event reversals may be sufficient to alter significance. We therefore recommend the combined reporting of FI and FQ metrics with *P*-values and clinically significant outcome metrics (i.e., the MCID) to ensure that clinicians are able to effectively interpret the robustness of outcomes reported in RCTs assessing knee cartilage restoration techniques.

## Acknowledgments and Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Ethical Approval

IRB approval was not required for this manuscript as only publicly available data was included in the investigation.

## ORCID iD

Avanish Yendluri  <https://orcid.org/0009-0005-3383-8138>



## References

1. Lesage C, Lafont M, Guihard P, Weiss P, Guicheux J, Delplace V. Material-assisted strategies for osteochondral defect repair. *Adv Sci*. 2022;9(16):e2200050.
2. Mirza MZ, Swenson RD, Lynch SA. Knee cartilage defect: marrow stimulating techniques. *Curr Rev Musculoskelet Med*. 2015;8(4):451-6.
3. Steadman JR, Rodkey WG, Briggs KK. Microfracture: its history and experience of the developing surgeon. *Cartilage*. 2010;1(2):78-86.
4. Mithoefer K, McAdams T, Williams RJ, Kreuz PC, Mandelbaum BR. Clinical efficacy of the microfracture technique for articular cartilage repair in the knee: an evidence-based systematic analysis. *Am J Sports Med*. 2009;37(10):2053-63.
5. Gao L, Goebel LKH, Orth P, Cucchiari M, Madry H. Subchondral drilling for articular cartilage repair: a systematic review of translational research. *Dis Model Mech*. 2018;11(6):dmm.034280. doi:10.1242/dmm.034280.
6. Steinwachs MR, Guggi T, Kreuz PC. Marrow stimulation techniques. *Injury*. 2008;39(suppl 1):S26-31.
7. Vonk LA, Roël G, Hernigou J, Kaps C, Hernigou P. Role of matrix-associated autologous chondrocyte implantation with spheroids in the treatment of large chondral defects in the knee: a systematic review. *Int J Mol Sci*. 2021;22(13):7149. doi:10.3390/ijms22137149.
8. Brittberg M. Cell carriers as the next generation of cell therapy for cartilage repair. *Am J Sports Med*. 2010;38:1259-71. doi:10.1177/0363546509346395.
9. Mistry H, Connock M, Pink J, Shyangdan D, Clar C, Royle P, et al. Autologous chondrocyte implantation in the knee: systematic review and economic evaluation. *Health Technol Assess*. 2017;21(6):1-294.
10. Abraamyan T, Johnson AJ, Wiedrick J, Crawford DC. Marrow stimulation has relatively inferior patient-reported outcomes in cartilage restoration surgery of the knee: a systematic review and meta-analysis of randomized controlled trials. *Am J Sports Med*. 2022;50(3):858-66.
11. Gudas R, Stankevicius E, Monastyreckiene E, Pranys D, Kalesinskas RJ. Osteochondral autologous transplantation versus microfracture for the treatment of articular cartilage defects in the knee joint in athletes. *Knee Surg Sports Traumatol Arthrosc*. 2006;14(9):834-42.
12. Haien Z, Jiachang W, Qiang L, Yufeng M, Zhenwei J. Osteochondral autologous transplantation compared to microfracture for treating osteochondral defect: an updated meta-analysis of randomized controlled trials. *J Knee Surg*. 2018;31(4):341-7.
13. Arshi A, Fabricant PD, Go DE, Williams RJ, McAllister DR, Jones KJ. Can biologic augmentation improve clinical outcomes following microfracture for symptomatic cartilage defects of the knee? a systematic review. *Cartilage*. 2018;9(2):146-55.
14. Leopold SS, Porcher R. Editorial: threshold p values in orthopaedic research—we know the problem. What is the solution? *Clin Orthop Relat Res*. 2018;476(9):1689-91.
15. Sterne JAC, Smith GD. Sifting the evidence—what's wrong with significance tests? *Phys Ther*. 2001;81(8):1464-9.
16. Feinstein AR. The unit fragility index: an additional appraisal of “statistical significance” for a contrast of two proportions. *J Clin Epidemiol*. 1990;43(2):201-9.
17. Davey MS, Hurley ET, Doyle TR, Dashti H, Gaafar M, Mullett H. The fragility index of statistically significant findings from randomized controlled trials comparing the management strategies of anterior shoulder instability. *Am J Sports Med*. 2023;51(8):2186-92.
18. Fackler NP, Ehlers CB, Callan KT, Amirhekmat A, Smith EJ, Parisien RL, et al. Statistical fragility of single-row versus double-row anchoring for rotator cuff repair: a systematic review of comparative studies. *Orthop J Sports Med*. 2022;10(5):23259671221093391.
19. Ehlers CB, Curley AJ, Fackler NP, Minhas A, Rodriguez AN, Pasko K, et al. The statistical fragility of single-bundle vs double-bundle autografts for ACL reconstruction: a systematic review of comparative studies. *Orthop J Sports Med*. 2021;9(12):23259671211064626.
20. Parisien RL, Trofa DP, Cronin PK, Dashe J, Curry EJ, Eichinger JK, et al. Comparative studies in the shoulder literature lack statistical robustness: a fragility analysis. *Arthrosc Sports Med Rehabil*. 2021;3(6):e1899-904.
21. Megafu MN, Megafu EC, Nguyen JT, Mian HS, Singhal SS, Parisien RL. The statistical fragility of orbital fractures: a systematic review of randomized controlled trials. *J Oral Maxillofac Surg*. 2023;81(6):752-8.
22. Megafu M, Mian H, Megafu E, Singhal S, Lee A, Cassie R, et al. The fragility of statistical significance in distal femur fractures: systematic review of randomized controlled trials. *Eur J Orthop Surg Traumatol*. 2023;33(6):2411-8.
23. Cordero JK, Lawrence KW, Brown AN, Li X, Hayden BL, Parisien RL. The fragility of tourniquet use in total knee arthroplasty: a systematic review of randomized controlled trials. *J Arthroplasty*. 2023;38(6):1177-83.
24. Fackler NP, Karasavvidis T, Ehlers CB, Callan KT, Lai WC, Parisien RL, et al. The statistical fragility of operative vs nonoperative management for Achilles tendon rupture: a systematic review of comparative studies. *Foot Ankle Int*. 2022;43(10):1331-9.
25. Bragg JT, Ruelos VCB, McIntyre JA, Puzitiello RN, Pagani NR, Menendez ME, et al. Reverse fragility index comparing rates of rerupture after open Achilles tendon repair versus early functional rehabilitation: a systematic review of randomized controlled trials. *Am J Sports Med*. Published online June 12, 2023. doi:10.1177/03635465231178831.
26. Ruelos VCB, Masood R, Puzitiello RN, Moverman MA, Pagani NR, Menendez ME, et al. The reverse fragility index: RCTs reporting non-significant differences in failure rates between hamstring and bone-patellar tendon-bone autografts have fragile results. *Knee Surg Sports Traumatol Arthrosc*. 2023;31(8):3412-9.
27. Khan MS, Fonarow GC, Friede T, Lateef N, Khan SU, Anker SD, et al. Application of the reverse fragility index to statistically nonsignificant randomized clinical trial results. *JAMA Netw Open*. 2020;3(8):e2012469.
28. Tignanelli CJ, Napolitano LM. The fragility index in randomized clinical trials as a means of optimizing patient care. *JAMA Surg*. 2019;154(1):74-9.

29. Garcia MVF, Ferreira JC, Caruso P. Fragility index and fragility quotient in randomized clinical trials. *J Bras Pneumol*. 2023;49(1):e20230034.
30. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev*. 2021;10(1):89.
31. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, *et al*. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898.
32. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J*. 2007;7(5):541-6.
33. Somerson JS, Bartush KC, Shroff JB, Bhandari M, Zelle BA. Loss to follow-up in orthopaedic clinical trials: a systematic review. *Int Orthop*. 2016;40(11):2213-9.
34. Baer BR, Fremez SE, Gaudino M, Charlson M, Wells MT. On clinical trial fragility due to patients lost to follow up. *BMC Med Res Methodol*. 2021;21(1):254.
35. Khan M, Evaniew N, Gichuru M, Habib A, Ayeni OR, Bedi A, *et al*. The fragility of statistically significant findings from randomized trials in sports surgery: a systematic survey. *Am J Sports Med*. 2017;45(9):2164-70.
36. Wen HJ, Yuan LB, Tan HB, Xu YQ. Microfracture versus enhanced microfracture techniques in knee cartilage restoration: a systematic review and meta-analysis. *J Knee Surg*. 2022;35(7):707-17.
37. Sciarretta FV, Ascani C, Sodano L, Fossati C, Campisi S. One-stage cartilage repair using the autologous matrix-induced chondrogenesis combined with simultaneous use of autologous adipose tissue graft and adipose tissue mesenchymal cells technique: clinical results and magnetic resonance imaging evaluation at five-year follow-up. *Int Orthop*. 2024;48(1):267-277. doi:10.1007/s00264-023-05921-8.
38. Kim JH, Heo JW, Lee DH. Clinical and radiological outcomes after autologous matrix-induced chondrogenesis versus microfracture of the knee: a systematic review and meta-analysis with a minimum 2-year follow-up. *Orthop J Sports Med*. 2020;8(11):2325967120959280.
39. Brittberg M, Recker D, Ilgenfritz J, Saris DBF; SUMMIT Extension Study Group. Matrix-applied characterized autologous cultured chondrocytes versus microfracture: five-year follow-up of a prospective randomized trial. *Am J Sports Med*. 2018;46(6):1343-51.
40. Dahiru T. P-value, a true test of statistical significance? a cautionary note. *Ann Ib Postgrad Med*. 2008;6(1):21-6.
41. Grech V, Eldawlatly AA. The value—and its historical underpinnings—pro and con. *Saudi J Anaesth*. 2023;17(3):391-3.
42. Thiese MS, Ronna B, Ott U. P value interpretations and considerations. *J Thorac Dis*. 2016;8(9):E928-31.
43. Chavalarias D, Wallach JD, Li AHT, Ioannidis JPA. Evolution of reporting p values in the biomedical literature, 1990-2015. *JAMA*. 2016;315(11):1141-8.
44. Stern BZ, Poeran J. Statistics in brief: the fragility index. *Clin Orthop Relat Res*. 2023;481(7):1288-91.
45. Checketts JX, Scott JT, Meyer C, Horn J, Jones J, Vassar M. The robustness of trials that guide evidence-based orthopaedic surgery. *J Bone Joint Surg Am*. 2018;100(12):e85.
46. Parisien RL, Constant M, Saltzman BM, Popkin CA, Ahmad CS, Li X, *et al*. The fragility of statistical significance in cartilage restoration of the knee: a systematic review of randomized controlled trials. *Cartilage*. 2021;13(suppl 1):147S-155S.
47. Parisien RL, Trofa DP, Dashe J, Cronin PK, Curry EJ, Fu FH, *et al*. Statistical fragility and the role of p values in the sports medicine literature. *J Am Acad Orthop Surg*. 2019;27(7):e324-9.
48. Pearsall C, Constant M, Saltzman BM, Parisien RL, Levine W, Trofa D. The fragility of statistical significance in sham orthopaedic surgery: a systematic review of randomized controlled trials. *J Am Acad Orthop Surg*. 2023;31:e994-1002. doi:10.5435/JAAOS-D-23-00245.
49. Megafu M, Megafu E, Mian H, Singhal S, Nietsch K, Yendluri A, *et al*. The statistical fragility of outcomes in calcaneus fractures: a systematic review of randomized controlled trials. *Foot (Edinb)*. 2023;57:102047.
50. Chan JP, Vrla M, Thompson C, Trofa DP, Li X, Wang D, *et al*. Statistical fragility of randomized controlled trials evaluating platelet-rich plasma use for knee osteoarthritis: a systematic review. *Orthop J Sports Med*. 2023;11(8):23259671231187894.
51. Megafu MN, Mian HS, Hassan MM, Parsons BO, Li X, Parisien RL. The fragility of statistical findings in distal biceps tendon repairs: a systematic review of randomized controlled trials. *J Shoulder Elbow Surg*. 2023;32(8):e379-86.
52. Mian H, Megafu M, Megafu E, Singhal S, Richardson NG, Tornetta P, *et al*. The statistical fragility of the distal fibula fracture literature: a systematic review of randomized controlled trials. *Injury*. Published online March 20, 2023. doi:10.1016/j.injury.2023.03.022.
53. Parisien RL, Trofa DP, O'Connor M, Knapp B, Curry EJ, Tornetta P 3rd, *et al*. The fragility of significance in the hip arthroscopy literature: a systematic review. *JB JS Open Access*. 2021;6(4):e21.00035. doi:10.2106/JBJS.OA.21.00035.
54. Constant M, Trofa DP, Saltzman BM, Ahmad CS, Li X, Parisien RL. The fragility of statistical significance in patellofemoral instability research: a systematic review. *Am J Sports Med*. 2022;50(13):3714-8.
55. Doyle TR, Hurley ET, Davey MS, Klifto C, Mullett H. The statistical fragility of the management options for reverse shoulder arthroplasty: a systematic review of randomized control trial with fragility analysis. *JSES Rev Rep Tech*. 2023;3(3):279-84.
56. Doyle TR, Davey MS, Hurley ET. The statistical fragility of management options for acute Achilles tendon ruptures—a systematic review of randomized control trial with fragility analysis. *J ISAKOS*. 2022;7(4):72-81.
57. Lawrence KW, Okewunmi JO, Chakrani Z, Cordero JK, Li X, Parisien RL. Randomized controlled trials comparing bone-patellar tendon-bone versus hamstring tendon autografts in anterior cruciate ligament reconstruction surgery are statistically fragile: a systematic review. *Arthroscopy*. Published online August 4, 2023. doi:10.1016/j.arthro.2023.07.039.
58. Baer BR, Gaudino M, Fremez SE, Charlson M, Wells MT. The fragility index can be used for sample size calculations in clinical trials. *J Clin Epidemiol*. 2021;139:199-209.