

Statistical Fragility of Randomized Controlled Trials Evaluating Platelet-Rich Plasma Use for Knee Osteoarthritis

A Systematic Review

Justin P Chan,[†] MD, Michael Vrla,[†] MD, Claire Thompson,[‡] BA, David P Trofa,[§] MD, Xinning Li,^{||} MD, Dean Wang,^{*†} MD, and Robert L Parisien,[¶] MD

Investigation performed at University of California, Irvine, Irvine, California, USA

Background: Numerous studies have been published on the use of platelet-rich plasma (PRP) for knee osteoarthritis (OA), with conflicting results.

Purpose: To determine the fragility index (FI) and fragility quotient (FQ) of randomized controlled trials (RCTs) that evaluated the use of PRP to treat knee OA.

Study Design: Systematic review.

Methods: RCTs evaluating the efficacy of PRP injections for knee OA from 2000 to 2020 were included for analysis according to PRISMA guidelines. The FI was determined by calculating the number of outcome event reversals required to change the statistical significance. The associated FQ was determined by dividing the FI by the sample size.

Results: Our initial search resulted in 41,149 studies, of which 8 RCTs (678 patients, 72 outcome events) were included in the analysis. One study failed to report PRP formulation details, whereas 87.5% of studies reported using either leukocyte-rich or leukocyte-poor PRP. The platelet concentration was reported in 25% of the included trials. The overall FI of the 72 outcome events was 8.5. Accounting for sample size, the associated FQ was determined to be 0.14, suggesting that the reversal of 14% of outcome events was required to change outcome significance. There were 51 statistically significant outcomes, of which the FI and FQ were 12 and 0.164, respectively.

Conclusion: Comprehensive fragility analysis suggested that the published literature evaluating the efficacy of PRP use for knee OA may lack statistical stability. We recommend the reporting of both an FI and FQ in addition to *P* value analysis to provide a clear and thorough understanding of the statistical integrity of studies reporting on PRP use for knee OA.

Keywords: cartilage preservation; knee osteoarthritis; orthobiologics; platelet-rich plasma

Modern medicine is characterized in large part by the incorporation of evidence-based guidelines for treatment indications and techniques. As clinical research continues to advance, comprehensive evaluation of the available literature is critically important to the understanding of significant findings. The standard method of reporting statistical analyses in the literature is in the form of *P* values, with the conventional threshold set at $\alpha = .05$, correlating to a 5% chance of the observed outcome occurring by chance.³ If the *P* value is less than .05, the null hypothesis is rejected and the observed difference is determined to be statistically significant. The *P* value is interpreted in a dichotomous

fashion, representing either significance or no significance. Despite this relative ease of interpretation, it may prove more difficult to reconcile when similar studies offer conflicting significance of outcomes data. Statistical significance may also not necessarily equate to clinical significance.

The apparent reversal of significance can be determined via careful calculation of the fragility index (FI). First proposed by physician and epidemiologist Alvan Feinstein in 1990, the FI is a measure of the number of outcome event reversals necessary to alter statistical significance.⁹ The FI concept has since been applied across medical specialties with recent adoption in musculoskeletal medicine and orthopaedic surgery.^{5,10,14,15,20,21} To account for differing sample sizes, the FI is appropriately supplemented with the fragility quotient (FQ). The FQ is a measure of

quantitative significance and is determined by dividing the FI by the sample size.

Knee osteoarthritis (OA) is a common source of musculoskeletal pain, with a global prevalence estimated at 3.8%.⁶ As the population ages, knee OA is predicted to become an even more frequent cause of disability, resulting in an increasing burden on individuals and a financial burden to the health care systems and societies.⁶ Total joint arthroplasty is a well-established surgical method for the treatment of end-stage OA but may only be considered in severe OA in the setting of failed nonoperative management. Physical therapy, knee unloader braces, and nonsteroidal anti-inflammatories are also mainstays of treatment. Intra-articular steroid or hyaluronic acid injections are frequently employed as well, but have demonstrated inconclusive clinical effectiveness.¹¹⁻¹³

In the absence of consistently effective nonoperative treatments for knee OA, there has been a recent increase in scientific interest with regard to the development of new biologic modalities. In particular, platelet-rich plasma (PRP) is an autologous blood product containing high concentrations of growth factors, which have been hypothesized to enhance cartilage healing and reduce inflammation.¹ There have been an increasing number of studies in the last decade evaluating PRP for knee OA, but there remains a lack of consensus due to factors such as study heterogeneity, differing PRP formulations, and study bias.^{2,7,24}

In this study, we applied the FI and FQ to the literature on the use of PRP injections for the treatment of knee OA. The purpose of this study was to evaluate the fragility of randomized controlled trials (RCTs) evaluating the efficacy of PRP for knee OA with utilization of FI and FQ statistical analysis. We hypothesized that the overall FI and FQ would demonstrate significant statistical fragility with no difference appreciated between leukocyte-rich (LR) and leukocyte-poor (LP) formulations of PRP.

METHODS

This systematic review was conducted according to PRISMA guidelines. A comprehensive literature search of the PubMed database from 2000 to 2020 was performed using the search terms (*platelet rich plasma OR platelet rich products OR orthobiologics*) AND (*knee pain OR knee osteoarthritis*). Peer-reviewed RCTs pertaining to the use of

TABLE 1
The 2019 Impact Factor of the Included Journals

| Journal | Impact Factor |
|--|---------------|
| <i>American Journal of Sports Medicine</i> | 5.810 |
| <i>International Journal of Molecular Sciences</i> | 4.556 |
| <i>Arthroscopy</i> | 4.325 |
| <i>Journal of Orthopaedics and Traumatology</i> | 2.767 |
| <i>Clinical Rehabilitation</i> | 2.599 |
| <i>World Journal of Orthopedics</i> | 0.798 |

PRP in knee OA in select journals with high impact factors were included for analysis (Table 1). The literature search was conducted independently by 2 reviewers (J.P.C and M.V.). Duplicate records were removed. Disagreements were resolved by referral to a third reviewer (D.W.).

Included were RCTs that reported dichotomous comparative data with associated *P* value analysis. The type of outcome measure (primary, secondary, not specified) was documented. The particular formulation of PRP utilized was recorded as LR-PRP, LP-PRP, or not specified. Loss to follow-up (LTF) data were evaluated for all studies. Fragility analysis was performed by manipulating the reported outcome events in a 2 × 2 contingency table until a reversal of significance occurred, with statistical significance defined as *P* < .05 (Figure 1). For example, if a particular outcome was initially reported as statistically significant, the number of outcome events required to raise *P* to ≥ .05 was determined. Conversely, if the outcome was initially reported as nonsignificant, the number of outcome events required to decrease *P* to < .05 was determined. The corresponding number indicates the number needed to reverse a particular outcome event and was recorded as the FI for that event.

All overturned outcome events were calculated in this manner, with the median value representing the FI for the entire study population. The associated FQ was determined for each outcome event by dividing the FI by the sample size. In addition, the total FQ for all outcome events was determined. The reported *P* value was recorded for each outcome event and verified for accuracy using the two-tailed Fisher exact test. Interquartile ranges (IQRs) were calculated to provide a more comprehensive understanding and interpretation of the reported variability and dispersion as the difference between the 25th and 75th percentiles.

*Address correspondence to Dean Wang, MD, Department of Orthopaedic Surgery and Department of Biomedical Engineering, University of California, Irvine, 101 The City Drive S, Pavilion III Building 29A, Orange, CA 92868, USA (email: deanwangmd@gmail.com).

†Department of Orthopaedic Surgery, University of California, Irvine, Irvine, California, USA.

‡New York University Grossman School of Medicine, New York, New York, USA.

§Department of Orthopaedic Surgery, Columbia University Medical Center, New York, New York, USA.

||Department of Orthopaedic Surgery, Boston University School of Medicine, Boston, Massachusetts, USA.

*Department of Orthopaedic Surgery and Sports Medicine, Mount Sinai Health System, New York, New York, USA.

Final revision submitted March 18, 2023; accepted April 13, 2023.

Final revision submitted March 18, 2023; accepted April 13, 2023.

One or more of the authors has declared the following potential conflict of interest or source of funding: D.P.T. has received grant support from Arthrex and education payments from Arthrex and Smith & Nephew. D.W. has received grant support from Arthrex, education payments from Arthrex and Smith & Nephew, consulting fees from DePuy/Medical Device Business Services and Ipsen Bioscience, nonconsulting fees from Vericel, and hospitality payments from Linvatec. R.L.P. has received grant support and education payments from Arthrex. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

| | Outcome A | Outcome B |
|--------------------|-----------|-----------|
| Treatment A | 24 | 24 |
| Treatment B | 11 | 31 |
| P Value | .03 | |

| | Outcome A | Outcome B |
|--------------------|-----------|-----------|
| Treatment A | 22 | 26 |
| Treatment B | 11 | 31 |
| P Value | .08 | |

Figure 1. Example of a study with an outcome initially reported as statistically significant (top 2 × 2 contingency table). The number of outcome events required to change *P* to nonsignificant status (ie, the fragility index) was 2 (bottom 2 × 2 contingency table).

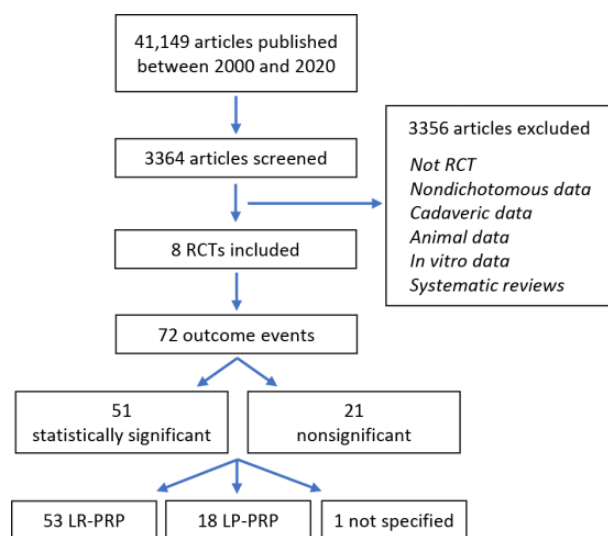


Figure 2. PRISMA study identification flowchart. LP, leukocyte poor; LR, leukocyte rich; PRP, platelet-rich plasma.

RESULTS

Of 41,149 initial studies, 3364 studies were screened, with 8 RCTs^{4,16-19,22,23,25} ultimately meeting both inclusion and exclusion criteria and thus included for analysis (Figure 2). There were 72 total outcome events, with 51 (70.8%) initially reported as statistically significant ($P < .05$) and 21 (29.2%) initially reported as not significant ($P \geq .05$). Outcomes initially reported as significant were found to be vastly more stable than those initially reported as not significant. Of the 51 outcomes initially reported as

TABLE 2
Fragility Data Based on Trial and Outcome Characteristics

| Characteristic | Events | Fragility Index (IQR) | Fragility Quotient (IQR) |
|-------------------------|--------|-----------------------|--------------------------|
| All Trials | 72 | 8.5 (4.5-14.5) | 0.14 (0.066-0.218) |
| Outcome | | | |
| Primary | 27 | 8 (5-14) | 0.111 (0.056-0.156) |
| Secondary | 45 | 9 (4-16.5) | 0.164 (0.071-0.282) |
| Reported <i>P</i> value | | | |
| <.05 | 51 | 12 (5-18) | 0.164 (0.078-0.273) |
| ≥.05 | 21 | 5 (1.5-7) | 0.093 (0.022-0.148) |
| PRP formulation | | | |
| LR-PRP | 53 | 12 (6.5-17.5) | 0.164 (0.090-0.264) |
| LP-PRP | 18 | 3.5 (2-6) | 0.061 (0.031-0.125) |
| Not specified | 1 | 6 | 0.111 |

^aAbbreviations: IQR, interquartile range; LP, leukocyte poor; LR, leukocyte rich; PRP, platelet-rich plasma.

statistically significant, the median number of events required to reverse significance (FI) was 12 (IQR 5-18) (Table 2). The associated FQ for statistically significant outcomes was 0.164 (IQR 0.078-0.273). Of the 21 outcomes initially reported as not statistically significant, the median number of events required to reverse significance (FI) was only 5 (IQR 1.5-7) with a mean *P* value of .51. The associated FQ for initially nonsignificant outcomes was 0.093 (IQR 0.022-0.148). Primary outcomes ($n = 27$) were slightly more fragile than secondary outcomes ($n = 45$) with an FI of 8 (IQR 5-14) and 9 (IQR 4-16.5), respectively. The associated FQ also supported slightly increased fragility with primary versus secondary outcomes with values of 0.111 and 0.164, respectively.

Of the 8 RCTs analyzed, 4 (50%) represented data from LR-PRP formulations^{16,22,23,25} and 3 (37.5%) from LP-PRP formulations,^{4,18,19} with 1 study (12.5%) that did not specify leukocyte characteristics of the PRP used.¹⁷ The platelet concentration was reported in 2 (25%) of the included trials.^{4,19} Further fragility analysis of RCTs utilizing LR-PRP was found to be more stable than those utilizing LP-PRP. In evaluation of 53 LR-PRP and 18 LP-PRP outcome events, the FI was determined to be 12 (IQR 6.5-17.5) and 3.5 (IQR 2-6), respectively. Similarly, the FQ was 0.164 for LR-PRP and 0.061 for LP-PRP.

The overall FI, incorporating 72 outcome events from all 8 RCTs, was 8.5 (IQR 4.5-14.5). Accounting for sample size, the overall FQ was 0.14 (IQR 0.066-0.218), suggesting the reversal of 14% of outcome events is required to alter trial significance. All 8 RCTs reported loss to follow-up data.

DISCUSSION

Our analysis of all outcome events after PRP in the treatment of knee OA suggests a lack of statistical stability in the included RCTs, with an FI of 8.5 and an associated FQ of 0.14, suggesting the reversal of only 14% of events

required to alter study significance and, thus, the conclusions found within the studies. When comparing LR-PRP and LP-PRP formulations, we found that the reported LR-PRP data were substantially more stable than the LP-PRP data.

The fragility findings of this review may be of particular concern, as outcomes of RCTs represent the best available data and are frequently utilized to inform clinical practice guidelines. Our findings provide further evidence in support of recent fragility analyses revealing statistical fragility in the greater orthopaedic literature.^{5,10,20,21} Furthermore, 1 of the 8 RCTs reported an LTF greater than the overall FI.¹⁸ If all the missing data were to skew in the same direction, it may be possible in this instance to realize a reversal of statistical significance by merely maintaining study follow-up. As such, it is important for academic and clinical professionals to appreciate the limitations of *P* value analysis by clearly distinguishing between statistical significance and clinical applicability. Therefore, we support previous recommendations for the presentation of the *P* value in conjunction with FI and FQ analysis to enhance understanding.

In evaluation of trial results initially reported as significant, we found they were considerably more stable than those reported as nonsignificant.^{20,21} The FI and FQ of the 51 outcomes reported as significant were 12 and 0.164, respectively. Comparatively, the FI of the 21 nonsignificant outcomes was only 5 with an associated FQ of 0.093. The latter represents the greater ease of reversal of potential false negative findings as compared to false positives. However, when we analyzed the trials individually, we found that the number of outcomes associated with significant and nonsignificant findings may explain the discrepancy.^{20,21} We found no consistent trend for the initial reported significance; rather, we found trials with more reported outcome events were generally more stable. In all included studies, increased statistical robustness via the FI correlated with cohort size. However, when normalizing with respect to sample size, the FQ for significant outcomes was 0.164 and provides further support of increased stability as compared with nonsignificant outcomes with an FQ of 0.093.

Statistical fragility is not a novel concept with growing research supporting the necessity of complementary statistics in the medical literature. Checketts et al evaluated 72 orthopaedic clinical trials cited as providing strong evidence in the American Academy of Orthopaedic Surgeons Clinical Practice Guidelines.⁵ They identified an FI of only 2 with an associated FQ of 0.022. The authors additionally noted an FI of 0 in 16 of the included trials as they demonstrated a reversal of statistical significance by calculating *P* values via the Fisher exact test. This highlighted alarming statistical fragility with 22% of all trials representing significant vulnerability to reversal with the simple use of an alternative statistical test for *P* value analysis.⁵ This further underscores the importance of reporting complementary statistics accompanying *P* values in order to provide a more complete picture of that statistical integrity of comparative trials.

Our findings regarding statistical fragility in RCTs on PRP for knee OA indicate relatively higher stability compared with findings in other orthopaedic subspecialties. In

an analysis of 132 outcome events in 40 eligible RCTs pertaining to spine surgery interventions, Evaniew et al reported an FI of 2 with 65% of trials demonstrating an FI less than or equal to the reported LTF.⁸ Similarly, an analysis of orthopaedic oncology literature conducted by Forrester et al identified a median FI of 2 which was less than number LTF in 60% of outcomes.¹⁰ In addition, Khan et al analyzed the results from 48 RCTs from sports medicine and arthroscopic surgery literature and identified an FI of 2 (IQR 1-2.8).¹⁴ Furthermore, Khormaei et al performed a systematic evaluation of pediatric orthopaedic literature published over a 10-year period between 2006 and 2016.¹⁵ They analyzed 116 outcome events from 17 RCTs and reported an FI of 3.¹⁵ These findings further emphasize the advantage of FI and FQ metrics as easily identifiable statistical complements to the *P* value, providing a more complete understanding of trial stability. This is of particular importance in the treatment of knee OA as there remains conflicting evidence and a lack of consensus regarding the clinical efficacy of PRP.

Limitations

Our study is not without limitations. We carefully evaluated RCTs examining the use of PRP for knee OA, but our findings may not be utilized to recommend for or against treatment. Furthermore, the included RCTs utilized differing PRP formulations with variable comparisons to corticosteroid or hyaluronic acid injections, so we therefore cannot suggest specific management recommendations. Rather, our fragility analysis provides an assessment of the statistical robustness of the highest level of evidence evaluating the management of knee OA with PRP formulations across all relevant RCTs reported in the literature. Given the lack of fragility analyses in the evaluation of PRP for knee OA, as well as across the greater orthopaedic literature, specific fragility thresholds have yet to be determined and require further study for future guidance on trial robustness and clinical decision-making.

CONCLUSION

Comprehensive fragility analysis suggests the published literature evaluating the efficacy of PRP for knee OA may lack statistical stability. We therefore recommend the reporting of both an FI and FQ in addition to *P* value analysis to provide a clear and thorough understanding of the statistical integrity of studies reporting on PRP use for knee OA.

REFERENCES

1. Ahmad Z, Howard D, Brooks RA, et al. The role of platelet rich plasma in musculoskeletal science. *JRSM Short Rep*. 2012;3(6):1-9. doi:10.1258/shorts.2011.011148
2. Bendich I, Rubenstein WJ, Cole BJ, Ma CB, Feeley BT, Lansdown DA. What is the appropriate price for platelet-rich plasma injections for knee osteoarthritis? A cost-effectiveness analysis based on evidence from level I randomized controlled trials. *Arthroscopy*. 2020;36(7):1983-1991.e1. doi:10.1016/j.arthro.2020.02.004

3. Biau DJ, Jolles BM, Porcher R. *P* value and the theory of hypothesis testing: an explanation for new researchers. *Clin Orthop Relat Res*. 2010;468(3):885-892. doi:10.1007/s11999-009-1164-4
4. Buendía-López D, Medina-Quirós M, Fernández-Villacañas Marín MÁ. Clinical and radiographic comparison of a single LP-PRP injection, a single hyaluronic acid injection and daily NSAID administration with a 52-week follow-up: a randomized controlled trial. *J Orthop Traumatol*. 2018;19(1). doi:10.1186/s10195-018-0501-3
5. 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. doi:10.2106/JBJS.17.01039
6. Cross M, Smith E, Hoy D, et al. The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis*. 2014;73(7):1323-1330. doi:10.1136/annrheumdis-2013-204763
7. Dai WL, Zhou AG, Zhang H, Zhang J. Efficacy of platelet-rich plasma in the treatment of knee osteoarthritis: a meta-analysis of randomized controlled trials. *Arthroscopy*. 2017;33(3):659-670.e1. doi:10.1016/j.arthro.2016.09.024
8. Evaniew N, Files C, Smith C, et al. The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey. *Spine J*. 2015;15(10):2188-2197. doi:10.1016/j.spinee.2015.06.004
9. Feinstein AR. The unit fragility index: an additional appraisal of "statistical significance" for a contrast of two proportions. *J Clin Epidemiol*. 1990;43(2):201-209. doi:10.1016/0895-4356(90)90186-S
10. Forrester LA, Jang E, Lawson MM, Capi A, Tyler WK. Statistical fragility of surgical and procedural clinical trials in orthopaedic oncology. *JAAOS Glob Res Rev*. 2020;4(6): e19.00152. doi:10.5435/jaaosglobal-d-19-00152
11. Jevsevar D, Donnelly P, Brown GA, Cummins DS. Viscosupplementation for osteoarthritis of the knee: a systematic review of the evidence. *J Bone Jt Surg - Am Vol*. 2014;97(24):2047-2060. doi:10.2106/JBJS.N.00743
12. Jevsevar DS. Treatment of osteoarthritis of the knee: evidence-based guideline, 2nd edition. *J Am Acad Orthop Surg*. 2013;21(9):571-576. doi:10.5435/JAAOS-21-09-571
13. Jüni P, Hari R, Rutjes AWS, et al. Intra-articular corticosteroid for knee osteoarthritis. *Cochrane Database Syst Rev*. 2015;2015(10). doi:10.1002/14651858.CD005328.pub3
14. Khan M, Evaniew N, Gichuru M, et al. The fragility of statistically significant findings from randomized trials in sports surgery: a systematic survey. *Am J Sports Med*. 2017;45(9):2164-2170. doi:10.1177/0363546516674469
15. Khormaei S, Choe J, Ruzbarsky JJ, et al. The fragility of statistically significant results in pediatric orthopaedic randomized controlled trials as quantified by the fragility index: a systematic review. *J Pediatr Orthop*. 2018;38(8):e418-e423. doi:10.1097/BPO.0000000000001201
16. Kon E, Engebretsen L, Verdonk P, Nehrer S, Filardo G. Clinical outcomes of knee osteoarthritis treated with an autologous protein solution injection: a 1-year pilot double-blinded randomized controlled trial. *Am J Sports Med*. 2018;46(1):171-180. doi:10.1177/0363546517732734
17. Lisi C, Perotti C, Scudeller L, et al. Treatment of knee osteoarthritis: platelet-derived growth factors vs. hyaluronic acid. A randomized controlled trial. *Clin Rehabil*. 2018;32(3):330-339. doi:10.1177/0269215517724193
18. Louis ML, Magalon J, Jouve E, et al. Growth factors levels determine efficacy of platelets rich plasma versus hyaluronic acid in the treatment of osteoarthritis: a randomized double blind noninferiority trial compared with viscosupplementation. *Arthroscopy*. 2018;34(5):1530-1540.e2. doi:10.1016/j.arthro.2017.11.035
19. Montañez-Heredia E, Irizar S, Huertas PJ, et al. Intra-articular injections of platelet-rich plasma versus hyaluronic acid in the treatment of osteoarthritic knee pain: a randomized clinical trial in the context of the Spanish national health care system. *Int J Mol Sci*. 2016;17(7). doi:10.3390/ijms17071064
20. Parisien RL, Dashe J, Cronin PK, Bhandari M, Tornetta P. Statistical significance in trauma research: too unstable to trust? *J Orthop Trauma*. 2019;33(12): e466-e470. doi:10.1097/BOT.0000000000001595
21. Parisien RL, Trofa DP, Dashe J, et al. Statistical fragility and the role of *P* values in the sports medicine literature. *J Am Acad Orthop Surg*. 2019;27(7):e324-e329. doi:10.5435/JAAOS-D-17-00636
22. Sánchez M, Fiz N, Azofra J, et al. A randomized clinical trial evaluating plasma rich in growth factors (PRGF-Endoret) versus hyaluronic acid in the short-term treatment of symptomatic knee osteoarthritis. *Arthroscopy*. 2012;28(8):1070-1078. doi:10.1016/j.arthro.2012.05.011
23. Tavassoli M, Janmohammadi N, Hosseini A, Khafri S, Esmaeilnejad-Ganji SM. Single- and double-dose of platelet-rich plasma versus hyaluronic acid for treatment of knee osteoarthritis: a randomized controlled trial. *World J Orthop*. 2019;10(9):311-326. doi:10.5312/wjo.v10.i9.310
24. Trams E, Kulinski K, Kozar-Kaminska K, Pomianowski S, Kaminski R. The clinical use of platelet-rich plasma in knee disorders and surgery—a systematic review and meta-analysis. *Life*. 2020;10(6):94. doi:10.3390/life10060094
25. Vaquerizo V, Plasencia MÁ, Arribas I, et al. Comparison of intra-articular injections of plasma rich in growth factors (PRGF-Endoret) versus durolane hyaluronic acid in the treatment of patients with symptomatic osteoarthritis: a randomized controlled trial. *Arthroscopy*. 2013;29(10):1635-1643. doi:10.1016/j.arthro.2013.07.264