




The Statistical Fragility of Platelet-Rich Plasma as Treatment for Plantar Fasciitis: A Systematic Review and Simulated Fragility Analysis

Arjun Gupta, BS^{1,2}, Carlos Ortiz-Babilonia, BS^{1,3}, Amy L. Xu, BS¹, Davis Rogers, MD¹, Ettore Vulcano, MD⁴, and Amiethab A. Aiyer, MD¹

Abstract

Background: Plantar fasciitis (PF) is the most common cause of heel pain and can be a source of extensive physical disability and financial burden. Platelet-rich plasma (PRP) offers a potentially definitive, regenerative treatment modality that, if effective, could change the current paradigm of PF care. However, randomized controlled trials (RCTs) on the clinical benefits of PRP for refractory PF offer inconsistent conclusions, potentially because of the broader limitations of using *P* value thresholds to declare statistical and clinical significance. In this study, we use the Continuous Fragility Index (CFI) and Quotient (CFQ) to appraise the statistical robustness of data from RCTs evaluating PRP for treatment of PF.

Methods: RCTs comparing outcomes after PRP injection vs alternative treatment in patients with chronic PF were evaluated. Representative simulated data sets were generated for each reported outcome event using summary statistics. The CFI was determined by manipulating each data set until reversal of significance ($\alpha=0.05$) was achieved. The corresponding CFQ was calculated by dividing the CFI by the sample size.

Results: Of 259 studies screened, 20 studies (59 outcome events) were included in this analysis. From these simulations, the median CFI for all events was 9, suggesting that varying the treatment of 9 patients would be required to reverse trial significance. The corresponding CFQ was 0.177. Studies with reported *P* value $<.05$ were more statistically fragile (CFI=10, CFQ=0.122) than studies with reported *P* value $>.05$ (CFI=5, CFQ=0.179). Of 36 outcome events reporting lost to follow-up data, 10 events (27.8%) lost ≥ 9 patients.

Conclusion: Our findings suggest that, on average, the statistical fragility of RCTs evaluating PRP for nonoperative PF therapy is at least comparable to that of the sports medicine literature. However, several included studies had concerning low simulated fragility scores. Orthopaedic surgeons may benefit from preferentially relying on studies with higher CFI and CFQ values when evaluating the utility of PRP for chronic PF in their own clinical practice. Given the importance of RCT data in clinical decision making, fragility indices could help give context to the stability of statistical findings.

Level of Evidence: Level I, systematic review.

Keywords: statistical fragility, plantar fasciitis, platelet rich plasma, fragility index, fragility quotient, orthobiologics, randomized controlled trials

Introduction

Approximately 1 of every 10 adults aged 50 and older suffers from plantar fasciitis (PF).⁵⁴ With an aging population, the overall prevalence of PF is expected to increase.³⁴ PF is estimated to account for 8% of all running-related injuries and has been associated with increased social isolation, reduced functional capabilities, and poor perception of

health status.^{26,53} Nonoperative treatment of PF typically involves taping and stretching exercises, whereas intraligamentary steroid injection, shockwave therapy, or custom orthoses may be warranted in refractory cases.⁴² More recently, platelet-rich plasma (PRP) injection has emerged as a newer, potential treatment option for chronic PF. PRP is an autologous blood product that is enriched for platelets



ex vivo. The mechanism of action underlying the healing properties of local PRP injection is related, at least in part, to the increased concentrations of growth factors and secretory proteins, which are thought to enhance the recruitment, proliferation, and differentiation of cells comprising regenerative tissues.¹⁵ Musculoskeletal applications of PRP have grown in the last decade, garnering much interest from the orthopaedic community.⁶¹ The clinical indications are extremely diverse, ranging from foot and ankle pathologies to spinal disorders and wound healing.¹¹ However, there is a great deal of controversy surrounding PRP, in large part driven by the sudden influx of clinical trials with conflicting findings.^{19,39}

The chronic pathology, lack of definitive treatment options, and high frequency of recurrence mean that PF fasciitis is effectively a lifelong physical disability for many patients. Management for symptomatic PF often continues for years, imposing a tremendous financial and personal burden onto patients.⁵⁶ Although not currently part of standard care for PF, PRP offers a potentially definitive, regenerative treatment modality that, if effective, could change the current paradigm of PF care. However, as with other applications of PRP, data from randomized controlled trials (RCTs) on the clinical benefits of PRP for the treatment of refractory plantar fasciitis is conflicting and offers inconsistent conclusions.^{5,16,21,27,43,49,55}

Even though RCTs are considered the highest level of original research available,⁴⁰ they have faced scrutiny in the last 2 decades because of a disturbingly frequent lack of reproducibility.^{9,20,24} Some researchers have even presented the possibility that the majority of research findings are significantly biased and, therefore, false.²³ This has been attributed in large part to the nearly ubiquitous threshold for declaring statistical significance, $P < .05$, which can lead to inaccurate or misleading portrayal of outcome importance.^{46,60} Indeed, a difference in outcomes with $P = .04$ is not practically very different from one with $P = .06$, although these 2 findings may be presented and interpreted with disparate levels of enthusiasm and gravity. Multiple solutions have been proposed to overcome these limitations, including lowering the threshold to $P < .005$ or eliminating it altogether.^{3,25,60} However, many of these solutions either do not directly quantify the robustness of statistically significant findings or cannot be easily deployed because of practical limitations, and orthopaedic surgeons' reliance on significant P values in RCT data continues to play a

major role in their clinical decision making.⁴ Therefore, orthopaedic surgeons are in serious need of a simpler, better method to appraise RCT findings, especially on the topic of PRP for chronic PF.

More recently, the fragility index (FI) was proposed as an adjunct to P values in RCTs.⁵⁹ This metric was first described by Feinstein¹² as a measure to assess the reliance of statistical significance on seemingly unimportant quantitative differences in outcomes. The FI is defined as the number of outcome event reversals necessary to alter significance based on a given significance level, typically $\alpha = 0.05$.⁵⁹ The associated fragility quotient (FQ) normalizes this value by dividing the FI by the sample size.² By using the FI and FQ alongside P values, authors can provide readers with discrete measures of how robust statistically significant findings truly are. Until recently, however, a major limitation of the FI was that it could only be applied to dichotomous outcomes.⁵⁹ In 2021, Caldwell et al⁷ described a method to calculate a Continuous Fragility Index (CFI) and Fragility Quotient (CFQ) for continuous outcomes based on the same principles as Feinstein's metric. The CFI is defined as the minimum number of patients whose intervention must change (moved from experimental to control arm, or vice versa) to alter study significance.⁷

The FI and FQ have already been used by various fragility analyses evaluating statistical stability of RCTs in orthopaedic surgery.^{10,13,14,30,31,37,48} To our knowledge, there has been no such fragility analysis for peer-reviewed RCTs involving PRP injections in the treatment of PF. The aim of the present investigation was to systematically identify RCTs reporting outcomes on the utilization of PRP for chronic PF and evaluate their statistical robustness by applying the CFI and CFQ. We hypothesize that, commensurate with other similar studies in orthopaedic surgery, reported measures of statistical significance will be easily overturned. Furthermore, we sought to contribute to the growing literature for the relatively new CFI and CFQ, because sufficient benchmark data using this new metric are currently lacking.

Material and Methods

Study Selection

We conducted a systematic search in accordance with the Preferred Reporting Items and Meta-Analyses (PRISMA) guidelines.⁴¹ Six online databases were searched: PubMed,

¹Department of Orthopaedic Surgery, The Johns Hopkins University School of Medicine, Baltimore, MD, USA

²Department of Orthopaedic Surgery, Rutgers New Jersey Medical School, Newark, NJ, USA

³Department of Orthopaedic Surgery, University of Puerto Rico Medical Sciences Campus, San Juan, Puerto Rico

⁴Department of Orthopaedic Surgery, Columbia University Mount Sinai Medical Center, Miami, FL, USA

Corresponding Author:

Amiethab A. Aiyer, MD, Department of Orthopaedic Surgery, The Johns Hopkins Hospital, 1800 Orleans St, Baltimore, MD 21287, USA.
Email: aaiyer2@jhmi.edu

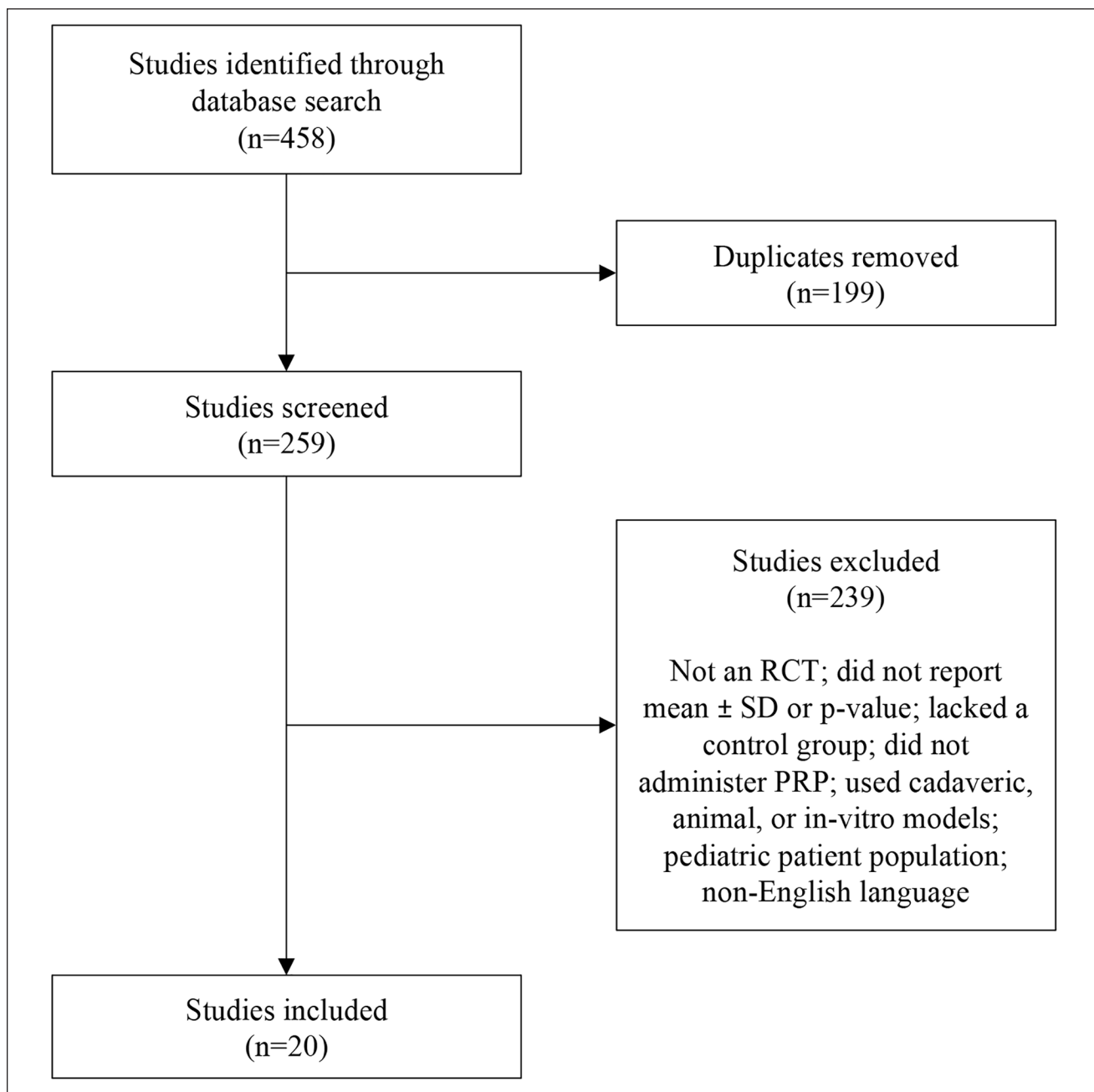


Figure 1. Preferred Reporting Items and Meta-Analyses (PRISMA) flow diagram for studies reporting on platelet-rich plasma in plantar fasciitis.

Embase, Cochrane, Web of Science, Scopus, and Clinicaltrials.gov. Standardized Medical Subject Headings and Emtree terms were used with keywords to identify RCTs reporting PRP therapy for chronic PF. Exclusion criteria for this study were incorrect study design; lacking a control group; use of cadaveric, animal, or in vitro models; patient age <18 years; or published in a non-English language. Articles that did not report data as mean \pm SD or did not report *P* values were also excluded.

After duplicate removal from 458 studies, 2 independent reviewers performed initial screening of 259 extracted articles using titles and abstracts, as well as subsequent full-text review of 75 articles, to identify studies that met all inclusion criteria. During both initial screening and full-text review, a third reviewer acted as a tiebreaker to resolve disagreements. Ultimately, 20 articles reporting results from RCTs were included in this systematic review and fragility analysis (Figure 1).

Table 1. Summary Characteristics of Included Studies.

First Author	Year	Control Treatment	FU Duration, wk	Sample Size, n	Total Lost FU, n	Outcome Events, n	CFI, Median (IQR)	CFQ, Median (IQR)
Gogna	2016	LDRT	24	40	NR	3	5 (5-7)	0.125 (0.125-0.175)
Gonnade	2018	Phon+KT	24	54	10	3	21 (18.5-21)	0.389 (0.343-0.389)
Shafaat	2020	Steroid	24	120	0	2	30 (29-31)	0.25 (0.242-0.258)
Acosta-Olivo	2017	Steroid	16	28	4	2	4 (3.5-4.5)	0.143 (0.125-0.161)
Goel	2021	ESWT	24	60	NR	4	10.5 (7.75-12)	0.175 (0.129-0.2)
Haddad	2021	ESWT	24	104	1	1	30 (N/A)	0.288 (N/A)
Jain	2015	Steroid	48	46	NR	3	2 (1.5-3)	0.043 (0.033-0.065)
Jain	2018	Steroid	24	80	NA	6	10 (10-12.25)	0.125 (0.125-0.153)
Khurana	2021	Steroid	24	118	29	2	33 (32-34)	0.28 (0.271-0.288)
Kim	2014	DP	24	20	1	3	4 (4-4)	0.2 (0.2-0.2)
Mahindra	2016	Steroid	12	50	NR	2	4 (3.5-4.5)	0.08 (0.07-0.09)
Malahias	2019	PPP	24	36	0	2	6 (5.5-6.5)	0.167 (0.153-0.181)
Omar	2012	Steroid	6	30	NR	2	4.5 (4.25-4.75)	0.15 (0.142-0.158)
Peerbooms	2019	Steroid	52	82	33	4	6.5 (3.75-9.5)	0.079 (0.046-0.116)
Sherpy	2016	Steroid	12	50	NR	2	10 (9.5-10.5)	0.2 (0.19-0.21)
Tabrizi	2020	Steroid	24	31	1	3	5 (3-7)	0.161 (0.097-0.226)
Tiwari	2013	Steroid	24	60	NR	1	10 (N/A)	0.167 (N/A)
Uğurlar	2018	ESWT, DP, steroid	144	79	0	9	15 (14-16)	0.19 (0.179-0.203)
Breton	2022	Steroid	24	42	12	2	4 (3.5-4.5)	0.102 (0.086-0.117)
Vahdatpour	2016	AWB	12	34	0	3	7 (4.5-7.5)	0.206 (0.132-0.221)

Abbreviations: AWB, autologous whole blood; CFI, Continuous Fragility Index; CFQ, Continuous Fragility Quotient; DP, dextrose prolotherapy; ESWT, extracorporeal shockwave therapy; FU, follow-up; IQR, interquartile range; KT, kinesiotaping; LDRT, low-dose radiation therapy; NR, not reported; Phon, phonophoresis; PPP, platelet-poor plasma.

Data Extraction

From each of the 20 included studies, the following variables were extracted: control treatment, duration of follow-up, and all primary and secondary outcomes. All extracted outcome measures were graded on continuous scales, and none were dichotomous. In their fragility analysis for continuous outcomes, Caldwell et al⁷ validated a method to create representative, synthetic data sets using mean \pm SD and sample size, eliminating the need for raw data collection in large meta- or fragility analyses. In brief, summary statistics (sample size n , mean, and SD) are input to create a simulated candidate data set for each arm consisting of a normally distributed list of n random numbers. If the mean and SD of the candidate data set are not within a specified tolerated range of the input values, the data set is rejected, and the process repeats with a new candidate data set. These methods were applied to the present analysis. For each outcome event, the following data were collected for both the control and PRP groups: mean \pm SD, sample size, and number of patients lost to follow-up. The original reported P value was also recorded for analyses comparing the control and intervention arms. The included studies and all extracted data, including first author and publication year, are presented in Table 1.

Calculation of Statistical Fragility

The Continuous Fragility Index (CFI) and Continuous Fragility Quotient (CFQ) were calculated, as recently proposed by Caldwell et al⁷, to analyze statistical fragility of continuous outcomes. This entails moving a single patient from one intervention arm to the other intervention arm within a simulation. In the case of outcomes with an initial $P < .05$, the patient is selected such that moving them would make the means of the 2 arms converge slightly. In the case of outcomes with an initial $P > .05$, the patient is selected such that moving them would make the means of the 2 arms diverge slightly. This process repeats until the P value flips across a specified alpha threshold ($\alpha = 0.05$).

The method to calculate CFI is fundamentally different from that of the traditional fragility index (FI) for dichotomous outcomes. The FI is defined as the number of patients whose outcome must change to alter significance, whereas the CFI is defined as the number of patients whose intervention must change to alter significance.^{7,59} The intuitive way to understand the FI is “the number of patients who, while undergoing the same treatment, hypothetically had a different outcome.” Analogously, the intuitive way to understand the CFI is “the number of patients who, while having the same outcome, hypothetically underwent

Table 2. Fragility Index and Quotient Data Based on Trial Characteristics.

Characteristic	Events	Patients, n	Lost FU, n	CFI, Median (IQR)	CFQ, Median (IQR)
All trials	59	3539	255	9 (4.5-14)	0.177 (0.125-0.203)
Outcome					
VAS pain	20	1243	83	9 (4.75-14.25)	0.172 (0.121-0.213)
VAS function	1	36	0	7 (N/A)	0.194 (N/A)
AOFAS	9	624	66	9 (5-10)	0.150 (0.100-0.225)
PF thickness	6	296	22	6 (5-10)	0.128 (0.097-0.198)
FFI disability	7	423	45	14 (4-16)	0.200 (0.114-0.203)
FFI activity	6	369	35	12.5 (6.5-14.75)	0.185 (0.166-0.197)
FADI	1	28	4	3 (3-3)	0.107 (0.107-0.107)
R&M	4	220	0	10 (6.5-12.25)	0.181 (0.133-0.209)
HTI	1	60	0	12 (12-12)	0.200 (0.200-0.200)
FAI core scale	2	160	0	10 (10-10)	0.125 (0.125-0.125)
FHSQ	2	80	0	6.5 (5.25-7.75)	0.157 (0.145-0.168)
Reported <i>P</i> value					
<i>P</i> < .05	20	1329	169	5 (3.75-14.5)	0.122 (0.062-0.241)
<i>P</i> > .05	39	2210	86	10 (5-14)	0.179 (0.133-0.201)

Abbreviations: AOFAS, American Orthopaedic Foot & Ankle Society score; CFI, Continuous Fragility Index; CFQ, Continuous Fragility Quotient; FADI, Foot and Ankle Disability Index; FAI, Foot and Ankle Instrument; FFI, Foot Function Index; FHSQ, Foot Health Status Questionnaire; FU, follow-up; HTI, Heel Tenderness Index; PF, plantar fascia; R&M, Roles and Maudsley scale; VAS, visual analog scale.

a different treatment.” The CFI simulation construct is modeled to increase linearly with sample size, increase logarithmically with mean difference, and decrease exponentially with SD.⁷ The corresponding CFQ for each outcome event was calculated as its CFI divided by the sample size. Additionally, we further expanded the CFI calculation to determine fragility for initially nonsignificant ($P > .05$) outcome events, using a similar approach to Caldwell et al’s and following the basic principles of fragility outlined in the seminal article by Feinstein.¹² CFI and CFQ values were reported as median and interquartile range (IQR). Data analysis was performed using R, version 3.6.1, software (The R Foundation for Statistical Computing, Vienna, Austria).

Results

A total of 20 studies met all inclusion criteria and were included in this analysis. The mean sample size was 52.7 patients (range: 20-120), with an average follow-up duration of 29.5 weeks (range: 6-144). Across the 12 studies that reported lost to follow-up data, the mean number of patients lost to final follow-up was 7.6 (range: 0-33). Control group treatments varied, with 13 studies using local steroid injection,^{1,5,27,28,32,35,43,45,49,50,52,55,57} 3 using extracorporeal shockwave therapy,^{16,21,57} 2 using dextrose prolotherapy,^{33,57} 1 using autologous whole blood,⁵⁸ 1 using low-dose radiation therapy,¹⁷ 1 using phonophoresis with kinesiotaping,¹⁸ and 1 using platelet-poor plasma.³⁶ Each study yielded an average of 2.95 outcome events (range: 1-9) suitable for analysis.

There were 59 total outcome events recorded across all included studies, of which 20 (33.9%) were originally reported as statistically significant and 39 (66.1%) were nonsignificant (Table 2). When grouped by outcome measure, the median CFI for all events was 9 (IQR: 4.5-14) and the median CFQ was 0.177 (IQR: 0.125-0.203). For initially significant events, the median CFI was 5 (IQR: 3.75-14.5), whereas initially nonsignificant events had a median CFI of 10 (IQR: 5-14). Across all outcome events, only 36 (61.0%) reported lost to follow-up data, with 10 events (27.8%) representing ≥ 9 patients lost. A detailed report of CFIs and CFQs grouped by clinical outcome can be found in Table 2.

Pain graded on a visual analog scale (VAS) was the most commonly reported outcome (18/20 studies), followed by American Orthopaedic Foot & Ankle Society (AOFAS) scores (9/20 studies), plantar fascia thickness (6/20 studies), and Foot Function Index (FFI) disability subscale scores (5/20 studies). Conflicting findings were documented for most of these categories. Eight studies reported improved pain after PRP injection,^{5,16,21,27,43,45,49,55} whereas 2 studies reported worse pain,^{32,52} and 8 studies reported no difference compared to control.^{17,18,28,33,35,36,57,58} For AOFAS scores, 5 studies reported greater improvement compared with control^{12,27,32,35,45,49} and 4 studies reported no difference,^{1,16,17,28} with none documenting lower scores in the PRP group. All 6 studies investigating plantar fascia thickness after treatment found no significant difference between the 2 intervention arms.^{5,17,18,28,50,58} Three of 5 studies found no difference in FFI disability subscale scores,^{18,33,57} whereas 1 reported better (lower) scores in

the PRP group⁴⁵ and 1 reported worse (higher) scores.⁵² Of note, the FFI disability category had the highest CFI and CFQ values, indicating a greater degree of statistical stability than other outcome measures. The Heel Tenderness Index category had an equivalent CFQ, although this represented data from a single study.¹⁶ The FFI activity category had the second highest CFI, but findings from 4 studies were still conflicting.^{33,45,52,57}

Discussion

Our systematic review and fragility analysis of RCTs involving PRP injection for refractory PF found that the overall CFI was 9, with an associated CFQ of 0.177. This suggests that, on average, changing the intervention of 9 patients would have been required to reverse trial significance. Notably, this is a key difference between the classic FI for dichotomous outcomes and the CFI for continuous outcomes. The FI is defined as the number of patients whose outcome must change to alter significance, whereas the CFI is defined as the number of patients whose intervention must change to alter significance.^{7,59} Furthermore, outcome events with significant P values ($P < .05$) actually suffered from greater statistical instability (CFI = 5, CFQ = 0.122) than events with nonsignificant P values (CFI = 10, CFQ = 0.179).

Fragility analyses have underscored the limitations of using P values, especially when reported without consideration for statistical robustness. Arbitrary α thresholds, inappropriate statistical methods, and variable sample sizes make misrepresentation of data a likely pitfall when using P values to declare significance.⁴⁷ Furthermore, P values themselves are inherently influenced by non-outcome-related factors such as effect size, sample size, and data dispersion.⁵⁹ Despite these recognized limitations, the proportion of abstracts and articles reporting P values has more than doubled from 1990 to 2015 in MEDLINE and PubMed Central (PMC) databases, with at least 1 P value $\leq .05$ appearing in 96% of studies.⁸ Evidence also suggests that significant P values heavily influence orthopaedic surgeons' perceptions of study importance and value.⁴ Interestingly, our data showed that outcomes events with significant P values ($P < .05$) actually suffered from greater instability than events with nonsignificant P values. Inclusion of 95% CIs, quartile spreads, or SDs alongside measures of average is one method to illustrate uncertainty. However, these measures do not reflect the clinical importance of observed differences between treatment arms, instead describing spread of the data. Effect size is an important measure that simply describes the difference in means between treatment arms. Although this does not directly quantify the robustness of statistically significant findings, effect size does provide key insight into clinical significance of observed differences between groups.⁵¹ It is

a common misconception that a lower P value equates to "more significant" differences, and therefore P values can be reported alone without mean values; however, the P value is heavily influenced by various other factors. For example, in database studies with a large enough sample size, even small, clinically meaningless differences between groups can yield a P value as low as $< .001$.²⁹ Effect size is therefore another important adjunct to P values, allowing readers to understand the magnitude of difference between groups and judge its clinical importance. Power calculations offer some insight into the robustness of statistically significant findings, although these often require complex calculations and cannot be performed for studies wherein "expected" values are not known. Thus, there is a need for a simple metric that describes how easily statistical significance can be overturned in terms of number of patients. Given the reliance of physicians on RCT data to make clinical decisions, we believe there is strong justification for including fragility indices alongside P values to better inform readers about the strength of statistical findings.

Previous fragility analyses appraising the statistical stability of RCT findings from various orthopaedic subspecialties have generated concerning results, with very low median FIs ranging from 2 to 5 and FQs ranging from 0.0323 to 0.092.^{10,13,14,30,31,37,48} However, the quantitative results from the present study cannot be directly compared to these values because the CFI and FI are inherently different in their derivation. This was demonstrated by Caldwell et al,⁷ who calculated the CFI based on summary statistics from a sports medicine fragility analysis by Khan et al³⁰ and found the novel CFI to be significantly higher ($P < .0001$) for continuous outcomes than the previously reported FI for dichotomous outcomes. To our knowledge, only 2 groups have written about the CFI and CFQ. Aside from Caldwell et al,⁷ who reported a mean CFI of 9 (IQR: 1.9-13.3) for sports medicine and arthroscopy literature, Ho et al²² retrospectively reported a CFI of 3 for a single RCT investigating vagal nerve electrical stimulation to enhance upper limb function after stroke. Additionally, our study is the first to include fragility analyses of continuous outcomes that were initially nonsignificant ($P > .05$), for which we are not aware of any other benchmark data. Nonetheless, our median CFI for only initially significant findings was 5 (IQR: 3.75-14.5), which was generally consistent with these 2 previous CFI articles. Although it is difficult to draw definitive conclusions in the absence of consensus guidelines regarding a threshold CFI value for strong vs weak evidence, the median CFI of 9 and median CFQ of 0.177 in our study suggest that although RCTs investigating PRP for PF are not impressively robust, their statistical fragility on average is at least comparable to that of the sports medicine literature.

The number of patients lost to follow-up may also help put fragility indices into context, as first suggested by Walsh

et al.⁵⁹ Although this representation is more intuitive for the dichotomous FI, important conclusions can be made with the CFI as well. In our fragility analysis, the mean number of patients lost to final follow-up was 7.6, although only 36 (61.0%) outcome events from 12 studies reported data on this. Of these, over a quarter (27.8%, n=10) lost a greater or equal number of patients to follow-up than would have needed to switch intervention arms to reverse event significance. This suggests that, given the fragility of RCTs reporting PRP utilization for chronic PF, loss to follow-up may be a major factor contributing to diminished statistical power and lower CFI. This is also true for the CFQ, which is normalized to sample size and would benefit from inclusion of more patients.

There is no clear consensus regarding the utility of PRP in treating chronic, refractory PF in adult patients. Conflicting findings were documented for most of the included outcome variables, with most studies reporting improved outcomes after PRP injection or no difference. Two studies reported worse pain,^{32,52} of which 1 also reported greater disability.⁵² The only consensus was among 6 studies investigating plantar fascia thickness, all of which found no difference between the 2 intervention arms.^{5,17,18,28,50,58} The FFI disability subscale, Heel Tenderness Index category, and FFI activity subscale had the highest CFI and CFQ values, indicating a greater degree of statistical stability compared to other outcome measures, although actual findings were still conflicting. The greater statistical stability of certain outcome measures may reflect inherently better reliability and/or validity compared to other questionnaires or metrics. The FFI in particular has been praised for excellent inter- and intraevaluator reliability, providing a quantifiable, patient-centered measure of foot health that is used extensively in clinical practice, even in multiple languages.^{6,38} Comparing the CFQs of different outcome measures in a multistudy fragility analysis may therefore provide insights into which metrics are most appropriate for a given clinical context. In this case, our findings suggest that the FFI may be a superior method for evaluating outcomes after plantar fasciitis treatment. There was also considerable variability in statistical fragility across studies, with CFQs ranging from 0.043 (4.3 per 100 patients) to 0.389 (38.9 per 100 patients). This likely reflects differences in methodology, including the presence of selection biases, variable PRP preparation or dosing/administration techniques, different control treatments, small sample sizes resulting in low power, lack of appropriate controls, or utilization of outcome measures with poor reliability/validity. These factors, among others, can have a considerable impact on a study's sensitivity for detection of clinically meaningful differences between treatment arms. The variability in results and statistical fragility across 20 RCTs suggest that no distinct conclusions can be drawn at this time about the benefits of PRP injection over current standard of care for chronic PF. Disagreement across studies also highlights the importance

of including fragility metrics to better distinguish studies with strong, clinically impactful findings. In our analysis, studies with the highest CFI and CFQ values, which offer more statistically robust evidence than others, may be more valuable for orthopaedic surgeons looking to evaluate the use of PRP in their clinical practice.

Limitations

There are several limitations to this study. Because of the challenges of acquiring raw data from a sufficient number of RCTs for meaningful fragility analysis, we used a previously validated method to generate synthetic data sets using reported summary statistics.⁷ Therefore, studies and outcome events for which mean \pm SD values were unavailable were excluded. Additionally, normal Gaussian distribution was assumed for all outcome measures reported as mean \pm SD in order to generate parametric data sets, although we recognize that some authors may report mean \pm SD for nonparametric data. For future studies, these 2 limitations can be overcome by collecting raw data. There was also considerable variability in follow-up duration across studies, which undermines the precise interpretation of the findings when the data sets are combined. Subsequent fragility analyses would benefit from restricting their inclusion criteria to studies with a specified follow-up period. Furthermore, included articles covered a limited range of conservative therapies, and we were not able to compare PRP to other common treatment modalities such as physical therapy.

The CFI and CFQ are not without limitations of their own. Currently, the CFI calculation algorithm utilizes Welch *t* test to iteratively measure significance, commensurate with the normally distributed synthetic data sets that were generated for analysis. However, some included studies originally reported *P* values from nonparametric statistical tests and cannot be used to calculate CFI. Additionally, at least 1 study adjusted its analyses to compensate for baseline differences. To account for this, we excluded any endpoint data if baseline values were significantly different and comparisons of mean change were not provided (ie, only baseline and endpoint values were provided). Furthermore, the CFI alone does not account for sample size, which is critical to interpretation of fragility metrics and should therefore be accompanied by the CFQ.^{2,44} One major limitation of conducting a fragility analysis is that every study does not contribute an equivalent number of outcome events. A study with many events would disproportionately influence the overall mean/median CFI compared to a study with few events. One way to potentially control for this may be to normalize the CFIs and CFQs for each study to the number of outcome events contributed by that study, although it is unclear how this would impact interpretation of results. Finally, there are currently very few CFI studies for comparison. Like the FI, the CFI is also an arbitrary

value, with no consensus on what constitutes a “strong” vs “weak” study.⁵⁹ However, the absence of a discrete threshold for declaring study strength may be warranted with fragility indices, leaving interpretation and judgment to the reader’s discretion. We hope that adoption of these newer continuous fragility metrics will contribute to a growing body of literature, which may help create context for the CFI and CFQ.

It is also worth noting that fragility analyses cannot compensate for inherent weaknesses in study design. For example, selection biases, potentially confounding variables, and lack of appropriate controls will inherently diminish the quality of evidence produced by a study. In meta-analyses, combining data from studies that are not fundamentally comparable can lead to misleading conclusions. Publication biases also influence the type of research that ultimately reaches broader audiences; often certain “hot topics” are met with seasonal enthusiasm that makes them more likely to be published in high-impact journals. Such factors cannot be corrected using any statistical method. Fragility indices are simply meant to provide an adjunct to the commonly used *P* value (much like the 95% CI provides an adjunct to the mean), allowing readers to understand how easily a significant finding may be overturned.

Conclusions

Given the importance of RCTs for guiding clinical practice, fragility indices can provide valuable information about the robustness of statistical analyses and give context to the commonly utilized *P* value. However, fragility analyses must be interpreted alongside the innate strengths, weaknesses, and clinical relevance of each individual study. Based on the wide range of CFI and CFQ values across RCTs in this systematic review, we conclude that while all studies investigating PRP for PF are not impressively robust, their statistical fragility is at least comparable to that of the sports medicine literature. At least in the absence of consensus guidelines, CFIs <9 may be considered weak, whereas CFIs >9 may be considered strong. Increased adoption of the newer continuous fragility metrics may help create further context for the simulated CFI and CFQ. Practicing orthopaedic surgeons should consider fragility metrics as a potential tool to help interpret the clinical importance of RCT findings.

Ethical Approval

Ethical approval was not sought for the present study because it was a systematic review and meta-analysis of previously published data.


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. ICMJE forms for all authors are available online.

Funding

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

ORCID iDs

Arjun Gupta, BS,  <https://orcid.org/0000-0002-7062-7584>

Carlos Ortiz-Babilonia, BS,  <https://orcid.org/0000-0002-9983-3365>

Amy L. Xu, BS,  <https://orcid.org/0000-0002-3651-2361>

References

- Acosta-Olivo C, Elizondo-Rodriguez J, Lopez-Cavazos R, Vilchez-Cavazos F, Simental-Mendia M, Mendoza-Lemus O. Plantar fasciitis—a comparison of treatment with intral-lesional steroids versus platelet-rich plasma (a randomized, blinded study). *J Am Podiatr Med Assoc.* 2017;107(6):490-496. doi: 10.7547/15-125
- Ahmed W, Fowler RA, McCredie VA. Does sample size matter when interpreting the fragility index? *Crit Care Med.* 2016;44(11):e1142-e1143. doi: 10.1097/ccm.0000000000001976
- Benjamin DJ, Berger JO, Johannesson M, et al. Redefine statistical significance. *Nat Hum Behav.* 2018;2(1):6-10. doi: 10.1038/s41562-017-0189-z
- Bhandari M, Montori VM, Schemitsch EH. The undue influence of significant p-values on the perceived importance of study results. *Acta Orthop.* 2005;76(3):291-295.
- Breton A, Leplat C, Picot MC, et al. Prediction of clinical response to corticosteroid or platelet-rich plasma injection in plantar fasciitis with MRI: a prospective, randomized, double-blinded study. *Diagn Interv Imaging.* 2022;103(4):217-224. doi: 10.1016/j.diii.2021.10.008
- Budiman-Mak E, Conrad KJ, Mazza J, Stuck RM. A review of the Foot Function Index and the Foot Function Index – Revised. *J Foot Ankle Res.* 2013;6(1):5. doi: 10.1186/1757-1146-6-5
- Caldwell JE, Youssefzadeh K, Limpisvasti O. A method for calculating the fragility index of continuous outcomes. *J Clin Epidemiol.* 2021;136:20-25. doi: 10.1016/j.jclinepi.2021.02.023
- Chavalarias D, Wallach JD, Li AH, Ioannidis JP. Evolution of reporting P values in the biomedical literature, 1990-2015. *JAMA.* 2016;315(11):1141-8. doi: 10.1001/jama.2016.1952
- Colquhoun D. The reproducibility of research and the misinterpretation of p-values. *R Soc Open Sci.* 2017;4(12):171085.
- 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
- Everts P, Onishi K, Jayaram P, Lana JF, Mautner K. Platelet-rich plasma: new performance understandings and therapeutic considerations in 2020. *Int J Mol Sci.* 2020;21(20):7794. doi: 10.3390/ijms21207794
- 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

13. Forrester LA, Jang E, Lawson MM, Capi A, Tyler WK. Statistical fragility of surgical and procedural clinical trials in orthopaedic oncology. *J Am Acad Orthop Surg Glob Res Rev.* 2020;4(6):e19.00152. doi: 10.5435/JAAOSGlobal-D-19-00152
14. Forrester LA, McCormick KL, Bonsignore-Opp L, et al. Statistical fragility of surgical clinical trials in orthopaedic trauma. *J Am Acad Orthop Surg Glob Res Rev.* 2021;5(11):e20.00197. doi: 10.5435/JAAOSGlobal-D-20-00197
15. Foster TE, Puskas BL, Mandelbaum BR, Gerhardt MB, Rodeo SA. Platelet-rich plasma: from basic science to clinical applications. *Am J Sports Med.* 2009;37(11):2259-2272.
16. Goel N, Talwar J, Agarwal S, Krishna LG, Rustagi A. A comparative study between intralesional platelet rich plasma injection and extracorporeal shockwave therapy for the treatment of plantar fasciitis. *J Arthrosc Joint Surg.* 2021;8(3):246-252. doi: <https://doi.org/10.1016/j.jajs.2021.04.003>
17. Gogna P, Gaba S, Mukhopadhyay R, Gupta R, Rohilla R, Yadav L. Plantar fasciitis: a randomized comparative study of platelet rich plasma and low dose radiation in sportspersons. *Foot (Edinb)* 2016;28:16-19. doi: 10.1016/j.foot.2016.08.002
18. Gonnade N, Bajpayee A, Elhence A, et al. Regenerative efficacy of therapeutic quality platelet-rich plasma injections versus phonophoresis with kinesiotaping for the treatment of chronic plantar fasciitis: a prospective randomized pilot study. *Asian J Transfus Sci.* 2018;12(2):105-111. doi: 10.4103/ajts.AJTS_48_17
19. Gupta S, Paliczak A, Delgado D. Evidence-based indications of platelet-rich plasma therapy. *Expert Rev Hematol.* 2021;14(1):97-108. doi: 10.1080/17474086.2021.1860002
20. Hacke C, Nunan D. Discrepancies in meta-analyses answering the same clinical question were hard to explain: a meta-epidemiological study. *J Clin Epidemiol.* 2020;119:47-56.
21. Haddad S, Yavari P, Mozafari S, Farzinnia S, Mohammadsharifi G. Platelet-rich plasma or extracorporeal shockwave therapy for plantar fasciitis. *Int J Burns Trauma.* 2021;11(1):1-8.
22. Ho AK. The fragility index for assessing the robustness of the statistically significant results of experimental clinical studies. *J Gen Intern Med.* 2022;37(1):206-211.
23. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA.* 2005;294(2):218-228. doi: 10.1001/jama.294.2.218
24. Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q.* 2016;94(3):485-514.
25. Ioannidis JPA. The proposal to lower P value thresholds to .005. *JAMA.* 2018;319(14):1429-1430. doi: 10.1001/jama.2018.1536
26. Irving DB, Cook JL, Young MA, Menz HB. Impact of chronic plantar heel pain on health-related quality of life. *J Am Podiatr Med Assoc.* 2008;98(4):283-289. doi: 10.7547/0980283
27. Jain K, Murphy PN, Clough TM. Platelet rich plasma versus corticosteroid injection for plantar fasciitis: a comparative study. *Foot (Edinb).* 2015;25(4):235-237. doi: 10.1016/j.foot.2015.08.006
28. Jain SK, Suprashant K, Kumar S, Yadav A, Kearns SR. Comparison of plantar fasciitis injected with platelet-rich plasma vs corticosteroids. *Foot Ankle Int.* 2018;39(7):780-786. doi: 10.1177/1071100718762406
29. Kelley K, Preacher KJ. On effect size. *Psychol Methods.* 2012;17(2):137-152. doi: 10.1037/a0028086
30. 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
31. Khormae 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
32. Khurana A, Dhankhar V, Goel N, Gupta R, Goyal A. Comparison of midterm results of Platelet Rich Plasma (PRP) versus steroid for plantar fasciitis: a randomized control trial of 118 patients. *J Clin Orthop Trauma.* 2021;13:9-14. doi: 10.1016/j.jcot.2020.09.002
33. Kim E, Lee JH. Autologous platelet-rich plasma versus dextrose prolotherapy for the treatment of chronic recalcitrant plantar fasciitis. *PM R.* 2014;6(2):152-158. doi: 10.1016/j.pmrj.2013.07.003
34. Kowal P, Goodkind D, He W. *An Aging World: 2015, International Population Reports.* US Census Bureau; 2016.
35. Mahindra P, Yamin M, Selhi HS, Singla S, Soni A. Chronic plantar fasciitis: effect of platelet-rich plasma, corticosteroid, and placebo. *Orthopedics.* 2016;39(2):e285-e289. doi: 10.3928/01477447-20160222-01
36. Malahias MA, Mavrogenis AF, Nikolaou VS, et al. Similar effect of ultrasound-guided platelet-rich plasma versus platelet-poor plasma injections for chronic plantar fasciitis. *Foot (Edinb).* 2019;38:30-33. doi: 10.1016/j.foot.2018.11.003
37. Maldonado DR, Go CC, Huang BH, Domb BG. The fragility index of hip arthroscopy randomized controlled trials: a systematic survey. *Arthroscopy.* 2021;37(6):1983-1989. doi: 10.1016/j.arthro.2021.01.049
38. Martinez BR, Staboli IM, Kamonseki DH, Budiman-Mak E, Yi LC. Validity and reliability of the Foot Function Index (FFI) questionnaire Brazilian-Portuguese version. *Springerplus* 2016;5(1):1810. doi: 10.1186/s40064-016-3507-4
39. Martínez-Martínez A, Ruiz-Santiago F, García-Espinosa J. Platelet-rich plasma: myth or reality? *Radiologia (Engl Ed).* 2018;60(6):465-475. doi: 10.1016/j.rx.2018.08.006
40. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg.* 2012;10(1):28-55. doi: 10.1016/j.ijsu.2011.10.001
41. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4(1):1. doi: 10.1186/2046-4053-4-1
42. Morrissey D, Cotchett M, Said J'Barí A, et al. Management of plantar heel pain: a best practice guide informed by a systematic review, expert clinical reasoning and patient values. *Br J Sports Med.* 2021;55(19):1106-1118. doi: 10.1136/bjsports-2019-101970
43. Omar AS, Ibrahim ME, Ahmed AS, Said M. Local injection of autologous platelet rich plasma and corticosteroid in

- treatment of lateral epicondylitis and plantar fasciitis: randomized clinical trial. *Egypt Rheumatol*. 2012;34(2):43-49. doi: <https://doi.org/10.1016/j.ejr.2011.12.001>
44. Parisien RL, Dashe J, Cronin PK, Bhandari M, Tornetta P 3rd. Statistical significance in trauma research: too unstable to trust? *J Orthop Trauma*. 2019;33(12):e466-e470. doi: 10.1097/bot.0000000000001595
 45. Peerbooms JC, Lodder P, den Ouden BL, Doorgeest K, Schuller HM, Gosens T. Positive effect of platelet-rich plasma on pain in plantar fasciitis: a double-blind multicenter randomized controlled trial. *Am J Sports Med*. 2019;47(13):3238-3246. doi: 10.1177/0363546519877181
 46. Peng R. The reproducibility crisis in science: a statistical counterattack. *Significance*. 2015;12(3):30-32. doi: <https://doi.org/10.1111/j.1740-9713.2015.00827.x>
 47. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: "P" values, statistical significance and confidence intervals. *Perspect Clin Res*. 2015;6(2):116-117. doi: 10.4103/2229-3485.154016
 48. Ruzbarsky JJ, Khormae S, Daluiski A. The fragility index in hand surgery randomized controlled trials. *J Hand Surg Am*. 2019;44(8):698.e1-698.e7. doi: 10.1016/j.jhsa.2018.10.005
 49. Shafaat H, Aziz T, Butt A, Iqbal S, Aziz U, Aziz T. Comparison of platelet rich plasma with local steroid injection in the management of chronic plantar fasciitis. *Pakistan Armed Forces Med J*. 2020;70(2):442-446.
 50. Sherpy NA, Hammad MA, Hagrass HA, Samir H, Abu-ElMaaty SE, Mortada MA. Local injection of autologous platelet rich plasma compared to corticosteroid treatment of chronic plantar fasciitis patients: a clinical and ultrasonographic follow-up study. *Egypt Rheumatol*. 2016;38(3):247-252. doi: <https://doi.org/10.1016/j.ejr.2015.09.008>
 51. Sullivan GM, Feinn R. Using effect size—or why the P value is not enough. *J Grad Med Educ*. 2012;4(3):279-282. doi: 10.4300/jgme-d-12-00156.1
 52. Tabrizi A, Dindarian S, Mohammadi S. The effect of corticosteroid local injection versus platelet-rich plasma for the treatment of plantar fasciitis in obese patients: a single-blind, randomized clinical trial. *J Foot Ankle Surg*. 2020;59(1):64-68. doi: 10.1053/j.jfas.2019.07.004
 53. Taunton JE, Ryan MB, Clement DB, McKenzie DC, Lloyd-Smith DR, Zumbo BD. A retrospective case-control analysis of 2002 running injuries. *Br J Sports Med*. 2002;36(2):95-101. doi: 10.1136/bjism.36.2.95
 54. Thomas MJ, Whittle R, Menz HB, Rathod-Mistry T, Marshall M, Roddy E. Plantar heel pain in middle-aged and older adults: population prevalence, associations with health status and lifestyle factors, and frequency of healthcare use. *BMC Musculoskelet Disord*. 2019;20(1):337. doi: 10.1186/s12891-019-2718-6
 55. Tiwari M, Bhargava R. Platelet rich plasma therapy: a comparative effective therapy with promising results in plantar fasciitis. *J Clin Orthop Trauma*. 2013;4(1):31-35. doi: 10.1016/j.jcot.2013.01.008
 56. Tong KB, Furia J. Economic burden of plantar fasciitis treatment in the United States. *Am J Orthop (Belle Mead NJ)*. 2010;39(5):227-231.
 57. Uğurlar M, Sönmez MM, Uğurlar Ö Y, Adıyeye L, Yıldırım H, Eren OT. Effectiveness of four different treatment modalities in the treatment of chronic plantar fasciitis during a 36-month follow-up period: a randomized controlled trial. *J Foot Ankle Surg*. 2018;57(5):913-918. doi: 10.1053/j.jfas.2018.03.017
 58. Vahdatpour B, Kianimehr L, Ahrar MH. Autologous platelet-rich plasma compared with whole blood for the treatment of chronic plantar fasciitis; a comparative clinical trial. *Adv Biomed Res*. 2016;5:84-84. doi: 10.4103/2277-9175.182215
 59. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. *J Clin Epidemiol* 2014;67(6):622-628. doi: 10.1016/j.jclinepi.2013.10.019
 60. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat*. 2016;70(2):129-133. doi: 10.1080/00031305.2016.1154108
 61. Wu PI, Diaz R, Borg-Stein J. Platelet-rich plasma. *Phys Med Rehabil Clin N Am*. 2016;27(4):825-853. doi: 10.1016/j.pmr.2016.06.002