



Statistical Significance and Sports Medicine Trials.



CENTER FOR HEALTH SCIENCES

Jake Xavier Checketts B.S.¹, Austin Johnson B.S.¹, Jonathan Riddle B.S.¹, Sheridan Evans BS¹, Jared Scott BS^{1,2} Keith Fishbeck DO², Marshall Boose DO², Jake Whitener Do^e, Jaclyn Jones DO², Clint Basener DO², Chad Hanson DO², Matthew Vassar PhD¹.
 1. Oklahoma State University College of Osteopathic Medicine 2. OSU-CHS Orthopaedic Surgery Residency Program

INTRODUCTION

Lowering the threshold for statistical significance in medical research from a *P* value of .05 to .005 was recently proposed to reduce misinterpretation of study results.^{1,2} *P* values less than .05 but greater than .005 would be reclassified as “suggestive”, whereas *P* values less than .005 would be considered significant. What effect this proposal would have on orthopaedic sports medicine literature is currently unclear. We evaluated primary endpoints in randomized clinical trials (RCTs) published in the 3 highest ranked orthopaedic sports medicine journals, as determined by Google Scholar’s H-5 index, to determine how the newly proposed threshold could affect the interpretation of previously published sports medicine RCTs.

METHODS

We searched PubMed from January 01, 2016 to December 31, 2017 for RCTs published in the *American Journal of Sports Medicine*, *Arthroscopy*, and *Knee Surgery, Sports Traumatology, Arthroscopy*. All RCTs were screened by at least 2 authors. We extracted *P* value data for primary endpoints, since RCTs are most often powered for these endpoints. If a study had multiple primary endpoints, or evaluated the primary endpoint from multiple domains, all *P* values for these endpoints were include. Data were extracted blinded and in duplicate fashion. Discrepancies were resolved by consensus. We used Google Forms for data extraction and STATA 13.1 for the data analysis.

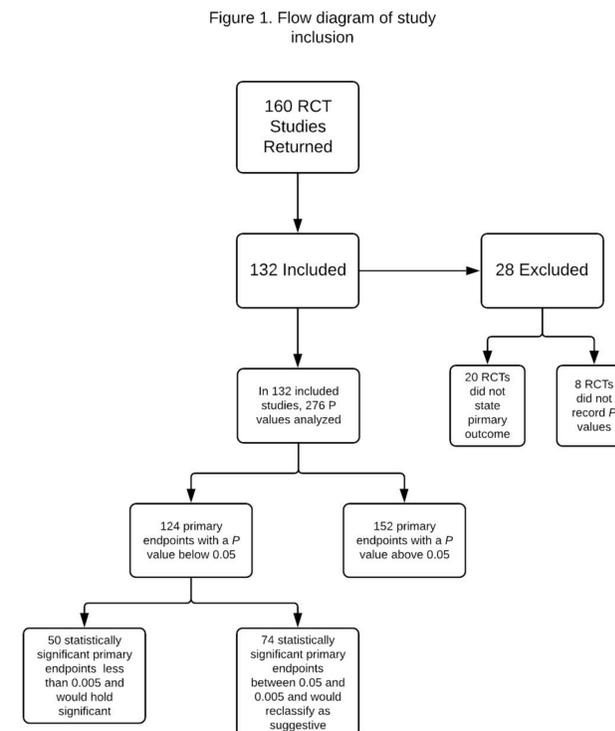
Table 1. Characteristics of included clinical trials (n = 132) or endpoints (n = 276).

Characteristic	No. (%)
Journal (n = 132)	
<i>American Journal of Sports Medicine</i>	43 (32.6%)
<i>Arthroscopy</i>	21 (15.9%)
<i>Knee Surgery, Sports Traumatology, Arthroscopy</i>	68 (51.5%)
Intervention (n = 132)	
Drug	
Procedure	32 (24.2%)
Anesthesia/Analgesia (Nerve Blocks/Pain Management)	15 (11.4%)
Surgery	65 (49.2%)
Education	3 (2.3%)
Other	17 (12.9%)
*Funding Source (n = 132)	
Industry	16 (12.1%)
Public	14 (10.6%)
Private	5 (3.8%)
Hospital	3 (2.3%)
Mixed	12 (9.1%)
Not Mentioned	75 (56.8%)
None	7 (5.3%)
Number of trial centers (n = 132)	
Multicenter	4 (3.0%)
Single center	128 (97.0%)
Location (n = 132)	
Multinational	1 (0.8%)
Single country	131 (99.2%)
Type of endpoint (n = 132)	
Subjective	85 (64.4%)
Objective	47 (35.6%)
Sample size (median, [IQR])	
	64 [46.5-101]

*In some of the trials, multiple funding sources were identified.

Table 2. Statistical characteristics of included clinical trials (n = 151)

Effect from lowering the statistical significance threshold	No. %
Primary Endpoints (n = 276)	
Reclassified as Significant (<.005)	50 (18%)
Reclassified as Suggestive (>0.005 - <0.05)	74 (26.8%)
Remain Non-significant	152 (55.07%)



RESULTS

Our final sample consisted of 132 included studies with 276 primary endpoints recorded. Of the endpoints included in the studies, only 13.6% (18/132) of the studies have endpoints in which all *P* values are below the new threshold of .005. 14.4% (19/132) of the studies have endpoints in which some would meet the new *P* value threshold of .005, and some would not meet this new threshold and 71.9% (95/132) of the studies have no endpoints in which the *P* value(s) was less than .005. Within the 160 studies, we identified individual endpoints with reported *P* value data. 1/9% endpoints had a *P* value less than .05 and 193 (55%) had a *P* value greater than .05. Overall, 40.3% (50/124) of the previously statistically significant primary endpoints were less than .005, while 59.7% (74/124) would be reclassified as suggestive. The most common type of endpoints were outcome scores, with 50% (66/132) of the studies using this method of assessment. Surgery was the primary intervention type for 49.2% (65/132) of the included trials. The majority of trials did not mention any funding source (56.8% 75/132). The majority of included trials, (125/132), had randomized groups.

CONCLUSION

Of statistically significant endpoints RCTs published in the 3 highest impact orthopaedic sports medicine journals, only 17% (59/350) would maintain their statistical significance with a *P* value threshold of less than .005, and only 8% of studies would maintain their overall significance with all *P* values falling below the new threshold. A .005 threshold for significance may address the shortcomings of *P* values, such as underpowered RCTs³, spurious false positive results⁴, and *P* hacking (when researchers analyze data multiple ways until a significant effect is found)⁵. However, as our results indicate, doing so would heavily alter the significance of orthopaedic sports medicine RCTs and this proposed threshold should be further evaluated, and cautiously interpreted.