Evaluation of Effect of Lowering the Threshold of Statistical Significance From 0.05 to 0.005 in Foot and Ankle Randomized Controlled Trials

Yoshiharu Shimozono, Noboru Funakoshi¹, Masahiko Kobayashi², Shuichi Matsuda³

¹JAPAN, ²Kyoto Shimogamo Hospital, ³Grad Sch of Med, Kyoto Univ/Dept Ortho Surg

INTRODUCTION: Although P-values have long been the gold standard for determining statistical significance in evidencebased medicine, the importance of P-values are often overemphasized in the medical literature. Recently, it has been suggested that suggested shifting the p value threshold from the commonly accepted 0.05 convention to 0.005, and Pvalues between 0.05 and 0.005 would be labeled "suggestive". If this new threshold were accepted, only P-values less than 0.005 would be considered statistically significant. By increasing the stringency of statistical significance, the risk of false positives decreases across the medical literature. The purpose of this study was to determine how changing the Pvalue threshold of statistical significance from 0.05 to 0.005 could affect the statistical significance of findings in previously published foot and ankle randomized controlled trials (RCTs).

METHODS:

The authors searched PubMed from 2019 to 2024 for RCTs published in the top 10 ranked in orthopedic journals based on impact factors, only for foot and ankle. The top three foot and ankle related journals were also included. The authors then extracted P-value data for primary endpoints because 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 included.

RESULTS:

A total of 281 primary endpoints were identified in 123 trials. Analysis of primary endpoints found that 44.1% (124/281) had a p-value less than 0.05, and were classified as statistically significant under the current threshold, whereas 55.9% (157/281) had a p-value greater than 0.05. Of those endpoints that were previously considered statistically significant, 41.9% (52/124) were less than 0.005, whereas 58.1% (72/124) were between 0.005 and 0.05; therefore, they would be reclassified as suggestive rather than statistically significant under the proposed threshold. Overall, when analyzing the 281 primary endpoints, we found that only 52 (18.5%) endpoints were less than 0.005 and would hold statistical significance with the proposed threshold. The protocol changes also caused only 31 studies in the included 123 trials to have statistically significant primary endpoints.

DISCUSSION AND CONCLUSION:

Of the endpoints currently considered statistically significant, 58% are no longer statistically significant when the new P-value protocol is used. Our results suggest that changing the threshold for statistical significance from 0.05 to 0.005 would significantly alter the significance of the published foot and ankle RCTs. Because P-values from RCTs can often influence the ways physicians choose interventions, it is important to implement a methodology that decreases the incidence of bias and misrepresentation of these results. However, lowering the P-value could increase the required sample size and, consequently, increase study costs. It is recommended that this proposed threshold be further evaluated and interpreted cautiously.