Confidence Crisis of Results in Biomechanics Research

Duane Knudson, Ph.D.

Department of Health and Human Performance

601 University Drive

Texas State University

San Marcos, TX 87666 USA

512-245-2561

dknudson@txstate.edu

This is an accepted manuscript of an article published by Taylor & Francis in *Sports Biomechanics* 16, 425-433 on September 5, 2017 available online:

http://wwww.tandfonline.com/doi:10.1080/14763141.2016.1246603

Abstract

Many biomechanics studies have small sample sizes and incorrect statistical analyses, so reporting of inaccurate inferences and inflated magnitude of effects are common in the field. This review examines these issues in biomechanics research and summarizes potential solutions from research in other fields to increase the confidence in the experimental effects reported in biomechanics. Authors, reviewers, and editors of biomechanics research reports are encouraged to improve sample sizes and the resulting statistical power, improve reporting transparency, improve the rigor of statistical analyses used, and increase the acceptance of replication studies to improve the validity of inferences from data in biomechanics research. The application of sports biomechanics research results would also improve if a larger percentage of unbiased effects and their uncertainty were reported in the literature.

Key words: effect size, error, false positive, power, replicability

Introduction

Original research, peer-review, and replication in science are supposed to systematically advance human knowledge. This advancement in scientific knowledge is a complex, irregular phenomenon given unexpected discoveries, conflicting results, new paradigms, and technological advancements that all must be mediated through the social system of peer review in a field. Recently there has been an increasing 'crisis of confidence' in the accuracy and robustness of experimental results in a variety of scientific fields due to bias from errors in experimental design and statistical analysis (Begley & Ellis, 2012; Buchanan & Lohse, 2016; Button et al., 2013; Hoekstra, Morey, Rouder, & Wagenmakers, 2014; Ioannidis, 2005, 2012; Pashler & Harris, 2012). These problems contribute to uncertainty in the true effect and the magnitude of effect in research reports. The robustness or consistency of effects across similar studies are a problem in some scientific fields because of the rarity of replication studies to support and refine the magnitude of effects (Klein et al., 2014; Makel, Plucker, & Hegarty, 2012).

To what extent are these problems present in biomechanics research? In this review I examine the errors in sampling and statistical analyses in biomechanics research, as well as summarize solutions recently proposed to increase the confidence in inferences about experimental effects reported in the field.

Sample Sizes and Effect Sizes in Biomechanics

For results of scientific research to be robust and representative of a population of interest, sampling procedures and sample sizes need to be adequate. Problems in defining and obtaining adequate sampling include access to participants, cost, obtaining informed consent, unclear expectations for variability that must be assumed for sample size calculations, and adjusting sample size calculations for complex and multiple statistical tests on dependent variables (Batterham & Atkinson, 2005). Consequently, pressures to minimize cost of research often result in studies with small sample sizes that introduce a number of biases including low statistical power, inconsistent results (poor replicability), and inflated effect sizes (Button et al., 2013; Ioannidis, 2005, 2008; Maxwell, 2004; Schweizer & Furley, 2014). Small sample sizes are the primary contributor to the large 'false positive' rate in biomedical and psychological research, which is likely 10 times higher than the $p < 0.05$ incorrectly assumed by many researchers (Ioannidis, 2005; Simmons, Nelson, & Simonsohn, 2011). Studies with small samples and their related problems are also common in biomechanics.

Several studies have reported typical sample sizes in biomechanics research (Knudson, 2011, 2012; Knudson & Bahamonde, 2012). Knudson (2011) reported the number of authors and sample sizes from the 2009 volumes of three applied biomechanics journals. Mean sample sizes ranged between 15 and 42, with median sample sizes between 12 and 18 for these three journals. These typical, small sample sizes in biomechanics research reports have not changed over the last 20 to 25 years (Knudson, 2012; Knudson and Bahamonde, 2012).

Fraley and Vazire (2014) found the median sample sizes (70-180) of reports in top psychology journals resulted in low statistical power (50%) to detect typical effect sizes ($d = 0.4$) in that field. Given the typical samples sizes of biomechanics research are substantially lower than in psychology, it is likely that most biomechanics research has even lower statistical power and biased effect sizes than in psychology.

This inflation of the size of effects in biomechanics from small sample sizes can be seen in the effect sizes reported in the literature. Figure 1 illustrates the distribution of 130 effect sizes of statistically significant overall effects reported in ten recent meta-analyses using biomechanical data (Barton, Levinger, Menz, & Webster, 2009; Bogaerts, Desmet, Slagmolen, & Peers, 2016; Brown, Brughelli, & Hume, 2014; Cheung, Chung, & Ng, 2011; Fernando et al., 2013; Fransz, Huurnink, Kingma, Verhagen, & van Dieën, 2013; Hall, Barton, Remy, & Morrissey 2013; Mills, Hunt, Leigh, & Ferber, 2013; Napier, Cochrane, Taunton, & Hunt, 2015; Wicke, Keeley, & Oliver, 2013). The median effect size for 'significant' effects in these biomechanics research studies (0.7) was between a medium and a large effect (Knudson, 2009). Less than 10% of the reported effect sizes were at a generous definition (0.3) of a small effect. One would logically expect more true small effect sizes, even though these extracted data do not include overall nonsignificant effects in these recent meta-analyses. This is especially true since these meta-analyses are what could be described as best case scenarios in the field, were there is greater external funding examining major health problems. More inflated effect sizes may be more likely in sports biomechanics were many studies may only have funds and access to small samples of athletes.

One reason for the dearth of small effects in most biomedical sciences like biomechanics is the 'publication bias' or 'file drawer' effect. Scientific journal editorial policies tend to be biased against accepting reports with nonsignificant effects (Csada, James, & Espie, 1996), so there are likely many unpublished studies relegated to researcher's file drawers. Current publication and promotion incentives encourage researchers to 'salami slice' or submit 'smallest publishable unit' articles with small sample sizes which are more likely to have inflated effect sizes (Button et al., 2013). The small sample sizes used in biomechanics has likely inflated the magnitude of the statistically significant effects reported in the field, which we will see in the next section is made worse by errors in statistical analysis that inflate the reporting of ‘false positive’ effects.

Errors in Statistical Analysis in Biomechanics

Several authors have pointed out consistent errors in the statistical analyses used in many biomechanics research reports. Many of these scholars have emphasized the high rate of ‘false positives’ likely in the biomechanics research results because of numerous statistical tests of dependent variables from the same data set that do not correct for inflation of the experiment-wise type I error rate (James & Bates, 1997; Knudson, 2005, 2009; Morris, 1981; Mullineaux, Bartlett, & Bennett, 2001). Knudson (2005) reported a majority (73 to 81%) of applied biomechanics original research reports used multiple uncorrected statistical tests inflating the experiment-wise type I error rate. These fishing expeditions with uncorrected multiple statistical tests can even identify physically meaningless or illogical variables as having statistically significant effects (Austin, Mamdani, Huurlink, & Hux, 2006; Knudson 2007).

Recent research has also shown that extracting data from biomechanical curves, which naturally have a relationship with other points on the curve, for statistical testing also inflates the ‘false positives’ observed with traditional null-hypothesis significance testing (Pataky, Vanrenterghem, & Robinson; 2015, 2016). Unfortunately, small samples and errors in statistical testing also bias attempts at magnitude-based inference testing

(Gelman & Carlin, 2014), although recent research indicates this bias may be less than the bias in null-hypothesis significance testing (Hopkins & Batterham, 2016).

Perhaps the most pernicious statistical errors reside in the gray area between research design/sample size and statistical analyses. Flexibility in data collection and statistical analysis, the so called ‘researcher degrees of freedom’ also contribute to the high rate of ‘false positive’ effects in the scientific literature (Simmons et al., 2011). One common example is ‘p-hacking,’ the progressive statistical testing of results by researchers to determine when to terminate sampling/data collection. A recent survey of health and exercise scientists reported a significant bias of scholars towards ‘p-hacking’ when presented with scenarios to collect additional data in the hope of reaching an observed p value less than 0.05 (Buchanan & Lohse, 2016). Simmons and coworkers (2011) simulations show that researchers who statistically test two moderately correlated dependent variables double the false positive findings detected. They also show that researchers who perform preliminary statistical testing of sample sizes of 20 and detecting a significant effect, increase false positive findings 50% compared to studies using tests of an additional 10 samples. Biomechanics researchers testing small samples that reach statistical significance would also be more likely to report inflated and erroneous effects (Button et al., 2013) because of the greater influence of outlier observations in small samples relative to larger sample sizes.

These errors in sampling and statistical analysis in a large percentage of biomechanics research represent a threat to the internal validity or accuracy of the results and knowledge generation in the field. The problem of untrue and biased effects has persisted in the scientific literature and biomechanics because some scholars do not keep up with recent advances in research design and statistics, the low rates of replication of research results, (Ioannidis, 2005, 2012; Nosek, Spies, & Motyl, 2012; Pashler & Harris, 2012), and weaknesses and inconsistency in peer-review (Knudson, Morrow, & Thomas, 2014).

Proposed Solutions

Fortunately for researchers interested in improving the accuracy of the size of experimental effects and inferences in biomechanics, research and recommendations from other fields on these problems can be implemented. The following four solutions are based on this research and recent recommendations (Button et al., 2013; Nosek et al., 2012; Simmons et al., 2011). It is the collective responsibility of authors, peer reviewers, and journal editors in to strive to improve the accuracy of the accumulated findings of published biomechanics research. These recommendations should also be included in graduate training in biomechanics.

Improve Sample Size and Statistical Power

Authors of biomechanics research should establish adequate sample size/termination of data collection *a priori*. Researchers should perform statistical power calculations to establish adequate sample sizes before a study using reasonable estimates of variability and the effect sizes expected. Adequate samples sizes are necessary because small samples bias results and inflate type I error with multiple statistical tests of dependent variables. Journal editorial policies should normally require authors to justify the adequacy of the sample size and address how representative the sample is of the population of interest. Authors interested in support for establishing adequate sample sizes have several resources to justify their decisions (Abramson, 2011; Hopkins, 2006; Lenth, 2001).

If inadequate resources result in an underpowered study, this limitation should be noted and the results interpreted as preliminary. The major decisions on sample size and statistical power of a study should normally be made prior to termination of data collection and statistical analysis.

Improve Reporting Transparency

Biomechanics authors should clearly disclose all methods, variables, and findings of their study. Readers of research reports need to know exactly how data were collected and all variables that were measured that influence the researcher degrees of freedom in the analysis and interpretation of the results. Treatment of missing or excluded data should also be clearly disclosed with its potential effects on the statistical analyses. Improvements in computing power and cost of data storage may one day allow archiving of the raw data of published studies, improving transparency, and the future integration of results of similar studies.

Improve Statistical Analyses

Biomechanics authors, reviewers, and editors should ensure all accepted papers follow journal submission requirements, current standards of statistical analysis (Curran-Everett & Benos 2004; Greenland et al., 2016; International Committee of Medical Journal Editors 1997; Kirk 2001; Lang & Altman, 2013; Lang & Secic, 2006), and standards for data reporting and interpretation (Knudson, Elliott, & Hamill, 2014). Statistical analyses should account for inflation of the type I error rate when testing multiple dependent variables from the same data set (Knudson, 2009).

The controversy between null-hypothesis significance testing and magnitude-based statistical inference (Hopkins, Marshall, Batterham, & Hanin, 2009; Wasserstein & Lazar, 2016; Welsh & Knight, 2015) means that uniform standards for statistical analysis can be difficult to determine. Authors need to justify the statistical analysis that is used in a study on current statistical standards or other strong rationale. Whatever statistical analyses are used for inference from the data, the reports should report and interpret the magnitude (e.g., Cohen's *d*, Glass' delta, η^2, or adjusted R^2) of effects using appropriate standard deviations as estimates for population or athlete variability (Hopkins, Marshall, Batterham, & Hanin, 2009; Knudson, 2009). Use of author and reviewer checklists for ensuring these important statistical issues are addressed (Nosek et al., 2012) should be used in biomechanics journals to facilitate these improvements in the literature.

Increase Acceptance of Replication Studies

Biomechanics authors, reviewers, and journal editors should support research that replicates findings that may be particularly important to the field. Direct replication of an important study could be to support effects and improve the estimate of the magnitude of the effect, especially if the original study had a small sample size and resulting low statistical power. Direct replications of the original studies are valuable since conceptual replications can only provide indirect, general support or extension of when findings are similar. Failed conceptual replications have limited logic in order to refute original findings (Nosek et al., 2012) that may be in question from design and analysis flaws. Meta-analysis is also a valuable methodology to combine similar research, however missing information and serious flaws in some research precludes their inclusion in these analyses.

Replication can also be expensive, so collaboration, meaningful dependent variables, and the magnitude of effects that are worthy of replication in a field should be explored (Nosek et al., 2012). Nosek and coworkers also propose more expansive reforms to improve the accuracy of research results including peer-review focused on soundness rather than importance/novelty and initiatives to open up data collection, methods, and reporting. Hopefully editorial teams will modify biomechanics journal policies to be more open to replication studies in the future. This would support identification of meaningful biomechanical effects and refinement of the magnitude of these effects.

Conclusion

Application of biomechanics research naturally depends on the accuracy of the results of research reports. Unfortunately, given many biomechanics studies utilize small sample sizes and incorrect statistical analyses, there is a substantial portion of biomechanics with overestimated size of effects, some of which are 'false positives' of null-hypothesis significance testing or meaningless effects given their magnitude. These weaknesses in many biomechanics research reports makes it difficult for readers to know if reported effects are true or the size reported. The limited replication in sports biomechanics also means that these biased and sometimes false effects are often not corrected for many years. Biomechanics authors and reviewers can improve the accuracy of results in the field by improving sample sizes and the resulting statistical power, greater reporting transparency, improving the rigor of statistical analyses used, and increasing the acceptance of direct replication studies. Correct application of sports biomechanics research results (Knudson, Elliott, & Hamill, 2014) would improve if there were a larger percentage of unbiased effects and their uncertainty were reported in the literature.

References

Abramson, J. H. (2011). WINPEPI updated: Computer programs for epidemiologists, and their teaching potential. *Epidemiologic Perspectives & Innovations, 8*, 1, 1. Retrieved from http://epi-perspectives.com/content/8/1/1

Austin, P. C., Mamdani, M. M., Huurlink, D.N., & Hux, J. E. (2006). Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. *Journal of Clinical Epidemiology*, 59, 964-969.

Barton, C. J., Levinger, P., Menz, H. B., & Webster, K. E. (2009). Kinematic gait characteristics associated with patellofemoral pain syndrome: A systematic review. *Gait & Posture, 30*, 405-416. doi:10.1016/j.gaitpost.2009.07.109

Batterham, A. M., & Atkinson, G. (2005). How big does my sample need to be? A primer on the murky world of sample size estimation. *Physical Therapy in Sport, 6*, 153-163. doi:10.1016/j.ptsp.2005.05.004

Begley C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature, 483*, 531-533,

Bogaerts, S., Desmet, H., Slagmolen, P., & Peers, K. (2016). Strain mapping in the Achilles tendon – A systematic review. *Journal of Biomechanics, 49*, 1411-1419. doi:10.1016/j.jbiomech.2016.02.057

Brown, S. R., Brughelli, M., & Hume, P. A. (2014). Knee mechanics during planned and unplanned sidestepping: A systematic review and meta-analysis. *Sports Medicine, 44*, 1573-1588. doi:10.1007/s40279-014-0225-3

Buchanan, T. L., & Lohse, K. R. (2016). Researchers’ perceptions of statistical significance contribute to bias in health and exercise science. *Measurement in Physical Education and Exercise Science*, *20*, 131-139. doi:10.1080/1091367X.2016.1166112

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafo, M. R. (2013). Power failure: Why small sample size undermines the reliability of *Neuroscience*, *14*, 365-376. doi:10.1038/nm3475

Chueng, R. T. C., Chung, R. C. K., & Ng, G. Y. F. (2011). Efficacies of different external controls for excessive foot pronation: a meta-analysis. *British Journal of Sports Medicine, 45*, 743-751. doi:10.1136/bjsm.2010.079780

Curran-Everett, D., & Benos, D.J. (2004). Guidelines for reporting statistics in journals published by the American Physiological Society. *American Journal of Physiology: Heart and Circulatory Physiology*, 287, H447-H449.

Csada, R. D., James, P. C., & Espie, R. H. M. (1996). The "File Drawer Problem" of Non-Significant Results: Does It Apply to Biological Research? *Oikos, 76*, 591-593.

Fernando, M., Crowther, R., Lazzarini, P., Sangla, K., Cunningham, M., Buttner, P., & Golledge, J. (2013). Biomechanical characteristics of peripheral diabetic neuropathy: A systematic review and meta-analysis of findings from the gait cycle, muscle activity and dynamic barefoot plantar pressure. *Clinical Biomechanics, 28*, 831-845. doi:10.1016/j.clinbiomech.2013.08.004

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLoS ONE, 9*(10), e109019. doi:10.1371/journal.pone.0109019

Fransz, D. P., Huurnink, A., Kingma, I., Verhagen, E. A. L. M., & van Dieën, J. H. (2013). A systematic review and meta-analysis of dynamic tests and related force plate parameters used to evaluate neuromusculoskeletal function in foot and ankle pathology. *Clinical Biomechanics, 28*, 591-601. doi:10.1016/j.clinbiomech.2013.06.002

Gelman. A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (Sign) and type M (Magnitude) errors. *Perspectives on Psychological Science, 9*, 641-651. doi:10.1177/1745691614551642

Greeland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, *P* values, confidence intervals, and power: A guide for misinterpretations. *European Journal of Epidemiology, 31*, 337-350. doi:10.1007/s10654—16-0149-3

Hall, J. P. L., Barton, C., Remy, C., & Morrissey, D. (2013). The biomechanical differences between barefoot and shod distance running: A systematic review and preliminary meta-analysis. *Sports Medicine, 43*, 1335-1353. doi:10.1007/s40279-013-0084-3

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review, 21*, 1157-1164. doi:10.3758/s13423-013-0572-3

Hopkins, W. G. (2006). Estimating sample size for magnitude-based inferences. *Sportscience, 10*, 63-70.

Hopkins, W. G., & Batterham, A. M. (2016). Error rates, decisive outcomes and publication bias with several inferential methods. *Sports Medicine*, Advance online publication. doi:10.1007/s40279-016-0517-x

Hopkins, W. G., Marshall, S. W., Batterham, A. M., and Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine in Science in Sports & Exercise, 41*, 3-13. doi:10 .1249/MSS.0b013e31818cb278

International Committee of Medical Journal Editors (1997). Uniform requirements for manuscripts submitted to biomedical journals. *Journal of the American Medical Association*, 277, 927-934.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), 696-701.

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*, 640-648. doi:10.1097/EDE.0b013e31818131e7

Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives in Psychological Science, 7*, 645-654. doi:10.1177/1745691612464056

Ioannidis, J. P. A. (2014). How to make more published research true. *PLoS Medicine, 11*(10). e1001747 doi:10.1371/journal.pmed.1001747

James, C. R., & Bates, B. T. (1997). Experimental and statistical design issues in human movement research. *Measurement in Physical Education and Exercise Science*, 1, 55-69.

Kirk, R. E. (2001). Promoting good statistical practices: some suggestions. *Educational and Psychological Measurement*, 61, 213-218.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahnik, S., Bernstein, M. J. . . . Nosek, B. A. (2014). Investigating variation in replicability: A “many labs” replication project. *Social Psychology, 45*(3), 142-152. doi:10.1027/1864-0335/a000178

Knudson, D. (2005). Statistical and reporting errors in applied biomechanics research. In Q. Wang (Ed.) *Proceedings of XXIII International Symposium on Biomechanics in Sports: Volume 2* (pp. 811-814). Beijing: China Institute of Sport Science.

Knudson, D. (2007). Analysis of multiple comparison errors using ground reaction force data. In H-J. Menzel & M. H. Chagas (Eds.), *Proceedings of the XXVth International Symposium on Biomechanics in Sports* (pp. 176-179). Belo Horizontae: Federal University of Mineas Gerais.

Knudson, D. (2009). Significant and meaningful effects in sports biomechanics research. *Sports Biomechanics, 8*, 97-105. doi: 10.1080/14763140802629966

Knudson, D**.** (2011). Authorship and sampling practice in selected biomechanics and sports science journals. *Perceptual & Motor Skills*, *112*, 838-844. doi:10.2466/17.PMS.112.3.838-844

Knudson, D. (2012). Twenty- year trends of authorship and sampling in applied biomechanics research. *Perceptual & Motor Skills*, *114*, 16-20. doi:10.2466/11.PMS.114.1.16-20

Knudson, D**.**, & Bahamonde, R. (2012). Twenty-five year trends of authorship and sampling in ISBS proceedings. In E. Bradshaw, A. Burnett, & P. A. Hume (Eds.), *eProceedings of the 30th Conference of the International Society of Biomechanics in Sports Volume 4*. (pp. 381-384). Melbourne: Australian Catholic University.

Knudson, D., Elliott, B., & Hamill, J. (2014). Proposing application of results in sport and exercise research reports. *Sports Biomechanics, 13,* 195-203. doi:10.1080/14763141.2014.959551

Knudson, D., Morrow, J., & Thomas, J. (2014). Advancing kinesiology through improved peer review. *Research Quarterly for Exercise and Sport*, *85*, 127-135. doi:10.1080/02701367.2014.898117

Lang, T. A., & Altman, D. G. (2013). Basic statistical reporting for articles published in biomedical journals: The “statistical analyses and methods in the published literature” or the SAMPL guidelines. In Smart, P., Maisonneuvre, H., & Polderman, A. (Eds.) *Science Editors Handbook* 2nd ed.

Lang, T., & Secic, M. (2006). How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers. 2nd ed. Philadelphia: American College of Physicians.

Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *American Statistician, 55*, 187-193.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*, 537-542. doi:10.1177/1745691612460688

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, *9*, 147-163. doi:10.1037/1082-989X.9.2.147

Mills, K., Hunt, M. A., Leigh, R., & Ferber, R. (2013). A systematic review and meta-analysis of lower limb neuromuscular alterations associated with knee osteoarthritis during level walking. *Clinical Biomechanics*, *28*, 713-724. doi:10.1016/j.clinbiomech.2013.07.008

Morris, H. H. (1981). Statistics and biomechanics: selected considerations. In J.M. Cooper & B. Haven (Eds), *Proceedings of the CIC Symposium: Biomechanics*. (pp. 216-225). Bloomington, IN: Indiana State Board of Health.

Mullineaux, D. R., Bartlett, R. M., & Bennett, S. (2001). Research design and statistics in biomechanics and motor control. *Journal of Sports Sciences,* **19**, 739-760.

Napier, C., Cochrane, C. K., Taunton, J. E., & Hunt, M. A. (2015). Gait modifications to change lower extremity gait biomechanics in runners: a systematic review. *British Journal of Sports Medicine, 49*, 1382-1388. doi:10.1136/bjsports-2014-094393

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives in Psychological Science*, *7*, 615-631. doi:10.1177/1745691612459058

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives in Psychological Science, 7*, 531-536. doi:10.1177/1745691612463401

Pataky, T. C., Vanrenterghem, J., & Robinson, M. A. (2015). Zero- vs. one-dimensional, parametric vs. non-parametric, and confidence interval vs. hypothesis testing procedures in one-dimensional biomechanical trajectory analysis. *Journal of Biomechanics*, 48, 1277-1285. doi:10.1016/j.jbiomech.2015.02.051

Pataky, T. C., Vanrenterghem, J., & Robinson, M. A. (2016). The probability of false positives in zero-dimensional analyses of one-dimensional kinematic, force and EMG trajectories. *Journal of Biomechanics*, 49, 1468-1476. doi:10.1016/j.jbiomech.2016.03.032

Schweizer, G., & Furley, P. (2016). Reproducible research in sport and exercise psychology: The role of sample sizes. *Psychology of Sport and Exercise*, *23*, 114-122. doi:10.1016/j.psychsport.2015.11.005

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359-1366. doi:10.1177/0956797611417632

Wasserstein, R. L., & Lazar, N.A. (2016). The ASA's statement on *p*-values: Context, process, and purpose. *The American Statistician, 70*, 129-133. doi:10.1080/00031305.2016.1154108

Welsh, A. H, & Knight, E. J. (2015). "Magnitude-based inference": A statistical review. *Medicine in Science in Sports & Exercise*, *47*, 874-884. doi:10.1249/MSS.0000000000000451

Wicke, J., Keeley, D. W., & Oliver, G. D. (2013). Comparison of pitching kinematics between youth and adult baseball pitchers: a meta-analytic approach. *Sports Biomechanics*, *12*, 315-323. doi:10.1080/14763141.2013.838692

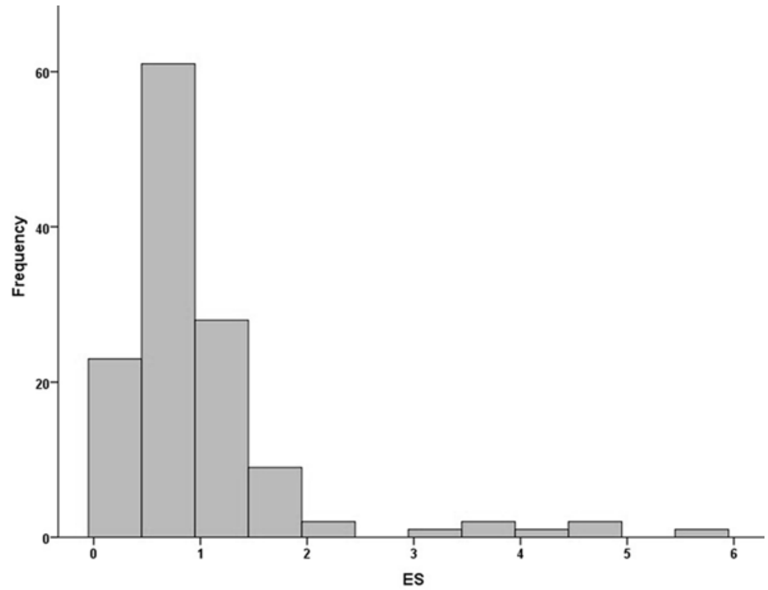

Figure 1. Distribution of statistically significant effect sizes reported in ten recent biomechanics meta-analyses (see text for references).