STATISTICAL AND REPORTING ERRORS IN APPLIED BIOMECHANICS RESEARCH

Duane Knudson
California State University-Chico, Chico, CA, USA

Applied biomechanics research reports in the 2004 volumes of the *ISBS Proceedings* (n=94) and the *Journal of Applied Biomechanics* (n = 11) were analyzed for statistical errors. There was no significant difference in the distribution of ratings of quality based on statistical errors. The percentages of various statistical errors in reporting data were also quite similar. Both sources of applied biomechanics research had unacceptably large percentages of papers with errors in reporting statistical testing and results that limit readers' ability to interpret the findings. Improvements need to be made in the reporting and peer review of applied biomechanics papers in *ISBS Proceedings* in order to achieve the mission of ISBS.

KEY WORDS: type I error, sampling, reliability, effect size

INTRODUCTION: The primary objectives of the International Society of Biomechanics in Sports (ISBS) are expanding the knowledge of biomechanics in sports and providing a forum for biomechanists, coaches, and athletes to interact. This mission to promote applied biomechanics in sports requires that ISBS presentations and proceedings be based on rigorous scientific methodology and statistical analysis, rather than anecdotal observations. Biomechanics research has been criticized for use of small samples, weak experimental designs, and inappropriate statistical analyses (Bates, 1989; Morris, 1981; James & Bates, 1997; Mullineaux, Bartlett, & Bennett, 2001).

There are many barriers for the results of applied biomechanics research to be robust for use in practice. Major barriers to internal and external validity of biomechanics research are the rationale (logic and literature) and design of the study, sampling, biofidelity of the model/variables, the reliability of dependent variables, data and statistical analysis. Traditional experimental designs comparing group effects can also be confounded by variability and unique responses of individual subjects (James & Bates, 1997) or repeated measures (Stergiou & Scott, 2005).

To initially address the issue of the rigor of sport biomechanics literature the issue of statistical or data-reporting errors was examined. The biomechanics literature is full of studies reporting multiple univariate statistical tests of many correlated dependent variables, even though it is well known that this inflates the experiment-wise type I error rate, resulting in the discussion of differences or associations that are not likely meaningful (Atkinson, 2002; Morris, 1980; Holmes, 2004; Schutz, Smoll, & Gessaroli, 1983). The sports biomechanics literature should avoid focusing coaches and athlete's attention on unimportant biomechanical factors. The purpose of this study was to compare the statistical errors reported in the 2004 volumes of the *ISBS Proceedings* (ISBSP) and the *Journal of Applied Biomechanics* (JAB). It was hypothesized that there would be no differences in statistical errors and quality between these two sources of applied biomechanics research.

METHODS: All full papers published in the 2004 volumes of the ISBSP and the JAB were classified as descriptive, correlation, comparison, modeling, technical note, or review papers. Only correlation and comparison papers were analyzed for statistical and reporting errors. Six common statistical and reporting errors were recorded and papers were rated on a five-point ordinal rating scale (Table 1). For example, "poor reporting of test statistics" would be indicated for a study not reporting observed test statistics, df, interactions, or post hoc tests as required by the APA (2001) or journal policies (Bond et al. 1995). Quality ratings were primarily based on the number or statistical errors. The distribution of overall ratings between ISBS and JAB were compared with Chi squared with statistical significance accepted at $p<0.05$. Descriptive statistics on quality ratings, errors, and sample size were also calculated.

RESULTS: Ninety-four papers in the ISBSP and eleven papers in JAB met the inclusion

criteria. The small number of papers in JAB was due to the high percentage of technical note (18.2) and modeling papers (42.4) in 2004. Two ISBSP comparison papers had to be removed because critical results pages were missing in the proceedings. The percentage of various quality ratings of the papers are reported in Table 2. There was no significant difference (Yates' corrected $\chi^2_5 = 0.72$, $p = 0.95$) in the distribution of paper quality between ISBSP and JAB.

Table 1 Study Rating Scale and Common Statistical Errors.

Score	Characteristics	Statistical Errors
4 Excellent	Good logic and literature rationale appropriate design & statistics assumptions addressed report size of effect & application	none
3 Good	logic and literature rationale appropriate design & statistics weak meaningfulness & application	assumptions unaddressed
2 Typical	weak logic & literature rationale weak design, sample & statistics overgeneralize results	above likely plus poor reporting of test statistics
1 Poor	poor logic & literature rationale poor design, sample, & statistics some methodology omitted overgeneralize results	above likely plus multiple tests inflating type I errors mix tested and untested observations
0 Very Poor	no logical rationale for study unreliable dependent variable(s) inappropriate design & statistics key methodology omitted data processing errors overgeneralize results	above likely plus inappropriate/missing statistical tests

* Papers were rated primarily on the tabulation of statistical errors. Paper characteristics were only considered in rare instances to decide between two rating levels.

The percentages of papers with various statistical and reporting errors are listed in Table 3. There was striking consistency in the distribution of statistical errors in papers published in ISBSP and JAB. The mean (SD) sample sizes for ISBSP comparison and correlation papers (with one large outlier removed from each group) were 18 (18) and 19 (23), respectively. The mean sample size for comparison and papers in JAB was 14 (7) and one correlation paper had a sample size of 10.

DISCUSSION: The reporting of applied biomechanics research must struggle with a trade-off between scientific rigor (e.g experimental control, statistics) and realism to the actual sport and athletes. This study examined the errors in reporting data and statistical tests as an indicator of scientific quality of applied biomechanics papers in ISBSP and JAB. Both ISBSP and JAB report a large percentage of papers with major statistical and reporting errors that threaten the integrity of the observations. This confirms the qualitative observations of this problem cited in the literature (Bates, 1989; Morris, 1981; James & Bates, 1997; Mullineaux, Bartlett, & Bennett, 2001). The most common errors (Table 3) were failing to report on statistical assumptions, poor reporting of test statistics, and the use of multiple comparisons. The latter problem can be addressed by multivariate ANOVA procedures, but use of these procedures is rare in ISBSP (9.3%) and JAB (11.1%) comparison studies with multiple statistical tests. This is a problem since a large percentage of comparison studies (ISBSP:

73.9 and JAB 72.3) inflated the type I error rate by making multiple comparisons of correlated dependent variables.

Table 2 Ratings of 2004 Papers in ISBS Proceedings and JAB (percent).

	Very Poor	Poor	Typical	Good	Excellent
ISBSP	14.3	27.5	38.4	18.7	1.1
JAB	0.0	36.3	45.5	18.2	0.0

*No significant (Yates' corrected $\chi^2_5 = 0.72$, $p = 0.95$) difference in the distribution of ratings.

The sample sizes comparison studies in applied biomechanics papers were small relative to the typical statistical standard of 30 (Clarke & Clarke, 1984). The mean sample size of comparison studies was 18 and 14 subjects. These sample sizes are common because of the high cost of biomechanical studies, and the small population of elite athletes or cadaver/tissue specimens often available. Many comparison studies utilize a repeated measures design, so these small sample sizes may have adequate statistical power for detecting differences with medium and large effect sizes. Correlation studies also had a mean sample size of 19 that was small given the purpose, number of variables, and intended generalization of these kinds of studies.

Table 3 Percentage of Papers with Statistical/Reporting Errors (percent).

	ISBSP	JAB
Assumptions	88.0	90.9
Reporting Test Stats	79.4	81.8
Mult. Tests/Inflated α	73.9	72.7
Mix tested/untested	40.2	27.3
Inappropriate Tests	21.7	18.2

The high percentage of papers in both JAB and ISBSP with significant errors in reporting statistical tests indicates there is a need for improvement in reporting and peer review in the applied biomechanics literature. The implied limit of four pages for papers in ISBSP may be a factor in limiting the reporting of critical information about statistical tests and data. It is recommended that authors be allowed to publish longer papers in ISBSP, possibly with an extra charge per page to cover the added cost.

This study is limited by the unknown reliability of the rating scale used, the inability to rate all factors affecting the internal and external validity of the biomechanics papers analyzed, and the small sample of papers in the 2004 volume of JAB which resulted in very low observed and expected frequencies. For example, small sample sizes often result in underpowered statistical tests that inflate the type II error rate. An applied biomechanics study with a small sample and multiple statistical tests can result in such high type I and II error rates that any conclusions from the statistical tests are in doubt. Some issues of statistical treatment and testing of experimental data are controversial (Sterne & Smith, 2001) and it takes considerable time to reach consensus and for these procedures to reach scientists using statistics. Many journals, however, have established statistical standards (Bond, Mintz, & McHugo, 1995) or recommendations for size of effect statistics (Thomas, Salazar, & Landers, 1991) for papers submitted for publication. This would also be a good policy to adopt in applied biomechanics publications like ISBSP, JAB, and Sports Biomechanics.

CONCLUSION: The 2004 volumes of ISBSP and JAB showed no significant differences in paper quality, nor was there differences the distribution of statistical and data reporting errors. Both sources of applied biomechanics research, however, showed unacceptably high

percentages of papers with major errors in conducting and reporting statistical tests. Improvements in peer review, statistical and reporting standards for biomechanics publications are needed so that sports biomechanics results can be applied in the field with athletes.

REFERENCES:

American Psychological Association (2001). *Publication Manual of the American Psychological Association* 5th ed. Washington, DC: American Psychological Association.

Atkinson, G. (2002). Analysis of repeated measurements in physical therapy research: multiple comparisons amongst level means and multi-factorial designs. *Physical Therapy in Sport*, 3, 191-203.

Bates, B.T. (1989). Comment on "the influence of running velocity and midsole hardness on external impact forces in heel-toe running. *Journal of Biomechanics*, 22, 963-965.

Bond, G.R., Mintz, J. & McHugo, G.J. (1995). Statistical guidelines for the *Archives of PM&R. Archives of Physical Medicine and Rehabilitation*, 76, 784-787.

Chen, A., & Zhu, W. (2001). Revisiting the assumptions for inferential statistical analyses: a conceptual guide. *Quest*, 53, 418-439.

Clarke, D.H. & Clarke, H.H. (1984). *Research processes in physical education* 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Holmes, T.H. (2004). Ten categories of statistical errors: a guide for research in endocrinology and metabolism. *American Journal of Endocrinology and Metabolism*, 286, E495-E501.

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

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.

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.

Schultz, R.W., Smoll, F.L. & Gessaroli, M.E. (1983). Multivariate statistics: a self-test and guide to their utilization. *Research Quarterly for Exercise and Sport*, 54, 355-263.

Stergiou, N. & Scott, M.M. (2005). Baseline measures are altered in biomechanical studies. *Journal of Biomechanics*, 38, 175-178.

Sterne, J.A.C. & Smith, G.D. (2001). Sifting the evidence—what's wrong with significance tests? *British Medical Journal*, 322, 226-23.

Thomas, J.R., Salazar, W. & Landers, D.M. (1991). What is missing in $p < .05$? Effect Size. *Research Quarterly for Exercise and Sport*, 62, 344-348.