ANALYISIS OF MULTIPLE COMPARISON ERRORS USING GROUND REACTION FORCE DATA

Duane Knudson

California State University-Chico, Chico, CA, USA

This technical note examined the effect of uncorrected multiple statistical comparisons that are often inappropriately reported in the applied biomechanics literature. Bilateral ground reaction forces (1000 Hz) were collected for a single subject performing twenty vertical jumps to create multiple comparisons of discrete kinetic variables. Nine of twenty comparisons were statistically significant using multiple t tests, while only seven were truly statistically significant when controlling for inflation of type I error. Most of the differences observed were not of practical significance to performance. The data confirmed the danger of inflation of type I errors in biomechanics and not interpreting the results based on meaningful effects or previous literature.

KEY WORDS: vertical jump, type I error, significance, Holm's procedure.

INTRODUCTION:

There are many challenges to conducting and analyzing biomechanical data for the results to be robust and meaningful to practice in sport. Issues related to the internal and external validity of biomechanics research are the rationale and design of the study, sampling, biofidelity of the model/variables, the reliability of dependent variables, and the statistical analyses performed.

Many biomechanical studies report inappropriate statistical analyses that compromise the credibility of the results (Morris, 1981; James & Bates, 1997; Mullineaux, Bartlett, & Bennett, 2001). One common error is the use of multiple univariate statistical tests on numerous dependent variables that inflates the experiment-wise type I error rate (Atkinson, 2002; Bender & Lange, 2001; Lundrook, 1998; Morris, 1981). Knudson (2005) reported that 73% of papers in the *Journal of Applied Biomechanics* and *ISBS Proceedings* made this error, despite many well-known techniques to correct for this error (Lundbrook, 1998). Austin et al. (2006) has recently shown that physiologically meaningless variables can be falsely identified as statistically significant by uncorrected multiple statistical comparisons.

The purpose of this study was to demonstrate the problems of multiple comparisons without correcting for inflation of the type I errors with common biomechanical data. It was hypothesized that uncorrected multiple comparisons would identify significant differences that would actually be nonsignificant and biomechanically meaningless.

METHOD:

One male intercollegiate jumping athlete gave informed consent to participate in the study. Following a warm-up the subject performed 20 standing vertical jumps on two Kistler 9286A force plates. Ground reaction forces (GRF) were colleted at 1000Hz and analyzed using *Bioware* software. The subject rested thirty seconds between each jump. The large number of jumps ensured stable mean kinetic variables (Rodano et al. 1996). The dominant limb was defined as the leg the athlete primarily used in one-legged jumps.

Twenty variables were derived from the GRF from each force plate, with variables temporally separated into two phases relative to the magnitude of the vertical GRF relative to half of body weight (BW): below (- phase) and above (+ phase) 0.5•BW. Peak force + phase, mean force + phase, time to peak force + phase, rate of force development (RFD) + phase, negative impulse to 0.5•BW, positive impulse to 0.5•BW, net impulse relative to 0.5•BW, duration of – and + phases, and flight time were calculated from the vertical forces. Jump height was calculated from the flight time for each force platform (Aragon-Vargas, 2000). Peak force + phase and peak force - phase, impulse + and - phase, and net impulse were calculated from both the antero-posterior (AP) and medio-lateral (ML) forces. This provided twenty dependent variables, both biomechanically meaningful and implausible for vertical jump performance that could be compared between sides of the body. Dependent t tests were performed without correcting for inflation of the EER and using the Holm's step-down procedure to control ERR at $P < 0.05$ (Lundbrook, 1998). Pearson correlations were also calculated between the jump GRF variables for the dominant leg. Descriptive statistics were calculated and reliability of the variables was documented using the coefficient of variation (CV). The practical significance of statistically identified differences was examined by effect sizes and qualitative comparisons with the literature on bilateral differences and the reliability of GRF data in vertical jumping.

RESULTS:

Descriptive data for the kinetic variables and results of the statistical testing are listed in Table 1. Uncorrected t tests indicated that nine of twenty variables were significantly ($P < 0.05$) different between limbs, however the true experiment-wise error rate (EER) for this statistical analysis would be 0.64 ($EER = 1 - (1 - 0.05)^{20}$) if all these comparisons were independent. Twenty-eight of the zero-order correlations between the vertical GRF variables (90 correlations) were statistically significant ($P < 0.05$). Since many of these variables were correlated with each other or not independent there is no simple formula to estimate the true EER (Bender & Lange, 2001).
Using Holm's correction for multiple comparisons, however, indicated that only seven comparisons between sides of the body were significantly different when holding the ERR to $P < 0.05$ (Table 1). The majority of the differences (four) were logical differences due to directional differences of laterally directed forces from each leg.

DISCUSSION:

Applied biomechanics research is often focused on what are hypothesized to be real and meaningful differences in dependent variables within and between athletes. The Holm's comparisons and partial correlations of the kinetic data for this athlete supported the hypothesis that uncorrected multiple comparisons of GRF variables will likely incorrectly identify variables as statistically significant when they are not. These errors can be because a variable is correlated with another variable or because they are type I errors. Since almost a third of the dependent variables were significantly correlated with each other, the EER for uncorrected t tests would likely be higher than the 0.64 that could be expected for unrelated variables. Inflation of the EER is clearly a problem for uncorrected multiple comparisons in one study regardless of whether the dependent variables are correlated or uncorrelated.
The meaningfulness of potential differences between limbs should also be examined in the context of reliability, effect size, previous research, and likely experimental error. Of the seven variables that remained significantly different across sides of the body using Holms correction, only mean peak vertical force seems to have plausible effect on jump height. The significantly greater mean peak vertical force in the dominant leg (4.1% and $d = 0.27$) than the non-dominant leg, however, was actually at the low end of meaningful bilateral differences in squats (McCurdy & Langford 2005) or typical variation in ground reaction forces in jumping (Cordova & Armstrong, 1996; Rodano et al. 1996; Vaverka & Janura, 1997). The significant difference in peak anterior-posterior force in the – phase does not seem meaningful because this variable was not significantly correlated with jump height. Previous research has reported larger variability in the kinetic variables of the eccentric phase of vertical jumping than in the concentric (Vaverka & Janura, 1997). The significant differences across side of the body in the medio-lateral direction have no logical mechanism of contribution to vertical jump height and also have large variability. Several statistically significant differences across sides of the body in this subject were not of practical significance based on the logic, effect size, or variability observed in the data. The variability of jump height ($CV = 2.3\%$), however, was consistent with previous research showing low variability ($2 < CV < 4\%$) in performance in repeated vertical jumps (Goodwin et al. 1999; Knudson & Corn, 1999).

Table 1 Mean Ground Reaction Force Data over Twenty Vertical Jumps

	Dominant				Non-Dominant			
	Mean	SD	CV		Mean	SD	CV	Sig
Vertical								
Peak Force (BW)	1.51	0.25	16.6		1.45	0.22	15.2	ε *
Time to Peak (ms)	292	80.9	27.7		304	96.5	31.7	
Mean Force (BW)	0.78	0.04	5.1		0.78	0.04	5.1	
RFD (BW/s)	3.79	1.54	40.6		3.58	1.75	48.9	
Impulse - (N·s)-54.7	15.5	28.3		-49.6	11.2	22.6		
Impulse + (N·s)	146	11.5	7.9		145	9.5	6.6	
Net Impulse (N·s)	91.3	18.9	20.7		95.2	16.0	16.8	
Duration - (ms)	593	161	27.1		540	98.1	18.2	
Duration + (ms)	724	90.8	12.5		719	101	14.0	
Jump height (m)	0.87	0.02	2.3		0.87	0.02	2.3	
Anterior-Posterior								
Peak Force - (N)	-50.8	20.1	39.6		-9.8	36.7	374	ε *
Peak Force + (N)	71.4	80.6	113		24.0	101	421	
Impulse - (N·s)-4.8	3.0	62.5		1.6	1.8	113	ε	
Impulse + (N·s)	2.9	4.0	138		-0.2	5.3	2650	ε
Net Impulse (N·s)	-1.8	5.1	283		1.3	5.7	438	
Medio-Lateral								
Peak Force - (N)	26.6	17.0	63.9		-34.3	16.6	48.4	ε *
Peak Force + (N)	81.8	16.3	19.9		-71.1	39.2	55.1	ε *
Impulse - (N·s)8.4	4.1	48.8		-8.5	2.3	27.1	ε *	
Impulse + (N·s)	14.6	7.5	51.4		-19.7	5.7	28.9	ε *
Net Impulse (N·s)	23.0	8.1	35.2		-26.9	6.9	25.7	ε *

Note: - and + in variable names refer to negative and positive phases relative to 0.5•BW in vertical ground reaction forces, while signs in the data refer to directions.
[ε]Apparently significant ($P < 0.05$) using uncorrected multiple dependent t tests.
[*]Significantly different using Holm's correction (EER $P < 0.05$).

CONCLUSION:

The study confirmed that uncorrected multiple comparisons of GRF variables can result in the false identification of differences as statistically significant. GRF data from one subject also confirmed that biologically implausible variables or variables correlated with meaningful variables can be falsely identified as significant if corrections for multiple statistical comparisons are not performed. Sport biomechanics papers should report results based on statistical analyses corrected for multiple comparisons, appropriate statistics documenting the meaningful of the differences (Thomas, Salazar, & Landers, 1991), and the relevant literature on variability and experimental error.

REFERENCES:

Aragon-Vargas, L.F. (2000). Evaluation of four vertical jump tests: methodology, reliability, validity, and accuracy. *Measurement in Physical Education and Exercise Science*, **4**, 215-228.

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.

Bender, R., & Lange, S. (2001). Adjusting for multiple testing—when and how? *Journal of Clinical Epidemiology*, **54**, 434-349.

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.

Cordova, M.L., & Armstrong, C.W. (1996). Reliability of ground reaction forces during a vertical jump: implications for functional strength assessment. *Journal of Athletic Training*, **31**, 342-345.

Goodwin, P.C., Koorts, K., Mack, R., Mai, S., Morrissey, M.C., & Hooper, D.M. (1999). Reliability of leg muscle electromyography in vertical jumping. *European Journal of Applied Physiology*, **79**, 374-378.

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.

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., & Corn, R. (1999). Intrasubject variability of the kinematics of the vertical jump. In R.H. Sanders & B. Gibson (Eds.) *Scientific Proceedings of the XVII International Symposium on Biomechanics in Sports*, (pp. 381- 384). Perth, Western Australia: Edith Cowan University.

Lundbrook, J. (1998). Multiple comparison procedures updated. *Clinical and Experimental Pharmacology and Physiology*, **25**, 1032-1037.

McCurdy, K., & Langford, G. (2005). Comparison of unilateral squat strength between the dominant and non-dominant leg in men and women. *Journal of Sports Science and Medicine*, **4**, 153-159.

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.

Rodano, R., Squadrone, R., Rabuffetti, M., & Mingrino, A. (1996). Lower limb kinetic variability in vertical jumping. In T. Bauer (Ed). *Proceedings of the XIIIth International Symposium on Biomechanics in Sports* (pp. 198-201). Thunder Bay, Ontario: Lakehead University.

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.

Vaverka, F., & Janura, M. (1997). Comparison of the force-time structure of the vertical jump between men and women. In J. Wilkerson, K. Ludwig, & W. Zimmerman (Eds). *Proceedings of the XVth International Symposium on Biomechanics in Sports* (pp. 75-80). Denton, TX: Texas Woman's University.