USAGE AND ACCEPTABILITY OF DATA NORMALIZATION IN BASEBALL PITCHING

Kevin Giordano,[12] Adam Barrack,[3] Kyle Wasserberger,[4] William Murrah,[5] Andy Karduna,[6] Lori Michener[3]

[1]Creighton University Department of Physical Therapy, Phoenix, AZ, USA
[2]Holy Cross Hospital Orthopedics Institute, Fort Lauderdale, FL, USA
[3]University of Southern California Department of Biokinesiology and Physical Therapy, Los Angeles, CA, USA
[4]Driveline Baseball Research & Development, Kent, WA, USA
[5]Auburn University Department of Educational Foundations, Leadership, and Technology, Auburn, AL, USA
[6]University of Oregon Department of Human Physiology, Eugene, OR, USA

To account for athletes of different sizes, kinetic values are commonly 'normalized' by dividing by mass and/or height. However, the creation of a ratio variable requires certain statistical assumptions to be met. The purpose of this study was to determine if elbow valgus torque predicted by pitching velocity is influenced by the normalization method using regression model comparison with normalized torque values. Both mass and mass*height normalization satisfied the correlation and zero intercept assumptions. Results did not agree between analysis methods that elbow valgus torque could be predicted with pitching velocity at the α = 0.05 level, indicating caution should be exercised before normalizing pitching kinetics data without confirming the assumptions for a ratio variable are met.

KEYWORDS: Assumptions, Biomechanics, Elbow, Kinetics, Statistics

INTRODUCTION: Anthropometrics influence pitching biomechanics data. Larger and taller pitchers have greater segment masses and longer external moment arms resulting in greater throwing arm kinetics. To compare data between subjects of different sizes, biomechanists commonly practice 'data normalization,' where they divide a variable (i.e., elbow valgus torque) by the anthropometric variable they want to account for. Most commonly, torque is divided by mass, but often also by height (Nicholson et al., 2020; Post et al., 2015) This is subsequently interpreted as " X amount increase in torque *per kg of body mass,"* to allow this new, ratioed variable to be compared across pitchers of multiple sizes. Logically, ratio normalization makes sense, however, there are inherent limitations in using a created ratio to compare across individuals (Allison et al., 1995; Curran-Everett, 2013).

For the new ratio variable to be valid, the following assumptions should be met: (1) the relationship of the variables in the ratio must have a linear regression line that passes through the origin, (2) the new 'normalized' variable must not correlate with the variable used to normalize, and (3) the interpretation and conclusion drawn from using the normalized variable must be the same as the interpretation and conclusion drawn using regression model comparison approach accounting for the covariate (Allison et al., 1995; Curran-Everett, 2013).

While studies have analyzed the use of different normalization parameters (mass, mass*height, mass*segment length, etc.), they have primarily focused on the correlation aspect in assumption #1 described above. Only one study (Hirsch et al., 2022) has analyzed the intercept and agreement with regression models. Further, an analysis in this manner has not been performed in baseball biomechanics, where normalization is a common practice. Therefore, the purpose of this study was to determine the influence of normalizing by mass on the prediction of elbow valgus torque by fastball pitching velocity. We hypothesized the normalization technique would not meet the defined statistical assumptions and the conclusions would differ between analysis techniques.

METHODS: Data were analyzed from a previously collected sample of Division I collegiate pitchers (n=97; 19.8 ± 1.3 yrs 188.0 ± 5.0 cm, 90.5 ± 7.3 kg). Participants were fitted with a wearable arm sensor designed to calculate elbow valgus torque during throwing (Motus Global, Massapequa, NY). Ball velocity was tracked using a calibrated radar gun. Participants were allowed to perform their own warmup to prepare for game effort pitching. Once participants deemed themselves ready, they threw five pitches at 75% effort and five pitches at 100% effort. Mean torque and velocity from the pitches at 100% effort were used for analysis.

Statistical Analysis: To assess the appropriateness of the normalization process, a regression model of elbow valgus torque regressed on mass allowing a free intercept term was compared to the same model with a fixed y-intercept at 0. Mass normalized torque (Nm/kg) was then regressed on mass to determine if the process of dividing elbow torque by mass accounted for the relationship between elbow valgus torque and body mass. This process was repeated for mass*height normalized torque (Nm/kg*m).

For the regression model comparison analyses, elbow valgus torque was the dependent variable. Elbow valgus torque was regressed on mass, then regressed on mass and velocity. ΔR^2 was reported to determine the effects of ball velocity on elbow valgus torque after accounting for mass. To account for both mass and height, elbow valgus torque was regressed on mass and height, then regressed on mass, height, and ball velocity, with ΔR^2 being used to determine the effects of ball velocity on elbow valgus torque after accounting for mass and height. These results were compared to regressing elbow valgus torque normalized to mass, and mass times height on ball velocity. Statistical significance was set at α = 0.05.

RESULTS: Body mass demonstrated a significant relationship with elbow valgus torque (β = 0.743, R^2 = 0.189, SSE = 11583 p < 0.001) when allowing a free intercept term. When fixing the intercept at 0, there was no significant difference (ΔSSE =38, p = 0.577). After dividing elbow valgus torque by body mass, there was no significant relationship (β = 0.001, R^2 = 0.008, p = 0.606). Body mass*height had a relationship with elbow valgus torque (β = 0.356, adj.R^2 = 0.223, SSE = 10951 p < 0.001). When compared to the same model with an intercept fixed at 0, there was no significant difference (ΔSSE =156, p = 0.516). After dividing elbow valgus torque by body mass*height, there was no longer a significant relationship (β >0.001, R^2 > 0.00, p = 0.876). Adding ball velocity to the regression model predicting elbow valgus torque after mass resulted in an insignificant effect of ball velocity on elbow valgus torque (β = 0.654, ΔR^2 = 0.024, p = 0.051). Similarly, regressing mass normalized torque on ball velocity resulted in a significant relationship (β = 0.008, R^2 = 0.037, p = 0.033). Adding ball velocity to the regression model predicted elbow valgus torque after mass and height resulted in an insignificant effect of ball velocity on elbow valgus torque (β = 0.621, ΔR^2 = 0.022, p = 0.051). Similarly, regressing mass*height normalized torque on ball velocity was significant (β = 0.004, R^2 = 0.033, p = 0.042).

DISCUSSION: The results of this study indicate that, in this sample of collegiate pitchers, two of three main assumptions made through normalizing elbow valgus torque to mass and mass*height were met. However, the results of the regression model comparison were not in agreement. When normalizing elbow valgus torque to mass and mass*height, ball velocity was a significant predictor of normalized elbow valgus torque. In contrast, using regression model comparison concluded that ball velocity does not significantly predict elbow valgus torque after accounting for mass, or mass and height at the α = 0.05 threshold. We recommend caution when using ratio normalization.

Interestingly, and contrary to our initial hypothesis, both mass and mass*height passed two of three criteria for normalization (Allison et al., 1995; Curran-Everett, 2013), yet our conclusions still differed between analysis methods. This highlights several statistical implications. First, when we normalize data using ratios, any residual error was lumped into our newly calculated variable. Our new variable was no longer *elbow valgus torque*, but rather it was now *elbow valgus torque*

divided by mass and/or height. Transforming variables makes assumptions that are mathematically satisfied at the α = 0.05 level, but not perfectly met, analogous to playing a game of telephone. A little bit of the variable gets changed each time a transformation is made, and in the end, it may not represent the original form. This is particularly the case when arbitrary thresholds of α = 0.05 are used. When p > 0.05, it does not mean that there is no difference, as it is often treated in the literature. It means that the null hypothesis cannot be rejected at the α = 0.05 threshold. In the present study, ball velocity significantly predicted mass normalized elbow valgus torque (p = 0.033) and mass*height normalized torque (p = 0.042), but when using the model comparison approach, ball velocity was not a significant predictor of elbow valgus torque (p = 0.051). Further, the effect sizes of ball velocity on mass normalized elbow valgus torque (0.037) and mass*height normalized elbow valgus torque (0.033) were roughly 50-55% higher than those found using the model comparison approach (0.024 and 0.022 respectively). On an absolute scale, this is likely not meaningful. However, the relative difference may be clinically meaningful, depending on the research question. By chance, these results falling on opposite sides of the p = 0.05 threshold elucidates that, even when assumptions for normalization are met, the wrong conclusion may be reached. The debate on the efficacy of the commonly used α = 0.05 threshold (Greenland, 2017) is outside the scope of this study, but the fact remains that it is predominant in sports biomechanics research.

The purpose of normalization should also be considered when selecting a normalizing method. Generally, investigators normalize data to control for the influence of a covariate (mass) that is inferred to have an impact on the relationship between the dependent (elbow valgus torque) and the independent variable of interest (ball velocity). This is done to determine the influence of the separate independent variable (ball velocity) on the dependent variable of interest. If this is the goal, the best approach is to use a regression model comparison analysis, because, mathematically, regression permits the simultaneous comparison of multiple variables without inheriting the error of imperfect ratio assumptions (Table 1). Further, it allows the dependent variable to remain in its original units, aiding in understanding and communication with stakeholders.

Table 1. Normalized Models (Left) and Their Analogous Model Comparison Analyses (Right)

Model	BV β	R^2	p		Model	Kg β	m β	BV β	R^2	ΔR^2 (p)
Nm/Kg ~ BV	0.0077	0.037	0.033	➡	Nm ~ Kg	0.743			0.189	0.024 (0.051)
					Nm ~ Kg + BV	0.680		0.654	0.213	
Nm/Kg*m ~ BV	0.0038	0.033	0.042	➡	Nm ~ Kg + m	0.546	58.7		0.223	0.022 (0.051)
					Nm ~ Kg + m + BV	0.500	56.6	0.621	0.247	

Model reads dependent variable regressed on independent variables. BV = ball velocity, Kg = kilograms body mass, m = meters body height. Table shows common model setup used in the field followed by a two model comparison that would replace the analogous normalized setup.

Importantly, normalization may not always be necessary. When evaluating how impactful a normalization can be, it is helpful to consider the variance structure of the normalizing variable in question. Specifically, the homogeneity of the sample will play a role in this interpretation. If the sample is homogeneous in mass, the influence of mass on ball velocity will be relatively small. In this case, the need to normalize is also diminished. Investigators also use normalization to create a universal metric for comparison across groups. In this case, even further attention should be given to whether the assumptions of normalization hold true. Our sample of pitchers had a body mass range of 72-104 kg with a standard deviation of 7.3 kg, however their ball velocity range

was less variable at 32.6-41.6 m/s with a standard deviation of only 1.5 m/s. Limited velocity variance is likely due to all participants coming from one of the top performing baseball conferences in the country. Limited ball velocity variance compared to body mass is likely to favor body mass's ability to predict elbow valgus torque. Therefore, future studies should assess the validity of normalization assumptions in more heterogeneous populations.

CONCLUSION: Data normalization is a common practice in baseball pitching biomechanics to control for a covariable affecting the dependent variable. Researchers should consider using a regression model comparison approach rather than creating 'normalized' ratio variables. Further, using model comparison simplifies the interpretation for coaches and clinicians. Interpreting the current research question using mass normalized data would read, "each 1 MPH increase in ball velocity would increase your elbow valgus torque 0.008 Nm for every kg of body mass." Interpreting the model comparison approach to an athlete or clinician would read, "All other factors equal, each 1 MPH increase of ball velocity would increase your elbow valgus torque .65 Nm," eliminating the need for extra math. This simple improvement in statistical practices may result in more reproducible data and aid in statistical interpretation, increasing the accessibility of coaches, clinicians, and athletes to baseball biomechanics data.

References

Allison, D., Paultre, F., Goran, M., Poehlman, E., & Heymsfield, S. (1995). Statistical considerations regarding the use of ratios to adjust data. *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity*, *19*(9), 644-652.

Curran-Everett, D. (2013). Explorations in statistics: the analysis of ratios and normalized data. *Advances in physiology education*, *37*(3), 213-219.

Greenland, S. (2017). Invited commentary: The need for cognitive science in methodology. *American journal of epidemiology*, *186*(6), 639-645.

Hirsch, S. M., Chapman, C. J., Frost, D. M., & Beach, T. A. (2022). Comparison of 5 Normalization Methods for Knee Joint Moments in the Single-Leg Squat. *Journal of applied biomechanics*, *38*(1), 29-38.

Nicholson, K. F., Hulburt, T. C., Beck, E. C., Waterman, B. R., & Bullock, G. S. (2020). The relationship between pitch velocity and shoulder distraction force and elbow valgus torque in collegiate and high school pitchers. *Journal of shoulder and elbow surgery*, *29*(12), 2661-2667.

Post, E. G., Laudner, K. G., McLoda, T. A., Wong, R., & Meister, K. (2015). Correlation of shoulder and elbow kinetics with ball velocity in collegiate baseball pitchers. *Journal of athletic training*, *50*(6), 629-633.