Comment on: 'Reliability of a commercially available and algorithm-based kinetic analysis software compared to manual-based software'

Letter to editor

Thomas Dos'Santos[1], Jason Lake[2], and Paul Comfort[1]

[1] Human Performance Laboratory, Directorate of Sport, Exercise, and Physiotherapy, University of Salford, Greater Manchester, United Kingdom

[2] Department of Sport and Exercise Sciences, University of Chichester, Chichester, United Kingdom

Thomas Dos'Santos

University of Salford,

Allerton Building, Frederick Road Campus,

Salford, Greater Manchester, United Kingdom,

M6 6PU.

Telephone: +447961744517

Email: t.dossantos@edu.salford.ac.uk

Word count: 1102 words

We read with interest the published study regarding the reliability and comparison between manual- and algorithm-based kinetic analysis software by Carroll et al. (2019) entitled: '***Reliability of a commercially available and algorithm-based kinetic analysis software compared to manual-based software***' DOI: 10.1080/14763141.2017.1372514, and we congratulate the authors for producing an informative and interesting study. However, we would like to raise an issue regarding the interpretation of their results for the differences in peak force (PF) between the two methods. We outline this issue in further detail in the following paragraphs of this letter.

The authors compared isometric mid-thigh pull (IMTP) kinetics between two analysis software methods: manual-based identification of contraction onset time (COT) using LabView analysis software (LabView); and an algorithm-based analysis program (NMP technologies Ltd). Unfortunately, the specific details regarding the COT identification for the algorithm-based analysis software was not provided which we feel is a notable limitation. Nevertheless, the authors reported that IMTP PF from the algorithm-based method was greater than the manual-based method obtained via LabView ($p < 0.001$, $d = 0.05$, 3072 ± 800 N vs. 3033 ± 809 N), and the rate of force development values 0-50 ms ($p < 0.001$, $d = 0.11$, 1959 ± 1407 N vs. 2121 ± 1618 N) and 0-200 ms ($p < 0.001$, $d = 0.56$, 3519 ± 2650 N vs. 5045 ± 2832 N) were also greater for the algorithm-based method.

It is somewhat surprising that there was a significant ($p < 0.001$), although not meaningful ($d = 0.05$) difference in PF between the two methods. The authors provide a potential explanation for this difference and state the following: 'these differences were likely a result of differences in COT identification between the two methods.' We respectfully disagree with this explanation because COT identification will not affect absolute gross PF (inclusive of body weight) during the IMTP, as shown in our previous work (Dos'Santos, Jones, Comfort, & Thomas, 2017). If net PF was examined (i.e. PF – BW or PF – force at COT), we could therefore understand how differences in COT could result in differences net PF between methods. However, the authors clearly state the following: 'we did not remove body mass from the force variables; therefore, the resulting values are absolute and not relative.' While we agree that not removing body weight from force variables results in absolute gross forces, dividing force variables by body mass would provide relative forces, not removing body mass (as they state). Nevertheless, we suggest it is incorrect to attribute the differences in PF due to differences in COT, based on their reported methods, and question whether they did in fact examine net PF.

In addition, the authors do not state whether any of the force-time data were low-pass filtered. We have previously shown that low-pass filtering can affect PF data, with lower low-pass filtering cut-off frequencies resulting in lower PF values (Dos' Santos, Lake, Jones, & Comfort, 2018). We question whether the authors have low-pass filtered their data, and if the LabVIEW software low-pass filtered the data or if different COFs were used between the two software. This could be a plausible explanation for the differences in PF values between methods. We would also like to express an issue regarding the scatter plot shown in Figure 2. There is a clear outlier in the data, and we find it difficult to comprehend how there can be an approximate 500 N difference in PF for one trial between the two methods. The authors report a strong, nearly perfect relationship (R^2 = 0.9816) in PF between the two methods; however, this method is inappropriate for assessing agreement because the strength of linear association is measured around a line of best fit, irrespective of whether the slope of the line differs from unity (proportional bias) or whether its intercept differs from zero (fixed bias) (Ludbrook, 1997, 2002). Consequently, for method comparative designs it is integral to establish whether fixed (i.e. greater or lower value by a constant amount across whole range of measurements) and proportional bias (i.e. greater or lower value that is proportional to the level of the measured variable) is present between the two methods (Ludbrook, 1997, 2002; Rankin & Stokes, 1998). Additionally, using a linear regression statistical approach assumes that the gold standard method is not prone to error (noise), when in fact both methods in the case of Carroll et al. (2019) are susceptible to error (i.e. signal noise) (Ludbrook, 1997, 2002).

A more suitable method to illustrate the differences and assess the agreement between methods would be using Bland-Altman plots and limits of agreement (LOA) (Bland & Altman, 1986; Ludbrook, 1997; Mundy & Clarke, 2019). Performing this would allow the researchers to assess the systematic bias and random error, while the upper and lower limits of the LOA could be interpreted to determine if they are of practical importance (Bland & Altman, 1986; Mundy & Clarke, 2019). Furthermore, it is also suitable to determine if heteroscedasticity is present by creating a scatter plot of the difference between the two measures versus the average of the two methods (Mundy & Clarke, 2019). This will permit the researchers to examine if the bias and variability is uniform throughout the whole range of the measurements. Additionally, it has been suggested that because ordinary least-squares regression makes the assumption that the gold standard method does not contain error (i.e. signal noise), that least-products regression be used instead (Ludbrook, 1997; Mullineaux, Barnes, & Batterham, 1999). This method assumes that both methods can contain error, considers this, and quantifies the magnitude of fixed and proportional bias. Performing this process is integral to establish the agreement in measurements between two different methods, and to ascertain if a specific method produces constantly greater (or lower) values, or greater (or lower) values which are proportional across the whole range of measurements. However, as with the LOA approach (that does not consider proportional bias), if heteroscedasticity is present, data should be log-transformed to remove this (Atkinson & Nevill, 1998; Mullineaux et al., 1999). If this works, then least products regression (or LOA) can be used. However, if log-transformation does not reduce heteroscedasticity then these methods cannot be used (Mullineaux et al., 1999).

Based on the above, we feel further clarification regarding the force-time data analysis procedures is required to better understand why there was differences in PF between the software. We would like to state that this letter is not intended to detract from the excellent study by Carroll et al. (2019) but merely intended to help improve the study with some additional but important information and clarify some aspects for the reader. This will enable the reader to effectively interpret the findings and help improve best practice for IMTP testing.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors

REFERENCES

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine, 26*(4), 217-238.

Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307-310.

Carroll, K. M., Wagle, J. P., Sato, K., DeWeese, B. H., Mizuguchi, S., & Stone, M. H. (2019). Reliability of a commercially available and algorithm-based kinetic analysis software compared to manual-based software. *Sports Biomechanics, 18*(1), 1-9.

Dos' Santos, T., Lake, J., Jones, P. A., & Comfort, P. (2018). Effect of low-pass filtering on isometric midthigh pull kinetics. *The Journal of Strength & Conditioning Research, 32*(4), 983-989.

Dos'Santos, T., Jones, P. A., Comfort, P., & Thomas, C. (2017). Effect of different onset thresholds on isometric mid-thigh pull force-time variables. *The Journal of Strength & Conditioning Research, 31*(12), 3463-3473.

Ludbrook, J. (1997). Special article comparing methods of measurement. *Clinical and Experimental Pharmacology and Physiology, 24*(2), 193-203.

Ludbrook, J. (2002). Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clinical and Experimental Pharmacology and Physiology, 29*(7), 527-536.

Mullineaux, D. R., Barnes, C. A., & Batterham, A. M. (1999). Assessment of bias in comparing measurements: a reliability example. *Measurement in Physical Education and Exercise Science, 3*(4), 195-205.

Mundy, P., & Clarke, N. D. (2019). Reliability, validity and measurement error. In P. Comfort, P. A. Jones, & J. J. McMahon (Eds.), *Performance Assessment in Strength and Conditioning* (First ed., pp. 23-32). Abdingdon, Oxon, United Kingdom: Routledge.

Rankin, G., & Stokes, M. (1998). Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clinical Rehabilitation, 12*(3), 187-199.