International Journal of Strength and Conditioning

Bailey, C. A. (2023). Statistical Considerations When Measuring Absolute Reliability And Variability Of Vector Data In Sport Performance. International Journal of Strength and Conditioning
https://doi.org/10.47206/ijsc.v3i1.163

Statistical Considerations When Measuring Absolute Reliability And Variability Of Vector Data In Sport Performance

Chris A. Bailey[1]
[1]Department of Kinesiology, Health Promotion, and Recreation, University of North Texas, USA
Corresponding Author: chris.bailey2@unt.edu

ABSTRACT

Vector-based data in sport performance include a magnitude and direction. Statistically speaking, they are interval in nature as they may be positive or negative. The coefficient of variation (CV) is a commonly reported measure of variability, but its use with vector data is questionable and may be contraindicated. Limits of agreement (LOA) and standard error of measurement (SEM) may be better alternatives for vector data such as acceleration. The purpose of this study was to demonstrate the issues with quantifying variability of vector data, while also evaluating the utility of commonly used measures. Acceleration data at three intervals from 0 to 27.4 m (0 to 90 ft) were calculated from publicly available sprint performance data from 310 athletes participating in the 2018 and 2019 Major League Baseball seasons. CV, LOA, and SEM were calculated to evaluate inter-season variability. Variability of the first two intervals was acceptable for all measures, but the final interval was unclear as the CV was quite large (50.78%), while the LOA and SEM were only slightly larger than the other interval values. The final interval includes both positive and negative acceleration, contraindicating the usage of the CV. LOA and SEM are more useful for vector data, showing that the final interval was more variable between trials than the others, but not to the extent portrayed by the CV. The CV likely should not be used with vector data unless it is known that the data does not cross zero. LOA and SEM are appealing alternatives for the CV and should be considered since they work with positive and negative data.

Keywords: coefficient of variation, consistency, interval, ratio

INTRODUCTION

It is recommended that both relative and absolute measures be used when evaluating reliability (Atkinson and Nevill 1998, Weir 2005). Concerning absolute reliability measurement in sport performance, the coefficient of variation (CV) appears to be quite popular. As of this writing, 1,016 results appear in Pubmed when searching for both “sport performance” and “coefficient of variation.” The CV may be more appropriately described as a measure of variation or dispersion as it produces a ratio of the standard deviation (sd) to the mean and is often represented as a percentage (Atkinson and Nevill 1998). It has the benefit of being easy to calculate, only requiring the mean and sd, and allows for comparisons between sample data (Ospina and Marmolejo-Ramos 2019). It also has the added benefit of only requiring one trial to compute, unlike other reliability assessments that require a minimum of two trials (Hopkins 2000).

Despite its popularity and ease of production, the CV should not be applied in every situation. Many of the situations where the CV may be inappropriate regularly occur in sport performance when dealing with vectors. Biomechanically, vectors have both a magnitude and a direction, and this is indicated by the value being positive or negative. If vector values can cross the zero point, they are classified as interval data (Vincent and Weir 2012). A requirement of the CV is that the data are measured on the ratio scale, not interval (Mitchell 1986, Velleman and Wilkinson 1993). Ratio data have a "true" zero point and no positive or negative direction (Vincent and Weir 2012), which means it is often scalar data from a sport performance standpoint. Thus, it is statistically appropriate to use the CV for data such as time and distance, but inappropriate for vector data such as force, acceleration, velocity, or power as they are vectors. This issue has been demonstrated previously using temperatures (Allison 1993). In a data set of any size that contains the same temperatures reported in Celsius, Fahrenheit, and Kelvin, different CV values will be produced for each temperature unit even though they are equivalent temperatures. Temperatures reported in Celsius will produce the largest CVs and Kelvin will produce the smallest (Allison 1993). This is due to how the CV is calculated. The ratio requires a measure of dispersion (sd) and a measure of central tendency (mean). When a data set has positive and negative values (interval), it will have a lower mean compared to a data set with similar magnitude values that are only positive (ratio). In the previous example, Celsius and Fahrenheit (interval) produce larger CVs since they can have positive and negative values, unlike Kelvin (ratio). Concerning sport performance data, even with similar dispersion, vector data may have larger measured CVs than scalar data due to the inclusion of a vector direction. Even though CVs computed on interval data may be unstable and lack any meaning (Velleman and Wilkinson 1993, Yao 1994) many still report poor CVs when working with acceleration and deceleration (negative acceleration) (Delaney, Cummins, Thornton, and Duthie 2018, Harper, Morin, Carling, and Kiely 2020, Varley, Fairweather, and Aughey 2012) as well as positive and negative force-time data (Byrne, Moody, Cooper, Kinsella, and Byrne 2017, Heishman, Daub, Miller, Freitas, Frantz, and Bemben 2020, Howarth, Cohen, McLean, and Coutts 2021, Nibali, Tombleson, Brady, and Wagner 2015). Some authors may avoid the issue in sprint kinematics by computing CVs on velocities at specific intervals instead of including acceleration data which may be negative as well as positive (Ashton and Jones 2019). In the area of strength asymmetry research, some avoid this issue by scalarizing data during analysis by excluding the direction and only considering the variable magnitude in the calculation (Fort-Vanmeerhaeghe, Bishop, Brusca, Aguilera-Castells, Vicens-Bordas, and Gonzalo-Skok 2020 , Padros-Mainer, Casajus, and Gonzalo-Skok 2019, Read, Oliver, Myer, and De Ste Croix 2017). This is not recommended as measures could switch directions between trials, which wouldn't be detected if only scalar versions of the data are used (Bailey, Sato, and McInnis 2021).

In many of these situations, there may be more appropriate options than the CV. Other options for measuring absolute reliability and variability include limits of agreement (LOA) and the standard error of measurement (SEM) (Atkinson and Nevill 1998, Bland and Altman 1986, Weir 2005). Bland and Altman's LOA demonstrate the amount of variability that is to be expected between trials to a 95% confidence interval. They are often demonstrated by plotting the between trial differences against the overall means but can also be reported simply as the lower limit value and upper limit value, as the LOA retains the original units of measure (Atkinson and Nevill 1998, Bland and Altman 1986).

The SEM estimates the noise between trials of a test and is an indicator of precision (Vincent and Weir 2012). Like LOA, the SEM retains the original units of measure for ease of interpretation. The SEM can be calculated in multiple ways, but commonly requires the standard deviation and a previously calculated reliability measure such as the intraclass correlation coefficient (ICC) (Atkinson and Nevill 1998, Thomas, Nelson, and Silverman 2015, Vincent and Weir 2012). Alternatively, it can be also calculated as the square root of mean square error from a repeated measures ANOVA between trials (Eliasziw, Young, Woodbury, and Fryday-Field 1994, Stratford and Goldsmith 1997, Vincent and Weir 2012). The resulting SEM is the value where 34% of the error variance will lie above and below the measured value, indicating the amount of precision within one standard deviation (Atkinson and Nevill 1998). This can be taken a step further to the 95% confidence level with the SEM minimum detectable change at the 95% confidence limit (SEM_{mdc95}), which indicates the amount of difference required to be considered an actual change because it is past what might be credited to measurement error (Stratford 2004, Vincent and Weir 2012). This is similar to the smallest worthwhile change (SWC) which indicates the smallest change that is likely to have a performance impact, but the SWC is based on the CV, which means it should not be applied to vector-based data (force, acceleration, power, etc.) unless it is known that only positive values appear in the data (Hopkins, Hawley, and Burke 1999, Malcata and Hopkins 2014). A potential issue with the SEM, and therefore SEM_{mdc95}, is that it assumes that data are homoscedastic, which means that the chance for error is the same regardless of measure magnitude. This is opposite of the CV assumption of heteroscedasticity, which is that the chance for error increases at higher or lower measure magnitudes. If proportional bias is present, a log transformation may be required making interpretation more difficult (Atkinson and Nevill 1998).

Previous literature has explained that the CV should not be applied to interval or vector-based data, but it appears that recommendation may not always be considered in sport performance. There are other options for assessing variability and absolute reliability in sport performance, which may have more readily applicable results as they retain the original unit of measure. Therefore, the purpose of this study was to demonstrate the issues associated with measuring variability of vector data while evaluating the utility of commonly used measures in a sample of Major League Baseball (MLB) players' sprint acceleration data.

METHODS

Design

The utility of the CV and several other measures of variability was evaluated using publicly available MLB sprint performance data from the 2018 and 2019 seasons. Sprint interval times were used to produce acceleration data from 0 to 27.432 m (0 to 90 ft) at three intervals. Data were reduced to only include athletes that participated in both seasons to facilitate a test-retest design. CV, LOA, SEM and SEM_{mdc95} were then calculated.

Subjects

This study initially included data from 525 major league baseball players from the 2018 and 2019 MLB seasons. The sample was reduced by excluding players that did not have data in each season, resulting in a final sample of 310 players. Data for this study were publicly available through MLB's Statcast database (https://baseballsavant.mlb.com/leaderboard/running_splits). As the data are a part of the public domain and it does not include private information, it is not considered human subjects research and does not require consent or approval from an institutional review board.

Procedures

Utilizing doppler radar and camera systems collecting data at 30 Hz in every MLB ballpark, Statcast tracks athlete's location, and many variables can be produced from this data (Lage, Ono, Cervone, Chiang, Dietrich, and Silva 2016, Pareek, Parkes, Leontovich, Krych, Conte, Steubs, Wulf, and Camp 2021). This data includes sprint times (s) for every 1.5 m (5 ft) from 0 to 27.4 m or 0 to 90 ft. Only "competitive" sprints qualify for this data set by MLB. This helps control for athletes not running with full effort on plays where they likely would not reach base safely. Competitive sprints included sprints of two or more bases on hits that were not home runs and sprints from home to first on softly hit balls where they must give a full effort in order to reach base safely. Athletes scoring from second base on an extra-base hit were also excluded as these are not always competitive. The average of the best 66% of each athletes' sprints were calculated and made publicly available. A minimum of 10 competitive sprints were required to be included in this data set.

Data Analyses

All data analyses were completed in R (R Core Team 2020). Individual sprint times at every 1.5 m (5 ft) were used to create three interval velocities (interval displacement (m) / interval time (s)) and then interval accelerations ((final interval velocity (m/s) – initial interval velocity (m/s)) / interval time (s)) in each sprint. Intervals included 0 to 9.1 m (0 to 30 ft) (a1), 9.1 to 18.3 m (30 to 60 ft) (a2), and 18.3 to 27.4 m (60 to 90 ft) (a3). Table 1 serves as an example of the data available for each athlete. It includes example data, interval positions, and equations for interval velocities and accelerations. For the purposes of this investigation, the mean data from the 2018 and 2019 MLB seasons were used in a test-retest design to evaluate inter-season variability of a vector measure (acceleration). It should be noted that measures separated by a year likely better represent stability, rather than reliability, but this does not impact the current study as the purpose is to demonstrate the difference in utility of different measures of variability and reliability in vector data, and not necessarily to report the reliability of MLB sprint data.

Table 1. Example data with interval positions indicated along with velocity and acceleration equations used.

Displacement (m)	Time (s)	Interval	Interval Velocity (m/s)	Interval Acceleration (m/s^2)
0	0			
1.5	0.55			
3.0	0.86			
4.6	1.14	1	v1 = 9.1 (m)/interval time (s)	a1 = (v1 - 0 m/s)/ interval time
6.1	1.39			
7.6	1.63			
9.1	1.85			
10.7	2.06			
12.2	2.26			
13.7	2.45	2	v2 = 9.1 (m)/interval time (s)	a2 = (v2-v1)/interval time
15.2	2.65			
16.8	2.84			
18.3	3.02			
19.8	3.21			
21.3	3.39			
22.9	3.57	3	v3 = 9.1 (m)/interval time (s)	a3 = (v3-v2)/interval time
24.4	3.76			
25.9	3.94			
27.4	4.13			

Statistical Analyses

Statistical analyses and all data visualization were also completed in R (R Core Team 2020). Density plots of data distributions at each interval were produced to demonstrate the amount of overlap between seasons and dispersion of data. As described earlier, the CV and SEM have opposite assumptions on proportional bias and the selection of an appropriate measure will likely be dictated by test results (Atkinson and Nevill 1998). Both measures were included for the purposes of this study, but the assumption of homogeneity of variance was evaluated with a Levene's test. Data from each season were used to calculate summary statistics as well as CV %, LOA, intraclass correlation coefficient (ICC), SEM, and SEM_{mdc95}. ICC was calculated via a function in the irr package that utilized an ANOVA model, type of consistency, and 2 raters (season) (Gamer, Lemon, Fellows, and Singh 2019), while other statistics on the calculated acceleration interval data were produced via the equations below.

$$CV\% = \frac{sd}{mean} * 100$$

$$LOA = mean\ difference\ between\ trials \pm sd(difference\ between\ trials) * 1.96$$

$$\mathrm{SEM} = sd * \sqrt{(1 - ICC)}$$

$$SEM_{mdc95} = SEM * \sqrt{2} * 1.96$$

Finally, Bland-Altman LOA plots were included to demonstrate the usage and interpretation of LOA at specific intervals.

RESULTS

Homogeneity of variance was not violated at any of the 3 acceleration intervals between the two seasons of the 310 baseball players, thus the SEM and/or LOA would be more appropriate than the CV with this sample. That being said, CVs are still shown for comparison purposes in this study. Results of the summary descriptive statistics and CV for each season and interval are shown in Table 2. The final interval, a3, appears to be the most variable with the largest CV values (mean of 50.8%). Figure 1 demonstrates the amount of overlap in each season's data distribution, the amount of dispersion for each interval, and that the 3rd interval demonstrates that both positive and negative acceleration values are included. Table 3 shows the results of all measures of variability and absolute reliability (CV, LOA, SEM, and SEM_{mdc95}). Figure 2 depicts the limits of agreement for each season's acceleration data during the final interval, where 95% of data likely vary between 0.34 m/s^2 below and 0.27 m/s^2 above actual values.

Table 2. Summary statistics for 2018 & 2019 MLB accelerations $(\frac{m}{s^2})$ at each interval, represented as mean ± standard deviation (coefficient of variation %)

Year	Interval 1 Acceleration	Interval 2 Acceleration	Interval 3 Acceleration
2018	2.85 ± 0.18 (6.4%)	2.51 ± 0.28 (11.0%)	0.41 ± 0.19 (47.1%)
2019	2.83 ± 0.18 (6.5%)	2.47 ± 0.28 (11.5%)	0.38 ± 0.21 (54.4%)
Total	2.84 ± 0.18 (6.5%)	2.49 ± 0.28 (11.3%)	0.40 ± 0.22 (50.8%)

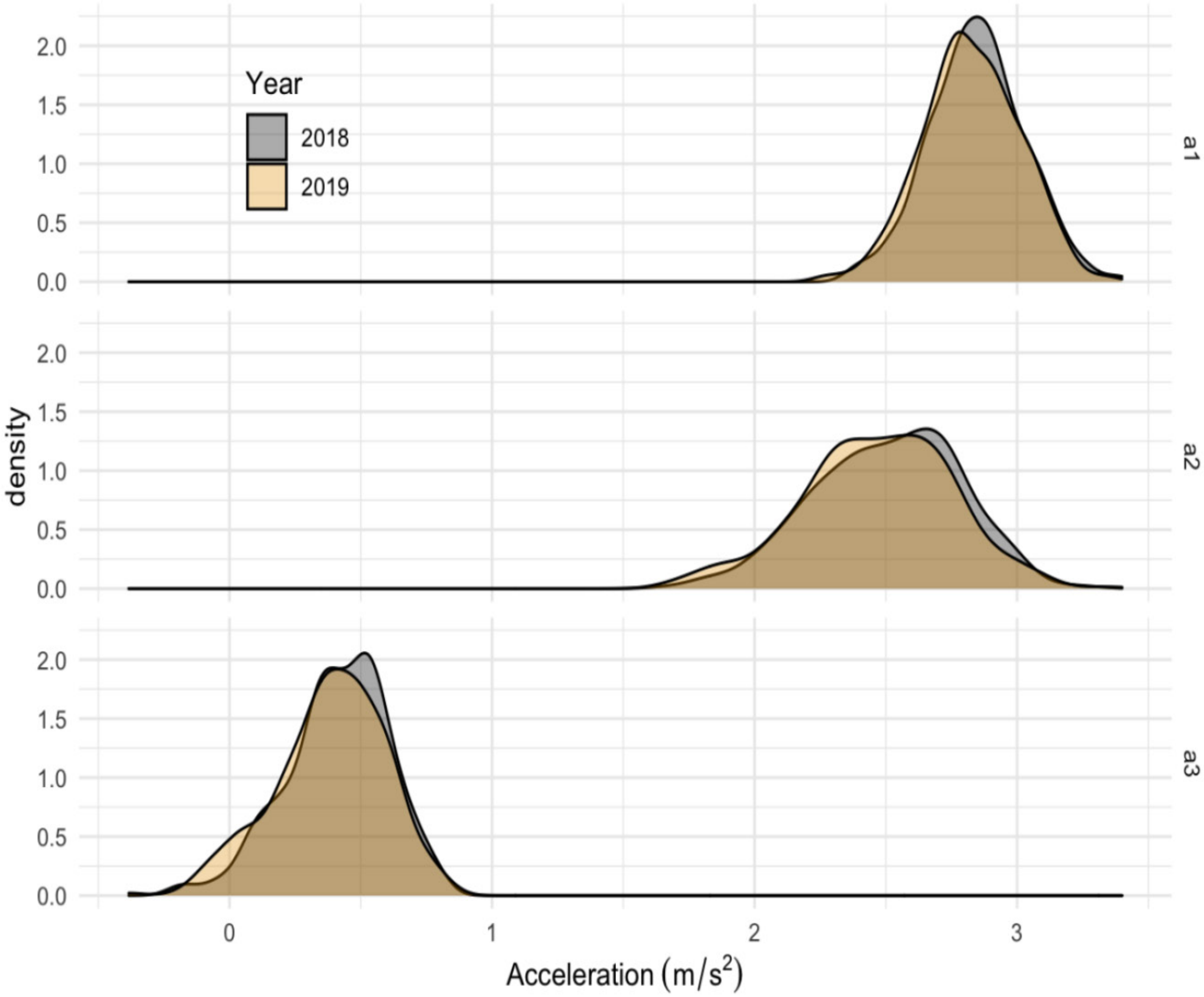

Figure 1. Density plots of data distributions at each interval (a1, a2, a3).

Table 3. Measures of Absolute Reliability, Variation, and Minimum Detectable Change

Interval	CV%	LOA	SEM	SEM_{mdc95}
a1	6.5%	[-0.16,0.11]	0.05	0.14
a2	11.3%	[-0.23,0.15]	0.07	0.20
a3	50.7%	[-0.34,0.27]	0.11	0.32

DISCUSSION

This study sought to demonstrate potential issues when applying statistical procedures designed for ratio (scalar) data on interval (vector) data. The primary finding was that the CV is an inappropriate measure when data include both positive and negative values and should not be used as a result. Additionally, LOA, SEM, and SEM_{mdc95} appear to be more appropriate options as they more appropriately demonstrate the amount of variation and absolute reliability when dealing with vector data.

When initially only considering the mean and sd of the data, it seems evident that the final interval is much more variable than the first two (50.8 % versus 6.5% or 11.3%). Moving out to two standard deviations in the final interval, it becomes apparent that there are some athletes that accelerate negatively (or decelerate) during this stage. This results in a smaller mean value that is approaching zero, causing a disproportionate increase in CV even when the standard deviation value remains somewhat similar. This is also evident in Figure 1. When Figure 1 is also considered along with the calculated CV values, the issue of variability becomes more ambiguous. Figure 1 visually demonstrates similar amounts of overlap between both seasons at all intervals and the amount of dispersion in a3 is similar to a1 despite having a CV that is nearly 8 times as large. As has been discussed previously, this CV issue is due to having a smaller mean caused by possessing negative as well as positive values (Mitchell 1986, Ospina and Marmolejo-Ramos 2019, Velleman and Wilkinson 1993). Similarly high CVs have been reported with acceleration and other research that deals with vector data with negative values (Byrne, Moody, Cooper, Kinsella, and Byrne 2017, Delaney, Cummins, Thornton, and Duthie 2018, Harper, Morin, Carling, and Kiely 2020, Heishman, Daub, Miller, Freitas, Frantz, and Bemben 2020, Howarth, Cohen, McLean, and Coutts 2021, Nibali, Tombleson, Brady, and Wagner 2015, Varley, Fairweather, and Aughey 2012).

When considering the statistics that have the ability work with vector (interval) data, the results appear to lead to more appropriate interpretation. The LOA demonstrates the amount that data can be expected to vary from one trial to the next, or in this case, one season to the next at each interval (Atkinson and Nevill 1998, Bland and Altman 1986). The LOA are the closest at a1 where 95% of data may vary between 0.16 m/s^2 below and 0.11 m/s^2 above the actual value. They are furthest apart at a3, where it may be 0.34 m/s^2 below and 0.27 m/s^2 above. This is depicted in the Bland-Altman plot in Figure 2. It is important to note that this measure does not say if any of the findings are acceptable or not. That decision is left to practitioners to determine what is clinically or practically relevant and this determination should be completed a priori (Giavarina 2015). It has also been argued that 95% is too strict and that some meaningful change may happen within those bands (Hopkins 2000). If practitioners felt the same, they could simply adjust the critical value used when calculating LOA. For example, instead of using 1.96 for a 95% confidence interval, a value of 1.64 could be used to create 90% confidence intervals instead (Harza 2017).

The SEM values found in this study were 0.05 m/s^2, 0.07 m/s^2, and 0.11m/s^2, which can be interpreted as 68% of the variability where 34% lies 0.05 m/s^2, 0.07 m/s^2, or 0.11m/s^2 above and below the measurement for each interval, respectively (Atkinson and Nevill 1998). That is useful, but confidence to the 95% level would be beneficial, which is the purpose of the SEM_{mdc95}, and it can be interpreted more directly. For example, with a SEM_{mdc95} of 0.32 m/s^2 at a3, one could say that differences in values need to be larger than 0.32 m/s^2 to be considered an actual change that is past what might be credited to measurement error (Stratford 2004). Knowing the SEM_{mdc95} may be particularly useful in athlete monitoring by helping practitioners discern between signal and noise.

A potential limitation of this study is that the trial data are separated by season, or approximately one year. This likely would not be suitable for normal test-retest reliability assessments as there are many opportunities for increased variability over the course of an entire year. That being said, the purpose of this study was to demonstrate the issues associated with using common variability and absolute reliability measures, such as the CV, and not necessarily evaluating the reliability of MLB acceleration data, which was used as an example.

CONCLUSION

When measuring absolute reliability or variability, the CV is an appealing option due to its ease of calculation, but it is contraindicated in many situations in sport performance as having ratio or scalar data is a requirement for the CV. As such, the CV likely should not be used with vector data, such as force, acceleration, or velocity, but if it is, practitioners should make sure that the data does not vary on

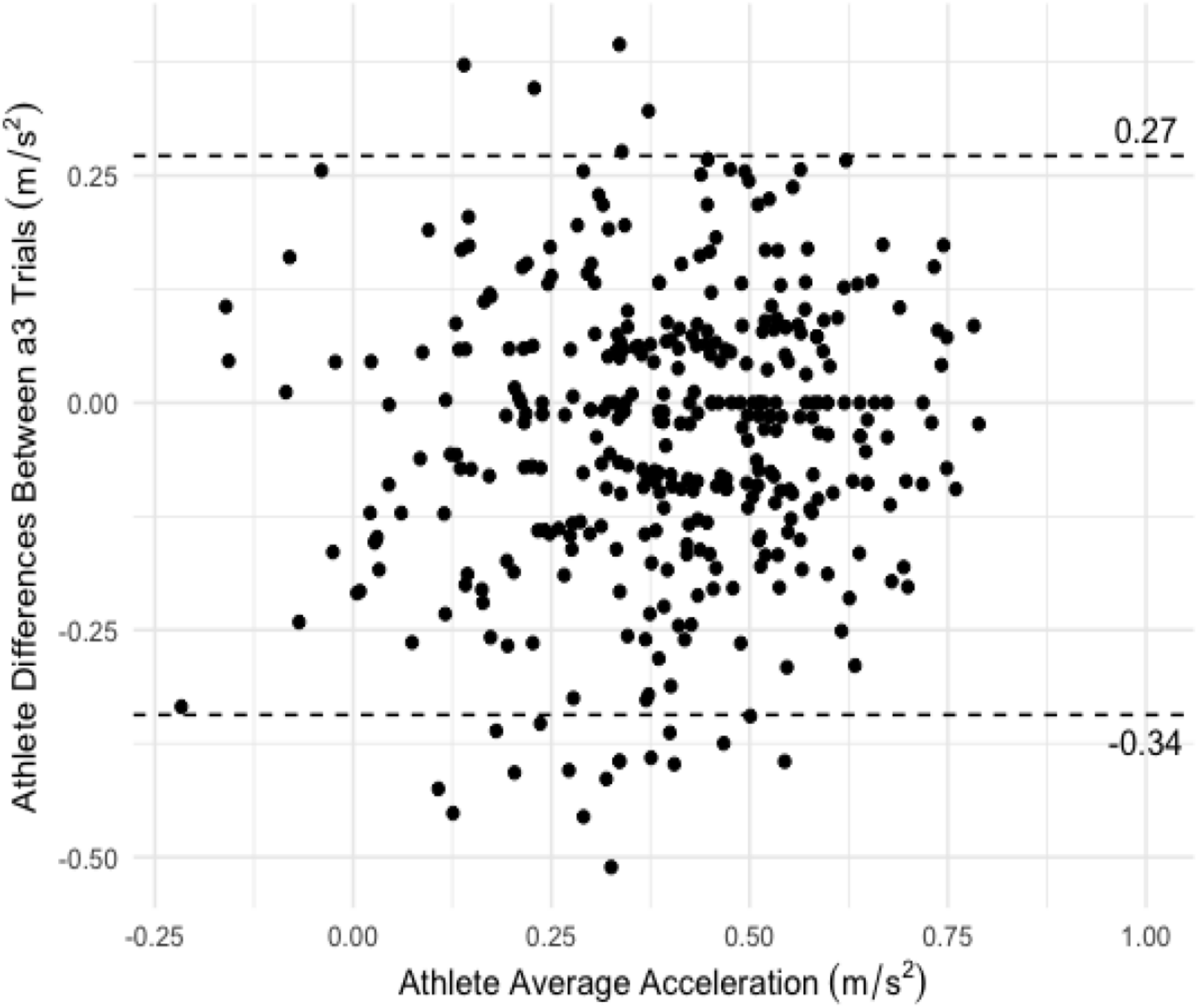

Figure 2. Bland-Altman plot with limits of agreement of acceleration (m/s2) during the final acceleration interval (a3). The x axis shows the average acceleration values between trials for each athlete and the y-axis shows the individual differences between trials.

both sides of zero. The SEM and LOA are appealing alternatives for the CV and should be considered as they can work with positive and negative data. They have the added benefit of staying in the initial unit of measure for easier interpretation. The SEM_{mdc95} should also be considered in athlete monitoring situations when practitioners wish to separate actual change from noise or error.

REFERENCES

1. Allison, DB. 1993. "Limitations of coefficient of variation as index of measurement reliability." Nutrition, 9:559-561.
2. Ashton, J, Jones, P. 2019. "The reliability of using a laser device to assess deceleration ability." Sports, 7: 191. 10.3390/sports7080191
3. Atkinson, G, and A M Nevill. 1998. "Statistical Methods for Assessing Measurement Error (Reliability) in Variables Relevant to Sports Medicine." Sports Med, 26: 217–38. 10.2165/00007256-199826040-00002
4. Bailey, C, Sato, K, McInnis, T. 2021. "A technical report on reliability measurement in asymmetry studies." J strength Cond Res, 35, 1779-1783. 10.1519/JSC.0000000000004024
5. Bland, J, and D Altman. 1986. "Statistical methods for assessing agreement between two methods of clinical measurement." Lancet, 1(8476): 307–310.
6. Byrne, PJ, Moody, J, Cooper, SM, Kinsella, S, Byrne, P. 2017. "The reliability of countermovement jump performance and the reactive strength index in identifying drop-jump height in hurling players." J

Exercise and Sports Medicine, 1:1-10. https://core.ac.uk/outputs/151392020
7. Delaney, J, Cummins, C, Thornton, H, Duthie, G. 2018. "Importance, reliability, and Usefulness of acceleration measures in team sports." J Strength Cond Res, 32:3485-3493. 10.1519/JSC.0000000000001849
8. Eliasziw, M, Young, SL, Woodbury, MG, Fryday-Field, K. 1994. "Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example." Phys Ther, 74: 777-778. 10.1093/ptj/74.8.777
9. Fort-Vanmeerhaeghe, A, Bishop, C, Brusca, B, Aguilera-Castells, J, Vicens-Bordas, J, Gonzalo-Skok, O. 2020. "Inter-limb asymmetries are associated with decrements in physical performance in youth elite team sports athletes." PLOS One, 15: 1–14. 10.1371/journal.pone.0229440
10. Gamer, M, Lemon, J, Fellows, I, and Singh, P. 2019. irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1. https://CRAN.R-project.org/package=irr
11. Giavarina D. 2015. "Understanding Bland Altman analysis." Biochemia medica, 25:141–151. 10.11613/BM.2015.015
12. Harper, D, Morin, JB, Carling, C, Kiely, J. 2020. "Measuring Maximal horizontal deceleration ability using radar technology: reliability and sensitivity of kinematic and kinetic variables." Sports Biomech, Online ahead of print. 10.1080/14763141.2020.1792968
13. Hazra A. 2017. "Using the confidence interval confidently." Journal of Thoracic Disease, 9: 4125–4130. 10.21037/jtd.2017.09.14
14. Heishman, A, Daub, B, Miller, R, Freitas, E, Frantz, B, Bemben, M. 2020. "Countermovement jump reliability performed with and without an arm swing in NCAA Division 1 intercollegiate basketball players." J Strength Cond Res, 34: 546-558. 10.1519/JSC.0000000000002812
15. Hopkins, W, Hawley, J, Burke, L. 1999. "Design and analysis of research on sport performance enhancement." Med Sci Sports Exerc, 31:472-485. 10.1097/00005768-199903000-00018
16. Hopkins, W. 2000. "Measures of Reliability in Sports Medicine and Science." Sports Med, 30:1-15. 10.2165/00007256-200030010-00001
17. Howarth, D, Cohen, D, McLean, B, Coutts, A. 2021. "Establishing the noise: Interday ecological reliability of countermovement jump variables in professional rugby union players." J Strength Cond Res, Online ahead of print. 10.1519/JSC.0000000000004037
18. Lage, M, Ono, JP, Cervone, D, Chiang, D, Dietrich, C, and Silva, CT. 2016. "StatCast Dashboard: Exploration of Spatiotemporal Baseball Data." IEEE Computer Graphics and Applications, 36, 28-37. 10.1109/MCG.2016.101
19. Malcata, RM, Hopkins, WG. 2014. "Variability of Competitive Performance of Elite Athletes: A Systematic Review." Sports Medicine, 44(12), 1763–1774. 10.1007/s40279-014-0239-x
20. Mitchell J. 1986. "Measurement scales and statistics: a clash of paradigms." Psychological Bulletin, 100:398–407. https://doi.org/10.1037/0033-2909.100.3.398
21. Nibali, M, Tombleson, T, Brady, P, Wagner, P. 2015. "Influence of familiarization and competitive level on the reliability of countermovement vertical jump kinetic and kinematic variables." J Strength Cond Res, 29: 2827-2835. 10.1519/JSC.0000000000000964
22. Ospina, R, Marmolejo-Ramos, F. 2019. "Performance of some estimators of relative variability." Frontiers in Applied Mathematics and Statistics, 5, 1-20. https://doi.org/10.3389/fams.2019.00043
23. Pardos-Mainer E, Casajus J, Gonzalo-Skok O. 2019. "Adolescent female soccer players' soccer specific warm-up effects on performance and inter-limb asymmetries." Biol Sport, 36: 199–207. 10.5114/biolsport.2019.85453
24. Pareek, A, Parkes, CW, Leontovich, AA, Krych, AJ, Conte, S, Steubs, JA, Wulf, CA, & Camp, CL. 2021. "Are Baseball Statistics an Appropriate Tool for Assessing Return to Play in Injured Pitchers? Analysis of Statistical Variability in Healthy Players." Orthopaedic journal of sports medicine, 9(11). 10.1177/23259671211050933
25. R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
26. Read, P, Oliver, J, Myer, G, De Ste Croix, M. 2017. "The effects of maturation on measures of asymmetry during neuromuscular control tests in elite male youth soccer players." Pediatric Exercise Science, 30: 1–23. 10.1123/pes.2017-0081
27. Stratford PW, Goldsmith, CH. 1997. "Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data." Phys Ther, 77:745-750. 10.1093/ptj/77.7.745
28. Stratford, PW. 2004. "Getting More from the Literature: Estimating the Standard Error of Measurement from Reliability Studies." Physiotherapy Canada, 56. 10.2310/6640.2004.15377
29. Thomas, J, Nelson, J, & Silverman, S. 2015. Research Methods in Physical Activity (7th Edition). Champaign, IL: Human Kinetics.
30. Varley, M, Fairweather, I, Aughey, R. 2012. "Validity and reliability of GPS for measuring instantaneous velocity during acceleration, deceleration, and constant motion." J Sport Sciences, 30:121-127. 10.1080/02640414.2011.627941
31. Velleman PF, Wilkinson L. 1993. "Nominal, ordinal, interval, and ratio typologies are misleading." Am Stat, 47:65–72. 10.1080/00031305.1993.10475938
32. Vincent, W, & Weir, J. 2012. Statistics in Kinesiology (4th Edition). Champaign, IL: Human Kinetics.
33. Weir J. 2005. "Quantifying test-retest reliability using the intra-class correlation coefficient and the SEM." J Strength Cond Res, 19:231-40. 10.1519/15184.1
34. Yao, L, Sayre, J. 1994. "Statistical Concepts in the interaction of serial bone densitometry." Investigative Radiology, 29:928-932. 10.1097/00004424-199410000-00011