

# The Reliability and Validity of Different Methods for Measuring Countermovement Jump Height

Bailey Cameron<sup>1</sup> , James Steele<sup>1</sup> , Lee Bridgeman<sup>1\*</sup> 

<sup>1</sup>Department of Sport and Health, Solent University, UK

\*Correspondence: [lee.bridgeman@solent.ac.uk](mailto:lee.bridgeman@solent.ac.uk)

Cite as: Bailey Cameron et al. (2024). The Reliability and Validity of Different Methods for Measuring Countermovement Jump Height. *SportRxiv*. DOI: 10.51224/SRXIV.430

Supplementary Materials: <https://osf.io/e7qg5/>

## Abstract

Previous research indicates the importance of the countermovement jump (CMJ) test to monitor lower-limb power and neuromuscular fatigue. While jump height (JH) can be measured using various equipment, this study compared the JH obtained from the Just Jump System (JJS), OptoJump and the My Jump Lab app against the Vald ForceDecks system using the impulse-momentum calculation method, which is regarded as the gold standard method to calculate JH. This study also assessed the one-week test-retest reliability of these pieces of equipment. The participants in this study were 20 ( $n = 12$  male and  $n = 8$  female) university sports students and staff (mean  $\pm$  SD; age:  $20.90 \pm 2.63$  years; stature:  $1.76 \pm 0.10$  m; mass:  $72.17 \pm 11.07$  kg). Participants completed a standardised warm-up and rested for three minutes before completing three CMJs on each piece of equipment in a randomised, counterbalanced order. The same protocols were used in the second session, with a different equipment testing order. Both My Jump Lab and OptoJump have high agreement levels (Mean bias and 95% CI = 2.32 cm [1.57 – 3.09] and 1.92 cm [1.23 – 2.59], respectively) with the gold measurement standard (ForceDecks using IM). However, a high mean bias for the JJS (Mean bias = 9.88 cm [9.26 – 10.46]) was reported. This study also found that all methods are reliable for assessing JH (Mean bias and [95% CI]: ForceDecks = 0.24 cm [-0.47 – 0.92], JJS = 0.74 cm [0.08 -1.42], My Jump Lab = 0.05 cm [-0.57 – 0.71] and Optojump = -0.14 cm [-0.77-0.49]). Overall, the equipment investigated in this study showed high levels of reliability, and only the JJS had low validity compared to the ForceDecks. Coaches should consider what data they want to collect, its validity and reliability, the purpose of the

testing and the cost of the equipment when deciding which system to purchase.

Keywords: *Jumping; ForceDecks; OptoJump; Just Jump System; My Jump Lab*

## Introduction

The measurement of jump height (JH), specifically during the countermovement jump (CMJ), is widely used as a non-invasive method to assess neuromuscular fatigue (NMF) (Akyildiz et al., 2022; Claudino et al., 2017; Hughes et al., 2019), profile lower body power (Nuzzo et al., 2008), and identify talent (Pino-Ortega et al., 2021). In football players, it was found that both CMJ height and maximum velocity during the movement were significantly different between 24 hours pre-match and 24 hours post-match testing ( $p < 0.001$ , Cohen's  $d = 1.578$ ;  $p < 0.001$ , Cohen's  $d = 1.210$ , respectively); significant differences in CMJ height were also observed for up to 48 hours post-match ( $p < 0.01$ , Cohen's  $d = 0.922$ ), indicating that NMF can be monitored using CMJ height with the greatest decrease in performance occurring 24 hours post-match (Akyildiz et al., 2022). CMJ height can also assess lower body power and its relationships with other variables. Nuzzo et al. (Nuzzo et al., 2011) reported a significant relationship between CMJ height and relative squat one repetition maximum ( $r = 0.69$ ,  $p < 0.05$ ). Additionally, a significant relationship was found between CMJ height and sprint times for 10m and 30m ( $r = 0.72$ ,  $p < 0.001$ ;  $r = 0.60$ ,  $p < 0.001$ , respectively). However, more research is required to link CMJ height directly to performance in competition (Wisløff et al., 2004). Thus, it is proposed that the CMJ, particularly JH as an outcome measure, is of interest to coaches and athletes.

Several methods exist for measuring CMJ height, including force platforms (FP), jump and reach tests, 3D motion-capture systems, and calculations from flight time using contact mats or photoelectric cells (Glatthorn et al., 2011; Whitmer et al., 2015). Because there are many ways to measure JH, assessing the reliability and validity of the equipment used is crucial. A dual FP system using the impulse-momentum (IM) calculation is widely regarded as the gold standard for measuring JH in the CMJ (Heredia-Jimenez & Orantes-Gonzalez, 2020; Xu et al., 2023). Force plates can also measure JH utilising a variety of calculation methods such as flight time (FT), displacement, take-off velocity (TV), work energy, and take-off velocity added to the centre of mass at take-off (COM) and the double-integration method (DI) which utilises a time-displacement curve (Chiu & Dæhlin, 2020; Linthorne, 2001; Xu et al., 2023). However, due to the number of different calculations available, there is debate over which method provides the most valid and reliable results (Chiu & Dæhlin, 2020). A comparison study that utilised five different calculation methods found that FT and TV had the highest systematic and random error occurrence. However, this study utilised a single-plate rather than dual FP systems, which are now commonplace (Chiu & Dæhlin, 2020). A systematic review examining different calculation methods for the CMJ and drop-jump found that the FP IM method was the most appropriate way to measure JH compared to jump and reach tests, motion capture and flight time calculations (Xu et al., 2023). It has also been reported that FPs,

specifically the VALD ForceDecks, have high reliability (ICC = 0.93) when using the IM method to calculate CMJ height (Merrigan et al., 2024). A further study by Collings et al. (2024) also reported a 5% relative difference between ForceDecks and an embedded laboratory FP system and excellent test-retest reliability (ICC = 0.97 [0.92-0.99]) for JH.

OptoJump is another way to measure JH; this is done via photoelectric cells. FT is recorded via beams, which connect as an athlete leaves the ground during the flight phase and are broken again upon landing, which allows JH to be calculated (Comyns et al., 2023; Dobbin et al., 2017). One study that compared OptoJump to FP data found that there was a large systematic bias and OptoJump produced significantly lower JH measurements (-31% to -27%;  $p < 0.001$ ), but this was in comparison to measurements from a single FP, so it cannot be assumed that the same error would be found when using a dual FP system utilising the IM calculation (Attia et al., 2017). Another study compared OptoJump to a FP and reported a strong relationship but a large difference ( $r = 0.90$ ;  $ES = 2.75$ ) between JH (Rago et al., 2018). Similar findings were presented in a study by Glatthorn et al. (2011), who found a consistent, systematic difference between OptoJump and a single FP system (-1.06 cm;  $p < 0.001$ ). Despite the underestimation of JH compared to the FP measurements, high test-retest reliability was observed in this study (ICC = 0.987), indicating that OptoJump is a piece of equipment with high reliability despite being less accurate than FP (Glatthorn et al., 2011).

Another method for JH measurement using FT is via contact mats (CM), such as the Just Jump System (JJS). It is widely reported in the literature that CMs consistently overestimate flight time by up to 105 milliseconds and thus also overestimate JH during the countermovement jump by up to 18 cm (Dobbin et al., 2017; Whitmer et al., 2015). During a comparison between the JJS and a FP system, it was found that the JJS produces significantly greater JH values ( $p < 0.001$ ,  $d = 1.39$ ) and the following correction equation was proposed: Criterion jump height =  $(0.8747 \times \text{alternative jump height}) - 0.0666$  (McMahon et al., 2016). When consecutive jumps were compared, high reliability was reported for CM (ICC = 0.999-1.000). However, this study did not assess inter-session reliability (Pueo et al., 2020). Based on this, it appears that although CMs are reliable, they produce inflated JH results.

The final method utilised in this study to calculate JH is the mobile My Jump Lab app. This uses video filmed via a smartphone or tablet and requires the researcher to select take-off and landing frames post-jump so JH can be calculated from FT (Bogataj et al., 2020). A previous version of the app, was investigated by Yingling et al. (2018) to assess the reliability relative to values collected from Vertec. The study found a moderate to excellent degree of consistency for JH (ICC = 0.813) and peak power (ICC = 0.926), yet poor absolute agreement was found for JH (ICC = 0.665; 95% CI [0.050-0.859]), however as this was a comparison to a jump and reach test, it is not representative of measurements for the CMJ (Yingling et al., 2018). A meta-analysis investigating My Jump Lab found near-perfect reliability ( $r = 0.986$ ) for JH but highlighted no current comparisons to the gold measurement standard to assess validity (Gençoğlu et al., 2023). Thus, this is an area that this research aims to address.

One of the objectives of this study was to compare the equipment used above to measure CMJ height and provide insights into the validity and reliability of each system while considering their practical benefits and limitations. This cannot be achieved in isolation without consid-

eration for the affordability of these pieces of equipment. Force plates such as the VALD ForceDecks are expensive (note all prices correct at the time of writing), with most systems costing around  $\sim$ £20,000, and although the costs of these portable systems have reduced over recent years, they still may not be accessible to teams and individuals working with a small budget. Optojump, which may not be as accurate as the gold standard method but has a high test-retest reliability, costs significantly less than force plates, at around  $\sim$ £3000. While the JJS, which may overestimate JH but is reliable, costs  $\sim$ £1000, and the cheapest method utilised in this study to measure JH is the My Jump Lab app, which is available as part of a suite of apps based on a subscription model (1 month =£4.99, 1 year = £34.99, Lifetime = £99.99 – prices correct at time of writing). Therefore, when considering cost, choosing equipment with high reliability yet lower validity compared to the gold standard method may be appropriate if coaches wish to monitor changes over time, such as measuring JH to assess NMF post-performance or adaptations during and after a training block.

According to a systematic review by Xu et al. (2023) more investigation is needed to determine the reliability and validity of JH measurement and calculation methods utilising apparatus with various settings or sampling frequencies. This study aims to estimate the validity and reliability of CMJ measurements using the previously mentioned equipment. Although it is acknowledged that cost may influence decisions, this study sets out to educate coaches about the benefits and drawbacks of each approach. Based on previous research, we expect that, when compared to the IM method, all equipment will have estimates suggesting high reliability but varying validity, with the JJS likely to exhibit the lowest concurrent validity.

## Method

### Participants

The participants in this study were 20 ( $n = 12$  male and  $n = 8$  female) university sport students and staff (mean  $\pm$  SD; age:  $20.90 \pm 2.63$  years; stature:  $1.76 \pm 0.10$  m; mass:  $72.17 \pm 11.07$  kg). Recruitment was done via a convenient sampling approach aimed at participants aged 18-35 free from any current lower-limb injury or major lower-limb surgery in the six months preceding testing. This project was conducted as part of an undergraduate dissertation conducted by the lead author and thus sample size was limited due to resource constraints. However, since we have made our data open and focused on estimation, communicating these results is justified as they may contribute to future meta-analyses. Further, our sample size aligns with typical sample sizes in the field (Lakens, 2022). We also utilised multiple measurements per method per session per participant, thus increasing the effective sample size.

### Procedures and protocols

Participants were asked not to consume caffeine or eat two hours before testing. In session one, participants completed a standardised warm-up consisting of five minutes on a cycle ergometer (Monark, Ergomedic 874e) at 100 W, ten bodyweight squats and three CMJs; participants then rested for three minutes. The equipment order was randomly assigned to each participant

in a counterbalanced design. Three CMJs were performed on each of the following: FDLite force plates (ForceDecks, VALD, Brisbane, Australia) sampling at 1000 Hz, Optojump Next (Microgate, Italy), Just Jump System (Probotics Inc, USA) and My Jump Lab (v.4.2.8, 2024) downloaded on a phone (iPhone 15 Pro Max, Apple, California) which recorded at 240 frames per second. Participants were given 30 seconds of rest between jumps and three minutes of rest between the different pieces of equipment. The jumps were standardised by ensuring a countermovement depth of  $\sim 90$  degrees, which was visually confirmed by the tester, and participants were instructed to keep their hands on their hips and jump as high as possible each time.

The ForceDecks system was used to measure JH using the IM method, and the plates were zeroed before the participants stepped on. Then, the participant was instructed to place one foot on each plate and stand still to record their weight before jumping. Optojump and the JJS recorded JH via FT. The correction factor for the JJS was applied to every jump before data analysis (Criterion jump height =  $[0.8747 \times \text{alternative jump height}] - 0.06660$  (McMahon et al., 2016). My Jump Lab also calculated JH using FT. Two researchers (BC and LB) selected and confirmed take-off and landing frames to ensure agreement. According to the My Jump Lab manufacturer's guidance, take-off is the first frame where both feet are in the air, and landing is the first frame where at least one foot touches the ground.

The second session was conducted at the same time ( $\sim 1$  hour) one week later to ensure results were not impacted by variations in circadian rhythm, and the same protocols were implemented with a randomised order of equipment again. This allows for reliability to be measured between the two sessions. Validity was assessed between equipment across both sessions, allowing us to determine consistency in agreement between methods.

## Data Analysis

The present analysis was not pre-registered as we had no a priori hypotheses and, given the limited sample size due to resource constraints, was considered exploratory. Inferential statistics were treated as highly unstable local descriptions of the relations between model assumptions and data in order to acknowledge the inherent uncertainty in drawing generalised inferences from single and small samples (Amrhein et al., 2019). For all analyses we opted to take an estimation-based approach typical when examining validity and reliability. That is, we provide point estimates and the uncertainty in them for the statistical parameters reported. Two sets of models were employed exploring the JHs recorded from CMJ trials; one to examine the agreement between the IM method using the Force Decks (i.e., gold-standard) and each of the other methods, and one to explore the test-retest reliability for each method. Given we had nested data whereby each participant provided three trials for each method on two separate testing sessions we adopted a mixed effects limits of agreement approach (Parker et al., 2020). This allowed us to estimate mean bias for each method compared with the gold-standard for agreement utilising both first and second session data, and any test retest bias for reliability, in addition to upper and lower limits of agreement providing 95% coverage probabilities for both agreement and test-retest reliability. Models were fit using the `lme4` package and using Restricted Maximum Likelihood Estimation. For each of the models fit

we used nonparametric case based bootstrapping resampling 10000 times at the individual participant level and refitting models in order to construct 95% quantile intervals for both the mean bias and limits of agreement estimates. Bootstrapping was performed using the `lmeresampler` package.

### Agreement

For estimation of bias and limits of agreement regarding the agreement between the gold-standard and each device we utilised the following mixed effects model structure:

$$\begin{aligned} D_{ist} &= \mu + \alpha_i + \gamma_s + \epsilon_{ist} \\ \alpha_i &\sim N(0, \sigma_\alpha^2) \\ \gamma_s &\sim N(0, \sigma_\gamma^2) \\ \epsilon_{ist} &\sim N(0, \sigma_\epsilon^2) \end{aligned}$$

Where  $D_{ist}$  is the difference between measurements taken between the two devices (i.e.,  $y_{ist2} - y_{ist1}$ ), where the device indexed by 1 is the gold-standard and device indexed by 2 is the comparison device, for participant  $i$  during session  $s$  and for trial  $t$ . Here  $\mu$  is the overall mean of the between device differences (i.e., the mean bias),  $\alpha_i$  is the random effect for the  $i^{th}$  participant,  $\gamma_s$  is the random effect for the  $s^{th}$  session which is nested within participant, and  $\epsilon_{ist}$  is the error term. The 95% limits of agreement can then be calculated as:

$$\mu \pm 1.96 \sqrt{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}$$

with the square root of the total variance providing an estimate of the standard deviation of the differences for use in the conventional Bland-Altman limits of agreement calculation. A model was fit for each device in comparison o the gold-standard.

### Reliability

For estimation of bias and limits of agreement regarding the test-retest reliability between each test session for each device we utilised the following mixed effects model structure:

$$\begin{aligned} D_{it}^* &= \mu^* + \alpha_i^* + \epsilon_{it}^* \\ \alpha_i^* &\sim N(0, \sigma_{\alpha^*}^2) \\ \epsilon_{it}^* &\sim N(0, \sigma_{\epsilon^*}^2) \end{aligned}$$

Where  $D_{it}^*$  is the difference between measurements taken between the two sessions for a given device (i.e.,  $y_{it2} - y_{it1}$ ), where the session indexed by 1 is the first test session and the session indexed by 2 is the second test session, for participant  $i$  and for trial  $t$  (note we use the superscript  $*$  to distinguish this from the agreement model). Here  $\mu^*$  is the overall mean of the between session differences (i.e., the mean bias),  $\alpha_i^*$  is the random effect for the  $i^{th}$  participant, and  $\epsilon_{it}^*$  is the error term. The 95% limits of agreement can then be calculated as:

$$\mu^* \pm 1.96\sqrt{\sigma_{\alpha^*}^2 + \sigma_{\epsilon^*}^2}$$

with the square root of the total variance providing an estimate of the standard deviation of the differences for use in the conventional Bland-Altman limits of agreement calculation. A model was fit for each device in order to examine its test-retest reliability.

### Open data, code, and materials

All data and code is presented in the supplementary materials (<https://osf.io/e7qq5/>). The `renv` package was used for package version reproducibility and a function based analysis pipeline using the `targets` package was employed (the analysis pipeline can be viewed by downloading the R Project and running the function `targets::tar_visnetwork()`). We cite all packages used in the analysis pipeline below.

We used R version 4.3.3 (R Core Team, 2024) and the following R packages: `glue` v. 1.7.0 (Hester & Bryan, 2024), `knitr` v. 1.47 (Xie, 2014, 2015, 2024), `lme4` v. 1.1.35.3 (Bates et al., 2015), `lmeresampler` v. 0.2.4 (Loy et al., 2024), `patchwork` v. 1.2.0 (Pedersen, 2024), `quarto` v. 1.4 ([quarto?](#)), `renv` v. 1.0.3 (Ushey & Wickham, 2023), `rmarkdown` v. 2.27 (Allaire et al., 2024; Xie et al., 2018, 2020), `tarchetypes` v. 0.9.0 (Landau, 2021a), `targets` v. 1.7.0 (Landau, 2021b), `tidyverse` v. 2.0.0 (Wickham et al., 2019).

## Results

Table 1: Mean  $\pm$  SD jump height (cm) for each piece of equipment in session 1 and 2

	<b>ForceDecks</b>	<b>OptoJump</b>	<b>Just Jump System</b>	<b>My Jump Lab</b>
Session 1 Jump Height (mean $\pm$ SD)	28.65 $\pm$ 7.35 cm	30.76 $\pm$ 6.97 cm	38.28 $\pm$ 7.66 cm	31.06 $\pm$ 6.79 cm
Session 2 Jump Height (mean $\pm$ SD)	28.88 $\pm$ 6.97 cm	30.62 $\pm$ 6.83 cm	39.02 $\pm$ 7.30 cm	31.11 $\pm$ 6.77 cm

Table 1 shows the mean  $\pm$  SD CMJ heights for each piece of equipment in both sessions.

### Agreement

The exact mean bias and limits of agreement along with their bootstrapped interval estimates can be seen for the agreement of each device in comparison to the gold-standard (impulse momentum method using the Force Decks) in Figure 1. Both the My Jump Lab App and the Optojump demonstrated very similar degrees of mean bias of around  $\sim 2$  cm overestimation, and a similar width to their limits of agreement ranging  $\sim 11$ - $12$  cm about the mean bias. The Jump Mat however had a much larger mean bias of  $\sim 10$  cm overestimation yet a similar range for their limits of agreement as the other devices.

## Agreement Between Methods

Mixed Effects Model Mean Bias and 95% Limits of Agreement

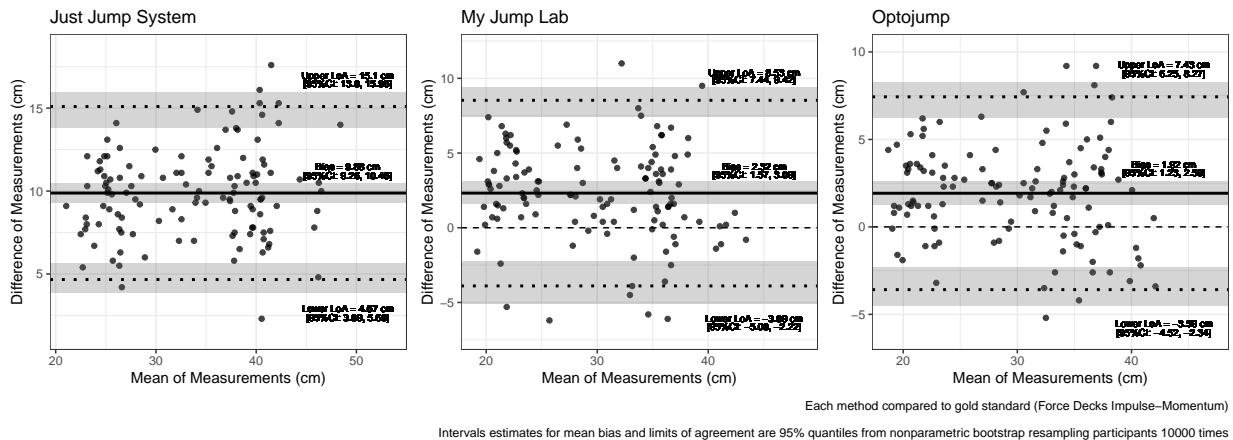


Figure 1: Mean bias (thick horizontal line) and 95% limits of agreement (dotted horizontal lines) along with 95% quantile interval estimates for agreement of each device in comparison to the gold-standard (Force Decks impulse-momentum method).

## Reliability

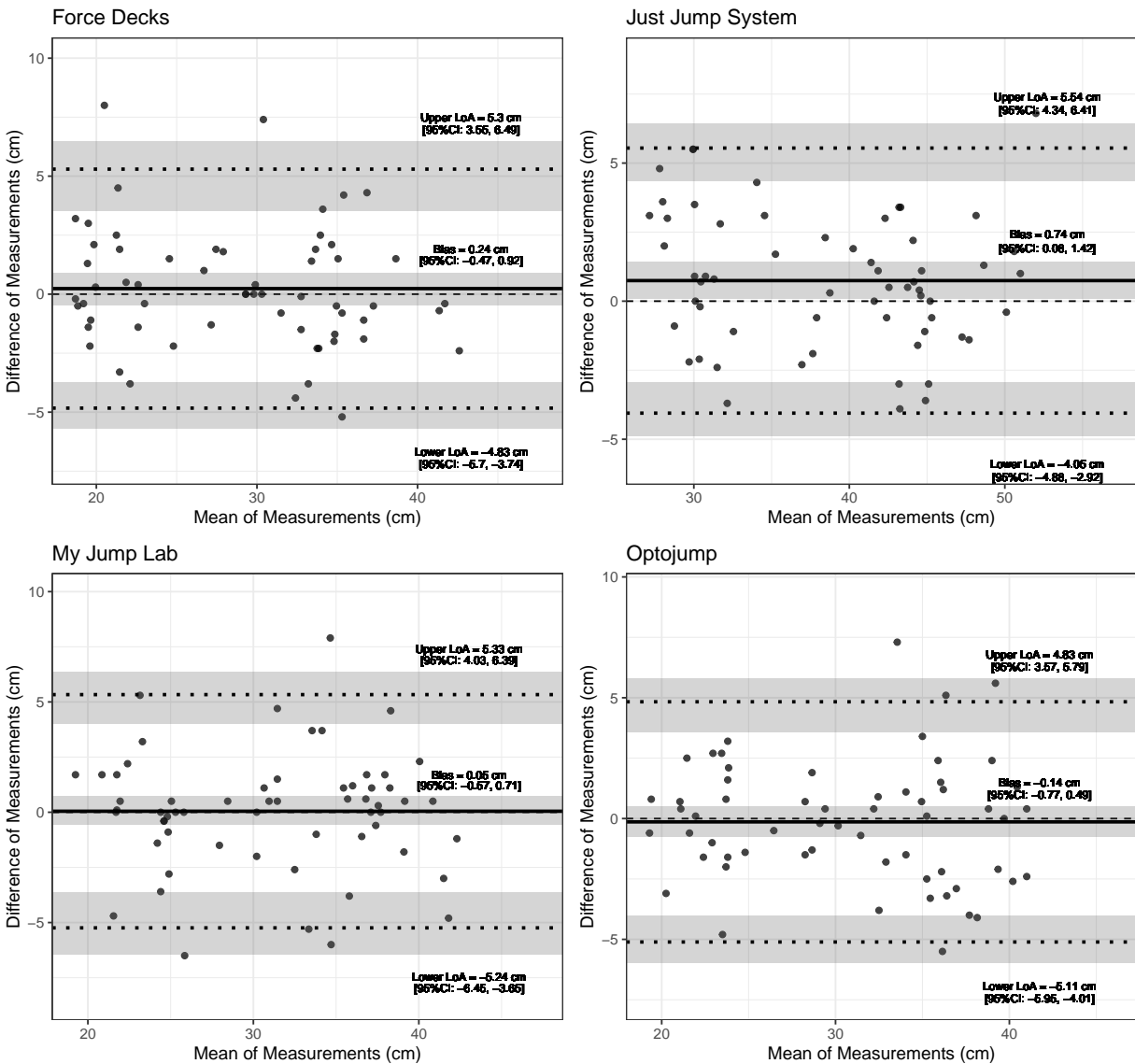
The exact mean bias and limits of agreement along with their bootstrapped interval estimates can be seen for the test-retest reliability of each device between sessions in Figure 2. All devices demonstrated minimal mean bias between sessions each typically less than 1 cm, and all demonstrated a similar width to their limits of agreement ranging  $\sim 10$  cm about the mean bias.

## Discussion

This study compared the validity and reliability of different measurement methods for CMJ jump height. The main finding of this study indicates that all methods are reliable (Mean bias [95% CI] and LoAs [95% CI]: ForceDecks = 0.24 cm [-0.47, 0.92], lower LoA = -4.83 cm [-5.70, -3.74], upper LoA 5.33 cm [3.55, 6.49]; JJS = 0.74 cm [0.08 -1.42], lower LoA = -4.05 cm [-4.88, -2.92], upper LoA 5.54 cm [4.34, 6.41]; My Jump Lab = 0.05 cm [-0.57, 0.71], lower LoA = -5.24 cm [-6.45, -3.65], upper LoA 5.33 cm [4.03, 6.39]; and Optojump = -0.14 cm [-0.77-0.49], lower LoA = -5.11 cm [-5.95, -4.01], upper LoA 4.83 cm [3.57, 5.79]). Both My Jump Lab and OptoJump have high agreement levels (Mean bias with 95% CI and LoA = 2.32 cm [1.57, 3.09], lower LoA = -3.89 cm [-5.08, -2.22], upper LoA 8.53 cm [7.44, 9.42]; and 1.92 cm [1.23, 2.59], lower LoA = -3.59 cm [-4.52, -2.34], upper LoA 7.43 cm [6.25, 8.27] respectively) with the gold measurement standard (ForceDecks using IM). There was, however, a high mean bias for the JJS (Mean bias = 9.88 cm [9.26, 10.46], lower LoA = 4.67 cm [3.89, 5.68], upper LoA 15.1 cm [13.8, 15.96]).



Test-Retest Reliability Across Methods  
 Mixed Effects Model Mean Bias and 95% Limits of Agreement



Intervals estimates for mean bias and limits of agreement are 95% quantiles from nonparametric bootstrap resampling participants 10000 times

Figure 2: Mean bias (thick horizontal line) and 95% limits of agreement (dotted horizontal lines) along with 95% quantile interval estimates for each the test-retest reliability of device between sessions.

## **ForceDecks**

The ForceDecks, using the IM calculation method, were used as the gold measurement standard for this study due to the findings of previous research that compared all calculation methods (Xu et al., 2023). There is extensive research into the validity and reliability of FP, with the accepted view being that they are both highly reliable and valid when measuring JH during a CMJ (Chiu & Dæhlin, 2020; Merrigan et al., 2024; Xu et al., 2023). The present study concurs with the previous findings as test-re-test reliability was high. One consideration when using FP in a practical context is the cost. It is the most expensive equipment used in this study and, therefore, may be beyond the financial reach of some teams. However, FP can produce more metrics than simply JH and FT (Lombard et al., 2017). This allows users to measure outcome variables and evaluate an athlete's strategies to achieve these. This may be particularly important when monitoring NMF as it has been shown that athletes can mitigate reductions in JH despite not fully recovering (Gathercole et al., 2015) by changing their jumping strategy (e.g., longer eccentric duration or change in CMJ depth). This means that when deciding what equipment to use when measuring CMJ, it is essential also to determine what data is required to assess the necessary components of fitness or NMF.

## **OptoJump**

Previous research found that OptoJump had lower validity compared to FP systems but high reliability (Attia et al., 2017; Glatthorn et al., 2011). The present study observed high reliability and, in contrast to earlier studies, found that OptoJump overestimated JH (Attia et al., 2017; Glatthorn et al., 2011). This may be because the previous studies compared the OptoJump system to other methods of measurement that were not the gold standard. For example, Glatthorn et al. (2011) used a single-plate system with a sampling frequency of 500Hz and used the FT calculation method, which is not representative of the gold standard of measurement and may have impacted the results, causing contrasting findings to the present study. The validity and reliability scores found in the present study suggest that while OptoJump is not the gold standard of measurement, it may be appropriate for monitoring changes over time. The measured value will be close to the true value for JH, and it is a relatively cheap piece of equipment to buy and is easily portable, which may make it more desirable for lower-level sports teams. However, OptoJump can only produce two metrics for each jump (FT and JH), and in cases such as monitoring NMF, other variables such as maximum velocity and countermovement depth may provide coaches with more information.

## **Just Jump System**

Contact mat systems have also been previously reported to have low validity and high reliability (McMahon et al., 2016; Pueo et al., 2020). Even with the correction factor applied, the present study still observed an overestimation for JH, resulting in low validity. To create the correction factor, McMahon et al. (2016) utilised a FP sampling at 600Hz and used FT to calculate JH for both the FP and the JSS. This may be why the correction equation did not exhibit high agreement with the gold standard in the present study. The present findings suggest that

creating a new correction factor that applies to the IM method is warranted. Whitmer et al. (2015) reported that a JM produced large overestimations of JH by up to 18 cm compared to FP data. The present study, although not of the same magnitude, also found large overestimations with a mean bias of 9.88 cm but agreed with the previous findings of high test-retest reliability (Nuzzo et al., 2011). In summary, while the JM overestimates both FT and JH, they are a reliable and relatively inexpensive method of determining changes to jump performance over time; the JSS is also practical for use in a field setting.

## **My Jump Lab**

It was previously reported that the My Jump Lab app has high reliability, but little research has been done into the validity of the app compared to the gold standard (Gençoğlu et al., 2023). The present findings indicate that the My Jump Lab app reliably measures JH (mean bias = 0.05 cm) but, as with the other equipment, overestimates JH compared to the ForceDecks system (mean bias = 2.32 cm). A potential issue with the app is the manual processing of the jumps, which involves visually identifying take-off and touch-down for each athlete, which is time-consuming, especially with a large squad of players. However, promisingly, the My Jump Lab app released a new update after the data collection for the present study. This update includes an 'AI mode' which automatically measures jump height as the participant jumps, removing the need to select take-off and touch-down frames. When compared with the ForceDecks, a near-perfect correlation ( $r = 0.968$ ,  $p = 0.001$ ) was found, as well as high intrasession reliability (SEM = 0.42 cm; CV, 1.21%), indicating that the AI mode is also an accurate and reliable measurement of JH for the CMJ (Şentürk et al., 2024). However, this study provided two-minute rest periods, which is not representative of how the equipment would be used as a screening process in team sports (Asmussen & Bonde-Petersen, 1974). Therefore, further research with shorter recovery periods that are more representative of what happens in team sport testing is suggested to be required to confirm these findings. The My Jump Lab app also advertises the ability to calculate metrics other than FT and JH, such as peak power and take-off velocity. However, there is currently limited research into the validity and reliability of these measures, indicating a future direction for investigation. The app is the cheapest method of measurement included in this study and is adaptable to field-based testing; the high reliability and validity reported suggest that coaches could use the app to measure outcome metrics accurately and consistently. However, with the same caveat as the other equipment, it only measures a small number of metrics compared to FP, which can measure both kinematics and kinetics.

## **Limitations**

One of the limitations of this study is that the jumps were unable to be measured by all equipment simultaneously due to the number of different methods investigated. If JH could be measured on all pieces of equipment simultaneously, the agreement between equipment would likely be higher as there is no variation in JH between the jumps for one participant on different equipment. Another limitation of the study is the small sample size; resource

and time constraints made it impossible to obtain a greater sample, thus warranting further investigation on a larger population.

## **Conclusion**

To conclude, all the equipment investigated in this study showed high levels of reliability. The FP using the IM calculation method remains the gold standard. However, the cost implications and which variables coaches require when assessing their team/athlete's performance must be considered. The My Jump Lab app had promising validity levels for JH and is the cheapest available option for JH measurement. OptoJump also had similar validity but remains relatively costly, considering only a small number of metrics can be calculated during jumping. However, it is acknowledged that OptoJump can also be used to assess other variables, such as running gait. The JSS had low validity but high reliability. In conclusion, coaches and athletes must consider what data they want to collect and the equipment's cost when deciding which measurement method to purchase. Overall, this study indicates that all of the investigated measurement methods are appropriate for monitoring changes in CMJ height over time, but the JSS may not be suitable for ascertaining the absolute JH measurement.

## **Future Recommendations**

The My Jump Lab app provides additional metrics from the jump, such as velocities, so it may be helpful to investigate the reliability and validity of these measures compared to FP systems, especially as these variables can be used to monitor NMF. Further research into the app's new AI feature is also needed to validate this new approach fully.

## **Contributions**

BC and LB conceived and designed the study, acquired the data, interpreted the results, and drafted and revised the manuscript. JS completed the data analysis and contributed to interpreting the results and manuscript revisions. All authors provided final approval of the version to be published.

## **Acknowledgements**

The authors wish to thank the participants for taking part in this study.

## **Data and Supplementary Material Accessibility**

All data and code is presented in the supplementary materials (<https://osf.io/e7qg5/>).

## References

- Akyildiz, Z., Ocak, Y., Clemente, F. M., Birgonul, Y., Günay, M., & Nobari, H. (2022). Monitoring the post-match neuromuscular fatigue of young Turkish football players [Journal {{Article}}]. *Scientific Reports*, *12*(1), 13835.
- Allaire, J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2024). *rmarkdown: Dynamic documents for R* [Manual].
- Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication [Journal {{Article}}]. *The American Statistician*, *73*(sup1), 262–270.
- Asmussen, E., & Bonde-Petersen, F. (1974). Storage of elastic energy in skeletal muscles in man [Journal {{Article}}]. *Acta Physiologica Scandinavica*, *91*(3), 385–392.
- Attia, A., Dhahbi, W., Chaouachi, A., Padulo, J., Wong, D., & Chamari, K. (2017). Measurement errors when estimating the vertical jump height with flight time using photocell devices: The example of Optojump [Journal {{Article}}]. *Biology of Sport*, *34*(1), 63–70.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bogataj, Š., Pajek, M., Andrašić, S., & Trajković, N. (2020). Concurrent validity and reliability of my jump 2 app for measuring vertical jump height in recreationally active adults [Journal {{Article}}]. *Applied Sciences*, *10*(11), 3805.
- Chiu, L. Z., & Dæhlin, T. E. (2020). Comparing numerical methods to estimate vertical jump height using a force platform [Journal {{Article}}]. *Measurement in Physical Education and Exercise Science*, *24*(1), 25–32.
- Claudino, J. G., Cronin, J., Mezêncio, B., McMaster, D. T., McGuigan, M., Tricoli, V., Amadio, A. C., & Serrão, J. C. (2017). The countermovement jump to monitor neuromuscular status: A meta-analysis [Journal {{Article}}]. *Journal of Science and Medicine in Sport*, *20*(4), 397–402.
- Collings, T. J., Lima, Y. L., Dutailis, B., & Bourne, M. N. (2024). Concurrent validity and test-retest reliability of VALD ForceDecks' strength, balance, and movement assessment tests [Journal {{Article}}]. *Journal of Science and Medicine in Sport*.
- Comyns, T. M., Murphy, J., & O'Leary, D. (2023). Reliability, usefulness, and validity of field-based vertical jump measuring devices [Journal {{Article}}]. *The Journal of Strength & Conditioning Research*, *37*(8), 1594–1599.
- Dobbin, N., Hunwicks, R., Highton, J., & Twist, C. (2017). Validity of a jump mat for assessing countermovement jump performance in elite rugby players [Journal {{Article}}]. *International Journal of Sports Medicine*, *38*(02), 99–104.
- Gathercole, R., Sporer, B., Stellingwerff, T., & Sleivert, G. (2015). Alternative countermovement-jump analysis to quantify acute neuromuscular fatigue [Journal {{Article}}]. *International Journal of Sports Physiology and Performance*, *10*(1), 84–92.
- Gençoğlu, C., Ulupınar, S., Özbay, S., Turan, M., Savaş, B. Ç., Asan, S., & İnce, İ. (2023). Validity and reliability of "My Jump app" to assess vertical jump performance: A meta-analytic review [Journal {{Article}}]. *Scientific Reports*, *13*(1), 20137.
- Glatthorn, J. F., Gouge, S., Nussbaumer, S., Stauffacher, S., Impellizzeri, F. M., & Maffiuletti,

- N. A. (2011). Validity and reliability of Optojump photoelectric cells for estimating vertical jump height [Journal {{Article}}]. *The Journal of Strength & Conditioning Research*, 25(2), 556–560.
- Heredia-Jimenez, J., & Orantes-Gonzalez, E. (2020). Comparison of three different measurement systems to assess the vertical jump height [Journal {{Article}}]. *Revista Brasileira de Medicina Do Esporte*, 26, 143–146.
- Hester, J., & Bryan, J. (2024). *glue: Interpreted string literals* [Manual].
- Hughes, S., Chapman, D. W., Haff, G. G., & Nimphius, S. (2019). The use of a functional test battery as a non-invasive method of fatigue assessment [Journal {{Article}}]. *PLoS One*, 14(2), e0212870.
- Lakens, D. (2022). Sample size justification [Journal {{Article}}]. *Collabra: Psychology*, 8(1), 33267.
- Landau, W. M. (2021a). *tarchetypes: Archetypes for targets* [Manual].
- Landau, W. M. (2021b). The targets R package: A dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, 6(57), 2959.
- Linthorne, N. P. (2001). Analysis of standing vertical jumps using a force platform [Journal {{Article}}]. *American Journal of Physics*, 69(11), 1198–1204.
- Lombard, W., Reid, S., Pearson, K., & Lambert, M. (2017). Reliability of metrics associated with a counter-movement jump performed on a force plate [Journal {{Article}}]. *Measurement in Physical Education and Exercise Science*, 21(4), 235–243.
- Loy, A., Steele, S., & Korobova, J. (2024). *Imeresampler: Bootstrap methods for nested linear mixed-effects models* [Manual].
- McMahon, J. J., Jones, P. A., & Comfort, P. (2016). A correction equation for jump height measured using the just jump system [Journal {{Article}}]. *International Journal of Sports Physiology and Performance*, 11(4), 555–557.
- Merrigan, J. J., Strang, A., Eckerle, J., Mackowski, N., Hierholzer, K., Ray, N. T., Smith, R., Hagen, J. A., & Briggs, R. A. (2024). Countermovement jump force-time curve analyses: Reliability and comparability across force plate systems [Journal {{Article}}]. *The Journal of Strength & Conditioning Research*, 38(1), 30–37.
- Nuzzo, J. L., Anning, J. H., & Scharfenberg, J. M. (2011). The reliability of three devices used for measuring vertical jump height [Journal {{Article}}]. *The Journal of Strength & Conditioning Research*, 25(9), 2580–2590.
- Nuzzo, J. L., McBride, J. M., Cormie, P., & McCaulley, G. O. (2008). Relationship between countermovement jump performance and multijoint isometric and dynamic tests of strength [Journal {{Article}}]. *The Journal of Strength & Conditioning Research*, 22(3), 699–707.
- Parker, R. A., Scott, C., Inácio, V., & Stevens, N. T. (2020). Using multiple agreement methods for continuous repeated measures data: A tutorial for practitioners [Journal {{Article}}]. *BMC Medical Research Methodology*, 20, 1–14.
- Pedersen, T. L. (2024). *patchwork: The composer of plots* [Manual].
- Pino-Ortega, J., Rojas-Valverde, D., Gómez-Carmona, C. D., & Rico-González, M. (2021). Training design, performance analysis, and talent identification—A systematic review about the most relevant variables through the principal component analysis in Soccer, Basketball, and

- Rugby [Journal {{Article}}]. *International Journal of Environmental Research and Public Health*, 18(5), 2642.
- Pueo, B., Penichet-Tomas, A., & Jimenez-Olmedo, J. (2020). Reliability and validity of the Chronojump open-source jump mat system [Journal {{Article}}]. *Biology of Sport*, 37(3), 255–259.
- R Core Team. (2024). *R: A language and environment for statistical computing* [Manual]. R Foundation for Statistical Computing.
- Rago, V., Brito, J., Figueiredo, P., Carvalho, T., Fernandes, T., Fonseca, P., & Rebelo, A. (2018). Countermovement jump analysis using different portable devices: Implications for field testing [Journal {{Article}}]. *Sports*, 6(3), 91.
- Şentürk, D., Yüksel, O., & Akyildiz, Z. (2024). The concurrent validity and reliability of the My Jump Lab smartphone app for the real-time measurement of vertical jump performance [Journal {{Article}}]. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 17543371241246439.
- Ushey, K., & Wickham, H. (2023). *renv: Project environments* [Manual].
- Whitmer, T. D., Fry, A. C., Forsythe, C. M., Andre, M. J., Lane, M. T., Hudy, A., & Honnold, D. E. (2015). Accuracy of a vertical jump contact mat for determining jump height and flight time [Journal {{Article}}]. *The Journal of Strength & Conditioning Research*, 29(4), 877–881.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wisløff, U., Castagna, C., Helgerud, J., Jones, R., & Hoff, J. (2004). Strong correlation of maximal squat strength with sprint performance and vertical jump height in elite soccer players [Journal {{Article}}]. *British Journal of Sports Medicine*, 38(3), 285–288.
- Xie, Y. (2014). knitr: A comprehensive tool for reproducible research in R. In V. Stodden, F. Leisch, & R. D. Peng (Eds.), *Implementing reproducible computational research*. Chapman and Hall/CRC.
- Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman and Hall/CRC.
- Xie, Y. (2024). *knitr: A general-purpose package for dynamic report generation in R* [Manual].
- Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman and Hall/CRC.
- Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook*. Chapman and Hall/CRC.
- Xu, J., Turner, A., Comfort, P., Harry, J. R., McMahon, J. J., Chavda, S., & Bishop, C. (2023). A systematic review of the different calculation methods for measuring jump height during the countermovement and drop jump tests [Journal {{Article}}]. *Sports Medicine*, 53(5), 1055–1072.
- Yingling, V. R., Castro, D. A., Duong, J. T., Malpartida, F. J., Usher, J. R., & Jenny, O. (2018). The reliability of vertical jump tests between the Vertec and My Jump phone application [Journal {{Article}}]. *PeerJ*, 6, e4669.