Reliability and Validity of Two Measurement Systems in the Quantification of Jump Performance

Marlene Mauch, Praxisklinik Rennbahn AG, Dr. rer. nat.
Hans-Joachim Rist, Praxisklinik Rennbahn AG, Dr. med.
Xaver Kaelin, Praxisklinik Rennbahn AG

Summary

There are different devices on the market for assessing strength and power in vertical jumping as a fundamental requisite of an athlete’s performance. The purpose of this study was to assess the reliability and validity of two instruments measuring force, power, velocity, and jump height in squat jumps. Myotest® (MYO) (Myotest SA, Switzerland) was compared with force plate measurements (Quattro Jump® [QUATTRO], Kistler, Switzerland & SPSport Software, Trins, Austria). Forty-three frontier-guards (age range 25–58 years) performed twice a series of five squat jumps (SJ) simultaneously using MYO device along with QUATTRO force plate. Reliability was analysed using ICC, CV and RMSE. Results for reliability for both devices show good results with ICCs ranging from 0.910 to 0.955, and CVs ranging from 2.33% to 6.59% for discrete outcome variables. The validity of the methods was investigated using the Limits of Agreement (LoA) method. MYO overestimated jump performance compared to QUATTRO with a bias of 4.38 cm (±2.59) for jump height, 1.82 Watt/kg (±4.08) for power, and 0.85 N/kg (±1.24) for force. For velocity the two methods displayed good agreement.

In conclusion, based on the variability of the measurements, coaches may use complemental variables in addition to jump data in the realm of performance testing and training control to better understand the performance of their athletes. In addition, on the basis of the results regarding the validity interchangeability of the two systems is limited.

Key Words
Measurement error, force, power, Myotest, force plate.

Schlüsselbegriffe
Messfehler, Kraft, Leistung, Myotest, Kraftmessplatte.
Introduction

Force and power are fundamental performance determinants in many sports. The evaluation of performance is essential in the analysis of an athlete’s training condition or training process (Abernethy and Wilson, 2000; Kraemer and Newton, 2000). Force and power of the lower limbs are mainly tested by protocols including squat jumps and counter-movement jumps. While the squat jump allows the measuring of the concentric strength of the lower limbs, the counter-movement jump provides information about the eccentric-concentric strength in the stretch-shortening cycle (Sale, 1991; Abernethy and Wilson, 2000).

There are different devices to assess force, power, velocity and jump height. The force plate – generally considered as the “golden standard” – uses force transducers (e.g. piezoelectric or strain gauge sensors) to measure the force transmitted to the plate by an individual. Force is defined as F=m [kg] x a [m/s²]. Power is calculated by the force and velocity. The velocity and subsequent jump height is determined by integration of the force time data to produce displacement time data which is integrated again to produce displacement time data from which jump height is determined. Thereby, jump height (H) is usually calculated as the peak in displacement time data. Displacement time data is derived by integrating the velocity time data (H=v²/2g). Force plate devices generally require a cost-intensive laboratory setup. Thus, other systems like accelerometers are becoming increasingly popular as tools for calculating force and power during exercise training or field testing (Abernethy and Wilson, 2000; Leard, Cirillo et al., 2007; Glatthorn, Gouge et al., 2011). These devices are attached to a person while performing athletic or jumping tasks. From the acceleration time data, integration is used to determine velocity time and displacement time data. Force and power is calculated by the differentiation of acceleration of the jump, the body mass of the individual and gravity. The body mass is pre-entered into the device. The jump height (H) is based on the flight time (H=9.81 x 100 x [Vemin after peak–Vmax]²/8).

The validity and reliability of accelerometers has recently been reported in the literature (Leard, Cirillo et al., 2007; Casartelli, Muller et al., 2010; Comstock, Solomon-Hill et al., 2011; Crewther, Kilduff et al., 2011). They are based on different populations regarding sport, training level and age and focused on various outcome variables. Unfortunately, most of the results refer to relative-based measures rather than absolute measures. There is still a lack of information about the reliability of both methods as well as about the validity of the test results when measuring jump performance with a force plate compared to an accelerometer. Knowing about the reliability of a measurement is relevant when comparing two methods.

Only a few studies described the presence and amount of bias when comparing the two devices. Knowing about the agreement between the two methods allows practitioners to conveniently interpret the results of the respective tests and the changes in the course of training or even to use the two devices interchangeably. Furthermore, most of the authors included well-trained and young athletes in the study, which will not allow conclusions for less well-trained amateurs.

Therefore the aim of this study was to assess the repeatability of two devices measuring jump performance: the force plate measurement on a Quattro Jump (QUATTRO) (Kistler, Switzerland) based on ground reaction forces and the accelerometer measurement using a Myotest® (MYO) (Myotest Inc., Switzerland) placed on the pelvis. Based on previous studies we hypothesised a good reliability of the devices.

Furthermore, the validity of the MYO device compared to the force plate measurements (QUATTRO) was assessed when measuring force, power, velocity and height in squat jumps. In this issue we hypothesised good relative agreement between the methods, but also the presence of a systematic bias in the outcome variables based on different calculations of the variables.

Methods

Experimental Approach to the Problem

To assess the reliability and validity of the two measurement devices, the subjects twice performed a series of five maximal-effort squat jumps. In between the two series the subjects took a rest of three minutes. The squat jump was considered a suitable task to answer our research questions, as the movement can be more precisely standardised and controlled than for example the counter-movement jump. The mean of the three best jumps (out of five) of each series was used to assess the reliability of squat jumps for both devices. Furthermore, the grand mean of all squat jumps performed with the MYO and QUATTRO, respectively, was used to compare the two methods. Therefore, both tests were conducted simultaneously. The devices represent the independent variable, the outcome force, power, height and velocity of the dependent variables.

Subjects

The study involved 43 border guards (2 women; 41 men) with a mean body height and mass of 176 cm (±6.78 cm) and 87.67 kg (±12.81 kg), yielding a BMI of 28.17 kg*m² (±3.75) across the subjects. The age ranged from 25 to 58 years with an average of 46.74 years (±7.27). They had widely spread activity levels, ranging from nothing to 17 hours of sport per week, with a mean of 284.61 ±458.9 minutes/week. The subjects, tested in 2010, were part of a comprehensive national project to promote physical activity for border guards (“Fit for Work”). Before testing, they all underwent a health check to preclude any medical risks in testing. Informed consent was obtained from all subjects before testing.

Procedures

Before testing, body height and mass were taken using a scale and levelling rule and participants underwent a five-minute warm-up on an ergometer. After that the test procedure was explained and two warm-up jumps were conducted. The participants were instructed to move to a half-squat position (90° bending of the knees) with the hands on the hips, to maintain this position for at least two seconds and then to extend the lower limbs explosively with the aim of jumping as high as possible.

The jumps were simultaneously recorded using the Myotest® (MYO) device, which was attached with an adjustable belt on the pelvis, and a force plate (QUATTRO) (Kistler, Switzerland) interfaced to a computer (see Figure 1). The output variables of the force plate measurement were calculated using SPSSport software (Trins,
Austria). Jump height determined by the force plate measurement is therefore based on the maximum jump velocity \( H = \frac{v^2}{2g} \), whereas the jump height assessed by the accelerometer is based on the flight time \( H = 9.81 \times 100 \times \left[ \frac{t_{\text{Vmin after peak}} - t_{\text{Vmax}}}{2} \right]^2 / 8 \).

**Statistical Analyses**

There are different methods discussed in the literature to assess reliability and validity between methods in the realm of sports medicine and sports science (Bland and Altman, 1986; Bland and Altman, 1990; Atkinson and Nevill, 1998; Weir, 2005). Based on these discussions, relative reliability for both measurement devices (MYO and QUATTRO) was assessed using the intraclass correlation coefficient (ICC), which has been applied in several recent studies for calculating data variability (Nuzzo, Anning et al., 2011). Although there are sceptics who criticise ICC (Bland and Altman, 1990), Atkinson and Nevill (Atkinson and Nevill, 1998) support the citation of the ICC in reliability studies but suggest that it should not be used as the sole statistic.

In addition to the ICC, some absolute measures of reliability, such as standard deviation (SD), coefficient of variation (CV) and the root mean square error (RMSE), were calculated. RMSE, also known as the within-standard deviation, provides a more comprehensive picture of the true reliability of the acquired data (Bland and Altman, 1996). Similar to SD, RMSE is an absolute measure given in the unit of the variable it refers to, and is therefore much easier to interpret.

Furthermore, to analyse the (lack of) agreement (validity) between the two methods, the differences between their outcome variables were tested using a paired t-test. The Limits of Agreement method (LoA) by Bland and Altman (Bland and Altman, 1986) was applied to illustrate the agreement between the two measurement methods. It allows for a comprehensive illustration of a systematic bias between the measured values. Therefore, the differences between the methods (bias) and the 95% LoA as the 1.96 times standard deviation were calculated for all variables and plotted against the mean of both methods. The significance level was set at \( p<0.05 \) for all statistical analyses.

**Results**

**Reliability**

The differences (mean, SD) between series 1 (M1) and series 2 (M2) of both devices are presented in Table 1. They range for jump height between 1.09 ±1.66 cm (MYO) and 0.44 ±2.65 cm (QUATTRO), respectively, and 0.10 ±3.65 Watt/kg, 0.35 ±0.93 N/kg and 1.46 ±9.27 cm/s (MYO) as well as 0.72 ±2.61 Watt/kg, 0.44 ±1.06 N/kg and 1.40 ±11.58 cm/s (QUATTRO) for power, force and velocity, respectively.
The ICC as an estimator of the relative reliability yielded high coefficients between 0.919 and 0.955 for MYO and 0.910 and 0.937 for QUATTRO.

As different units were compared in this study, the coefficient of variation (CV) was applied to get a dimensionless number of reliability for comparison. The CVs ranged from 2.33% to 6.59% depending on the variables assessed. Furthermore, the Root Mean Square Error (RMSE) was calculated to assess the measurement error and subsequently the reliability. The results are displayed in Table 1; for example, for MYO jump height the reliability is 3.85 cm, and for QUATTRO jump height it is 5.15 cm.

Validity (Agreement)

After testing the reliability of both measurement devices, the agreement between the two methods was assessed. Descriptive statistics for the two testing devices regarding jump height, power, force and velocity are presented in Table 2.

The differences between the two methods (bias) and the 95% Limits of Agreement (LoA) were calculated to illustrate the degree of agreement between the two methods (Figure 2).

For the jump height the mean difference (MYO minus QUATTRO) is 4.38 cm (±2.59) (Table 2). As the differences are normally distributed, it is expected that the differences for 95% of observations would be below (4.38 + [1.96 x 2.59]) and above (4.38 – [1.96 x 2.59], that is +9.46 and –0.70 cm LoA (see Figure 2, a). The mean difference for power is 1.82 Watt/kg (±4.08), resulting in +9.82 Watt/kg and –6.18 Watt/kg LoA (see Figure 2, b). For the force there are mean differences of 0.85 N/kg (±1.24) resulting in +3.28 N/kg and –1.58 LoA (see Figure 2, c). The differences for the two devices were statistically compared and indicated significant differences for jump height, power and force (see Table 2). The mean difference in velocity is 1.99 cm/s (±13.69), resulting in +28.82 cm/s and –24.84 cm/s LoA (see Figure 2, d). However, the difference between the two devices is not significant (see Table 2).

Discussion

The aim of this study was to assess the reliability and validity of the Myotest® and Quattro Jump® when measuring the force, power, velocity and jump height of a squat jump.

Regarding the reliability, the calculated ICCs between 0.910 and 0.955 (Table 1) suggest “good” repeatability for both devices in all the measurements. When calculating ICC,

<table>
<thead>
<tr>
<th>Difference (M1-M2)</th>
<th>ICC (1,1)</th>
<th>CV [%]</th>
<th>RMSE (sw)</th>
<th>Repeatability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Jump Height [cm]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYO</td>
<td>–1.09</td>
<td>1.66</td>
<td>0.942</td>
<td>3.88</td>
</tr>
<tr>
<td>QUATTRO</td>
<td>–0.44</td>
<td>2.63</td>
<td>0.910</td>
<td>6.59</td>
</tr>
<tr>
<td><strong>Power [Watt/kg]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYO</td>
<td>–0.10</td>
<td>3.65</td>
<td>0.919</td>
<td>4.99</td>
</tr>
<tr>
<td>QUATTRO</td>
<td>–0.72</td>
<td>2.61</td>
<td>0.937</td>
<td>4.43</td>
</tr>
<tr>
<td><strong>Force [N/kg]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYO</td>
<td>–0.35</td>
<td>0.93</td>
<td>0.955</td>
<td>2.34</td>
</tr>
<tr>
<td>QUATTRO</td>
<td>–0.40</td>
<td>1.06</td>
<td>0.920</td>
<td>3.13</td>
</tr>
<tr>
<td><strong>Velocity [cm/s]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYO</td>
<td>–1.46</td>
<td>9.27</td>
<td>0.947</td>
<td>2.33</td>
</tr>
<tr>
<td>QUATTRO</td>
<td>–1.40</td>
<td>11.58</td>
<td>0.920</td>
<td>3.09</td>
</tr>
</tbody>
</table>

Table 2: Descriptive statistics of the two measurements: Myotest (MYO) and Quattro Jump (QUATTRO).

<table>
<thead>
<tr>
<th>Overall (M1+M2)</th>
<th>Paired t-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>Max</td>
</tr>
</tbody>
</table>
it is suggested that an ICC close to 1 indicates "excellent" repeatability. Various categories for the ICC are provided ranging from "questionable" (0.7–0.8) to "high" (>0.9) (Vincent, 2005). Other authors reported similar ICCs (Bampouras, Relph et al., 2010; Casartelli, Muller et al., 2010; Nuzzo, Anning et al., 2011) for jump height, force or power ranging between ICC=0.87 and ICC=0.92. As described in the methods section, there is criticism of the use of ICC, suggesting the use of additional methods to supplement ICC. The most common methods of analysing absolute reliability are CV or RMSE. Both statistics have in common that they are unaffected by the range of measurements. Some scientists have chosen an analytical goal of the CV being 10% or below. A CV of 10% obtained on an individual actually means that, assuming the data are normally distributed, 68% of the differences between tests lie within 10% of the mean of the data (Strike, 1991). Therefore, the variability is not described for 32% of the individual differences. For example, a test-retest CV of 3.88% in MYO jump height (Table 1) with a grand mean of the two tests of 28.02 cm (Table 2) results in a variability of 1.09 cm for 68% of the differences. Similar results were found for the other variables for both devices and were validated by other studies (Bampouras, Relph et al., 2010; Nuzzo, Anning et al., 2011). In practice, if a person reaches a jump height of 35 cm in the first jumping series you can assume that the jump will be 35 ±1.09 cm in the second jumping series, which is something between 33.9 cm and 36.1 cm. If a coach wishes to detect a training effect of, for example, 5 cm, this device would be sensitive enough to detect this effect. However, this applies only to 68% of the differences between two series.

Therefore, it is suggested to calculate the RMSE in the unit of the respective variable and to report reliability, which is 2.77 x RMSE. Hereby, the difference between two measurements for the same subject is expected to be less than 2.77 x RMSE for 95% of pairs of observations. In other words, for MYO jump height (Table 1), for example, the reliability is 2.77 x 1.389 = 3.85 cm, and for QUATTRO jump height it is 2.77 x 1.860 = 5.15 cm. In our example with a person jumping 35 cm with MYO in the first series it is assumed they are jumping somewhere between 33.1 cm and 36.9 cm. A training effect in jump height should be greater than 3.85 cm to be detected with the MYO or an even greater 5.15 cm to be detected with QUATTRO. In general, QUATTRO shows a slightly increased measurement error compared to MYO (except for power).

Figure 2: LoA for data presented in Table 2. The difference between the methods are plotted against each individual’s mean for the two methods; a) jump height b) power c) force and d) velocity.
A potential systematic error between the two series of squat jumps was formally calculated (Table 1). Thereby, the mean differences between series 1 and series 2 for both devices throughout all measures were negative, indicating a higher performance in series 2 than in series 1. Although the participants completed a warm-up and several test trials, this could be due to a learning or post-activation facilitation effect. In conclusion, RMSE, as well as CV and ICC, show reasonable repeatability for both devices. Nonetheless, it is essential for the user to standardise the testing protocol including the briefing and to allow the athlete to become familiar with the testing task.

The second aim of this study was to compare the two measurement devices MYO and QUATTRO and to evaluate the validity of their related outcome variables. Due to the fact that the devices use different algorithms for their outcome variables, we assumed that the results may differ from each other: whereas QUATTRO is based on force measurements, MYO utilises accelerometer data multiplied by the known mass to determine force.

The descriptive statistics presented in Table 2 display a constant bias for jump height, power and force, supported by significant differences (p<0.05) between the methods tested with a paired t-test. These findings are supported by Crewther et al. (Crewther, Kilduff et al., 2011), who validated MYO in performing squat jumps with different loads regarding force and power using a Kistler force plate. In that study, systematic bias and relatively large random errors were revealed. Similarly, Casertelli et al. (Casertelli, Muller et al., 2010) as well as Comstock et al. (Comstock, Solomon-Hill et al., 2011) found that MYO overestimated jump height compared to the criterion method Optojump (photoelectric cells) or a force plate.

To analyse the absolute consensus between the two methods, Bland and Altman (Bland and Altman, 1986) suggest plotting the difference between the methods against each individual’s mean for the two methods (LoA). It is recommended that 95% of the data points should lie within these limits of agreement. In our study, the LoA indicates generally good agreement between the methods. Nevertheless, for jump height, power and force there is a constant bias evident. The extent of this bias is especially serious for jump height with 4.38 ±5.08 cm, given that 95% of differences with MYO are 9.46 cm above or 0.70 cm below QUATTRO. In contrast to this, the differences for velocity between MYO and QUATTRO are small and not significant.

Although the study presents good agreement for the two devices, the limits of agreement are still wide for practical use. A device should be sensitive enough to assess changes during a training process or to compare the performance within a group (e.g. volleyball team). If the measurements of two different devices agree, one can use it interchangeably. In our study the limits of agreement for jump height are 4.38 ±5.08 cm. Visualise a person jumping 35 cm on a QUATTRO force plate. When measuring the same person with MYO, the jump height could be 0.7 cm below or 9.46 cm above the QUATTRO result, which is a jump height between 34.3 cm and 44.5 cm. These wide limits of agreement could also be due to the variability of the participants, but shows the limiting comparability between the two methods, at least in a population with different physical activity levels.

The restricted agreement of the two methods could be due to the different algorithms of the devices used for calculating their output variables, which can magnify any errors. For both devices the velocity results from the first integration, whereas velocity is integrated again to determine jump height. Additionally, QUATTRO utilises maximum jump velocity to determine jump height, whereas MYO jump height is based on the flight time. Another likely reason for the different outcome variables could lie in the location of the device and with it the centre of mass, when measured by a force plate or by an accelerometer on the pelvis. MYO was mounted at the left side of the pelvis. Therefore, rotational effects of the hip or the placement of the device itself could be responsible for the difference to the force plate derived measures. It could be assumed that, as a result of the rotational effects, the pelvis does not exactly track the body’s centre of mass and therefore is not able to measure leg extensor performance. It remains unclear whether the measurement itself or rather the applied algorithm is producing the different results. The comparison of acceleration signals in different movements as well as kinematic analyses using 3D motion measurement could help to assess the actual jump height, and subsequently the agreement between the two methods could be more appropriately evaluated.

Conclusions

Our study indicates that both devices provide reliable results for assessing force, power, velocity and jump height. Out of this, the use of QUATTRO as well as MYO is legitimate for the evaluation of jump performance. However, the variability of the test results should be carefully considered when interpreting an athlete’s performance or the course of a training process. In practice, the test conditions and the protocol should be strictly standardised to minimise the test-retest variability. It is recommended that coaches use complementary variables, e.g. quadriceps and hamstrings power, in addition to jump data to better understand the performance of their athletes, for example using isokinetic strength test or a one-repetition maximum strength test.

Furthermore, the evaluation of the agreement between the two methods indicates that results have to be interpreted with caution, since MYO tends to overestimate jump height, power and force compared to the data obtained from a force plate (QUATTRO). On the basis of these results interchangeability of the two systems is limited. Nevertheless, MYO offers benefits regarding portability, cost-effectiveness and handling.

Grant Funding

None of the authors have received any payment or other financial support related to this work.

Corresponding Address:

Dr. rer. nat. Marlene Mauch, Praxisklinik Rennbahn AG, Kriegackerstrasse 100, 4132 Muttenz,
Phone +41 61 465 6464, marlene.mauch@rennbahnklinik.ch

References

Reliability and Validity of Two Measurement Systems in the Quantification of Jump Performance


