Reliability and agreement between 2 strength devices used in the newly modified and standardized Constant score

Morten Tange Kristensen, PT, PhD\textsuperscript{a,b,c,*}, Maria Aagesen, PT\textsuperscript{d}, Signe Hjerrild, PT\textsuperscript{d}, Pernille Lund Skov Larsen, PT\textsuperscript{d}, Bente Hovmand, PT, MSc\textsuperscript{d}, Ilija Ban, MD\textsuperscript{c,e}

\textsuperscript{a}Physical Medicine and Rehabilitation Research–Copenhagen (PMR-C), Copenhagen, Denmark
\textsuperscript{b}Department of Physical Therapy, Copenhagen University Hospital Hvidovre, Copenhagen, Denmark
\textsuperscript{c}Department of Orthopedic Surgery, Copenhagen University Hospital Hvidovre, Copenhagen, Denmark
\textsuperscript{d}Faculty of Physical Therapy, Metropolitan University College Copenhagen, Copenhagen, Denmark
\textsuperscript{e}Clinical Orthopaedic Research Hvidovre, Copenhagen, Denmark

Hypothesis: The new and standardized test protocol for the Constant score (CS) provides new methodology, but different devices are still used for shoulder strength testing. It was hypothesized that strength measurements using the IsoForceControl (IFC) dynamometer (MDS Medical Device Solutions, Oberburg, Switzerland) would provide results comparable with the IDO isometer (Innovative Design Orthopaedics, Redditch, UK).

Materials and methods: Sixty healthy subjects, aged 19 to 83 years, were studied, with 5 men and 5 women in each of 6 ten-year age groups. The IFC and IDO were used in randomized order with an 8-minute interval between testing. Subjects performed 3 successive trials with strong verbal encouragement, with 1 minute between trials. The best strength performance was used in the analysis. The rater and subjects were blinded to all results.

Results: The IFC produced 0.28-kg (0.62-lb) higher strength values on average than the IDO ($P = .002$). The intraclass correlation coefficient (ICC\textsubscript{2,1}) was 0.97 (95% confidence interval, 0.95-0.98), whereas the standard error of measurement and smallest real difference were 0.43 kg (0.95 lb) and 1.2 kg (2.63 lb), respectively. The total CS and strength reached mean values of 92.4 points (SD, 6.2 points) and 8.2 kg (SD, 2.6 kg) (18.0 lb [SD, 5.8 lb]), respectively, and were negatively associated with age ($r > -0.407$, $P \leq .001$). The strength values decreased ($P \leq .001$) by 1.3 CS points per decade, and women had strength values that were 8 CS points lower on average than those of men of the same age.

Conclusions: The relative (intraclass correlation coefficient) and absolute (standard error of measurement) reliability between the IFC and IDO is excellent, indicating that performances reported from settings using the IDO are comparable with those recorded with the IFC in other settings.

Level of evidence: Basic Science, Kinesiology.

© 2014 Journal of Shoulder and Elbow Surgery Board of Trustees.

Keywords: Shoulder; Constant score; standardization; strength devices; intrarater reliability; agreement
The Constant score (CS)\(^{10}\) is extensively used to evaluate outcome for patients with various shoulder disorders.\(^{1,7,20,21,26,28}\) Correspondingly, a large number of studies have examined or reviewed the psychometric properties of the CS (0-100 points) or parts of the score, for example, the shoulder strength part (0-25 points).\(^{1,3,6-8,14-16,19,22-25,27,29}\) Various test protocols have routinely been used at different centers, probably because of authors’ own interpretation of Constant’s original work but also because of a lack of an internationally accepted and standardized test protocol. Constant et al\(^{9}\) published a new guideline report in 2008 to solve some of the methodology problems associated with the CS, but still without including a standardized test protocol. A standardized CS test protocol in Danish and English was provided in 2013,\(^2\) based on this new guideline report and the original article by Constant and Murley.\(^{9,10}\) However, this protocol has not been validated, which is a requisite if it is to be recommended for use in daily clinical practice. The standardized protocol provides a thorough description of the CS including the assessment method of maximum shoulder strength (part D of the CS)\(^2\) but does not recommend use of a specific strength device. Strength devices differ, ranging from an unsecured mechanical spring balance to several digital devices that differ greatly in cost. The Isobex dynamometer and the IsoForceControl (IFC) dynamometer (both devices from MDS Medical Device Solutions, Oberburg, Switzerland) have been used in several previous studies and are, by many investigators, considered the gold standard.\(^{1,3,6-10,14-16,18,23,27}\) However, the IFC is expensive (approximately €1,600), which limits many centers from acquiring it. Several newer and cheaper devices designed specifically for the CS strength component are now available. One of these is the IDO isometer (Innovative Design Orthopaedics, Redditch, UK), a transportable digital dynamometer, costing approximately €350, that is designed for the CS strength component. Whether the results of strength assessment using the low-budget IDO are comparable with the results of the expensive IFC is unknown.

The primary aim of this study was to test the hypothesis that the relative reliability between strength measures using the IFC and IDO is excellent and that measurement error (absolute reliability) is low when using the newly standardized CS test protocol.\(^2\) The secondary aim was to examine whether the total CS and the strength part (part D) are related to age and sex in healthy subjects.

Material and methods

Sixty adult volunteers (30 women and 30 men, aged 19-83 years) from the Copenhagen area were tested within a 2-week period. The following inclusion criteria were used: age 18 years or older, ability to give informed consent, ability to speak and understand Danish, and no current shoulder problems. Subjects were recruited by telephone or E-mail, according to 6 prespecified age groups, with 10 subjects (5 women and 5 men) in each of the following 6 age groups: 18 to 29 years, 30 to 39 years, 40 to 49 years, 50 to 59 years, 60 to 69 years, and 70 to 90 years.\(^{15}\) All subjects received written and oral information about the project and signed an informed consent form before they were tested.

Strength devices

The IDO isometer is a transportable, digital CE-registered dynamometer designed to measure the average isometric muscle strength in kilograms or pounds over a period of 3 seconds (Fig. 1).

The IFC is a transportable, digital CE-registered dynamometer that registers strength values in newtons at the start and end of a self-selected preset period between 5 and 20 seconds, as well as maximum and average strength values (Fig. 2). The peak force produced over the period from 2 to 4 seconds in a 5-second period was used in our analysis. Values in newtons were converted to kilograms by dividing by 9.82, as 1 Newton equals 9.82 kg.

Procedures

All test procedures and instructions were standardized according to the newly published CS test protocol.\(^2\) The 2 strength devices are self-calibrating when started, but the agreement between devices was examined with a 5- and 10-kg weight load. The best of 3 recordings for each device showed a difference between devices of 0.02 kg and 0.04 kg. Pilot testing of the protocol and all procedures was performed in 10 subjects not included in the project. Two senior physiotherapy students were used as test raters. The first rater, who instructed all test subjects on the objective parts of the CS (parts C and D) including strength assessment, was blinded to all measurements and kept unaware of the results until all subjects had been tested. The second rater noted the results during testing and ensured that all subjects completed the subjective parts of the CS (parts A and B). Both raters were within 6 months of graduation, and the role of physiotherapy students as raters corresponds with a previous study with a focus similar to the present study.\(^{14}\)
To assess absolute reliability, the standard error of measurement (SEM) was calculated as $SD \times \sqrt{(1-ICC)}$, where $SD$ is the standard deviation of all strength scores (mean of IFC and IDO) from all subjects.\textsuperscript{13,30} The SEM is the estimated standard deviation of measurement error, or the difference between the observed values and the true values, and gives a clinician a result in the same unit as the measurement.\textsuperscript{13} To quantify the required change in strength that must be observed (exceeding measurement error and reflecting a real change at the individual level), the smallest real difference (SRD) was calculated at a 95% confidence level as $SEM \times 1.96 \times \sqrt{2}$.\textsuperscript{1} In addition, we calculated SEM% and SRD% as follows: $SEM\% = (SEM/mean) \times 100$ and $SRD\% = (SRD/mean) \times 100$, in which $mean$ is the mean of the IFC and IDO scores.

A Bland-Altman plot was used as a qualitative method\textsuperscript{5} to illustrate the magnitude of agreement between the 2 strength devices. The differences between the IFC and IDO were plotted against the respective average. A Pearson correlation coefficient was used to determine the correlation between age and the strength values and total CS, as well as to determine whether the numerical between-strength-device differences were correlated with the mean strength values from the IFC and IDO from all subjects (heteroscedasticity, with a significantly larger variability for higher-strength values).\textsuperscript{5}

According to recommendations by Eliasziw et al.,\textsuperscript{11} we needed to include approximately 50 subjects to yield 80% power at the 5% significance level to achieve an ICC set at 0.90 in our intra-device reliability study. To include 10 subjects in each of the 6 age groups, we included 60 subjects. Previous studies have shown that men, on average, present strength values that are 3.2 kg (7 lb) higher than those for women.\textsuperscript{14,28} A strength difference between sexes of at least 2.3 kg (5 lb) was considered an important difference in our study. With a standard deviation of the observations in each group set at 2.3 kg (5 lb), we required 16 subjects in each group.

Finally, we used simple and multiple linear regression analyses (enter method) to examine the influence of age and sex on the total CS and the strength part of the CS (part D).\textsuperscript{28} Age was entered as a continuous variable and men as the reference value for sex.

Data are presented as mean and standard deviation or as number of patients and percentage, as appropriate. The statistical analyses were conducted using SPSS software, version 19.0 (SPSS, Chicago, IL, USA). The level of significance was set at $P < .05$, and all statistical tests were 2 tailed.

## Results

Of the 60 subjects, 15 were students (bachelor’s or higher level), 31 were working, and 14 were retired; 55 used the right arm as the dominant arm during testing; 48 reported that they were physically active; 9 reported a previous injury in the dominant shoulder; and the mean age for all subjects was 49.4 years (SD, 18.5 years). The 15 students, with a mean age of 29.6 years (SD, 7.4 years), presented significantly ($P \leq 0.02$) higher total CS and strength values than the 14 retired subjects (mean age, 73.4 years [SD, 7.6 years]); otherwise, no significant differences were seen in the corresponding values between any of the

---

**Figure 2** IFC strength device.
abovementioned groups ($P > .1$). Ten subjects in total did not reach a maximum score in 1 or more of the subjective and objective parts of the CS (parts A-C). Thus, 3 patients achieved 13 out of the 15 possible points in the pain score (part A); 5 subjects achieved 10 ($n = 1$), 18 ($n = 1$), and 19 ($n = 3$) out of the 20 possible points for activities of daily living (part B); and 6 subjects achieved 34 ($n = 1$) and 38 ($n = 5$) out of the possible 40 points for range of motion (part C).

### Strength devices

Performances measured with the IFC and IDO were highly correlated ($r = 0.971$, $P < .001$), but the IFC produced adjusted strength values that were 0.28 kg (SD, 0.64 kg) (0.62 lb [SD, 1.4 lb]) higher on average than those with the IDO ($P = .001$). The ICC$_{2,1}$ between the IFC and IDO was 0.97 (95% confidence interval, 0.95-0.98), whereas the SEM and the SRD were 0.43 kg (0.95 lb) (SEM%, 4.6%) and 1.2 kg (2.63 lb) (SRD%, 12.9%), respectively. The Bland-Altman plot (Fig. 3) with adjusted strength values showed no signs of heteroscedasticity ($r = -0.165$, $P = .2$).

### Influence of age and sex

Seventeen of the 30 men presented raw strength values above the possible 25 points (part D), whereas all women had raw strength values below 19 points (Fig. 4, A). The total CS and adjusted strength values reached an average of 92.4 points (SD, 6.2 points) and 8.2 kg (SD, 2.6 kg) (18.0 lb [SD, 5.8 lb]), respectively, and were negatively associated with age ($r > -0.407$, $P \leq .001$), as shown in Figure 4.

Men presented higher ($P < .001$) total CS (mean, 96.0 points vs 88.8 points) and adjusted strength (mean, 10 kg [22.0 lb] vs 6.36 kg [14.0 lb]) values than women in general, as well as in most of the 6 age groups (Table I and Fig. 4). In comparison, subjects aged 70 years or older presented significantly ($P \leq .04$) lower CS and strength values than those in the 4 age groups from 19 to 59 years but not in those aged 60 to 69 years. Furthermore, the influence of age and sex on the total CS and strength performances when evaluated in both simple and multiple linear regression analyses was significant ($P \leq .001$) (Table II). Thus, women on average presented strength values that were 8 CS points lower than men ($\beta$ weight, $-0.701$), whereas the corresponding values for age showed a decrease.

![Figure 3](image-url)  
**Figure 3** Bland-Altman plot comparing strength values of IFC and IDO strength measurement devices. The plot shows a nonsignificant correlation ($r = -0.164$, $P = .2$) for the numerical difference between the IFC and IDO and the mean of the respective measurements (no heteroscedasticity) but a systematic bias across trials ($P = .001$).

![Figure 4](image-url)  
**Figure 4** Comparison of raw strength values according to sex (A) and CS (B) by age group.
The regression model was statistically stable and explained 66% \((r = 0.809)\) of the variation in strength performances.

### Discussion

This study, following the newly standardized guidelines for CS testing,\(^3\) found excellent relative reliability (ICC, 0.97) and very low measurement noise at a group level \((< 1 \text{ CS point})\) between the 2 strength devices examined: the IFC and the IDO. Thus, an increase or decrease by 1 or more CS strength points can be considered a true difference for a group of subjects, enabling comparison between data recorded with the IFC and IDO. Furthermore, we found that the total CS and both the raw and adjusted strength values were related to the age and sex of subjects.

### Strength devices

The systematic error between devices in this study of 0.28 kg \((0.62 \text{ lb})\) corresponds well to findings in previous studies evaluating whether different strength devices yield the same result.\(^3\) Some of the difference in this study could be explained by the different periods and values recorded by the 2 evaluated devices. The IDO records the average force for 3 seconds (the only available setting), whereas the IFC enables recording of both the mean and the peak force produced between 2 and 4 seconds of a 5-second trial. Thus, maximum strength values seem to be consistently higher than mean strength values\(^5\) and are recommended for use in the CS.\(^2\) Accordingly, we compared the maximum average force (IDO) with the maximum peak force (IFC) for 3 repetitions from each device, which resulted in very similar values. Thus, it seems as if results for a group of persons assessed with the low-cost IDO, though not able to measure the peak force, can be compared with results from another group assessed with the more advanced and costly IFC strength measurement device. Two of the previous studies evaluating the results of different strength devices did not report data for relative and absolute reliability,\(^3\) whereas our ICC values are very similar to the value reported for a comparison between 2 other strength devices (ICC, 0.96; 95% confidence interval,
0.91-0.99) also using maximum values of a pull force technique.\textsuperscript{15}

Influence of age and sex

Our finding that age and sex affect the total CS and strength measurements indicates a high degree of external validity. The study was not powered to evaluate differences between age groups, and our subgroups were smaller than those of previous studies.\textsuperscript{15,28} However, they still are comparable with those reported by Katolik et al\textsuperscript{16} in a larger group of patients (N = 441) divided into the same age groups and those reported by Walton et al\textsuperscript{29} in 108 patients aged between 50 and 89 years—both studies in patients without shoulder symptoms. In our study, with equally sized age groups of men and women (age range, 19-83 years), we found that the total strength values fell by 0.13 CS points per year (1.3 per decade) whereas the corresponding strength values were approximately 3.6 kg (8 CS points) lower for women than for men. Correspondingly, Walton et al reported that the CS fell by 0.3 points per year and that strength values, on average, were 7.5 CS points higher for men, whereas our CS and strength values in relation to age groups (Fig. 4) are quite similar to those shown by Katolik et al in Figures 2 and 3 in their study.\textsuperscript{15}

Overall, our findings support results from previous studies and further emphasize the importance of the strength component in the CS, in addition to the need for age and sex stratification, when using the CS in research studies.

Conclusions

The results of this study show that performances of the standardized strength test in the CS, carried out with the IFC and the IDO, are comparable at a group level because high relative reliability and very low measurement noise were found. It is recommended that an individual be tested with the same device to measure change over time. Our findings that age and especially sex affect strength values and thereby the total CS of healthy subjects are in accordance with results previously reported, which indicate the need for stratification of patients when using the CS in research projects.

Disclaimer

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

References


