J Shoulder Elbow Surg (2017) ■■, ■■–■■

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

ORIGINAL ARTICLE

# Shoulder strength value differences between genders and age groups

**Eudald Balcells-Diaz, MD[a,*], Pepus Daunis-i-Estadella, PhD[b]**

[a]*Department of Orthopaedics and Trauma Surgery, Hospital de Mollet, Barcelona, Spain*
[b]*Department of Computer Science, Applied Mathematics and Statistics, Universitat de Girona, Girona, Spain*

**Background:** The strength of a normal shoulder differs according to gender and decreases with age. Therefore, the Constant score, which is a shoulder function measurement tool that allocates 25% of the final score to strength, differs from the absolute values but likely reflects a normal shoulder. To compare group results, a normalized Constant score is needed, and the first step to achieving normalization involves statistically establishing the gender differences and age-related decline. In this investigation, we sought to verify the gender difference and age-related decline in strength.
**Methods:** We obtained a randomized representative sample of the general population in a small to medium-sized Spanish city. We then invited this population to participate in our study, and we measured their shoulder strength. We performed a statistical analysis with a power of 80% and a $P$ value < .05.
**Results:** We observed a statistically significant difference between the genders and a statistically significant decline with age.
**Conclusion:** To the best of our knowledge, this is the first investigation to study a representative sample of the general population from which conclusions can be drawn regarding Constant score normalization.
**Level of evidence:** Level III; Cross-Sectional Design; Epidemiology Study


**Keywords:** Shoulder function; outcome measure; age-gender decline; strength; Constant score; methodology; statistics

Outcome measures in shoulder surgery have become a standard in the communication of results among surgeons; similar to most fields in orthopedic and trauma surgery, outcome measures have replaced subjective opinions or investigator-based measurement systems.[26]

The Constant score (CS) is a shoulder function measurement tool developed by Christopher Constant while he was working on his Master of Surgery thesis published in 1986.[9] Since its development, the CS has become widely used, and in 1992, the European Society for Shoulder and Elbow Surgery recommended its use in publications and communications.[10] The CS scoring system includes 4 variables that are scored according to a detailed item grid with a sum of 100. In a perfect score of 100, 35 points are allocated to a self-reported subjective assessment, including the presence of pain and the ability to perform daily living activities, and 65 points are allocated to objective measurements, including 40 points allocated to range of motion and 25 points allocated to strength.[11] The 25 points allocated to the objective measure of strength remain controversial because the strength of a normal shoulder could differ on the basis of gender and decrease with age. Thus, a decreased CS

Comité d'Ètica d'Investigació Clínica approved this study: CEIC FHAG No. 20166011.

*Reprint requests: Eudald Balcells-Diaz, MD, Av Jaume Morató, 36, E-08440 Cardedeu, Barcelona, Spain.

E-mail address: ebalcells@gmail.com (E. Balcells-Diaz).

1058-2746/$ - see front matter © 2017 Journal of Shoulder and Elbow Surgery Board of Trustees. All rights reserved.
https://doi.org/10.1016/j.jse.2017.10.021

value in certain individuals could reflect normal functioning.[4,6,14,16]

Numerous peer-reviewed papers claim to have used a *normalized* or *adjusted CS* to report shoulder function[7,12,17,20,24]; however, to the best of our knowledge, the variation in shoulder strength in the general population according to age and gender has not been established using appropriate methodology. In his seminal thesis, Constant devoted an entire chapter to the effect of age on normal shoulder function. Although the methods used by Constant were appropriate during the time he performed his study, subsequent major advances in our understanding of the statistical sciences indicate that his design and therefore his results and conclusions are weak because of the poor sampling choices according to the current standards. Despite this limitation, his study strongly suggests that strength deteriorates with age.

The purpose of this study was to establish whether a variation in shoulder strength exists according to age and gender. Therefore, we proposed and tested the following 2 null hypotheses:

H0 1. There is no statistically significant difference in the distribution of the shoulder *strength* function between the genders.

H0 2. There is no statistically significant difference in the distribution of the shoulder *strength* function due to age.

## Materials and methods

The first step to testing our null hypotheses was to obtain a statistically representative sample of the general population. We sought advice from the Center for Applied Medical Statistics at the University of Cambridge (UK). This group suggested that we obtain subjects from a universal source and then randomly extract the desired sample size with a confidence interval to allow for nonresponders. To facilitate the sample size calculation, this group suggested that we perform a pilot study.

We chose the small to medium-sized city Manacor (37,963 inhabitants according to the 2008 population census) as a universally representative source because its population pyramid and distribution are similar to those in the entire country[11,18] (Spain). To perform the shoulder strength measurements in a pilot study before the sample size calculation, we followed the methodologic recommendations published in 2008 by a group of experts from the European Society for Shoulder and Elbow Surgery[10] and performed measurements in 79 random volunteer outpatients and coworkers at our hospital. We recorded each patient's age, gender, and shoulder strength in pounds. We chose pounds because the original CS describes the strength score as the score given to the equivalence of a maximum of 25 pounds (eg, a measurement of 12 results in a score of 12, whereas a measurement of 27 results in a score of 25). We excluded all volunteers who did not have a *normal* shoulder as defined by Constant in his previous publications.[9-11] In this pilot study, we obtained measurements from 30 men and 49 women, and after stratifying according to age and gender, we calculated the means and standard deviations (Table I).

**Table I** Pilot measurement data, stratified by age and gender

| Age (y) | Gender | $\bar{X}$ | Standard deviation |
|---|---|---|---|
| <31 | Male | 22.33 | 6.80 |
| | Female | 10.26 | 3.05 |
| | All | 13.88 | 7.14 |
| 31-39 | Male | 21.80 | 4.86 |
| | Female | 11.08 | 2.34 |
| | All | 15.33 | 6.36 |
| ≥40 | Male | 16.85 | 6.64 |
| | Female | 9.42 | 3.18 |
| | All | 12.65 | 6.17 |
| All | Male | 20.28 | 6.39 |
| | Female | 10.35 | 2.88 |
| | All | 14.10 | 6.61 |

**Table II** Sample size calculation for different statistical power levels

| | | *n* for different power | | |
|---|---|---|---|---|
| | Difference | 80% | 90% | 95% |
| Men* | | | | |
| <31 vs. 31-39 y | 0.53 | 2282 | 3055 | 3778 |
| <31 vs. ≥40 y | 5.48 | 22 | 29 | 36 |
| 31-39 vs. ≥40 y | 4.95 | 27 | 36 | 44 |
| Women† | | | | |
| <31 vs. 31-39 y | −0.82 | 194 | 260 | 321 |
| <31 vs. ≥40 y | 0.83 | 190 | 254 | 313 |
| 31-39 vs. ≥40 y | 1.66 | 48 | 64 | 79 |

*Difference* is the minimal expected difference between both groups; *n*, sample size.
* Common standard deviation = 6.39.
† Common standard deviation = 2.88.

Using these data, we calculated the sample size required for our 2-tailed finite population study with different power values to detect statistically significant differences similar to those observed in our pilot study at a $P$ value < .05 using the following formula[21,22]:

$$n = N\sigma^2 Z^2 / (N-1)e^2 + \sigma^2 Z^2$$

where $n$ is the sample size, $N$ is the size of our universal population, $Z$ is the upper 0.025 percentage point of the standard normal distribution, $e$ is the desired power, and $\sigma$ is the standard deviation.

Then, we performed the sample size calculation with different power levels (Table II). Comparing the age groups of <31 years vs. 31-39 years in men, the sample size required to identify a statistically significant difference was very large because the measurement values are similar between the 2 age groups. Figure 1 presents this information graphically in a multiple box plot diagram. A visual analysis of this box plot diagram suggests that the results for the age groups <31 years and 31-39 years are similar, and the differences could be irrelevant in a clinical setting. Thus, based on a Tukey pairwise comparison and given the clinical relevance of the differences, we decided to group the age analysis into the following two categories: <40 years and 40+ years. We recalculated the sample size using the new values as presented in Table III.

In men, 25 subjects were required in each age group to identify a statistically significant difference in strength values of at least 5.15