A review of the Constant score: Modifications and guidelines for its use

Christopher R. Constant, MA, LLM, MCh, FRCS, Christian Gerber, MD, Roger J. H. Emery, MS, FRCS(Ed), Jens Ole Søjbjerg, MD, Frank Gohlke, MD, and Pascal Boileau, MD, Saint-Genis-Laval, France, and London, England

INTRODUCTION

The requirement to assess surgical results and the means of quantifying them have always provoked intense debate. Since Codman’s introduction of “the end result idea,” the main aim in assessment has not changed, but our requirements are now more sophisticated and include observation of natural history, follow-up, and disability quantification.

The Constant score was devised by Christopher Constant with the assistance of the late Alan Murley during the years 1981-1986. The score was first presented in a university thesis in 1986 and the methodology published in 1987. This functional assessment score was conceived as a system of assessing the overall value, or functional state, of a normal, a diseased, or a treated shoulder. If universally accepted, it would further progress in clinical research in shoulder surgery and enhance the value of multicenter trials.

In this score, 35 points are allocated for subjective assessments of pain and activities of daily living and 65 points are available for objective measures of range of movement and shoulder strength. A young healthy patient can therefore have a maximum score of 100 points.

METHODOLOGY

Pain

Pain is allotted 15 points; the assessment is made on the most severe pain felt by the patient during ordinary activities over a 24-hour period. Thus, many patients will be recording the most severe pain at night. Previously, pain was graded as none, mild, moderate, or severe. This has been replaced by a visual analog scale. It has been proposed that a sliding cursor system with an ungraduated line marked at either end with “no pain” and “intolerable pain,” respectively, be used (Figure 1). Equivalent terms have been agreed on for the French language. The numeric score can be seen on the reverse side of the scale. It is important to stress that episodic severe pain (eg, as in dislocation) is not relevant to a functional assessment score.

Activities of daily living

The activities of daily living can score a total of 20 points. Undisturbed sleep is allotted 2 points; occasional disturbance, 1 point; and disturbance every night, 0 points. Eight points are allotted to work and recreational activities. This is scored from 1 to 4 on a fractional basis as a response to the following questions: “How much of your normal work does your shoulder allow?” and “How much of your normal recreational activity does your shoulder allow?” This would similarly be assessed on a visual analog scale with a sliding cursor (Figure 1). The terms “all” and “none” define the range. Activities of daily living also include the ability to functionally use the arm up to a certain level, and this part of the assessment is allotted up to a further 10 points. This is recorded by asking patients to which level they can use their hand comfortably, from below the waist (0 points) to above the head (10 points) (Table I).

Movement

The 40 points allotted to movement are divided equally into forward elevation, lateral elevation, functional external rotation, and functional internal rotation. All movements must be painless and active. These functional movements are composite and distinct from the standard assessment of range of motion (ie, as commonly used for diagnostic purposes). The range of functional active movement is assessed. Thus, if the subject can lift his or her arm to 140° with pain but gets to 110° without pain, the value chosen in the score is 110°.

Pain-free forward elevation and lateral elevation should be measured with a goniometer, with the patient seated to avoid spinal tilting. The reference points are the axis of the arm and the spinous processes of the
thoracic spine for abduction and for forward flexion. These movements are recorded on both arms simultaneously. Points are allotted incrementally as the subject reaches 31°/C14, 61°, 91°, 121°, and 151°. It is important to point out that 150° equals 8 points and not 10 points.

Functional external rotation consists of allotting 2 points each for 5 separate active maneuvers. These must be unassisted, and the hand should be placed behind and above the head without touching the head, with points given as follows: hand to the back of the head with the elbow forward, 2 points; hand to the back of the head with the elbow back, 2 points; hand to the top of the head with the elbow forward, 2 points; hand to the top of the head with the elbow back, 2 points; and full elevation, 2 points. Thus, there is a maximum total of 10 points.

External rotation is only considered to be of functional relevance if it is available in the planes of frontal and lateral elevation, as pure external rotation is not considered to be a functional movement. The same applies for internal rotation, which is tested in combination with extension and adduction. Internal rotation is similarly measured as an unassisted movement by use of the thumb as a pointer against the anatomic landmarks, behind the buttock (2 points), the sacroiliac joint (4 points), the level of the waist (6 points), the twelfth thoracic vertebra (8 points), and the interscapular level (10 points). Patients only reaching the lateral aspect of the thigh score 0 points.

Strength

The final 25 points have caused considerable confusion. The method used for strength evaluation within the Constant score was originally devised by Moseley. He described use of either an unsecured cable tensiometer or spring balance held at arm’s length in both 90° of forward flexion and 90° of abduction in the coronal plane. The maximum force that the patient can resist as the examiner pulls down on the device is then measured. Constant’s modification was to incorporate a cuff attachment for the arm or wrist in those with poor grip strength from other disease. Others have given the patients weights that were held in 90° of abduction. A definition of the exact maneuver to measure strength with specification of the exact position of the arm and wrist and the location of resistance has never been given. How long the patient would have to resist a load and how many repetitions should be made have also not been defined. Furthermore, no specifications were given for patients with additional elbow problems. Finally, the lack of fixation at both ends of the spring balance makes accurate measurement difficult. The strength measurement has therefore been unsatisfactory up to now.

Gerber has drawn attention to these deficiencies and has questioned the validity of a spring balance as a measuring device. In an attempt to resolve these difficulties, the Isobex isometric dynamometer (Cursor, Bern, Switzerland) was developed, and a range of normal values for the strength of lateral elevation in the scapular plane measured at the wrist has been defined. The Isobex device has been carefully designed and extensively evaluated (Figure 2). The figure obtained through the visual readout measures force integrated over a period of time and takes into account

<table>
<thead>
<tr>
<th>Level</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to waist</td>
<td>2</td>
</tr>
<tr>
<td>Up to xiphoid</td>
<td>4</td>
</tr>
<tr>
<td>Up to neck</td>
<td>6</td>
</tr>
<tr>
<td>Up to top of head</td>
<td>8</td>
</tr>
<tr>
<td>Above head</td>
<td>10</td>
</tr>
</tbody>
</table>

Table I Activities of daily living: Ability to position hand in space

Figure 1 Front (A) and back (B) of visual analog scale.
aspects such as fatigue. It calculates the mean of 50 readings taken over the 5 seconds that follow the excluded first-second B mode of the device. So far, the device has been expensive, but it is intended that this reliable and relatively sophisticated piece of equipment will be obtainable at a very reasonable price.

Other electronic devices, including the Kinedyne myometer (Kinetic, Smith & Nephew, Memphis, TN), Mecmesin Myometer (Mecmesin, Broadbridge Heath, England), and more recently, the EZ Force Myometer (T.A.G. Medical Products, Kibbutz Gaaton, Israel), have been proposed but as yet have not been evaluated with regard to shoulder function.

In the past, this section of the assessment has been incorrectly termed "power"; now a more appropriate term is considered to be "strength," as it measures the force generated at the end of a lever arm, whereas power is the rate of work, and work is the product of force and distance. This measurement is recorded at 90° of abduction in the scapular plane. The wrist is in a position of pronation so that the hand is facing downward. A strap is applied to the level of the wrist with the arm at maximum span. Thus, in cases of elbow disease, this span may be reduced as a result of a flexion contracture.

It is proposed that the strength component of the Constant score should be measured by use of either an Isobex device or a defined spring balance technique. Manual muscle testing is unacceptable methodology and is condemned. Similarly, Moseley's method with an unsecured spring balance should be rejected because of its complexity and lack of definition. The value used for the score should be the maximum of 3 repetitions, each separated by at least 1 minute. The highest value per measurement has been used to exclude mis-pulls and reduce the likelihood of including a less-than-maximal effort. It has been shown that data from true maximum voluntary efforts are extremely reproducible, whereas those from deliberate submaximal efforts is not so. A maximum value may be more likely to reflect best function in real terms rather than a mean value that has been reduced by a wide range of submaximal values from mis-pulls, earlier fatigue, and pain. Furthermore, patients who are unable to achieve the test position of 90° of shoulder abduction should be assigned a strength score of 0. In this way, some of the inconsistencies in the reporting of the score may be overcome.

**VALIDATION**

The Constant score has not been the subject of validation experiments until comparatively recently. In the original paper, this was limited to a study of interobserver error. Conboy et al, using a floor-based pulley and positioning as near as possible to 90° of abduction, calculated the interobserver SD as 8.86 and found, based on 95% confidence limits, that a single observer measuring a single subject will be within 17.7 points of the true score. Two- and three-way analyses of variance were used to calculate the interobserver and intraobserver error. The observers on average differed significantly in their total score (F = 4.22, P = .02). Intraobserver variability with analysis of variance on different occasions was not statistically significant (P = .08); the difference between observers was also not significant (P = .06).

Reliability changes according to a subject's pathology, and Conboy et al found the reliability of a single score to be lower in the arthritis and impingement groups. In their article, they acknowledged that the "power" measurement is difficult. The implication is that, with low reliability, when assessing a response to treatment in a single patient, the difference in a score would have to exceed this minimum significant difference to give a positive result, and for analysis of groups of patients, sample sizes would have to be considered carefully.

A major source of error is the measurement of strength. Intraobserver reliability and interobserver reliability have been shown to be good for the Isobex...
device in normal shoulders by use of the same test position, when the mean of 3 values was used rather than the maximum. A quick and simple alternative method using a fixed spring balance has been described and validated. The same standard test position was used with the subject standing, the arm in 90° of lateral elevation in the scapular plane, the elbow extended, and the forearm pronated. The highest value at 5 seconds obtained from 3 successive maximum pulls provided the strength score. The reading from the fixed spring balance was taken after 5 seconds, as this was the shortest test duration that could be reliably measured after the marked oscillations that occur at the beginning of a maximal effort had subsided. It is also of comparable length to the Isobex test cycles.

Comparison within subjects between the Isobex device and the standardized spring balance showed no significant difference ($P < .001$) between the mean of the maximum value of the 5 successive readings with each method in a group of 20 patients with pathologic shoulders attending a dedicated clinic. On analysis of a group of 50 healthy volunteers, there was a difference of only 1 or 2 points between the mean of the maximum value of the 3 successive readings with each of the methods, although this difference was highly significant ($P < .001$). (For male subjects, the mean value was 10.52 ± 2.05 kg by use of the 3-second Isobex mode, 10.18 ± 1.83 kg by use of the 5-second Isobex mode, and 11.45 ± 1.80 kg by use of the spring balance. For female subjects, the values were 6.20 ± 1.29 kg, 6.05 ± 1.11 kg, and 6.60 ± 1.13 kg, respectively.)

This study also demonstrated that the maximum value occurred within the first 3 repetitions in all of the subjects with the spring balance and in over 80% of subjects when the Isobex was used. The number of repetitions is important, as it demonstrates the effect of fatigue with each successive effort. For abnormal shoulders, the maximum value was obtained within the first 3 pulls in 92% of subjects by use of the fixed spring balance, in 81% of subjects by use of the Isobex, and in 62% of subjects by use of the unsecured spring balance. Normal male shoulders seem to have greater consistency in maintaining the upward force for the whole test duration, although the magnitude of that force may fall with subsequent tests. The fixed spring balance demonstrated fatigue in all subject groups far more clearly because it measures the force at a single point in time, namely after 5 seconds of maximum effort. This makes it more sensitive to the progressively earlier decline in the force generated within a single pull with each successive effort, despite the rest periods. Despite the fact that each method is measuring something slightly different, Isobex readings fortuitously coincide with those from an inexpensive fixed spring balance in pathologic shoulders, with only very small differences encountered in normal shoulders.

### Normal Values

The determination of normal values for age and sex is clearly important but will undoubtedly be difficult, as there are noticeable differences between geographically separated populations. It is conceivable that some shoulder centers will be inclined and able to establish their own reference values. This is extremely important because we are still forced to report results of surgery in absolute terms. All of us know, however, that $110^\circ$ of flexion and an abduction strength of 1 kg comprise a miserable result for a dominant arm of a 20-year-old truck driver but that the same findings are very acceptable for the nondominant arm of an 84-year-old housewife. The determination of age- and gender-adjusted normal function allows us, for the first time, to determine whether a certain procedure is more beneficial in younger or in older patients and in men or in women, because the results can be compared with a matched standard. This allows a much more meaningful interpretation of results in situations such as rotator cuff repair.

The Constant score is the first system to be used to study the effects of aging on the function of the shoulder. Constant studied 900 healthy individuals and gave normal values for different age groups (Table I). A “relative Constant score” is in general use. This is calculated by dividing the obtained score of the patients by the age- and gender-matched score of the Constant population. Although left and right shoulders and dominant and nondominant shoulders do not score significantly differently, comparison with the opposite side should be avoided, as too many patients in a given population have a problem with the contralateral shoulder.

### Discussion

The limitations of scoring in general must be appreciated before we can embrace a particular method. The problem is to describe a rather complex, multiparametric situation in simple terms (ie, in a numeric figure). Consequently, a scoring system is a conscious simplification, with all the inherent advantages and disadvantages of such an approach. The chosen score’s correct and wide use is dependent on its practicality and acceptability. The Constant score has suffered from imprecise terminology and poorly defined methods, which have left too much room for interpretation.

The Constant score purely assesses how well a shoulder functions by choosing a number of functional parameters. It is independent of the diagnosis of the shoulder disorder. With assessment of pain,
activities of daily living, elevation, abduction, and functional internal and external rotation, as well as strength, the most important parameters are included.

The Society Européenne pour la Chirurgie de l’Épaule et du Coude (SECEC) adopted this score in 1991 and charged its Research and Development Committee to study the score and issue guidelines. It was unanimously agreed that the score should be retained as a minimal data set for presentations and communications to the Society and to the Journal of Shoulder and Elbow Surgery, respectively. It was widely accepted that this score does not provide sufficient information for the assessment of certain conditions, particularly instability. However, at present, it is considered to be the most appropriate score for assessing overall shoulder function.

The wide discrepancy between normal values from different centers was too great to be explained by population differences, and it is apparent that there is a great deal of uncertainty as to the methodology. It was appreciated that many investigators involved in prospective trials will have already established their methodology and would find it difficult to compare results using the new guidelines. In these circumstances, explanation of the modification will be required in the description of methods. Many studies have presented their data by comparing the Constant score with mean values from subjects of similar age and sex, or even the contralateral side. This ratio is variably termed the adjusted, relative, and even ponderated Constant score! It is generally accepted that this presentation of data has benefits, but it is pointed out that it is necessary to realize that there will be significant differences compared with the normal values obtained from our population studies as a consequence of the change in methodology.

The major changes are precise definitions of methodology and the measurement of force at the end of a lever arm, now termed “strength.” This latter measurement is taken with the humerus in the plane of the scapula, as this position provides certain biomechanical advantages, namely optimum length-tension relationships for the humeral abductors, a relaxed inferior capsule, and maximum conformity between the glenoid and humerus.11

The Constant system allocates a large proportion of the score to strength. The measurement of force is complex and open to debate; it is inevitable that a compromise is required, particularly if this score is to be useful in the clinical setting. Isokinetic dynamometers and measurements of torque can have no role to play in this context.17,19,21,23 Furthermore, there is no proven superiority of isokinetic testing over adequate isometric testing.25 A portable isometric dynamometer that allows measurement of the average force exerted during either 3 or 5 seconds has become available. The committee is aware that many investigators have limited budgets and may have difficulties acquiring this device. Alternative methods, incorporating the principles of the Isobex device with particular reference to fixing the measuring device and taking a measurement over time, with a spring balance have been proposed, hopefully encouraging the smaller units and younger investigators to participate in shoulder research. Thus the measurement little more than pseudoscientific quantification of manual muscle testing.

The majority of other scores commonly used originate as presentation tools in publications usually on specific conditions. Many of them are complementary to the Constant score. For example, in describing instability,23,26 investigators may choose an additional score. Extremely elaborate and precise systems do not necessarily translate into a more discriminating score than simpler systems.14 For example, the Patte score for the analysis of shoulders with chronically painful disabilities,20 which assessed pain, “function,” muscular strength, and daily handicap, was very complicated and time-consuming. It has essentially been abandoned in favor of simpler systems with relatively fewer, adequate parameters that had a greater power of discrimination. Other systems are overly inclusive, as they attempt to produce a documentation form rather than an efficient scoring system. This is the case of the American Shoulder and Elbow Surgeons (ASES) evaluation form,2 which allows standardization of the examination and the documentation of the findings.

The University of California, Los Angeles (UCLA) shoulder scoring system is a true functional scoring system and does not mix diagnostic and functional assessment. It includes pain, motion, “function,” strength, and patient satisfaction, thus all of the most important rating criteria. The score seems relatively simple and straightforward in most of its definitions. The low total score of only 35 points, however, is not forgiving of any minor judgement error. Strength of the shoulder is assessed manually. The UCLA score is one of the few rating systems that includes the crucial

<table>
<thead>
<tr>
<th>Table II</th>
<th>Constant score normal values for different age and sex groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Male subjects</td>
</tr>
<tr>
<td></td>
<td>Right</td>
</tr>
<tr>
<td>21-30 y</td>
<td>97</td>
</tr>
<tr>
<td>31-40 y</td>
<td>97</td>
</tr>
<tr>
<td>41-50 y</td>
<td>86</td>
</tr>
<tr>
<td>51-60 y</td>
<td>94</td>
</tr>
<tr>
<td>61-70 y</td>
<td>83</td>
</tr>
<tr>
<td>71-80 y</td>
<td>76</td>
</tr>
<tr>
<td>81-90 y</td>
<td>70</td>
</tr>
<tr>
<td>91-100 y</td>
<td>60</td>
</tr>
</tbody>
</table>

Data are from Constant.6

*Significant difference compared with opposite sex.
factor strength in its assessment. Two types of criticism appear appropriate for the assessment of strength in this system. First, it would be desirable that the method of measurement be more precisely specified. The positions of the arm and hand are not defined: although it is obvious that whether a patient can lift his or her arm against gravity with the elbow fully extended or fully flexed constitutes a major difference, both can be assigned 3 points. The duration of resistance is not defined, and for a grade of 1, we do not know which muscle must contract to obtain 1 point. More importantly, however, it appears that the value of shoulder strength is relatively underrated, and differences in strength are not appropriately reflected in the score.

Despite these criticisms, the UCLA score has been found to be easy and quick to use and to allow interpretations comparable to those of other, more sophisticated scoring systems. In the absence of a reproducible strength measurement, its use may still not be justified. The reproducibility of this score has not been tested, to our knowledge. This is of major concern and is common to many systems. It is extraordinary that new evaluation and outcome measures are still being designed and published without full assessment of their validity and sensitivity.

A further issue that has yet to be assessed is the intimate relationship of the elbow to the shoulder particularly in patients with rheumatoid disease. The current modifications of the strength measure of the Constant score seem logical, reflecting the functional deficit caused by a painful elbow with loss of extension in terms of the strength of an effectively shortened span or lever arm. It is possible that the score may have validity for isolated conditions of the elbow, but this will require a study involving comparison with a recognized functional assessment such as the Morrey score.

**CONCLUSION**

The SECEC Research and Development Committee appreciates the concerns regarding and limitations of the Constant score but believes that we are already in possession of a score that is widely accepted in Europe and increasingly so in other parts of the world and is significantly better than any of the alternatives from the other side of the Atlantic.

This article makes recommendations and modifications, which are the result of considerable discussion and take into account many validation experiments from a number of centers. Undoubtedly, further changes will be made, and the Constant score may be superseded in the course of time, but it should not be thrown out until an alternative is produced that has been properly tested and shown to have a high interobserver and intraobserver reliability. Thus the ideal score must be not only reproducible but also easy and quick to use in a clinical or research setting without the need for sophisticated equipment.

**REFERENCES**