SPECIAL ISSUE - ORIGINAL ARTICLE

Inter-operator reliability and prediction bands of a novel protocol to measure the coordinated movements of shoulder-girdle and humerus in clinical settings

Pietro Garofalo · Andrea Giovanni Cutti ·
Maria Vittoria Filippi · Stefano Cavazza ·
Alberto Ferrari · Angelo Cappello · Angelo Davalli

Abstract A clinical motion analysis protocol was developed to measure the coordinated movements of shoulder-girdle and humerus (*girdle-humeral rhythm*—GD-H-R) during humerus flexion–extension (HFE) and ab-adduction (HAA), through an optoelectronic system. In particular, the protocol describes the GD-H-R with 2 angle–angle plots for each movement: girdle elevation–depression and protraction–retraction vs HFE, and vs HAA. Each of these plots is further divided in two subplots, one for the upward and one for the downward phases of the movement. By involving 11 participants and 2 operators, we measured the protocol's inter-operator reliability which ranged from very-good to excellent depending on the angle–angle plot (median values of the inter-operator coefficient of multiple correlation for the angle–angle plots higher than 0.94). We then computed the subjects' average control patterns, together with statistically meaningful prediction bands. $\pm 1SD$ confidence bands were also computed and their width ranged from $\pm 0.5°$ to $\pm 4.6°$. Based on these results we could conclude that the method is robust and able to identify even limited differences in the GD-H-R.

1 Introduction

As soon as patients are surgically treated for shoulder instability or rotator-cuff tears and the active mobilisation begins, an altered coordination between shoulder-girdle and humerus rotations can be observed (*video, on-line material*). In particular, an abnormal girdle-thoracic kinematics is generally evident during humerus elevations, involving range and/or timing of elevation–depression and protraction–retraction [29]. One of the targets of rehabilitation is therefore to remove these compensatory movements, recovering a normal coordination between shoulder-girdle and humerus, which will be referred herein as *girdle-humeral rhythm* (GD-H-R). The standard clinical rating scales for the assessment of shoulder impairment, i.e. Constant and ASES [9], record valuable information about shoulder overall mobility, pain, power, functionality and stability. However, they do not address *how* a movement is performed, nor describe the specific alterations of the GD-H-R. Quantitative 3D motion analysis with the extraction of focused parameters appears a possible solution to overcome these limitations, and especially for monitoring the evolution of the GD-H-R during the different periods of rehabilitation.

Since the shoulder-girdle is formed by the clavicle and scapula, the measure of the GD-H-R can be decomposed in the measure of the clavicle-humeral and scapulo-humeral rhythms [8, 12, 14, 18–20, 23, 24, 28, 30, 34, 35].

P. Garofalo · A. G. Cutti (✉) · M. V. Filippi · A. Ferrari ·
A. Davalli
I.N.A.I.L. Prostheses Center, Via Rabuina, 14,
40054 Vigorso di Budrio, Bologna, Italy
e-mail: ag.cutti@inail.it; agcutti@inail-starter.org

P. Garofalo · A. Ferrari · A. Cappello
Department of Electronics, Computer Science and Systems,
University of Bologna, Bologna, Italy

S. Cavazza
Ospedale San Giorgio, Ferrara, Italy

However, the measure of the scapulo-humeral rhythm is not always possible, e.g. due to clinical routine constraints which preclude the use of currently available tracking systems for the scapula [14, 25]. The need to complete the acquisitions of (1) both shoulders of a subject, (2) either muscular or slim, (3) non-invasively, (4) within a time interval comparable to the time required to complete a patient's anamnesis and a clinical scale (i.e. 30 min), (5) with no constraint on the maximal humeral elevation admissible, (6) with the measure of multiple repetitions of the activities in order to obtain motion cycles truly representative of the subject's kinematics, is a concrete example of such a combination of constraints, taken from the authors' clinical routine.

In these same conditions, however, even though the single clavicle-humeral and scapulo-humeral rhythms cannot be measured, at least the measure of the overall GD-H-R could still be possible, with the non secondary advantage for the overall GD-H-R of being the most evident and visually detectable clinical sign.

Despite its clinical relevance, the quantitative measure of the alterations of the overall GD-H-R has received little attention in the movement analysis literature: no protocols are available for this purpose with the exception of a preliminary proposal by these same authors [7, 10, 32].

The purposes of this work were therefore (1) to complete the definition of the protocol to measure the overall GD-H-R in routinely clinical settings, and (2) to determine important metric properties of the protocol for clinical applications. Concerning this second purpose, since different operators can potentially conduct the measurements on a subject on separate sessions, we verified that the protocol is robust to a change in operator by measuring its inter-operator reliability on a group of control subjects. Moreover, since it is important to know if a subject is recovering a "normal" GD-H-R with the rehabilitation, we measured the average GD-H-R of the controls along with statistically meaningful prediction bands. An example of clinical application of the controls' average response and prediction bands is finally provided in Sect. 3.3.

2 Materials and methods

2.1 Development of the protocol

The protocol was intended to measure the overall GD-H-R in routinely clinical settings. For this purpose, we developed the protocol assuming the six constraints detailed in the introduction plus one more: to include in the protocol only motor tasks of the Constant and ASES scales, in order to ease the cross analysis of clinical and kinematic data.

Given these seven constraints, we developed the protocol by addressing the following six issues: (1) choice of the measurement system, (2) identification of the segments of interest and formulation of the reference kinematic model of the shoulder, (3) definition of the segments' coordinate systems and angles, (4) positioning of the system's sensors on the body of a subject, (5) identification of the activities the subjects have to execute, (6) formulation of the outcome measure and specification of the data processing.

2.1.1 Measurement system

Since non-invasive methods were required, the protocol was developed for a system able to track in time the position of skin-mounted sensors, e.g. optoelectronic, electromagnetic or ultrasound. Since the reliability analysis of the protocol presented in Sect. 2.2 was performed with an optoelectronic system [5], hereinafter we will only explicitly refer to this type of system.

2.1.2 Segments of interest and reference kinematic model

Thorax, shoulder-girdle and humerus were considered as the three segments forming the shoulder. In particular, the shoulder-girdle was defined as the segment connecting the midpoint between the Incisura Jugularis (IJ) and C7, and the centre of the Glenohumeral Head (GH). Thorax, shoulder-girdle and humerus form an open kinematics chain (Fig. 1a, b). Since segment kinematics is of interest,

Fig. 1 Segments of interest with their anatomical landmarks (*dots*), coordinate systems and marker placement. X_{GRD}, Y_{GRD}, Z_{GRD} are the axes of the girdle anatomical frame. Y_{THX} is the vertical axis of the thorax anatomical frame. **a** Frontal plane; **b** sagittal plane; **c** positioning of the clusters of markers
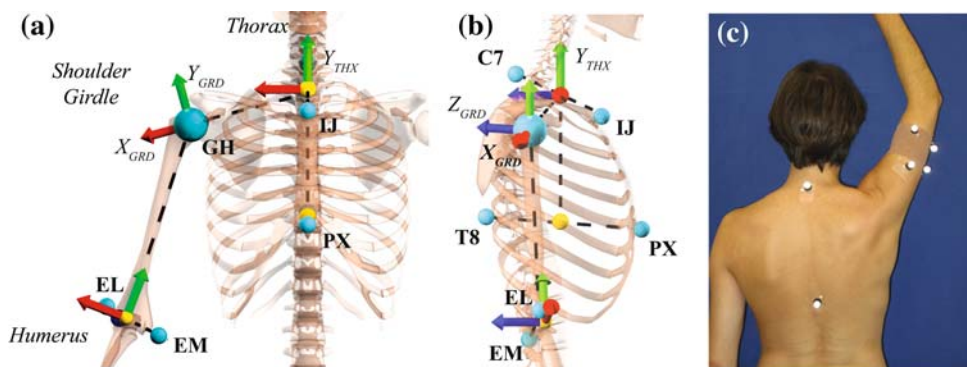
the girdle-thoracic motion was modelled with 2 degrees of freedom (DoFs), namely elevation–depression (ED) and protraction–retraction (PR). The humero-thoracic motion was instead modelled with 3 DoFs: flexion–extension (FE), ab-adduction (AA) and internal–external (IE) rotation.

2.1.3 Definition of the segments' frames and angles

To measure the girdle-thoracic and humero-thoracic angles, we defined anatomical frames for thorax, shoulder-girdle and humerus (1) consistently with the kinematic model, and (2) based on the identification of relevant anatomical landmarks.

The ISG standard [38] was followed for the anatomical frames of thorax and humerus (Table 1).

For the shoulder-girdle no standard is available. Therefore, for the shoulder-girdle the axes of the anatomical frame were defined as follows (Table 1; Fig. 1a, b): X_{GRD} from the midpoint of IJ and C7 to GH; Z_{GRD} perpendicular to X_{GRD} and the Y-axis of the thorax; Y_{GRD} perpendicular to X_{GRD} and Z_{GRD}.

To measure the ED and PR angles, the relative orientation of shoulder-girdle and thoracic frames was decomposed with the Euler angles sequence $YZ'X''$. It is important to notice that with the anatomical frame adopted for the shoulder-girdle, the third Euler angle of the sequence is always mathematically null, consistently with the mechanical model assumed.

To measure the FE, AA and IE angles, the relative orientation of the humerus and thoracic frames was decomposed with the Euler angles sequence $XZ'Y''$ for movements in the sagittal plane (obtaining FE, AA, IE),

and with the sequence $ZX'Y''$ for movements in the frontal plane (obtaining AA, FE, IE).

2.1.4 Sensors placement and anatomical calibrations

To link the thorax frame to the system's sensors, four markers are directly positioned over the anatomical landmarks IJ, Processus Xiphoideus (PX), C7, and T8. In case of visibility or placement problems (e.g. due to long hair or bra), two markers are positioned as apart as possible on the sternum, while the other two are positioned as close as possible to C7 and T8. The anatomical landmarks are then calibrated [3, 6, 36] with respect to this cluster of four markers.

To link the humerus frame to the system's sensors, the anatomical landmarks Epicondylus Lateralis (EL), Epicondylus Medialis (EM) and GH are calibrated relative to a cluster of four markers positioned on the humerus (Fig. 1c). Specifically, three of the four markers were positioned posteriorly, on a CO-PLUS (BSN Medical, UK) cuff wrapped around the humerus. The fourth marker is positioned at the insertion of the deltoid as proposed in Murray and Johnson [27]. The practice suggests that this configuration limits the deformation of the humerus cluster due to elbow flexion. GH was calibrated with respect to the humerus cluster by application of the regression equations described in Charlton [4] and Murray and Johnson [27]. These regression equations require the static calibration of the acromion (AC) relative to the thorax frame.

The link of the shoulder-girdle frame to the system's sensors comes as a consequence from the previous steps, since the shoulder-girdle frame is based on anatomical

Table 1 Definition of thorax, humerus and shoulder-girdle anatomical frames

Segment	Axes definition
Thorax	$Y_{THX} = ((IJ + C7)/2 - (PX + T8)/2)/\|(IJ + C7)/2 - (PX + 8)/2\|$
	$X_{THX} = Y_{THX} \wedge (T8 - PX)/\|Y_{THX} \wedge (T8 - PX)\|$: medio-lateral
	$Z_{THX} = X_{THX} \wedge Y_{THX}$: antero-posterior
	Origin $=$ IJ
Shoulder-girdle	$Y_{GRD} = (Z_{GRD} \wedge X_{GRD})/\|\cdot\|$: upward
	$X_{GRD} = (GH - (IJ + C7)/2)/\|(GH - (IJ + C7)/2)\|$: medio-lateral
	$Z_{GRD} = (X_{GRD} \wedge Y_{THX})/\|\cdot\|$: antero-posterior
	Origin $=$ IJ
Humerus	$Y_{H1} = (GH - E)/\|(GH - E)\|$: longitudinal
	$Z_{H1} = Y_{H1} \wedge (EM - EL)/\|Y_{H1} \wedge (EM - EL)\|$: antero-posterior
	$X_{H1} = (Y_{H1} \wedge Z_{H1})/\|\cdot\|$: medio-lateral
	Origin $=$ GH
	$E = (EL + EM)/2$

landmarks (GH, IJ and C7) already tracked through the markers on humerus and thorax.

2.1.5 Activities to be measured and number of repetitions

The protocol requires the subject under analysis to execute two activities common to both the ASES and Constant scales, i.e. humerus flexion–extension (HFE—sagittal plane) and ab-adduction (HAA—frontal plane). Before starting with the measurements, the subject has to familiarise with the movement. When confident, the subject is asked to repeat each movement at least five times, with an interval of relax after each repetition. Five repetitions are assumed to be sufficient to gather at least four cycles truly representative of the subject's GD-H-R, similarly to [2, 18].

2.1.6 Data processing and output of the protocol

The output of the protocol is a graphical representation of the GD-H-R during HFE and HAA. Specifically, the GD-H-R of HFE is described by four angle–angle plots, two for the upward phase of the movement (humerus moving cranially) and two for the downward phase (humerus moving caudally): ED vs FE and PR vs FE—upward phase, ED vs FE and PR vs FE—downward phase (e.g. see Fig. 3a). Similarly, the GD-H-R of HAA is described by: ED vs AA and PR vs AA—upward phase, ED vs AA and PR vs AA—downward phase (e.g. see Fig. 3b). Distinguishing the upward and the downward phases is important, because motion patterns can change between the two conditions [2, 18, 22].

To reach this representation of the GD-H-R, before being plotted FE, AA, ED and PR undergo two macrosteps: (1) a segmentation procedure to identify the repetitions of the movement and distinguish between the upward and downward phases, and (2) offset removal from ED and PR patterns.

For the seek of brevity, the details of the segmentation algorithm are reported in Annex 1 *(on-line material)*.

The algorithm to remove the offsets of ED and PR starts from the output of the segmentation algorithm and can be simply summarised in two steps. Firstly, the values of the ED angle in time, $ED(t)$, correspondent to the onsets of the upward phases are considered and their median value is computed ($\overline{\overline{ED}}$). Secondly, $\overline{\overline{ED}}$ is subtracted to $ED(t)$. Identical steps apply to PR.

2.2 Assessment of the protocol

The protocol was tested in vivo in two experiments that were executed simultaneously on a group of control subjects and with the involvement of two operators. The first

experiment aimed to assess the inter-operator reliability of the protocol. The second experiment aimed to compute the average GD-H-R of the controls along with statistically meaningful prediction bands.

2.2.1 Subjects and operators

Eleven able-bodied subjects (30 ± 3 years old, 9 males, 2 females) participated in the experiments after signing an informed consent. A physical examination excluded any pathology in the subjects' upper-limbs. The experiments also involved two operators (O_1 and O_2) with even familiarity in the application of the protocol (five patients each).

2.2.2 Set-up, procedure and preliminary data processing

A common set-up and procedure was used for both experiments. Specifically, the GD-H-R of a side of each subject was measured once by operator O_1 and O_2 through the application of the protocol, by using a Vicon MX 1.3 optoelectronic system (Oxford Metrics, UK) with a sampling frequency of 100 Hz. For each of the two acquisitions, therefore, each subject repeated five times the two movements HFE and HAA, but only the last four repetitions were used for the subsequent computations.

The side acquired for each subject was randomly selected as well as the order of the operators.

For each subject the acquisitions by O_1 and O_2 were from 10 to 30 min apart. The second operator was not aware of the position of the markers adopted by the first.

The first part of data processing was also common to the two experiments. For each subject and angle–angle plot, it was firstly selected the range of the angle on the X-axis common to both the four waveforms obtained by O_1 and by O_2. Then 80 equidistant values were selected in this common range and each of the eight waveforms was interpolated (using a cubic spline) and re-sampled on the 80 values.

2.2.3 Experiment 1: inter-operator reliability

The inter-operator reliability of the protocol in measuring the GD-H-R was quantified by means of two different parameters: (1) the coefficient of multiple correlation (CMC), similarly to Kavanagh et al. [15], and (2) the average inter-operator standard deviation (IOSD), as described by Meskers et al. [24]. The former was used since in recent years the CMC is becoming a standard for the measure of the repeatability of waveforms [13, 15, 16, 21, 26, 39]. The latter was used instead to enable the comparison of the results of this study with those reported in [24] about clavicle-humeral and scapulo-humeral rhythm.

2.2.3.1 Measure of the inter-operator reliability through CMC 2.2.3.1.1 Inter-operator reliability within each subject   The inter-operator reliability of the protocol was firstly quantified separately for each subject, for each of the eight angle–angle plots provided by the protocol. For what follows, it may be useful to recall that in each plot eight waveforms are reported, four measured by O_1 and four measured by O_2.

In details, for each plot we followed the two steps below:

(1)   We checked if the subject was highly repeatable in the execution of the movement, both when acquired by O_1 and by O_2 (intra-subject repeatability). To check for the intra-subject reliability we computed the similarity between the four waveforms acquired by each operator. The similarity was measured through the CMC named by Kadaba et al. *within-day CMC* [13], which will be referred to herein as *intra-subject CMC*. For the computation of the intra-subject CMC we applied the technique described in [16]. The intra-subject CMC is a single scalar value which combines the similarity for both operators, i.e. it tends toward 1 if both the waveforms of O_1 and of O_2 are similar, and toward 0 otherwise.

Only if the subject presented an intra-subject CMC higher that 0.95 we proceeded with step 2.

(2)   We evaluated the inter-operator reliability by comparing the similarity of the eight waveforms obtained by O_1 and O_2 considered altogether. In particular, the inter-operator reliability was evaluated, similarly to [15], through the second CMC proposed in [13], i.e.:

$$\mathrm{CMC} = \sqrt{1 - \frac{\sum_{i=1}^{O}\sum_{j=1}^{R}\sum_{k=1}^{A}\left(Y_{ijk} - \overline{Y}_k\right)^2 \Big/ A(OR-1)}{\sum_{i=1}^{O}\sum_{j=1}^{R}\sum_{k=1}^{A}\left(Y_{ijk} - \overline{Y}\right)^2 \Big/ (ORA-1)}} \tag{1}$$

where $k = 1,\ldots, A$ ($A = 80$): differentiate the 80 samples of the angle on the X-axis. $j = 1,\ldots, R$ ($R = 4$): differentiate the four repetitions of each movement, in order of execution. $i = 1,\ldots, O$ ($O = 2$): differentiate the waveforms obtained by operator O_1 from those of operator O_2. Y_{ijk} is the Y-axis angle correspondent to the k-th X-axis angle, of the j-th waveforms, obtained by the i-th operator; $\overline{Y}_k$ is the average among all the $R \times O$ waveforms of the subject at the k-th X-axis angle; $\overline{Y}$ is the grand mean of all the waveforms from all the operators.

This CMC will be referred hereinafter as inter-operator CMC.

As for the intra-subject CMC, when the waveforms are similar, CMC tends to 1. If the waveforms are dissimilar, CMC tends to 0. Thus, the inter-operator CMC measures the reliability of the waveforms by the two operators and is a combined measure over eight repetitions.

It may be observed that the inter-operator CMC can be lowered by a poor intra-subject repeatability, i.e. by the dispersion of the waveforms. More specifically, the inter-operator CMC can be lowered by (1) the low repeatability of the subject within each acquisition (intra-subject repeatability), and (2) the biological variability of the subject between the acquisitions of the two operators. To compensate for the first cause, through step (1) we considered only those subjects with excellent intra-subject repeatability. The effect on the inter-operator CMC of the (limited) biological variability of these subjects was ascribed to the inter-operator error. To compensate for the second cause, we chose a very limited time interval between the two acquisitions (30 min as worst case). This excluded large between-days variability. As before, the effect on the inter-operator CMC of the biological variability between the two acquisitions was entirely ascribed to the inter-operator error.

2.2.3.1.2 Inter-operator reliability among the subjects   Once assessed the inter-operator reliability for each plot and subject, statistical parameters were computed to describe the inter-operator reliability among the subjects of the study.

For each plot, the distribution of the inter-operator CMC among the subjects was firstly checked for normality by visual inspection of normality plots. Since the inter-operator CMC did not generally show a normal distribution among the subjects (ceiling effect), the median and interquartile distance (IQD) was computed for the subjects and a box and whiskers plot was created. For each plot, the protocol inter-operator reliability was then interpreted as follows, based on the median and IQD range of the inter-operator CMC and based on previous publications [13, 15, 39]:

– $0.65 < \mathrm{CMC} < 0.75$ moderate
– $0.75 < \mathrm{CMC} < 0.85$ good
– $0.85 < \mathrm{CMC} < 0.95$ very good
– $0.95 < \mathrm{CMC} < 1$ excellent

Finally, to access if certain plots were more reliable than others, we compared the inter-operator CMCs of the four angle–angle plots of HFE and of the four angle–angle plots of HAA through repeated measures non-parametric ANOVA (Friedman's test).

2.2.3.2 Measure of the inter-operator reliability through IOSD   In Meskers et al. [24], the inter-operator reliability was measured by computing for each subject the numerator

of the ratio of Eq. 1, and this quantity was called inter-operator variance. The average of the inter-operator variance was then computed over the subjects. The root square of this average, i.e. the average IOSD, was finally reported. For the seek of comparison, this same procedure was followed here.

2.2.4 Experiment 2: prediction bands and ± 1SD confidence bands

It is of clinical interest to know if the angle–angle patterns describing the GD-H-R of a subject are "normal" patterns. For instance, it can be of interest to know if a single upward phase of ED vs FE is "normal". A "normal" pattern will be assumed herein as a pattern which remains within the protocol's *minimal detectable difference band* (also called herein "prediction band" or MDDB) from the control subjects' average pattern. The *minimal detectable difference* [37] band for a given angle–angle plot is a $(1-\alpha)\%$ confidence band around the average pattern of the control subjects, with a clear statistical meaning: when a pattern of a new subject is outside of the prediction band, there the subject's pattern is different from the control average with a $(1-\alpha)\%$ probability. MDDB bands are alternative to the more common confidence bands used in motion analysis based on ±1SD of the population [33], with the remarkable advantage for the former of being meaningful from the inferential statistics viewpoint. As detailed below (Sect. 2.2.4.1), MDDBs are directly related to the standard errors of measurement (SEM) of the protocol [37].

For each of the eight angle–angle plots describing the GD-H-R, the control subjects' average and MDDB was computed as described in Sect. 2.2.4.1, starting from the kinematic data collected by both operator O_1 and O_2, altogether.

To compare the results from this study with previous works concerning the clavicle and scapulo-humeral rhythm, we also computed for each plot the ±1SD confidence bands as described in [33].

2.2.4.1 Computation of an average pattern and its prediction band (MDDB) For the description below, it is worth recalling that each angle–angle plot has 80 values in abscissa (see Sect. 2.2.2). Since 11 subjects were measured by two operators, and for each subject four repetitions were considered, 88 ordinate values exist for each abscissa.

To compute the average pattern and the MDDB for each angle–angle plot, each of the 80 angle values (of FE for HFE and AA for HAA) in abscissa was separately considered.

For the i-th abscissa, the average ordinate from its 88 ordinates was firstly computed, and named M_i.

For the i-th abscissa the MDDB's upper and lower values were then computed from its 88 ordinates as follows:

(a) A two-factors repeated measures ANOVA was executed, with the "ordinate angle" as dependent variable and with "operator" and "repetition" as the two independent variables, with two and four levels respectively (Supplementary Figure 1). The repeated measures ANOVA method allows to isolate the contribution due to the between-subjects variability and to the within-subject variability (MS_w). The within-subjects variability takes into account all the systematic and random errors associated with the application of the protocol, i.e. the variability of the data from repetition-to-repetition, operator-to-operator, a possible interaction of these two factors, and the total residual.

(b) The square root of MS_w was computed, thus obtaining the SEM of the protocol [1, 11, 37]:

$$\text{SEM}(i) = \sqrt{MS_w(i)} \tag{2}$$

The SEM is usually referred to as the "typical error" and, being based on the MS_w only, is a fixed characteristic of any measure, regardless of the sample of subjects under investigation [11].

(c) The upper and lower values of MDDB are computed from the SEM as follows [37]:

$$\text{MDDB}_{\text{up}(95\%)} = M_i + 1.96 \times \text{SEM} \times \sqrt{2} \tag{3}$$

$$\text{MDDB}_{\text{low}(95\%)} = M_i - 1.96 \times \text{SEM} \times \sqrt{2} \tag{4}$$

where

$$\text{MDDB}_{(95\%)} = 1.96 \times \text{SEM} \times \sqrt{2} \tag{5}$$

is defined as the minimal difference (MD) detectable through the protocol [37].

Since MS_w incorporated the measure of the repetition-to-repetition variability, operator-to-operator variability, their interaction and the residual, it can be stated that: a new single ordinate value for the i-th abscissa (1) obtained from a new subject, and (2) measured by a new operator, is different with a 95% probability from the controls average if it falls above the value indicated by $\text{MDDB}_{\text{up}(95\%)}$ or below the value indicated by $\text{MDDB}_{\text{low}(95\%)}$.

3 Results

3.1 Results for experiment 1—inter-operator reliability

The intra-subject variability was higher than 0.95 in 86/88 cases. Subject 3 presented an intra-subject variability of 0.89 in the upward phase of PR vs FE and subject 4

Fig. 2 Inter-operator CMCs among the control subjects for the different angle–angle plots describing the GD-H-R. **a** Box and whiskers plots with notches for all the eight angle–angle plots coming from the protocol; **b** Median, quartile values and IQD for the distributions reported in **a**
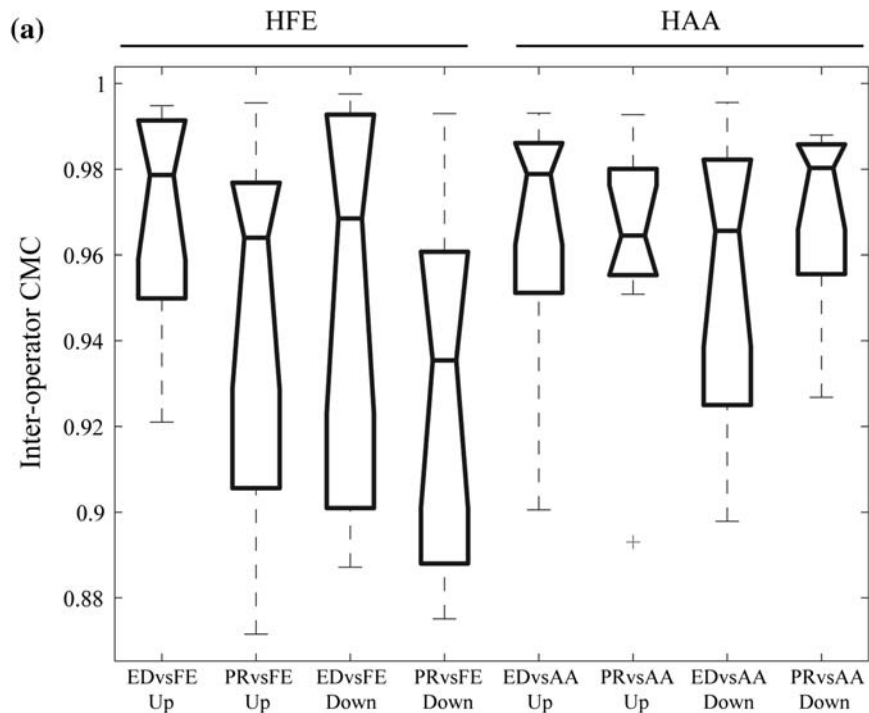
(b)	HFE				HAA			
	EDvsFE	PRvsFE	EDvsFE	PRvsFE	EDvsAA	PRvsAA	EDvsAA	PRvsAA
	Up	Up	Down	Down	Up	Up	Down	Down
Median	0.98	0.96	0.97	0.94	0.98	0.96	0.97	0.98
75 percentile	0.99	0.98	0.99	0.96	0.99	0.98	0.98	0.99
25 percentile	0.95	0.91	0.90	0.89	0.95	0.96	0.93	0.96
Interquartile Distance	0.04	0.07	0.09	0.07	0.03	0.02	0.06	0.03

presented an intra-subject variability of 0.92 in the downward phase of ED vs FE. These two cases were excluded from the computation of the inter-operator CMC.

Box and whisker plots with notches of the inter-operator CMCs of each angle–angle plot are reported in Fig. 2a. Numeric values for the medians and IQDs illustrated in Fig. 2a are reported in Fig. 2b.

The median values in the boxes are not generally centred and this suggests the non normality of the distributions, also confirmed by the inspections of the normality plots. CMCs median values among the eight angle–angle plots were all above 0.94. The lowest first quartile was 0.89. These results suggest that the inter-operator reliability varied, depending on the angle–angle plot, from "very good" to "excellent".

Friedman's tests confirmed that no statistically significant differences exist neither between the inter-operator CMCs of the angle–angle plots associated to HFE ($p = 0.07$), nor to those of HAA ($p = 0.52$).

The IOSD values for the eight angle–angles plots are reported in Table 2.

3.2 Results for experiment 2

The controls subjects' average pattern, MDDB and ±1SD bands for each of the eight angle–angle plots are reported in

Table 2 IOSD values describing the inter-operator reliability for each movement (HFE, HAA), phase of the movement (upward and downward) and angle–angle plot

	HFE				HAA			
	Upward		Downward		Upward		Downward	
	ED vs FE	PR vs FE	ED vs FE	PR vs FE	ED vs FE	PR vs FE	ED vs FE	PR vs FE
IOSD (deg)	1.78	2.00	2.14	2.28	1.69	1.72	1.74	1.63

Fig. 3. To ease the application of the protocol by other research groups, the numeric data required to plot the average plots, MDDBs and ±1SD confidence bands are provided as on-line material (*Microsoft Excel file*). MDDBs width ranged among the angle–angle plots between ±1.5° and ±7.9°. The SEM ranged therefore between ±0.6° and ±2.8°. The ±1SD bands ranged instead between ±0.5° and ±4.6°.

3.3 Clinical application

To illustrate an example of clinical application of the protocol, average patterns and MDDBs, let us consider the

case of a 34-year-old male patient surgically treated for rotator cuff tears. The patient was acquired with the protocol three times, i.e. after 42, 70 and 122 days from the surgery.

For this patient, the specific clinical questions were: (1) if the ED vs FE pattern measured during the first, second and third acquisition could be assimilated to a normal pattern, and (2) if the rehabilitation was effective in restoring a normal ED vs FE pattern.

Since the intra-subject repeatability of the patient for the ED vs FE plot was higher than 0.95, the repetition-to-repetition variability is the same of the population from

Fig. 3 Control subjects' average patterns, MDDBs (prediction bands) and ±1SD confidence bands for the eight angle–angle plots describing the GD-H-R. **a** The four angle–angle plots associated to HFE; **b** the four angle–angle plots associated to HAA
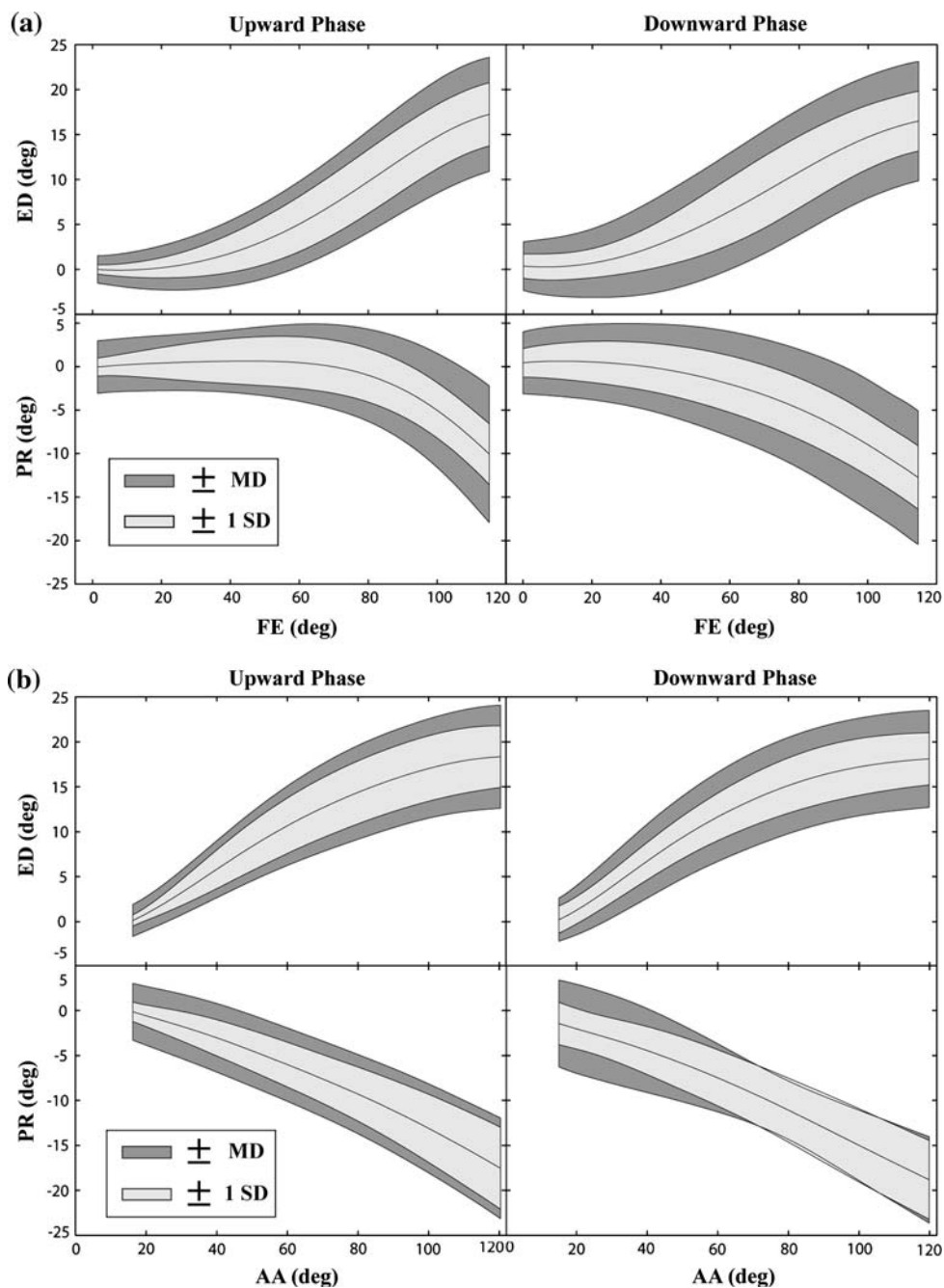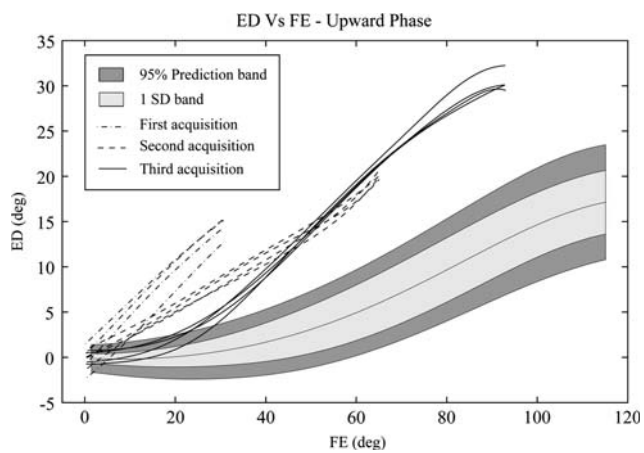
Fig. 4 Elevation–depression versus FE patterns of three acquisitions of a typical patient recovering from surgery for rotator cuff tear, during the HFE task. Controls' average and 95% MDDB are provided for statistical comparison

which the MDDB was computed. In all three acquisitions, therefore, the analysis could be performed comparing just one repetition of the movement to the average pattern and MDDB. However, to further decrease the probability of false positives, all four repetitions from each acquisition were compared to the average controls' pattern.

The ED vs FE patterns for all the three acquisitions are reported in Fig. 4.

To answer to the first clinical question, it should be noticed that the patterns of the first and second acquisitions felt above the MDDB almost immediately, i.e. from very small values of FE. This indicates a 95% probability of differences between the patients and the controls' average for almost the entire pattern. In the third acquisitions, instead, the patient's pattern remained within the MDDB until 30° of FE. This indicated that only from 30° to the end of the range of motion for FE the patient differed from the control's average with a probability of 95%.

For what concerns the second questions, based on (1) the previous considerations, (2) the fact that the shape of the pattern in the third acquisition is closer to the control's average shape, and (3) the increase in the range of motion of FE from the first to the third acquisition, it can be stated that the rehabilitation is having a positive influence on the patients but the recovered rhythm of FE and ED still remains statistically different from the normal pattern.

4 Discussion and conclusion

A motion analysis protocol was developed to measure the overall GD-H-R in clinical settings. Specifically, the protocol allows to measure the coordination between humerus

FE and shoulder-girdle ED and PR during HFE, and the coordination between humerus AA and shoulder-girdle ED and PR during HAA. For a single arm the protocol requires only eight markers, the calibration from three to seven anatomical landmarks, and the dynamic acquisition of two motor activities. No specialised equipment is required (e.g. a scapula locator). The girdle segment is not based on a scapula-tracker over the acromion, but only on the calibration of GH relative to the humerus cluster. Therefore, the validity of the measure is not a priori limited to 120° of humerus elevation [14, 25]. Overall, the protocol fulfils the clinical constraints declared in Sect. 1.

To the authors' knowledge, the protocol is original in the aims, in the description provided of the GD-H-R and in the definition of the shoulder-girdle coordinate system. This last represents an update of a previous proposal by these same authors [7–32]. With the new coordinate system, the third Euler angle of the sequence YZ'X'' used to compute the orientation of the shoulder-girdle relative to the thorax is always mathematically null. The Euler angles provided are therefore consistent with the mechanical model assumed for the 'joint' connecting the shoulder-girdle with the thorax. This was not the case with the previous proposal.

The protocol includes an offset removal step for the ED and PR patterns. The same offset is removed from all the four repetitions of ED (PR). This step was included since this common offset among the repetitions is strictly dependent on the specific anatomy and static posture of the subject considered, i.e. by the subject-specific position of the anatomical landmarks IJ, C7 and GH. Similarly to gait protocols [13], the main clinical interest was here to detect differences in the patterns describing the GD-H-R rhythm. Differences between subjects, if not compensated for, can increase the width of the MDDB and therefore decrease the sensitivity of the protocol in detecting differences in the angles patterns.

Since a common offset was removed from all the repetitions, it should be noticed that the variation of the initial value of ED and PR between the different repetitions (which can be of clinical interest and affects the value of the intra-subject variability), has not been removed and is therefore taken into account in the MDDBs.

The protocol requires the execution of activities included in the Constant scale. This was done since in preliminary assessments [7, 10, 32] the Constant resulted quite insensitive to the amplitude and pattern of patients' compensatory movements. Although extended clinical experimentations are needed, the information coming from the protocol can be intended as an integration of this popular clinical scale.

For what concerns the in vivo assessment of the protocol, the results from the first experiment confirmed that the

protocol has an inter-operator reliability ranging from "very-good" to "excellent", with no differences between the angle–angle plots considered for HFE and HAA. Unfortunately, at the moment no other studies are available in upper-extremity motion analysis which have assessed the inter-operator reliability by means of the inter-operator CMC. This is because the CMC and the inter-operator CMC is currently becoming a well accepted standard for this measurement. However, the inter-operator reliability results for the IOSD can be compared with previous results by Meskers et al. [24], and in particular with the IOSD they reported for the clavicle-humeral and scapulo-humeral rhythm. Meskers reported for these two rhythms IOSD values ranging from $2.2°$ to $5.6°$, with a mean value of $3.4°$. In the present study, IOSDs ranged from $1.6°$ to $2.3°$, with a mean value of $1.9°$. This demonstrates that the measure of the GD-H-R with the protocol presented here is more reliable than the measure of the scapulo-humeral and clavicle-humeral rhythm through the palpation method. A possible explanation of these results can lay in the fact that the protocol presented here does not require the intervention of an operator during the acquisitions, but only to set-up the procedure.

Results from the second experiment showed that the average patterns and MDDBs generally differed from the upward and the downward phase of each movement. This confirmed the need to consider the two phases separately in the analysis of the kinematics of a subject, as well as the need to always incorporate in an upper-extremity protocol a segmentation algorithm. The differences between the average patterns in the upward and downward phases are consistent with previous findings for the scapulo-humeral and clavicle-humeral rhythm [2, 18, 22].

Unfortunately, no other studies in upper-extremity motion analysis have ever reported any sort of prediction bands. The comparison with previous literature has therefore to be based on the comparison of the $\pm 1SD$ confidence bands. In particular, the best candidate for comparisons appears the paper by Meskers et al. [25], who have reported $\pm 1SD$ confidence bands for the scapulo-humeral rhythm measured with the tripod method, not considering just one observer, but three, similarly to the present study. Meskers reported the narrowest band for the scapula medio-lateral rotation, with 1SD ranging from about $5°$ to $10°$. In the present study, the widest band presents a $1SD = 4.6°$ (PR vs AA—upward phase). Given the typical alteration of the GD-H-R of patients recovering from surgery for shoulder-instability and rotator cuff tears (see Sect. 3.3), the widths of the MDDBs appear adequate to draw solid clinical conclusions, i.e. the protocol appears sensitive enough for the application. Further clinical experimentations are however required to draw definitive conclusions. In particular, further efforts will be intended to assess if the average patterns and the MDDBs do change between populations of different age and gender. At present no conclusions can be drawn to this regard and in the most conservative approach the patterns and bands reported here should be used for age-matched patients (i.e. between 20 and 40 years old).

Remarkably, the MDDBs are generally wider than the confidence bands based on the SD of the population. This suggests that $\pm 1SD$ confidence bands tend to underestimate the uncertainty of the average pattern and can lead to an increase of false positives. This is consistent with previous findings in gait analysis [17].

In computing the MDDBs, all the repetitions of all the subjects were considered, as well as the acquisitions of each subject by both operators. Usually the inter-operator variability is excluded in the computation of the confidence bands in motion analysis, excluding therefore an important source of potential errors.

The MDDBs computed allow to draw conclusion on a subject based on a new single measure taken with an operator with comparable experience to those who participated in these experiments. Since the MDDBs are based on subjects with a repetition-to-repetition repeatability of at least 0.95, they should be used to draw conclusion on a patient based on *a single measure* only if the patient presents a similar intra-subject variability (see Sect. 2.2.3.1.1). For those patients with less repetition-to-repetition repeatability, a simple solution is to compare all the repetitions of the movement with the controls' average and MDDB. This increases the probability of no difference between the patient's and average controls' pattern, thus reducing the probability of false positive.

The definition of confidence bands based on the minimal detectable difference is original but it is based on well known statistical methods and procedures [37]. The problem of estimating confidence bands with a clear inferential statistics meaning is receiving increasing interest in the literature [31, 40]. In particular, Schwartz and co-workers stressed the improper use of confidence bands based on $\pm 1SD$ for performing statistical tests on the data. The approach followed here, is similar to the method proposed by Schwartz et al. [31], with the advantage of (1) using a standard and well documented two-way analysis of variance with repeated measures, and (2) defining prediction bands which allows immediate graphical statistical tests. Compared to the bootstrap techniques, MDDBs have the advantage of explicitly excluding the between-subject variability from the computation of the prediction bands and to allow an angle-by-angle analysis of statistical difference. Moreover MDDBs are of very simple implementation.

References

1. Bland M (2000) An introduction to medical statistics. Oxford University Press, New York
2. Borstad JD, Ludewig PM (2002) Comparison of scapular kinematics between elevation and lowering of the arm in the scapular plane. Clin Biomech (Bristol, Avon) 17:650–659. doi:10.1016/S0268-0033(02)00136-5
3. Cappozzo A, Catani F, Della Croce U et al (1995) Position and orientation in space of bones during movement: anatomical frame definition and determination. Clin Biomech (Bristol, Avon) 10(4):171–178
4. Charlton IW (2003) A model for the prediction of the forces at the glenohumeral joint. Ph.D. thesis, Newcastle University, Newcastle upon Tyne
5. Chiari L, Della Croce U, Leardini A et al (2005) Human movement analysis using stereophotogrammetry. Part 2: instrumental errors. Gait Posture 21:197–211. doi:10.1016/j.gaitpost.2004.04.004
6. Cutti AG, Paolini G, Troncossi M, Cappello A, Davalli A (2005) Soft tissue artefact assessment in humeral axial rotation. Gait Posture 21(3):341–349
7. Cutti AG, Garofalo P, Filippi MV et al (2006) The evolution of compensation strategies in two patients with shoulder instability: a comparative study through quantitative motion analysis. In: Proceedings of ISG 2006, Chicago
8. De Groot JH (1997) The variability of shoulder motions recorded by means of palpation. Clin Biomech (Bristol, Avon) 12(7–8):461–472. doi:10.1016/S0268-0033(97)00031-4
9. Finch E, Brooks D, Statford PW et al (2002) Physical rehabilitation outcome measures. Lippincott Williams & Wilkins, Toronto
10. Garofalo P, Cutti AG, Filippi MV et al (2008) Limitations of the constant scale in the assessment of shoulder compensatory strategies. Gait Posture 28(Suppl 1):S29. doi:10.1016/j.gaitpost.2007.12.055
11. Hopkins WG (2000) Measures of reliability in sports medicine and science. Sports Med 30(1):1–15. doi:10.2165/00007256-200030010-00001
12. Johnson GR, Stuart PR, Mitchell S (1993) A method for the measurement of three dimensional scapular movement. Clin Biomech (Bristol, Avon) 8(5):269–273. doi:10.1016/0268-0033(93)90037-I
13. Kadaba MP, Ramakrishnan HK, Wootten ME et al (1989) Repeatability of kinematic kinetic and electromyographic data in normal adult gait. J Orthop Res 7:849–860. doi:10.1002/jor.1100070611
14. Karduna AR, McClure PW, Michener LA et al (2001) Dynamic measurements of three-dimensional scapular kinematics: a validation study. J Biomech Eng 123(2):184–190. doi:10.1115/1.1351892
15. Kavanagh JJ, Morrison S, James DA et al (2006) Reliability of segmental accelerations measured using a new wireless gait analysis system. J Biomech 39:2863–2872. doi:10.1016/j.jbiomech.2005.09.012
16. Lee RYW, Laprade J, Fung EHK (2003) A real-time gyroscopic system for three-dimensional measurement of lumbar spine motion. Med Eng Phys 25:817–824. doi:10.1016/S1350-4533(03)00115-2
17. Lenhoff MW, Santner TJ, Otis JC et al (1999) Bootstrap prediction and confidence bands: a superior statistical method for analysis of gait data. Gait Posture 9:10–17. doi:10.1016/S0966-6362(98)00043-5
18. Ludewig PM, Cook TM (2000) Alterations in shoulder kinematics and associated muscle activity in people with symptoms of shoulder impingement. Phys Ther 80(3):276–291
19. Magermans DJ, Chadwick EKJ, Veeger HEJ et al (2005) Requirements for upper extremity motions during activities of daily living. Clin Biomech (Bristol, Avon) 20(6):591–599. doi:10.1016/j.clinbiomech.2005.02.006
20. Marchese SS, Johnson GR (2000) Non-invasive measurement of the kinematics of the clavicle. In: Proceedings of the third conference of the international shoulder group. Delft University Press, Newcastle upon-Tyne, pp 61–65. ISBN 90-407-2268-4
21. Mayagoitia RE, Nene AV, Veltink PH (2002) Accelerometer and rate gyroscope measurement of kinematics: an inexpensive alternative to optical motion analysis systems. J Biomech 35:537–542. doi:10.1016/S0021-9290(01)00231-7
22. McClure PW, Michener LA, Sennett BJ, Karduna AR (2001) Direct 3-dimensional measurement of scapular kinematics during dynamic movements in vivo. J Shoulder Elbow Surg 10(3):269–277. doi:10.1067/mse.2001.112954
23. Mell AG, LaScalza S, Guffey P et al (2005) Effect of rotator cuff pathology on shoulder rhythm. J Shoulder Elbow Surg 14(1 Suppl S):58S–64S
24. Meskers CG, Vermeulen HM, de Groot JH et al (1998) 3D shoulder position measurements using a six-degree-of-freedom electromagnetic tracking device. Clin Biomech (Bristol, Avon) 13(4–5):280–292. doi:10.1016/S0268-0033(98)00095-3
25. Meskers CG, van de Sande MA, de Groot JH (2007) Comparison between tripod and skin-fixed recording of scapular motion. J Biomech 40(4):941–946. doi:10.1016/j.jbiomech.2006.02.011
26. Mills PM, Morrison S, Lloyd DG et al (2007) Repeatability of 3D gait kinematics obtained from an electromagnetic tracking system during treadmill locomotion. J Biomech 40:1504–1511. doi:10.1016/j.jbiomech.2006.06.017
27. Murray I, Johnson GR (2000) Definition of marker positions and technical frames for studying the kinematics of the shoulder. In: Proceedings of the third conference of the international shoulder group. Delft University Press, Newcastle upon-Tyne, pp 47–51. ISBN 90-407-2268-4
28. Pascoal AG, van der Helm FF, Pezarat Correia P et al (2000) Effects of different arm external loads on the scapulo–humeral rhythm. Clin Biomech (Bristol, Avon) 15(Suppl 1):S21–S24. doi:10.1016/S0268-0033(00)00055-3
29. Rockwood CA, Matsen FAIII, Wirth MA et al (2004) The shoulder. Saunders, Readfield
30. Roy JS, Moffet H, Hébert LJ, St-Vincent G, McFadyen BJ (2007) The reliability of three-dimensional scapular attitudes in healthy people and people with shoulder impingement syndrome. BMC Musculoskelet Disord 8:49
31. Schwartz MH, Trost JP, Wervey RA (2004) Measurement and management of errors in quantitative gait data. Gait Posture 20:196–203. doi:10.1016/j.gaitpost.2003.09.011
32. Stagni R, Fantozzi S, Cutti AG et al (2008) Kinematic analysis techniques and their application in biomechanics. In: Biomechanical systems technology, chap 2, vol 3. World Scientific, Singapore. ISBN 978-981-270-983-7, http://www.worldscibooks.com/medsci/6506.html
33. Stergiou N (2003) Innovative analyses of human movement. Human Kinetics Publishers, Champaign
34. van Andel CJ, Wolterbeek N, Doorenbosch CA et al (2008) Complete 3D kinematics of upper extremity functional tasks. Gait Posture 27(1):120–127. doi:10.1016/j.gaitpost.2007.03.002
35. van der Helm FF, Pronk GM (1995) Three-dimensional recording and description of motions of the shoulder mechanism. J Biomech Eng 117(1):27–40. doi:10.1115/1.2792267
36. Veeger HE, van der Helm FCT, Rozendal RH (1993) Orientation of the scapula in a simulated wheelchair push. Clin Biomech (Bristol, Avon) 8(2):81–90. doi:10.1016/S0268-0033(93)90037-I

37. Weir JP (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 19(1):231–240. doi:10.1519/15184.1

38. Wu G, van der Helm FC, Veeger HE et al (2005) ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human motion—part II: shoulder, elbow, wrist and hand. J Biomech 38(5):981–992. doi:10.1016/j.jbiomech.2004.05.042

39. Yavuzer G, Oken O, Elhan A et al (2008) Repeatability of lower limb three-dimensional kinematics in patients with stroke. Gait Posture 27(1):31–35. doi:10.1016/j.gaitpost.2006.12.016

40. Zoubir AM, Boashash B (1998) The bootstrap and its application in signal processing. IEEE Signal Process Mag 15(1):56–76. doi:10.1109/79.647043