

Article

Portable Sensors Add Reliable Kinematic Measures to the Assessment of Upper Extremity Function

Fredrik Öhberg, Tomas Bäcklund *, Nina Sundström and Helena Grip

Department of Radiation Sciences, Biomedical Engineering, Umeå University, 901 87 Umeå, Sweden; Fredrik.Ohberg@umu.se (F.Ö.); Nina.Sundstrom@vll.se (N.S.); Helena.Grip@vll.se (H.G.)

* Correspondence: Tomas.backlund@vll.se; Tel.: +46-90-7854025

Received: 17 January 2019; Accepted: 6 March 2019; Published: 12 March 2019



Abstract: Ordinal scales with low resolution are used to assess arm function in clinic. These scales may be improved by adding objective kinematic measures. The aim was to analyze within-subject, inter-rater and overall reliability (i.e., including within-subject and inter-rater reliability) and check the system's validity of kinematic measures from inertial sensors for two such protocols on one person. Twenty healthy volunteers repeatedly performed two tasks, finger-to-nose and drinking, during two test sessions with two different raters. Five inertial sensors, on the forearms, upper arms and xiphoid process were used. Comparisons against an optical camera system evaluated the measurement validity. Cycle time, range of motion (ROM) in shoulder and elbow were calculated. Bland–Altman plots and linear mixed models including the generalizability (G) coefficient evaluated the reliability of the measures. Within-subject reliability was good to excellent in both tests ($G = 0.80\text{--}0.97$) and may serve as a baseline when assessing upper extremities in future patient groups. Overall reliability was acceptable to excellent ($G = 0.77\text{--}0.94$) for all parameters except elbow axial rotation in finger-to-nose task and both elbow axial rotation and flexion/extension in drinking task, mainly due to poor inter-rater reliability in these parameters. The low to good reliability for elbow ROM probably relates to high within-subject variability. The sensors provided good to excellent measures of cycle time and shoulder ROM in non-disabled individuals and thus have the potential to improve today's assessment of arm function.

Keywords: inter-rater reliability; inertial sensor; kinematics; upper limb; arm function

1. Introduction

Gait analysis is well-established in clinical assessment of movement disabilities [1] but motion analysis in the evaluation of upper extremity function is less common but equally important in several neurological and musculoskeletal disorders, e.g., stroke [2–4], Parkinson's disease [5] and shoulder instability [6]. Information about the upper extremities' function can help to determine the extent of a problem and guide the clinician to design and evaluate an intervention. To fit the needs of today's health care, assessment procedures need to be valid, reliable and rater independent, time-effective and easy to perform.

Neurological dysfunctions affect the activities of daily living (ADL), hence standardized tests, e.g., the modified Ashworth scale of muscle spasticity [7], are important in the clinic. Drawbacks with such scales are that they do not give explicit information about the movement quality and may be insensitive to small and more specific changes [8]. One common ADL task is the drinking task, which is a natural and purposeful task to assess the function of the arm when grasping and lifting a glass of water, taking a sip and putting the glass down again. Recent studies point out that kinematic measures of the drinking task give additional and more sensitive information about movement function [9–11]. It is also important to measure movement quality, e.g., a study showed that persons with stroke had

excessive arm-trunk movements during a standing coordination test and that they could not perform fast movements with the same movement quality as a healthy control group [12]. To quantify such changes in movement capacity, kinematic measures might be valuable [13].

Arm function in persons with neurological dysfunction is also routinely evaluated using ordinal scales such as the Fugl-Meyer assessment [14,15]. The finger-to-nose task is part of the Fugl-Meyer assessment, and it aims to evaluate coordination and speed by measuring and comparing, for instance, total performance time between the affected and non-affected arm [16]. To individualize the rehabilitation program and track changes in a movement deficit over time, it is important to have other outcome measures apart from performance time that describes the quality of the movement. The challenge is to develop movement analysis methods that can be used in the clinical environment. Simple equipment such as goniometers, measuring the static angle in one plane, can be used to measure the range of motion in flexion/extension/rotation [15,17], but by introducing inertial motion sensors these and several other parameters can be quantified objectively and over time. Inertial motion sensors have been used to e.g., classify arm movement features during ADL activities [18] or to construct a portable rehabilitation system for patients/individuals with brain injury [19]. The sensors are easy to use, but to be applicable for evaluation of arm function in clinical practice, the within-subject and inter-rater reliability of the specific protocols also needs to be confirmed. One way to study reliability is to use the theory of generalizability. Here, multiple sources of measurement errors are used to characterize a measurement, as compared to classical test theory where instead undifferentiated random errors are modeled [20]. This gives a more detailed description of how different types of errors impact the overall reliability. It also enables calculation of the number of repetitions needed to optimize the reliability of a certain task. The question is whether kinematic outcome measures, based on inertial sensors, can be incorporated in a protocol that assesses the arm function in a reliable way.

The aim of this study was to analyze within-subject, inter-rater and overall reliability (i.e., including within-subject and inter-rater reliability) for joint rotations and temporal outcome measures of the finger-to-nose and drinking task using a portable inertial motion sensor system. The second aim was to do a preliminary validation of the system on one person against a gold standard optical 3D motion capture system.

2. Materials and Methods

2.1. Participants

Twenty healthy volunteers (39.8 ± 11.6 years (mean \pm standard deviation (SD)), seven women), height 1.74 ± 0.09 m and Body Mass Index 24.85 ± 2.4 kg/m² participated in this study. The participants had normal function in their back and upper extremities, i.e., no reported neurological conditions or reduced ability to perform daily routines and were recruited among personnel at University Hospital of Umeå, Sweden. The study was conducted in accordance with the Declaration of Helsinki, and the study was approved by the Regional Ethical Review Board in Umeå (Dnr 09–120 M). Written informed consent was obtained from each participant.

2.2. Measurements

The portable movement analysis system included five tri-axial inertial motion sensors (MoLabTM POSE, AnyMo AB, Umeå, Sweden) connected by flexible cables to a data acquisition unit, which wirelessly transmitted data to a laptop where it was registered with the software MoLabTM Measure. Each sensor was $60 \times 45 \times 10$ mm, included a 3D gyroscope, accelerometer, and magnetometer and weighted 14 g. Sampling frequency was 128 Hz (16-bit resolution, gyroscopes with full-scale range $\pm 1000^\circ/\text{s}$, accelerometers ± 8 g and magnetometers ± 4800 μT). Details about the signal processing of the inertial sensor data are described in an earlier study [21]. Measurement precision and accuracy of this system were previously validated against a gold standard optical system during standardized arm tasks (cone picking, throwing and two different coordination tasks) in ten healthy volunteers

and one participant with sequel after stroke. This study showed that the system's systematic error ranges between -1.2° and 2.0° [22]. A similar setup during leg movements in five healthy participants showed precision and accuracy of $2-3^\circ$ [21]. El-Gohran and McNames validated a similar sensor against a high-precision robot arm which attained an RMS angle error of about 3° for all measured angles during slow, normal and fast movements [23]. The measurement error was evaluated by comparisons of the portable system against a gold standard optical system (Qualisys Track Manager, Qualisys AB, Sweden) on one subject.

2.3. Study Design and Test Protocols

Each participant performed two tasks: finger-to-nose and drinking. In the finger-to-nose task, participants were asked to touch the nose with the index finger and then return the hand to the initial position on the table in front of him/her. Each movement was performed 10 times with eyes open at self-selected speed. In the drinking task, participants were asked to lift a glass filled with 2 dl of water, take a small sip, put the glass down and then return to the initial position at self-selected speed. The glass was placed along the midline 30 cm in front of the participant. To exclude dependency on the distance chosen, the test was also repeated placing the glass at 7 and 50 cm. Ten repetitions were performed at each distance at self-selected speed. Verbal instructions were recorded and played back to the participant before the tests started. Two different raters, with different background and experience in collecting movement data (one with long and one without any previous experience in using motion analysis systems) performed all measurements on each of the 20 participants. The rater who would start taking measurements was randomly chosen for each participant. Both tests were performed with each arm, in a randomized order, and with no rest in between.

The participant sat on a chair with hips and knees flexed 90° and feet firmly placed on the floor in front of an adjustable table. The upper arm was initially oriented vertically along the body with the elbow flexed 90° , the palm placed on the table and the forearm parallel with the body's sagittal plane. Two sensors were placed on each forearm, dorsally and with their center 5 cm proximally from the ulnar styloid processes, two on the upper arms, dorsally and with their center 10 cm proximally from the olecranon, and one on the upper body on the xiphoid process. All sensors were attached with custom made Velcro straps, illustrated in Figure 1. All sensors were removed from the participant before the subsequent rater started the second measurement session approximately five minutes after the first.

A few participants did less than ten repetitions by mistake, and data collected from one participant's non-dominant arm during one test session were excluded because one sensor came loose. In total, 776 and 2321 observations in the finger-to-nose and drinking task were analyzed.



Figure 1. The picture shows sensor placement on one arm and on the xiphoid process.

2.4. Data Processing

The software MoLab™ Analysis was used for data processing and analysis. Start and stop events for each repetition were manually determined from observation of joint angular curves and used for time normalization (Figure 2). Each body segment's orientation in a global space was defined by calculating the orientation of the motion sensor relative to a global coordinate system, using the gravity vector from the accelerometer, the compass direction from the magnetometer and the cross product between those two vectors. Quaternions were calculated for each time frame and then transposed into elbow and shoulder joint angles. Each joint angle was defined as the rotation of the distal segment relative to the proximal segment using the Cardan X-Y-Z convention; +X represented an extension, +Y adduction and +Z inward axial rotation, adapted from the ISB recommendations [24]. To determine the system's validity, the portable system's measurement error against the optical system was evaluated for one subject. Quaternions were calculated for each rigid marker cluster. The helical angle during movement in relation to the initial resting position was then compared between the systems for each body segment and task [25].

ROM was calculated for the shoulder and elbow joints for each repetition in the tests. As the shoulder is a ball-and-socket joint, ROM was given for *flexion-extension* (FE), *abduction/adduction* (AbAd) and *inward/outward axial rotation* (R). The elbow hinge joint allows flexion-extension and axial rotation (positive angle pronation, negative angle supination) of the forearm relative to the upper arm, hence FE and R were investigated. A temporal measure describing the time in seconds for each repetition, cycle time, was also calculated.

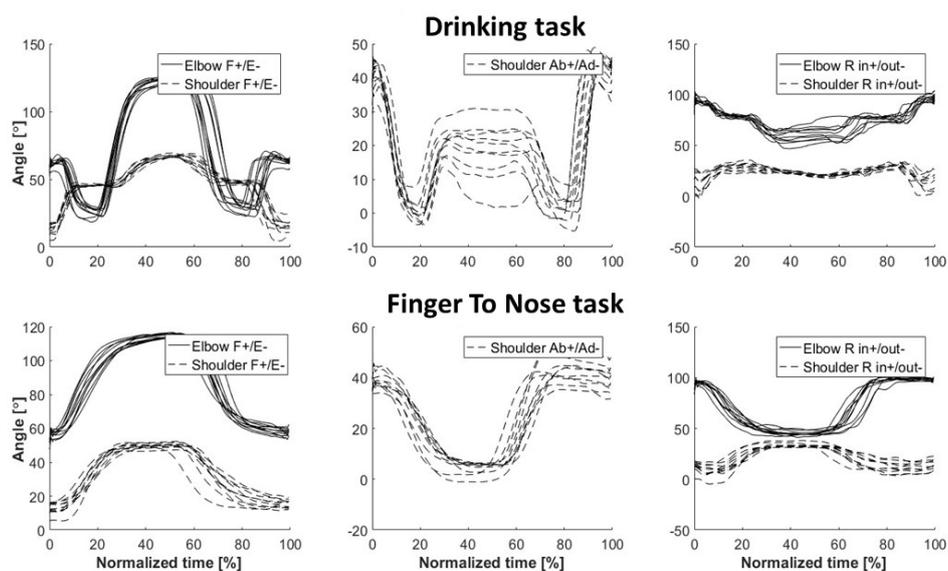


Figure 2. Examples of a subject's movement pattern during the finger-to-nose and drinking tasks for one rater. The patterns are illustrated using a normalized time window that was defined by the beginning and end of each repetition, giving 10 movement curves for each test. The illustration from the drinking task is collected from the test where the glass was placed 30 cm in front of the subject. The abbreviations are F (flexion; positive angles), E (extension; negative angles), Ab (abduction; positive angles), Ad (adduction; negative angles), R in (pronation (elbow); inward humeral rotation (shoulder); positive angles) and R out (supination (elbow); outward humeral rotation (shoulder); negative angles).

2.5. Statistics

R-studio (version 1.0.143) and R (version 3.3.3) were used for all statistical analyses and three different approaches were applied: linear regression, Bland–Altman plots and linear mixed models (LMM) including calculation of the generalizability (G) coefficient [26]. Linear regressions and Bland–Altman plots, analyzed for each outcome measure and test, were used to illustrate the agreement

between measures made by the two different raters (calculated using package ‘BlandAltmanLeh’, version 0.3.1). Bland–Altman plots were also used to compare the inertial and optical sensor systems to verify the system’s validity.

For the LMM, R-package ‘lme4’ version 1.1.12 was used. The response measure was set to ROM or cycle time. The random effect was a mix of subject, rater and repetition identifiers including all two-way interactions.

The dataset was stratified using the factor arm dominance since clinical investigations are likely to be performed mostly on one specific arm (dominant/non-dominant, injured/non-injured). Thus, estimates of the measurement variance (divided into six random factors) were calculated, i.e., between subjects (σ^2_{subj}); between repetitions (σ^2_{rep}); between raters (σ^2_{rat}); subject/rater interactions ($\sigma^2_{\text{subj_rat}}$); subject/repetition interactions ($\sigma^2_{\text{subj_rep}}$) and rater/repetition interactions ($\sigma^2_{\text{rat_rep}}$) as well as residuals ($\sigma^2_{\text{residual}}$).

For calculating the G-coefficients, the facets considered were rater and repetition in a crossed design with an absolute agreement. Equations (1) to (3) were applied for estimating reliability (i.e., within-subject), inter-rater reliability (between raters) and overall test reliability (total test situation including two raters, $n_{\text{rat}} = 2$, and 10 repetitions, $n_{\text{rep}} = 10$) respectively [26]. The term within-subject reliability, instead of within-subject variability, was used to avoid confusion with variance. For studying the effect of repeated measurements, Equation (3) with $n_{\text{rat}} = 1$ was used. As thresholds, $G > 0.90$ indicate excellent reliability, 0.80–0.89 good reliability and 0.70–0.79, acceptable reliability [27].

$$G_{\text{rep}} = \frac{\sigma^2_{\text{subj}} + \sigma^2_{\text{rat}} + \sigma^2_{\text{subj_rat}}}{\sigma^2_{\text{subj}} + \sigma^2_{\text{rat}} + \sigma^2_{\text{rep}} + \sigma^2_{\text{subj_rat}} + \sigma^2_{\text{subj_rep}} + \sigma^2_{\text{rat_rep}} + \sigma^2_{\text{residual}}} \quad (1)$$

$$G_{\text{rat}} = \frac{\sigma^2_{\text{subj}} + \sigma^2_{\text{rep}} + \sigma^2_{\text{subj_rep}}}{\sigma^2_{\text{subj}} + \sigma^2_{\text{rat}} + \sigma^2_{\text{rep}} + \sigma^2_{\text{subj_rat}} + \sigma^2_{\text{subj_rep}} + \sigma^2_{\text{rat_rep}} + \sigma^2_{\text{residual}}} \quad (2)$$

$$G_{\text{overall test}} = \frac{\sigma^2_{\text{subj}}}{\sigma^2_{\text{subj}} + \frac{\sigma^2_{\text{rat}}}{n_{\text{rat}}} + \frac{\sigma^2_{\text{rep}}}{n_{\text{rep}}} + \frac{\sigma^2_{\text{subj_rat}}}{n_{\text{rat}}} + \frac{\sigma^2_{\text{subj_rep}}}{n_{\text{rep}}} + \frac{\sigma^2_{\text{rat_rep}}}{n_{\text{rep}} * n_{\text{rat}}} + \frac{\sigma^2_{\text{residual}}}{n_{\text{rep}} * n_{\text{rat}}}} \quad (3)$$

The significance level was set at $\alpha = 0.05$.

3. Results

3.1. System Validity

The system’s validity was assessed on one subject during the finger-to-nose and drinking tasks. The systematic error ranged between 0.25° and 1.1° for the former and 0.45 to 0.84 for the latter test (Figure 3a,b).

3.2. Inter-Rater and Within-Subject Reliability

Inter-rater reliability of the outcome measures cycle time and ROM are displayed in Figures 4 and 5. R^2 was above 50% for all outcome measures except for elbow R, where the R^2 was low ($<10\%$, Figure 4). In addition, for elbow R the linear regression coefficients were near zero, indicating a large variation in elbow axial rotation between the test sessions. Bland–Altman analyses (Figure 5), show that average discrepancy (i.e., bias) between raters was close to zero for all outcome measures, there was a dependency on mean angle for some of the ROM measures. The 95% confidence intervals of the Bland–Altman analyses were approximately 1.2 s for the cycle times, 15° for shoulder ROMs and 25° to 90° for elbow ROMs.

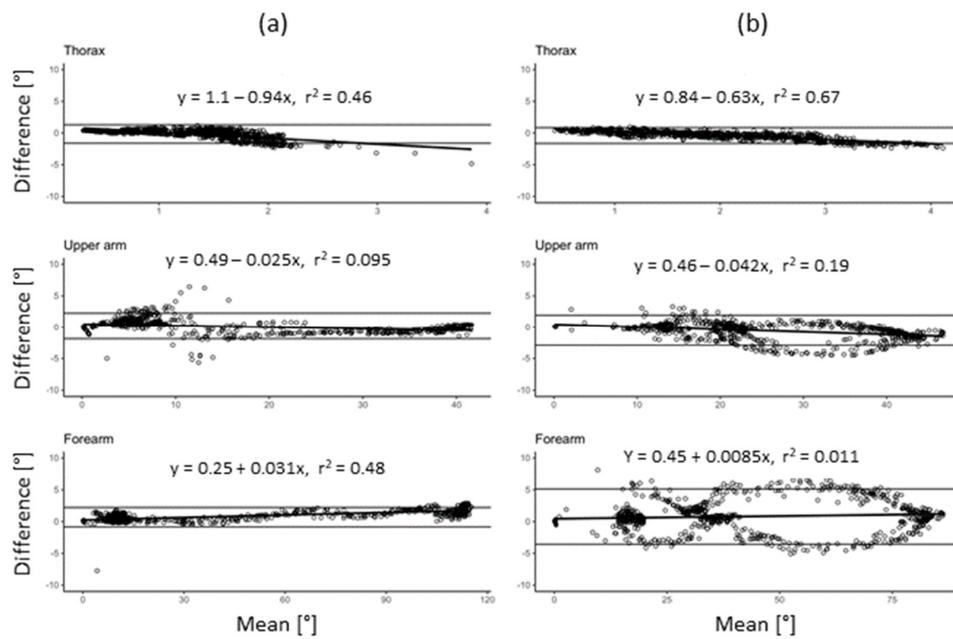


Figure 3. Bland–Altman analysis of the inertial and optical sensor systems. The data from one subject in one session of 10 repetitions. (a) finger-to-nose and (b) drinking task.

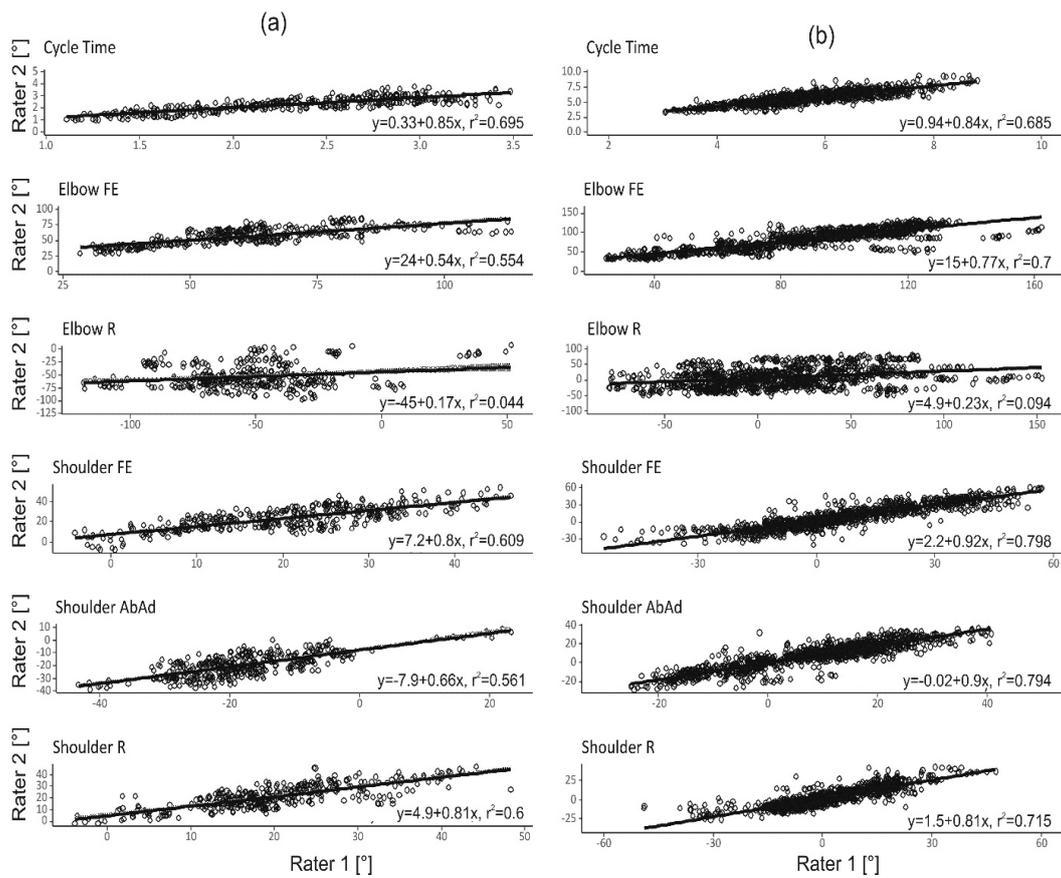


Figure 4. Illustration of inter-rater reliability between the two raters for the outcome measures of the (a) finger-to-nose and (b) drinking task. The regression lines are estimated using QR decomposition. Data of all 10 repetitions and 20 subjects are plotted.

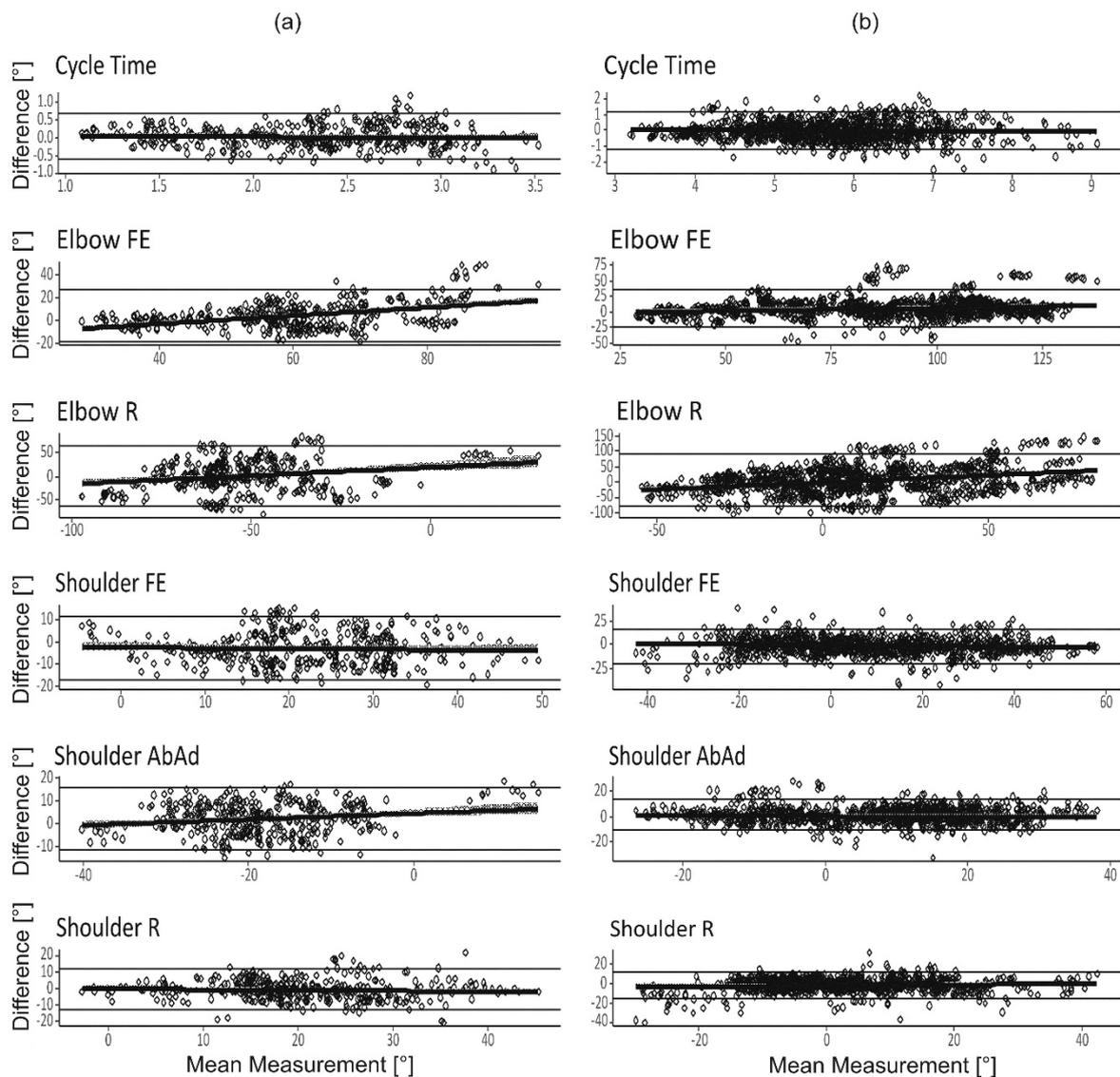


Figure 5. Bland–Altman analysis of the two raters for all outcome measures in the (a) finger-to-nose and (b) drinking task. Data of all 10 repetitions and 20 subjects are plotted.

Based on the two LMM models, variance components and G-coefficients were calculated. Within-subject reliability was good to excellent for all measures of both tests and arms (Tables 1 and 2). Results were the same for the distances 7 and 50 cm for the drinking task, although there was a small decrease in variability as distance increased (see Tables A1 and A2 and Figure A1 in Appendix A).

Table 1. Linear mixed effect model for the finger-to-nose task.

	Side	Reliability	Cycle Time	Elbow FE	Elbow R	Shoulder FE	Shoulder AbAd	Shoulder R	
G-coefficients	Dominant	Within-subject	0.89	0.91	0.97	0.89	0.92	0.91	
		Inter-rater	0.84	0.77	0.15	0.78	0.67	0.73	
		Overall	0.94	0.88	0.26	0.91	0.83	0.87	
	Non-dominant	Within-subject	0.91	0.94	0.96	0.90	0.92	0.94	
		Inter-rater	0.83	0.62	0.30	0.74	0.76	0.82	
		Overall	0.94	0.77	0.47	0.88	0.87	0.92	
Model and variance components	Dominant <i>n</i> = 397	Intercept	2.35 s ^{***} (0.12)	58.66 [°] *** (4.14)	−53.66 [°] *** (4.70)	22.42 [°] *** (2.81)	−20.08 [°] *** (1.72)	20.72 [°] *** (1.89)	
		σ^2_{rat}	0.00	13.71	0.00	3.99	0.00	0.29	
		σ^2_{rep}	0.00	0.00	5.90	0.76	0.62	0.14	
		σ^2_{subj}	0.28	186.75	113.81	107.42	48.16	60.17	
		$\sigma^2_{rep_rat}$	0.00	0.00	0.00	0.00	0.00	0.00	
		$\sigma^2_{subj_rat}$	0.03	37.48	629.23	17.02	18.76	16.16	
		$\sigma^2_{subj_rep}$	0.01	2.72	1.10	2.20	0.57	0.92	
		$\sigma^2_{residual}$	0.02	6.32	14.05	8.43	4.55	6.49	
		Non-dominant <i>n</i> = 379	Intercept	2.31 s ^{***} (0.13)	61.35 [°] *** (4.04)	−54.47 [°] *** (4.14)	21.87 [°] *** (2.64)	−16.26 [°] *** (2.69)	19.94 [°] *** (2.13)
			σ^2_{rat}	0.00	9.61	0.00	3.57	4.35	0.00
σ^2_{rep}	0.00		0.30	3.94	0.02	0.00	0.06		
σ^2_{subj}	0.29		173.26	148.50	87.90	85.66	79.07		
$\sigma^2_{rep_rat}$	0.00		0.00	0.00	0.14	0.09	0.00		
$\sigma^2_{subj_rat}$	0.03		90.53	338.92	20.29	19.94	12.80		
$\sigma^2_{subj_rep}$	0.00		1.89	2.41	1.24	1.60	1.42		
$\sigma^2_{residual}$	0.03		4.74	14.53	6.52	3.47	4.69		

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Statistical model: Response \sim (1 | "Subject") + (1 | "Rater") + (1 | "Repetition") + (1 | "Subject": "Repetition") + (1 | "Subject": "Rater") + (1 | "Repetition": "Rater"). The dataset was stratified using the factor "arm dominance". In the table, σ^2_{sub} , σ^2_{rat} , σ^2_{rep} , $\sigma^2_{rep_rat}$, $\sigma^2_{subj_rat}$, $\sigma^2_{subj_rep}$, and $\sigma^2_{residual}$ are the variance between subjects, among raters, among repetitions, of the interaction between repetitions and raters, subjects and raters, subjects and repetitions, and residuals respectively. Values within parentheses indicate the mean square error of each parameter estimate. The reliability estimates (defined by the G-coefficient, absolute agreement) for the different responses are subdivided into within-subject, inter-rater, and overall reliability. The number of subjects = 20 and the number of raters = 2.

Table 2. Linear mixed effect model for the drinking task at 30 cm distance.

	Side	Reliability	Cycle Time	Elbow FE	Elbow R	Shoulder FE	Shoulder AbAd	Shoulder R
G-coefficients	Dominant	Within-subject	0.92	0.88	0.97	0.85	0.80	0.90
		Inter-rater	0.81	0.47	0.002	0.67	0.72	0.78
		Overall	0.93	0.68	0	0.86	0.93	0.90
	Non-Dominant	Within-subject	0.90	0.96	0.96	0.87	0.90	0.87
		Inter-rater	0.84	0.42	0.14	0.72	0.81	0.60
		Overall	0.95	0.60	0.23	0.87	0.93	0.79
Model and variance components	Dominant <i>n</i> = 397	Intercept	5.65 s ^{***} (0.20)	90.76 [°] *** (2.77)	20.09 [°] *** (5.62)	5.68 [°] ** (1.76)	14.69 [°] *** (1.64)	−5.00 [°] ** (1.68)
		σ^2_{rat}	0.00	0.00	0.00	0.00	0.91	0.00
		σ^2_{rep}	0.00	2.38	0.00	0.16	0.00	1.04
		σ^2_{subj}	0.71	100.31	0.00	53.09	41.88	48.74
		$\sigma^2_{rep_rat}$	0.00	0.88	0.00	0.00	0.14	0.00
		$\sigma^2_{subj_rat}$	0.10	92.30	1260.94	16.06	4.53	10.01
		$\sigma^2_{subj_rep}$	0.01	0.00	2.08	1.22	0.57	0.77
		$\sigma^2_{residual}$	0.06	23.05	36.25	10.46	11.02	4.44
		Non-dominant <i>n</i> = 376	Intercept	5.68 s ^{***} (0.19)	97.79 [°] *** (2.96)	24.75 [°] *** (5.38)	2.94 [°] (2.37)	15.84 [°] *** (1.74)
σ^2_{rat}	0.00		0.00	0.00	0.22	0.00	0.00	
σ^2_{rep}	0.00		0.03	3.67	0.00	0.00	0.00	
σ^2_{subj}	0.62		99.10	124.19	90.38	53.51	39.08	
$\sigma^2_{rep_rat}$	0.00		0.40	0.00	0.82	0.09	0.21	
$\sigma^2_{subj_rat}$	0.05		133.03	832.17	24.28	6.91	19.92	
$\sigma^2_{subj_rep}$	0.01		1.36	9.89	4.21	0.68	1.33	
$\sigma^2_{residual}$	0.07		7.43	23.90	11.52	5.94	7.22	

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Statistical model: Response \sim (1 | "Subject") + (1 | "Rater") + (1 | "Repetition") + (1 | "Subject": "Repetition") + (1 | "Subject": "Rater") + (1 | "Repetition": "Rater"). The dataset was stratified using the factor "arm dominance". In the table, σ^2_{sub} , σ^2_{rat} , σ^2_{rep} , $\sigma^2_{rep_rat}$, $\sigma^2_{subj_rat}$, $\sigma^2_{subj_rep}$, and $\sigma^2_{residual}$ are the variance between subjects, among raters, among repetitions, of the interaction between repetitions and raters, subjects and raters, subjects and repetitions, and residuals respectively. Values within parentheses indicate the mean square error of each parameter estimate. The reliability estimates (defined by the G-coefficient, absolute agreement) for the different responses are subdivided into within-subject, inter-rater, and overall reliability. The number of subjects = 20 and the number of raters = 2.

Inter-rater reliability was acceptable to good for all measures except elbow R, shoulder AbAd (dominant arm) and elbow FE (non-dominant arm) for the finger-to-nose task. For the drinking task, inter-rater agreement was acceptable to good for all measures except the elbow ROMs, shoulder FE (dominant arm) and shoulder R (non-dominant arm).

The overall reliability of the tests, including both raters and 10 repetitions, was excellent for cycle time and good to excellent for all shoulder ROMs of both tests, except shoulder R for the non-dominant arm of the drinking task where it was acceptable. For the finger-to-nose task, elbow FE had an acceptable to good, and elbow R poor overall reliability. For the drinking task, reliability of all kinematic elbow measures was poor.

Figure 6 shows the G-coefficient as a function of the number of repetitions for one single rater. For both tests, three repeated measurements were enough to reach 95% of the final value obtained after 10 repetitions (Figure 6), but generally, the elbow ROMs did not reach an acceptable level of reliability regardless of the number of repetitions.

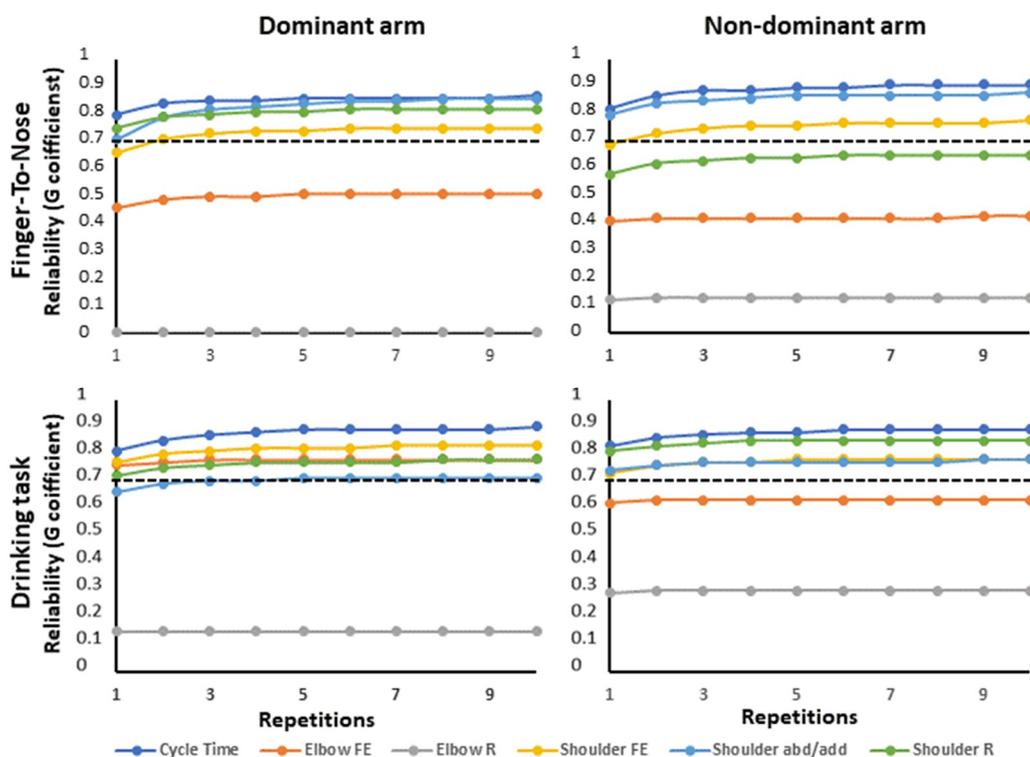


Figure 6. Illustrates the overall reliability (i.e., G-coefficient) estimated with a different number of repetitions for the finger-to-nose and drinking tasks based on one single rater. The results are subdivided into dominant arm (**left**) and non-dominant arm (**right**) for the different outcome measures. The dotted line marks the level for acceptable reliability, $G = 0.7$.

4. Discussion

This study showed that objective kinematic outcome measures can be reliably collected with a portable sensor system. Such measures may offer added information to the assessment of movement function in the upper extremities. The system's validity against a gold standard optical reference system was verified, however, the reliability depended on movement direction and type of joint: elbow ROMs had a higher inter-rater variability, while cycle time and shoulder ROMs were reliably assessed.

Both ROM and cycle time are clinically relevant outcome measures, e.g., ROM is often affected in neurological disorders [13,28], while a prolonged cycle time is used to indicate impaired coordination [15,29]. Within generalizability theory [26], G-coefficient is analogous to the reliability coefficient, intra-class correlation (ICC), in classical test theory. In contrast to ICC, G-coefficient

accounts for different sources of systematic and random errors and estimates them individually [20]. By incorporating different variance components into the error variance when calculating G, specific effects of factors, e.g., measurement repetitions within subjects and differences between raters, can be estimated. Based on G theory it is also possible to determine the optimal study design [20]. In our study, this was applied to demonstrate the number of repetitions needed to achieve the specified overall reliability.

4.1. Within-Subject Reliability

Within-subject reliability was good to excellent (0.89–0.97 and 0.80–0.96) for all outcome measures of the finger-to-nose and drinking task, respectively. Thus, healthy subjects perform repetitive arm movements (during one single session) with a consistent movement pattern. Drinking task is widely used in clinical practice, but the performance varies between applications. To evaluate whether the distance chosen was important for the overall reliability, three distances (close to far) were investigated. Previous studies show a variety in the distance utilized [10,11,30] or distances inadequately described [31]. Results indicate a small decrease in variability as distance increases (Figure A1, Appendix A), which enhances the importance of choosing a standardized distance to be able to compare between studies and occasions.

4.2. Rater Dependency

In general, inter-rater reliability for the finger-to-nose task was acceptable for all outcome measures except elbow R. For the drinking task it was poor for all elbow ROMs, and good or close to acceptable for all other measures. Both Bland–Altman plots and LMM analyses show a small systematic difference between raters (Figure 5 and σ^2_{rat} in Table 2), indicating that standardization of sensor placement was adequate. The largest source of error was the random variance due to the interaction between rater and subject ($\sigma^2_{\text{sub_rat}}$ Table 2), especially for elbow R. This was also shown by the low R^2 in the linear regression analysis (Figure 4) and large confidence intervals in the Bland–Altman plots (Figure 5). We speculate that high rater/subject interaction for elbow ROMs could have different explanations. (1) Difficulties in completely standardizing the test. Despite recorded instructions which informed participants to start with the hand in a pronated position, it was observed that starting position varied between forearm supination and pronation for some participants. (2) The elbow must be flexed to a certain degree to reach the nose or glass, but the amount of rotation may vary while still being able to fulfill the test. This well-known variability was described previously [32].

For all outcome measures, test-retest variations of the participants must be considered a source affecting the subject/rater variation. Since each rater only assessed each subject on one occasion, test-retest variation due to natural subject variations could not be estimated, i.e., the same rater tests the same participant at a later occasion.

Cycle time had higher within-subject, inter-rater, and overall reliability than all kinematic outcome measures. This finding was not surprising since goal-directed movements can be performed along different trajectories. Hence different combinations of joint angles can be used to reach the same target [32], which results in a larger variability in joint rotation compared to temporal measures.

4.3. Proposed Protocol Design for Increased Reliability

The simplicity of performing multiple measurements when using portable motion sensors encourages the strategy of applying repeated measurements to improve estimation reliability. Figure 6 shows the overall improvement in measurement reliability as a function of repetitions. It is evident that only a few repetitions are needed, due to low within-subject variability.

Even though the overall reliability was high the accuracy could be improved if the number of raters assessing each participant had been increased, i.e., the ratio $\sigma^2_{\text{sub_rat}}/n_{\text{rat}}$ will decrease as n_{rat} increases since $\sigma^2_{\text{sub_rat}}$ is large (Equation (3), Tables 1 and 2). This is also evident from higher overall reliability in Tables 1 and 2 (two raters) compared to Figure 2 (one rater). Unfortunately, in a clinical

situation, this is seldom possible due to practical limitations. However, often it is the relative change over time or comparison with a normal range, rather than the absolute value, that is meaningful for clinical diagnoses/treatment.

4.4. Kinematic Assessment of the Finger-To-Nose Task

Recent studies, also applying different versions of the finger-to-nose task, found total movement time to be the strongest indicator of stroke impairment (mean difference between the control and stroke groups was 2.6 s) [13] and to discriminate between mild and moderate-to-severe impairment post-stroke (cut-off level in total movement time was 10.6 s) [33]. This shows the clinical usefulness of this parameter, and since cycle time had excellent overall reliability and low between-subject variance ($\sigma^2_{\text{subj}} = 0.29$ s (non-dominant arm), Table 1) the current setup is well suited for detecting impairment post-stroke. The high reliability is consistent with a previous study, where the mean cycle time during a finger-to-nose task was 2.44 s with a corresponding ICC = 0.82 [34].

In order to quantify subtle changes during e.g., rehabilitation it is important, however, to extend the information extracted from the finger-to-nose task beyond cycle time [13]. Shoulder ROMs, reliably assessed in the present study, have been shown to be valuable in the treatment and diagnosis of neurological disorders, e.g., in a group of patients with stroke impairments, shoulder FE/AbAd were used to show improvements after training with a specific rehabilitation device (change in FE: $\sim 35^\circ$ and Ab-Ad: $\sim 23^\circ$) [35]. Shoulder ROMs were also used to investigate upper limb mobility in patients with different types of spinal cord injuries [36].

Further, elbow FE significantly separated between persons with mild and moderate post-stroke symptoms (group mean; mild symptoms 74.3° vs. moderate symptoms 59.9°) when measured with an optical measurement system [13]. This strengthens the clinical value of adding elbow FE to the clinical protocols. Unfortunately, elbow FE only had poor to acceptable overall reliability, stressing the need for further standardization of the measurement protocols when using inertial measurement unit (IMU) systems.

4.5. Kinematic Assessment of the Drinking Task

Like the finger-to-nose task, cycle time was the most repeatable parameter with highest inter-rater reliability. Alt Murphy et al. demonstrated that a decreased movement time of 2.5 to 5 s marks a clinical improvement post-stroke [29], which is easily detectable with our measurement setup (95% confidence interval for Bland–Altman analysis ± 1.2 s). Also, movement time discriminated between persons with a spinal injury on C6 level and healthy controls (group mean difference 2.6 s) [11], stressing the clinical value of measuring cycle time.

Kinematic measures assessed by opto-electronic motion capture systems have also been found to give additional and more sensitive information about upper extremity movement function during the drinking task [9–11], emphasizing the need to be able to reliably quantify these parameters also in a clinical setting. A pilot study on post stroke patients indicated decreased ROM in shoulder FE (group mean difference 20.5°) and increased ROM in shoulder AbAd (group mean difference 15.2°) [28]. Such minor changes in ROM are hard to see by visual inspection but may be reliably detected with the present sensor system (overall reliability including one rater and 10 repetitions, $G = 0.75$ – 0.88 , Figure 6).

4.6. Methodological Considerations and Strengths

High validity of the sensor system was confirmed in the current test situations. A deviation of 0.25° and 1.1° from the optical reference system is within the range expected and small compared to the kinematic measures assessed.

Since repeatability and measurement errors represent two important but different aspects of overall reliability, G-coefficients, as well as Bland–Altman analyses, were applied [26,37]. The test re-test variability was evaluated using two different raters, as opposed to the same rater evaluating all

participants at a later occasion. Since the variability among raters was high, we believe that the precision would improve even further if one single rater had performed the test. We, therefore, recommend that, if possible, the same clinician re-evaluate the patient during a rehabilitation process. If this is not possible, the variability due to different raters needs to be considered. Also, it is important that the test protocol is standardized. The high reliability in most outcomes was probably related to the strictly followed protocol when mounting the sensors, and standardization of verbally recorded instructions.

To get a more complete picture of the complex function of the upper extremities, future studies should focus on additional measures e.g., velocity and movement smoothness. It is also important to include patients with various diseases and healthy elderly people, rather than only healthy young individuals, as individual variability may differ with condition which in turn affects the applicability of a chosen measurement technique. Variability in kinematical measures can also be used on its own to detect deficiencies in motor function [38].

5. Conclusions

In conclusion, the reliability of the kinematic measures depended on movement direction and joint. The range in elbow joint axial rotation had poor reliability while the range in elbow flexion-extension was close to acceptable. Standardized measurement procedures are important to minimize confounding factors. Still, the high variability could also be due to natural movement variations in this joint. The within-subject variability may serve as a baseline when assessing the upper extremities in future patient groups. Cycle time and shoulder range of motion showed good to excellent reliability and are suitable to incorporate in the drinking and finger-to-nose tasks. The inertial sensors provided good to excellent measures of cycle time and shoulder ROM in non-disabled individuals and thus have the potential to improve today's assessment of arm function.

Author Contributions: F.Ö. had the main responsibility and contributed to all parts of the study. T.B. was the corresponding author and contributed to the analysis and interpretation of data, drafting and finalizing the manuscript. N.S. and H.G. contributed to the analysis and interpretation of data and to revising the manuscript. All authors read and approved the final manuscript.

Funding: This study was supported by the University Hospital of Umeå and financial grants from the European Research and Development Fund, Objective 2, Northern Sweden.

Acknowledgments: The authors also thank Dunia Kadhim for her assistance with data collection in this study.

Conflicts of Interest: Helena Grip and Fredrik Öhberg are currently involved in the startup company AnyMo AB which is manufacturing the system used in this study. The authors declare no other conflict of interest.

Appendix A

Table A1. Linear mixed effect model for the drinking task at 7 cm distance.

	Side	Reliability	Cycle Time	Elbow FE	Elbow R	Shoulder FE	Shoulder AbAd	Shoulder R	
G-coefficients	Dominant	Within-subject	0.88	0.97	0.93	0.93	0.82	0.88	
		Inter-rater	0.72	0.46	0.15	0.73	0.81	0.74	
		Overall	0.88	0.64	0.26	0.87	0.90	0.89	
	Non-dominant	Within-subject	0.88	0.97	0.93	0.90	0.88	0.86	
		Inter-rater	0.75	0.49	0.25	0.50	0.59	0.67	
		Overall	0.90	0.67	0.40	0.69	0.77	0.86	
Model and variance components	Dominant <i>n</i> = 399	Intercept	5.11 *** (0.18)	56.22 *** (4.80)	−12.61 * (4.96)	29.97 *** (2.72)	−8.33 *** (1.98)	12.15 *** (2.45)	
		σ^2_{rat}	0.00	20.72	0.00	3.32	0.00	2.33	
		σ^2_{rep}	0.01	0.97	0.48	0.00	7.53	0.36	
		σ^2_{subj}	0.58	167.39	128.74	101.04	57.90	86.98	
		$\sigma^2_{\text{rep_rat}}$	0.00	0.00	0.00	0.41	0.00	0.00	
		$\sigma^2_{\text{subj_rat}}$	0.15	166.03	717.93	25.69	9.82	17.05	
		$\sigma^2_{\text{subj_rep}}$	0.01	1.55	4.79	1.19	1.22	2.13	
		$\sigma^2_{\text{residual}}$	0.08	10.10	60.34	8.84	5.92	12.41	
		Non-dominant <i>n</i> = 377	Intercept	5.11 *** (0.17)	60.41 *** (4.53)	−13.04 ** (4.40)	29.66 *** (2.53)	−4.29 * (2.17)	15.13 *** (1.99)
			σ^2_{rat}	0.00	17.59	0.00	2.75	1.59	0.00
σ^2_{rep}	0.01		0.31	1.51	0.00	0.51	0.00		
σ^2_{subj}	0.51		154.37	146.22	67.17	56.97	64.87		
$\sigma^2_{\text{rep_rat}}$	0.00		0.00	0.32	0.26	0.10	0.00		
$\sigma^2_{\text{subj_rat}}$	0.10		134.95	433.00	54.84	30.80	19.13		
$\sigma^2_{\text{subj_rep}}$	0.00		1.45	8.57	1.65	2.91	0.68		
$\sigma^2_{\text{residual}}$	0.07		8.45	32.93	11.19	9.16	13.18		

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Statistical model: Response $\sim (1 \mid \text{“Subject”}) + (1 \mid \text{“Rater”}) + (1 \mid \text{“Repetition”}) + (1 \mid \text{“Subject”}:\text{“Repetition”}) + (1 \mid \text{“Subject”}:\text{“Rater”}) + (1 \mid \text{“Repetition”}:\text{“Rater”})$. The dataset was stratified using the factor “arm dominance”. In the table, σ^2_{sub} , σ^2_{rat} , σ^2_{rep} , $\sigma^2_{\text{rep_rat}}$, $\sigma^2_{\text{subj_rat}}$, $\sigma^2_{\text{subj_rep}}$, and $\sigma^2_{\text{residual}}$ are the variance between subjects, among raters, among repetitions, of the interaction between repetitions and raters, subjects and raters, subjects and repetitions, and residuals respectively. Values within parentheses indicate the mean square error of each parameter estimate. The reliability estimates (defined by the G-coefficient, absolute agreement) for the different responses are subdivided into within-subject, inter-rater, and overall reliability. The number of subjects = 20 and the number of raters = 2.

Table A2. Linear mixed effect model for the drinking task at 50 cm distance.

	Side	Reliability	Cycle Time	Elbow FE	Elbow R	Shoulder FE	Shoulder AbAd	Shoulder R	
G-coefficients	Dominant	Within-subject	0.91	0.87	0.98	0.86	0.91	0.86	
		Inter-rater	0.78	0.51	0.02	0.72	0.79	0.70	
		Overall	0.90	0.73	0.04	0.90	0.92	0.87	
	Non-dominant	Within-subject	0.89	0.94	0.97	0.87	0.85	0.86	
		Inter-rater	0.83	0.25	0.24	0.67	0.73	0.76	
		Overall	0.94	0.39	0.39	0.85	0.89	0.92	
Model and variance components	Dominant <i>n</i> = 395	Intercept	6.32 *** (0.21)	106.47 *** (2.58)	17.23 ** (5.91)	−8.03 *** (2.36)	15.69 *** (1.83)	−3.65 * (1.75)	
		σ^2_{rat}	0.00	3.17	0.00	0.00	0.00	0.58	
		σ^2_{rep}	0.00	0.02	0.00	0.13	0.00	1.25	
		σ^2_{subj}	0.77	74.92	25.21	100.03	61.76	46.20	
		$\sigma^2_{rep_rat}$	0.00	0.00	1.48	0.00	0.00	0.00	
		$\sigma^2_{subj_rat}$	0.17	51.27	1342.60	20.93	10.28	12.11	
		$\sigma^2_{subj_rep}$	0.03	0.00	4.31	0.00	0.49	0.38	
		$\sigma^2_{residual}$	0.06	18.56	24.63	18.96	6.63	8.23	
		Non-dominant <i>n</i> = 377	Intercept	6.28 *** (0.19)	114.38 *** (0.19)	29.04 *** (3.86)	−11.99 *** (6.40)	16.16 *** (2.56)	−4.74 ** (1.77)
			σ^2_{rat}	0.00	20.71	0.00	1.20	0.00	0.08
σ^2_{rep}	0.00		0.23	3.11	0.00	0.00	0.30		
σ^2_{subj}	0.62		37.73	302.47	97.08	53.01	48.27		
$\sigma^2_{rep_rat}$	0.00		0.00	0.80	0.00	0.12	0.00		
$\sigma^2_{subj_rat}$	0.07		94.76	935.13	30.81	11.60	6.90		
$\sigma^2_{subj_rep}$	0.02		3.66	5.39	2.83	2.25	0.18		
$\sigma^2_{residual}$	0.07		6.41	22.83	16.34	8.68	8.61		

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Statistical model: Response \sim (1 | "Subject") + (1 | "Rater") + (1 | "Repetition") + (1 | "Subject": "Repetition") + (1 | "Subject": "Rater") + (1 | "Repetition": "Rater"). The dataset was stratified using the factor "arm dominance". In the table, σ^2_{sub} , σ^2_{rat} , σ^2_{rep} , $\sigma^2_{rep_rat}$, $\sigma^2_{subj_rat}$, $\sigma^2_{subj_rep}$, and $\sigma^2_{residual}$ are the variance between subjects, among raters, among repetitions, of the interaction between repetitions and raters, subjects and raters, subjects and repetitions, and residuals respectively. Values within parentheses indicate the mean square error of each parameter estimate. The reliability estimates (defined by the G-coefficient, absolute agreement) for the different responses are subdivided into within-subject, inter-rater, and overall reliability. The number of subjects = 20 and the number of raters = 2.

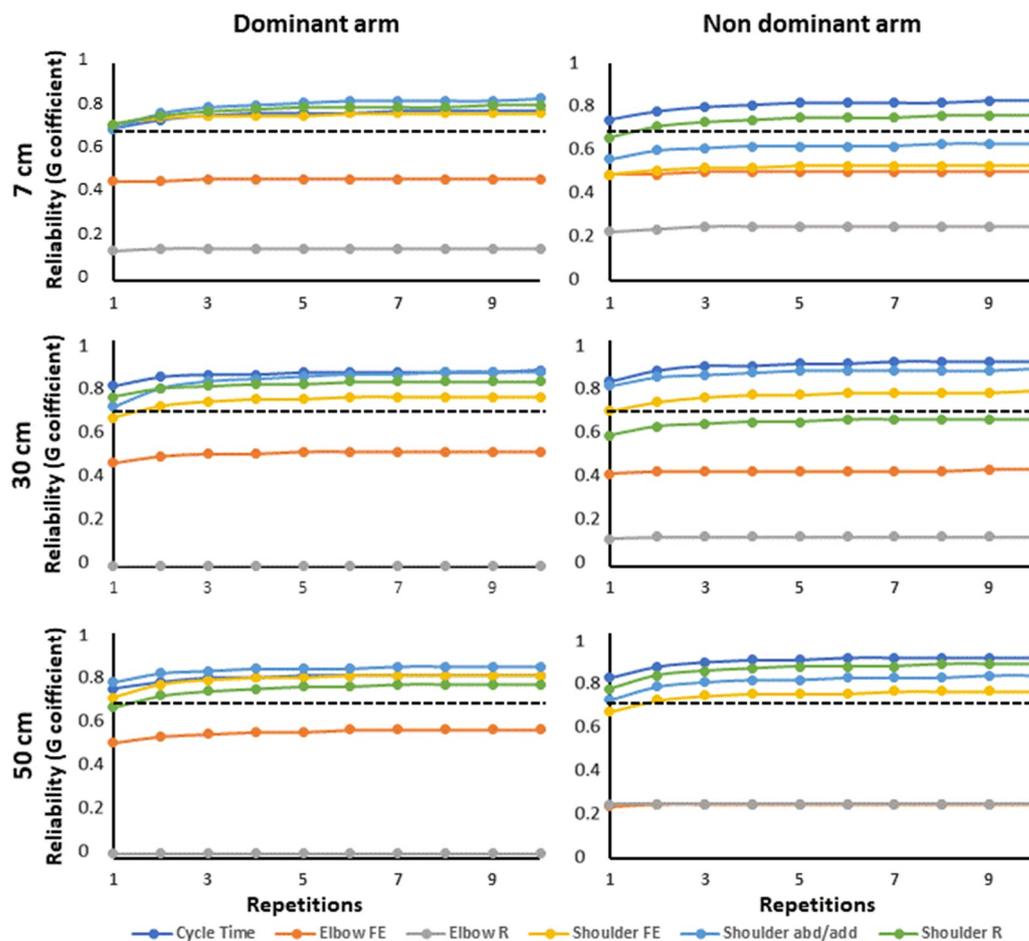


Figure A1. Illustrates the overall reliability (i.e., G-coefficient) at different distance to the glass, with different number of repetitions (assuming there is only one rater present) for the drinking task, with data from one rater. The results are subdivided into dominant arm (**left**) and non-dominant arm (**right**) for the different outcome measures. The dotted line marks the level for acceptable reliability; $G = 0.7$.

References

1. Wren, T.A.; Gorton, G.E., 3rd; Ounpuu, S.; Tucker, C.A. Efficacy of clinical gait analysis: A systematic review. *Gait Posture* **2011**, *34*, 149–153. [[CrossRef](#)] [[PubMed](#)]
2. Paulis, W.D.; Horemans, H.L.; Brouwer, B.S.; Stam, H.J. Excellent test-retest and inter-rater reliability for Tardieu Scale measurements with inertial sensors in elbow flexors of stroke patients. *Gait Posture* **2011**, *33*, 185–189. [[CrossRef](#)] [[PubMed](#)]
3. van der Pas, S.C.; Verbunt, J.A.; Breukelaar, D.E.; van Woerden, R.; Seelen, H.A. Assessment of arm activity using triaxial accelerometry in patients with a stroke. *Arch. Phys. Med. Rehabil.* **2011**, *92*, 1437–1442. [[CrossRef](#)] [[PubMed](#)]
4. McIntyre, A.; Viana, R.; Janzen, S.; Mehta, S.; Pereira, S.; Teasell, R. Systematic review and meta-analysis of constraint-induced movement therapy in the hemiparetic upper extremity more than six months post stroke. *Top. Stroke Rehabil.* **2012**, *19*, 499–513. [[CrossRef](#)] [[PubMed](#)]
5. Felix, K.; Gain, K.; Paiva, E.; Whitney, K.; Jenkins, M.E.; Spaulding, S.J. Upper Extremity Motor Learning among Individuals with Parkinson's Disease: A Meta-Analysis Evaluating Movement Time in Simple Tasks. *Parkinson's Dis.* **2012**, *2012*, 589152. [[CrossRef](#)] [[PubMed](#)]
6. El-Zayat, B.F.; Efe, T.; Heidrich, A.; Wolf, U.; Timmesfeld, N.; Heyse, T.J.; Lakemeier, S.; Fuchs-Winkelmann, S.; Schofer, M.D. Objective assessment of shoulder mobility with a new 3D gyroscope—A validation study. *BMC Musculoskelet. Disord.* **2011**, *12*, 168. [[CrossRef](#)] [[PubMed](#)]

7. Bohannon, R.W.; Smith, M.B. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys. Ther.* **1987**, *67*, 206–207. [[CrossRef](#)] [[PubMed](#)]
8. McCrea, P.H.; Eng, J.J.; Hodgson, A.J. Biomechanics of reaching: Clinical implications for individuals with acquired brain injury. *Disabil. Rehabil.* **2002**, *24*, 534–541. [[CrossRef](#)] [[PubMed](#)]
9. Alt Murphy, M.; Murphy, S.; Persson, H.C.; Bergstrom, U.B.; Sunnerhagen, K.S. Kinematic Analysis Using 3D Motion Capture of Drinking Task in People With and Without Upper-extremity Impairments. *J. Vis. Exp.* **2018**. [[CrossRef](#)] [[PubMed](#)]
10. Alt Murphy, M.; Willen, C.; Sunnerhagen, K.S. Kinematic variables quantifying upper-extremity performance after stroke during reaching and drinking from a glass. *Neurorehabil. Neural Repair* **2011**, *25*, 71–80. [[CrossRef](#)] [[PubMed](#)]
11. de los Reyes-Guzman, A.; Gil-Agudo, A.; Penasco-Martin, B.; Solis-Mozos, M.; del Ama-Espinosa, A.; Perez-Rizo, E. Kinematic analysis of the daily activity of drinking from a glass in a population with cervical spinal cord injury. *J. Neuroeng. Rehabil.* **2010**, *7*, 41. [[CrossRef](#)] [[PubMed](#)]
12. Ustinova, K.I.; Goussev, V.M.; Balasubramaniam, R.; Leven, M.F. Disruption of coordination between arm, trunk, and center of pressure displacement in patients with hemiparesis. *Motor Control* **2004**, *8*, 139–159. [[CrossRef](#)] [[PubMed](#)]
13. Johansson, G.M.; Grip, H.; Levin, M.F.; Hager, C.K. The added value of kinematic evaluation of the timed finger-to-nose test in persons post-stroke. *J. Neuroeng. Rehabil.* **2017**, *14*, 11. [[CrossRef](#)]
14. Gladstone, D.J.; Danells, C.J.; Black, S.E. The fugl-meyer assessment of motor recovery after stroke: A critical review of its measurement properties. *Neurorehabil. Neural Repair* **2002**, *16*, 232–240. [[CrossRef](#)] [[PubMed](#)]
15. Fugl-Meyer, A.R.; Jaasko, L.; Leyman, I.; Olsson, S.; Steglind, S. The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. *Scand. J. Rehabil. Med.* **1975**, *7*, 13–31. [[PubMed](#)]
16. Bergeron, D.; Vermette, A.; De La Sablonniere, J.; Cayer, A.M.; Laforce, R.; Bouchard, R.W. Finger-to-Nose Test Findings in Alzheimer's Disease. *J. Alzheimers Dis.* **2017**, *55*, 1335–1337. [[CrossRef](#)]
17. de Jong, L.D.; Nieuwboer, A.; Aufdemkampe, G. The hemiplegic arm: Interrater reliability and concurrent validity of passive range of motion measurements. *Disabil. Rehabil.* **2007**, *29*, 1442–1448. [[CrossRef](#)]
18. Totty, M.S.; Wade, E. Muscle Activation and Inertial Motion Data for Noninvasive Classification of Activities of Daily Living. *IEEE Trans. Bio-Med. Eng.* **2018**, *65*, 1069–1076.
19. Perez, R.; Costa, U.; Torrent, M.; Solana, J.; Opisso, E.; Caceres, C.; Tormos, J.M.; Medina, J.; Gomez, E.J. Upper limb portable motion analysis system based on inertial technology for neurorehabilitation purposes. *Sensors* **2010**, *10*, 10733–10751. [[CrossRef](#)]
20. Shavelson, R.J.; Webb, N.M.; Rowley, G.L. Generalizability theory. *Am. Psychol.* **1989**, *44*, 922–932. [[CrossRef](#)]
21. Öhberg, F.; Lundström, R.; Grip, H. Comparative analysis of different adaptive filters for tracking lower segments of a human body using inertial motion sensors. *Meas. Sci. Technol.* **2013**, *24*, 12. [[CrossRef](#)]
22. Ertzgaard, P.; Öhberg, F.; Gerdle, B.; Grip, H. A new way of assessing arm function in activity using kinematic Exposure Variation Analysis and portable inertial sensors—A validity study. *Man. Ther.* **2016**, *21*, 241–249. [[CrossRef](#)] [[PubMed](#)]
23. El-Gohary, M.; McNames, J. Human Joint Angle Estimation with Inertial Sensors and Validation with A Robot Arm. *IEEE Trans. Bio-Med. Eng.* **2015**, *62*, 1759–1767. [[CrossRef](#)] [[PubMed](#)]
24. Wu, G.; van der Helm, F.C.; Veeger, H.E.; Makhsous, M.; Van Roy, P.; Anglin, C.; Nagels, J.; Karduna, A.R.; McQuade, K.; Wang, X.; et al. ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: Shoulder, elbow, wrist and hand. *J. Biomech.* **2005**, *38*, 981–992. [[CrossRef](#)] [[PubMed](#)]
25. Soderkvist, I.; Wedin, P.A. Determining the movements of the skeleton using well-configured markers. *J. Biomech.* **1993**, *26*, 1473–1477. [[CrossRef](#)]
26. de Vet Henrica, C.W.; Terwee Caroline, B.; Mokkink Lidwine, B.; Knol, D.L. *Measurement in Medicine: A Practical Guide*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2011; p. x. 338p.
27. Salvia, J.; Ysseldyke, J.E.; Witmer, S. *Assessment in Special and Inclusive Education*, 13th ed.; Cengage Learning: Boston, MA, USA, 2017.
28. Molina Rueda, F.; Rivas Montero, F.M.; Perez de Heredia Torres, M.; Alguacil Diego, I.M.; Molero Sanchez, A.; Miangolarra Page, J.C. Movement analysis of upper extremity hemiparesis in patients with cerebrovascular disease: A pilot study. *Neurologia* **2012**, *27*, 343–347. [[CrossRef](#)]

29. Alt Murphy, M.; Willen, C.; Sunnerhagen, K.S. Responsiveness of upper extremity kinematic measures and clinical improvement during the first three months after stroke. *Neurorehabil. Neural Repair* **2013**, *27*, 844–853. [[CrossRef](#)]
30. Santos, G.L.; Russo, T.L.; Nieuwenhuys, A.; Monari, D.; Desloovere, K. Kinematic Analysis of a Drinking Task in Chronic Hemiparetic Patients Using Features Analysis and Statistical Parametric Mapping. *Arch. Phys. Med. Rehabil.* **2018**, *99*, 501–511. [[CrossRef](#)]
31. Lee, J.A.; Hwang, P.W.; Kim, E.J. Upper extremity muscle activation during drinking from a glass in subjects with chronic stroke. *J. Phys. Ther. Sci.* **2015**, *27*, 701–703. [[CrossRef](#)]
32. Bernstein, N.A. *The Co-Ordination and Regulation of Movements*; Pergamon Press: London, UK, 1967.
33. Rodrigues, M.R.; Slimovitch, M.; Chilingaryan, G.; Levin, M.F. Does the Finger-to-Nose Test measure upper limb coordination in chronic stroke? *J. Neuroeng. Rehabil.* **2017**, *14*, 6. [[CrossRef](#)]
34. Maurel, N.; Diop, A.; Gouelle, A.; Alberti, C.; Husson, I. Assessment of upper limb function in young Friedreich ataxia patients compared to control subjects using a new three-dimensional kinematic protocol. *Clin. Biomech.* **2013**, *28*, 386–394. [[CrossRef](#)]
35. Jeon, H.J.; An, S.; Yoo, J.; Park, N.H.; Lee, K.H. The effect of Monkey Chair and Band exercise system on shoulder range of motion and pain in post-stroke patients with hemiplegia. *J. Phys. Ther. Sci.* **2016**, *28*, 2232–2237. [[CrossRef](#)]
36. Eriks-Hoogland, I.E.; de Groot, S.; Post, M.W.; van der Woude, L.H. Passive shoulder range of motion impairment in spinal cord injury during and one year after rehabilitation. *J. Rehabil. Med.* **2009**, *41*, 438–444. [[CrossRef](#)]
37. Muller, R.; Buttner, P. A critical discussion of intraclass correlation coefficients. *Stat. Med.* **1994**, *13*, 2465–2476. [[CrossRef](#)]
38. Thies, S.B.; Tresadern, P.A.; Kenney, L.P.; Smith, J.; Howard, D.; Goulermas, J.Y.; Smith, C.; Rigby, J. Movement variability in stroke patients and controls performing two upper limb functional tasks: A new assessment methodology. *J. Neuroeng. Rehabil.* **2009**, *6*, 2. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).