Experimental validation of an upper limb benchmarking framework in healthy and post-stroke individuals: a pilot study

Valeria Longatelli[1,2], Clara B. Sanz-Morère[3,4], Diego Torricelli[4], Paula Martos Hernández[3], Eleonora Guanziroli[5], Jesús Tornero[3], Franco Molteni[5], José L. Pons[6], Alessandra Pedrocchi[1,2], Marta Gandolla[2,7]

***Abstract*— In the context of neurorehabilitation, there have been rapid and continuous improvements in sensors-based clinical tools to quantify limb performance. As a result of the increasing integration of technologies in the assessment procedure, the need to integrate evidence-based medicine with benchmarking has emerged in the scientific community. In this work, we present the experimental validation of our previously proposed benchmarking scheme for upper limb capabilities in terms of repeatability, reproducibility, and clinical meaningfulness. We performed a prospective multicenter study on neurologically intact young and elderly subjects and post-stroke patients while recording kinematics and electromyography. 60 subjects (30 young healthy, 15 elderly healthy, and 15 post-stroke) completed the benchmarking protocol. The framework was repeatable among different assessors and instrumentation. Age did not significantly impact the performance indicators of the scheme for healthy subjects. In post-stroke subjects, the movements presented decreased smoothness and speed, the movement amplitude was reduced, and the muscular activation showed lower power and lower intra-limb coordination. We revised the original framework reducing it to three motor skills, and we extracted 14 significant performance indicators with a good correlation with the ARAT clinical scale. The applicability of the scheme is wide, and it may be considered a valuable tool for upper limb functional evaluation in the clinical routine.**

***Index Terms*— Arm, Benchmark, Functional evaluation, Hemiplegia, Neurological disorders, Neurorehabilitation, Performance evaluation.**

I. Introduction

In the context of neurorehabilitation, there has been a rapid and continuous improvement in clinical tools to quantify body function and dysfunction following neurological conditions, such as stroke [1]. The assessment of the motor functions and the influences of deficits on daily life activities are important to reveal movement limitations and drive interventions for improving functional restoration [2]. In both acute and chronic stages, motor recovery is still possible following proper rehabilitation treatments. A detailed evaluation is, therefore, fundamental for all phases of neurorehabilitation. The assessment is necessary both in the early phase, to diagnose the extent of the injury, and in subsequent phases, to determine the effectiveness of different treatment approaches and to inform the clinician about the patient's progress [3], [4]. In both cases, assessment can be of great help to identify the most suitable therapy tailored to the patient's needs (e.g., tuning training parameters) [3], [4]. Moreover, due to rising healthcare costs, assessment has an important socio-economic role, as hospitals and insurance companies offer their services based on clinically meaningful thresholds on standardized assessment scales [4].

Upper limbs movements require multiple degrees of freedom coordination and control to allow a successful interaction with the environment [5]. The sensorimotor impairments following a stroke can result in reduced adaptability to task demands, inefficient movement trajectories, higher energy and force-consumption, or loss of inter-joint coordination [6].

Nowadays, the clinical assessment of motor impairments continues to be largely based on visual and physical inspection guided by criteria-based ordinal scales [1]. This approach is part of the so-called evidence-based medicine [3], which stands on the International Classification of Functioning, Disability and Health validated clinical scales as the major outcome for clinical trials. Such techniques have minimal costs and exploit the capability of the visual system of an expert human evaluator to identify human motor abilities. From a statistical point of view, properly validated clinical scales are reliable and sensitive for measuring gross changes in motor performance [7]. However, most of them exhibit floor and ceiling effects and rely on broad ordinal scales [6]. They are also less sensitive to smaller and more specific changes [8]. Finally, standard clinical scales cannot quantify specifically the variegate relevant aspects that characterize arm movement.

The use of technologies, such as kinematics and electromyography (EMG) sensors, can provide more objective and repeatable methods to support clinical evaluation. In the last years, many research groups have used quantitative measures to assess the upper limb performance of people with neurological

[1]Neuroengineering and Medical Robotics Laboratory, Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, Italy. [2]WE-COBOT Laboratory, Interdepartmental Laboratoy, Politecnico di Milano, Milan, Italy. [3]Center for Clinical Neuroscience, Hospital Los Madroños, Brunete, Madrid, Spain. [4]Neural Rehabilitation Group, Cajal Institute, Spanish National Research Council (CSIC), Madrid, Spain. [5]Villa Beretta Rehabilitation Center, Valduce Hospital, Costa Masnaga, Italy. [6]Shirley Ryan AbilityLab, Chicago, USA. [7]Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy. Marta Gandolla and Alessandra PEdrocchi hold shares in AGADE srl and AllyArm srl. Correspondence to *marta.gandolla@polimi.it*

conditions [9]. Kinematic measurements allow investigating spatio-temporal parameters of a motion act, while EMG measurements allow analyzing the behavior of muscles that causes the altered movement patterns. While kinematics is usually considered in clinical practice, EMG is usually neglected, due to several technical challenges related to data acquisition, analysis and interpretation [10]. Schwarz and colleagues [2] characterized upper limb movement behavior with a core set of kinematic metrics in subjects with and without stroke-related upper limb impairments when performing a large set of activities of everyday life. They highlighted the usefulness of kinematics to assess the spatio-temporal aspects of upper limb movement behavior for total task performance, as well as for task subphases. Two limitations could be identified in this study. First, they involved only five healthy individuals, and this sample size is not sufficient to derive a normative reference of outcome measures. Secondly, this study exploited inertial measurement units, which are not as accurate as optoelectronic motion capture systems [11]. Murphy and co-workers [12] determined a set of clinically-useful and sensitive kinematic variables to quantify upper-extremity motor control during reaching and drinking from a glass in a cohort of stroke patients and healthy people. They identified a subset of kinematics outcome measures that can be efficiently used to discriminate participants with different deficits in motor performance. They used an optoelectronic system, but they did not use marker triads. Thus they did not measure axial rotations, which are relevant to describe arm movements. Another similar study [13] acquired the kinematics of the drinking task in people with spinal cord injuries and healthy volunteers. They used more markers (18), which allowed them to measure also axial rotations. They could discriminate between different levels of patients' residual ability through kinematics variables, also suggesting the possibility of using them as therapeutic recommendations to be integrated into the clinical setting. However, the control group comprised only 8 participants, which could limit the generalizability of their results. Recently, robotic devices have been exploited as an alternative to external sensors to perform quantitative and repeatable assessments of upper limb functions [4], [14], [15]. The main example is the Kinarm Exoskeleton (BKIN Technologies Ltd, Canada), a bilateral robotic exoskeleton that has been widely used to measure patient-specific impairments in cognitive, motor, and sensory functions [16]. Very recently, they applied this device to assess a large cohort of 351 neurologically-intact subjects together with a statistical approach to estimate the recovery of neurologically-impaired individuals [1]. Despite its huge potential, this robotic platform only allows horizontal bi-dimensional movements performed using the exoskeleton. A device-restraint planar task could not be representative of movement tasks in daily living [2].

As a result of the increasingly frequent integration of technologies in the assessment procedure, the need to integrate evidence-based medicine with benchmarking has emerged in the scientific community [3]. The objective of benchmarking is to measure and compare the performance of different technologies or protocols using specific indicators and a reproducible methodology [17]. Benchmarks have long been established in the robotics and automotive industry, but it has not been widely adopted yet in the neurorehabilitation field [18]. Considering lower limb locomotion functions, we have been witnessing the widespread of standardized gait analysis. Even if gait analysis needs a dedicated instrumented motion lab, and expert personnel, it is a common and universally accepted assessment methodology when a deeper analysis is desired. As for the upper limbs, we do not have any correspondence like the instrumented gait analysis, and here is where our work wants to give a contribution.

In addition, some ongoing researches are adopting the benchmarking methods promoted by the EUROBENCH project [19] for benchmarking muscle fatigue [20], human-robot interaction [21], muscle synergies [22] and kinematics [23] when using lower limb exoskeletons.

For the upper limb, instead, we recently developed a benchmarking framework for evaluating motor capabilities in clinical and research settings [3]. It includes these elements: 1) a taxonomy for motor skills and motor abilities, 2) a list of performance indicators (PIs) to quantify each motor ability, 3) the required sensor networks to extract the PIs, and 4) a standardized protocol that should be followed to obtain comparable results.

This work presents the first experimental validation of this benchmarking scheme. The validation approach stands on three fundamental requirements: repeatability (i.e., "achievement of comparable results by the same team, measurement procedure, and locations on multiple trials" [24]), reproducibility (i.e., "obtention of comparable results by different teams, measuring systems, and locations" [24]), and clinical meaningfulness (i.e., "ability to constitute a relevant decision-making support system for clinicians in the neurorehabilitation context" [3]). First, we tested repeatability by performing a Test-Retest analysis on healthy individuals. Second, we investigated reproducibility on a cohort of healthy people in different locations using different instrumentation. In this way, we also determined the normative ranges of performance indicators (PIs) of healthy people that can be used as a standard for comparison. Finally, to investigate the clinical meaningfulness, we deployed the benchmarking scheme on a cohort of healthy elderly subjects and post-stroke patients, evaluating the impact of age and neurological conditions on the outcome of the scheme.

II. Methods

We performed a prospective multicenter study on neurologically-intact and post-stroke subjects. The study took place between November 2021 and October 2022 at the Villa Beretta Rehabilitation Institute (Italy) and the Center for Clinical Neuroscience - Hospital Los Madroños (Spain). It was approved by the ethical committees of Politecnico di Milano (Parere n. 13/2021) and Hospital Universitario Severo Ochoa – Leganés (Código A1366).

A. Participants

The sample size consisted of 60 participants divided into four groups (i.e., *Young* α, *Young* β, *Elderly*, and *Patients*).

The group *Young* α included 15 healthy people aged between 18 and 35 years, recruited at the Villa Beretta Rehabilitation Institute. Participants from groups *Young* β, *Elderly*, and *Patients* were recruited at the Center for Clinical Neuroscience - Hospital Los Madroños. Participants of the group *Young* β and *Elderly* were healthy individuals aged between 18 and 35 years, and between 60 and 85 years, respectively. Subjects were neurologically and orthopedically intact and excluded if they had any pathology affecting arm mobility, cognitive disorders, or symptomatic cardiovascular conditions. Finally, the group *Patients* included in-patients of the clinical center with the following inclusion criteria: i) diagnosis of ischaemic or hemorrhagic stroke causing functional limitation of the upper limb, ii) age between 60 and 85 years, iii) ability to passively extend the shoulder from 0° to 40°, iv) ability to actively extend the elbow from 90° to 120° (180° corresponds to fully extended elbow), and v) ability to understand verbal instructions. Patients were excluded if they met at least one of the following exclusion criteria: global aphasia, severe unilateral spatial neglect, Box and Block Test (BBT) $>$1, Ashworth scale score $\geq$4, total or severe impairment of visual acuity, instability of clinical parameters or presence of severe comorbidities, inability to sit down for more than 10 minutes or inability to comply with the protocol. All participants gave written informed consent before inclusion. We assured that all groups have at least 12 participants, as recommended by pilot studies guidelines [25].

B. Experimental protocol

All experiments were performed by experienced clinicians during a single day of measurements per subject. The experimental procedure strictly followed the benchmarking protocol described in our previous work [3]. All participants performed eight repetitions of six motor skills: anterior reaching at rest position height (ARR), anterior reaching at shoulder height, moving objects at rest position height, moving objects at shoulder height (MOS), bringing hand to mouth without object, and with object (HMO). With stroke patients unable to extend the elbow until 180°, the clinician manually sets the position of the target points in the relative direction at a distance from the acromion of the evaluated arm equal to the total arm length, calculated as the sum of the Euclidean distance between the acromion and lateral elbow markers, and the lateral elbow and the ulnar styloid markers. We instructed participants to reach each target as accurately as possible with the wrist. The object was a 0.5 liter water bottle with an ergonomic grip, representing a typical object of daily life. Healthy participants performed the movements with the dominant side, whereas stroke participants used the paretic limb.

Participants of the group *Young* β were additionally assessed on two different days within two weeks by two different assessors to investigate the benchmarking scheme repeatability in terms of Test-Retest and inter-rater reliability. Before retest, all participants were checked for inclusion and exclusion criteria, even if we did not expect any change in two weeks' time span for young healthy volunteers.

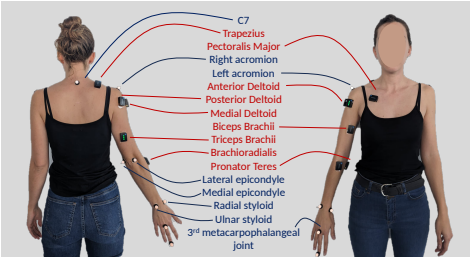

Fig. 1: Anatomical landmarks (blue), and EMG electrodes placement (red) that define the experimental set-up for the benchmarking scheme for a right upper limb evaluation.

C. Experimental set-up

We followed the upper limb benchmarking scheme previously developed [3] and recorded kinematics and EMG. The kinematics was recorded using optoelectronic systems, which represent the gold standard. In the study α, the experimental setup was represented by the SMART-DX 7000, BTS Bioengineering (Italy), and the wearable EMG system FREEEMG 1000, BTS Bioengineering (Italy). In the other groups (*Young* β, *Elderly* group, and *Patients* group), we used the optoelectronic system Vicon Vero and the wearable EMG device Trigno Avanti, Delsys (USA).

To position markers, we followed the guidelines of the International Society of Biomechanics [26], and adapted the model proposed by Rab and colleagues [27]. We placed eight reflective markers on the subject's trunk and the dominant upper limb, specifically on the right and left acromion, 7th cervical vertebra (C7), lateral and medial epicondyles of the elbow, ulnar and radial styloids, and 3rd metacarpophalangeal joint of the medium finger (Figure 1) to build an 8-degree of freedom (DOF) kinematic model of the upper limb, as described in [3].We considered the wrist as the end-effector. Precisely, it was defined as the mid-point between the ulnar and the radial styloid. The corresponding position of this point on the table was marked as the rest position for each subject. Each motor skill started and ended in this position. Moreover, we placed on the table one marker to define each target point and two markers on the object at opposite sites.

Nine bipolar EMG surface electrodes were placed on the following muscles according to the SENIAM (Surface ElectroMyoGraphy for the Non-Invasive Assessment of Muscles) guidelines [28]: trapezius descendens, pectoralis major, anterior deltoid, medial deltoid, posterior deltoid, triceps brachii (long head), biceps brachii (long head), brachioradialis, and pronator teres. The setup procedure lasted less than 5 minutes per participant.

D. Signal pre-processing

Kinematic data were acquired at 250 Hz in the group *Young* α, and at 100 Hz in the groups *Young* β, *Elderly*, and *Patients*. EMG was recorded at 1000 Hz and 2000 Hz, respectively. In terms of acquisition frequencies, we relied on the set-up already in use at the two centres, to preserve the standard clinical setting to show that the instrumentation already used for other scopes can be useful for the proposed benchmarking

scheme as well. The PIs calculated separately for the two centers are comparable (see Supplementary Materials(1) for details). Data from all studies were then post-processed with the same method. After interpolation to fill in missing data, kinematic data were low-pass filtered with a 3rd-order Butterworth filter at 15 Hz. With EMG signals, a standard pre-processing was applied, including high-pass filtering with a 3rd-order Butterworth filter at 20 Hz, rectification, and low-pass filtering with a 3rd-order Butterworth filter at 4 Hz. As for signals normalization process, given that patients often are not able to generate as large a contraction (in terms of EMG signal levels) as the they will generate in an active physical situation, and given the high number of muscles we recorded, we excluded the maximum voluntary contraction recording. The envelopes were then normalized to the 80% of each muscle's maximum observed during the session [29], [30], thus obtaining signals ranging from 0 to 1.

Each motor skill was subdivided into the constituting motor primitives, as described in [3]. To this aim, the onset of each sub-movement was derived from the EMG signal through the Teager-Kaiser operator [31] summed across all EMG channels as suggested by [32]. Movement offset, instead, was detected when the velocity of the wrist midpoint was less than 2% of the maximum velocity during that primitive [12].

E. Outcome measures

We computed the kinematic and EMG PIs suggested in [3] for each participant. For each motor skill, we computed the global PIs as the median across all repetitions and all motor primitives, excluding the "idle" motor primitive where the subject is not moving. For the *Patients* group, we also collected the Action Research Arm Test (ARAT) before the instrumented analysis.

F. Statistical analysis

Given the reduced sample size, we followed a non-parametric approach. First, we investigated the repeatability of the scheme by comparing data from the Test-Retest experiment on the group *Young* β. We computed the Kendall τ as a non-parametric measure of the degree of agreement to quantify the inter-rater reliability for each PI [33], [34]. We followed the guidelines from Cicchetti et al., who suggested the following interpretation when dealing with the clinical significance of the level of agreement: $\tau < 0.40$ is poor, $0.40 \leq \tau < 0.60$ is fair, $0.60 \leq \tau < 0.75$ is good, while $\tau \geq 0.75$ is excellent [35], [36]. For each PI, we defined the Minimum Detectable Change (MDC), which is the smallest change in score that is likely to reflect a true change rather than a measurement error [37]. To compute it, we calculated the Standard Error of Measurement with the formula: $StandardErrorOfMeasurement = GroupedInterquartileRange * \sqrt{1 - KendallTau}$ [38]. To define the $GroupedInterquartileRange$, we adapted the grouped formula for standard deviations suggested by Deeks et al. [39]. Then, the MDC was computed as follows: $MDC = StandardErrorOfMeasurement * 1.96 * \sqrt{2}$ [40]. Please note that in the case of absolute agreement, τ=0, and as a consequence, MDC is equal to zero. This means that any variation of the analyzed PI can be considered a true change.

Then, we extracted the most relevant PIs following two criteria. We excluded PIs with Kendall $\tau < 0.60$, considered poorly repeatable [35], [36]. Then, we computed the Spearman correlation coefficient between PIs, considering kinematics and EMG separately. Precisely, the correlation analysis was performed separately for the ten motor abilities that describe the benchmarking scheme: accuracy, efficacy, efficiency, movement amplitude, muscular effort, intra-limb coordination, planning predictability, power, smoothness, and speed [3]. Coefficients $\leq$ of -0.50 or $\geq$ 0.50 were defined as significant [2]. In this case, we selected only the PI with the higher correlation with the others, excluding the others. We considered only the PIs that respected the two criteria in all motor skills. In the EMG domain, we performed the features extraction considering the couple of antagonist muscles biceps and triceps, which were mostly involved in all motor skills.

The reproducibility of the protocol in different locations was evaluated by comparing data from group *Young* α and data from the Test of groups *Young* β with the Mann-Whitney U test. If data were resulting from being samples from the same population, we defined the Normality Range for each PI as the union of the interquartile ranges of the two groups. Precisely, the minimum was defined as the lowest 1st quartile and the maximum as the highest 3rd quartile between the two groups.

The effect of age and stroke was investigated by a multiple comparisons of groups *Young* β, *Elderly* group, and *Patients* group with the Kruskall-Wallis test. Post-hoc comparisons with Bonferroni correction were used to identify statistically significant differences between the three groups. The age match between *Elderly* and *Patients* groups was verified with the Mann-Whitney U test. We investigated the clinical meaningfulness of the benchmarking scheme in terms of: i) ability to distinguish healthy subjects from neurological ones, and ii) the correlation of the scheme results with the ARAT scale. For this final aim, we computed the correlation of PIs with the ARAT scale with the Kendall τ coefficient [41], [42].

For all tests, the significance threshold was set to 0.05. The statistical analyses were performed using Matlab 2022b.

III. Results

For the sake of simplicity, this section focuses on the results obtained on anterior reaching at rest position height (ARR), hand to mouth with object (HMO), and move object at shoulder height (MOS). After a first preliminary analysis, we selected these three most significant motor skills with an increasing level of difficulty according to clinicians' opinions. Results on other motor skills can be found in the Supplementary Material.

A. Participants results

60 subjects (15 per group) completed the benchmarking protocol and were included in the analyses.

Table I shows the demographic characteristics of each group. *Elderly* and *Patients* groups were not significantly different in terms of age (p-value = 0.13). Eight patients had an

ischemic stroke, while seven were hemorrhagic. The median time since the acute event was 8.00 [4.25 - 9.75] months. Median ARAT score for patients was 46, with interquartile range equals to 30. Full patients' description is given in Supplementary Material(2).

	Young α	**Young β**	**Elderly**	**Patients**
Age	23 (3.75)	24 (4.75)	69 (5.50)	73 (11.50)
Sex (M/F)	5/10	4/11	7/8	7/8
Dominance (R/L)	13/2	14/1	14/1	15/0
Test side (R/L)	13/2	14/1	14/1	8/7

TABLE I: Demographic characteristics of participants. M = male; F = female; R = right; L = left; Age is given in years in terms of median (interquartile range).

B. Repeatability results and features extraction

14 subjects of the group *Young* β completed the Test-Retest protocol on two different days. In the EMG domain, three couples of muscles showed good repeatability: trapezius and pectoralis, anterior and posterior deltoid, and biceps and triceps. Medial deltoid, brachioradialis, and pronator teres showed poor repeatability in most PIs for all motor skills, confirmed by EMG signals visual inspection. Therefore, we considered them non-repeatable, and excluded them from further analyses.

We excluded 21 PIs (15 kinematics and 5 EMG) that had $\tau < 0.60$ in at least one motor skill. The repeatable PIs for each motor ability are the following. *Accuracy*: end point error ($0.66<\tau<0.89$), area index ($0.64<\tau<0.92$). *Efficacy*: success rate (τ=1), number of movement stops (τ=1). *Efficiency*: described by three repeatable kinematics and one EMG PIs (i.e., movement time, path traveled, path length ratio, and waveform length). *Intra-limb coordination*: joint angle correlation ($0.60<\tau<0.77$), co-contraction index ($0.65<\tau<0.86$). *Movement amplitude*: three kinematics PIs (i.e., maximum reached distance, elevation angle ROM, and elbow flexion/extension ROM). *Muscular effort*: integrated EMG ($0.63<\tau<0.85$), root mean square ($0.61<\tau<0.91$), and activation level of the EMG signal ($0.64<\tau<0.89$). *Planning predictability*: time to peak velocity ($0.62<\tau<0.70$). *Power*: mean frequency ($0.60<\tau<0.8$), median frequency ($0.61<\tau<0.82$), and power spectrum ratio ($0.60<\tau<0.77$). *Smoothness*: five kinematics PIs (i.e., number of velocity peaks, movement arrest period ratio, normalized dimensionless jerk, spectral arc length, and mean acceleration) and one EMG PI (i.e., slope sign change) had good or excellent repeatability in all motor skills in the motor ability. *Speed*: mean velocity ($0.76<\tau<0.93$), peak velocity (peak velocity: $0.81<\tau<0.96$).

From this set of PIs coupled with the Spearman correlation analysis, we extracted 13 kinematics PIs and 4 EMG PIs that can be suggested for the instrumented assessment of upper limb capabilities. Table II shows the results of the MDC for each PI and motor skill. Kinematic PIs were generally characterized by lower MDC in ARR than HMO and MOS, due to a lower Standard Error of Measurement. It demonstrates that ARR is a highly standardizable and repeatable motor skill to be included in an assessment procedure considering the kinematics domain. With the EMG PIs, instead, we did not observe any trend related to the different motor skills. This optimized set of PIs was used for the subsequent analyses.

C. Reproducibility results and normative pattern definition

All selected PIs were estimated as reproducible with different lab equipment for all motor skills (p-values > 0.05). Results are shown in Supplementary Material(1).

Table II shows the normative range within which a neurologically-intact and young person should lie for the two motor skills. Normative ranges of the other motor skills are shown in Supplementary Material(3,4). Note that the end point error takes into consideration that the markers are on the dorsal side of the wrist's subject while he/she reaches the target marker with the palm of the hand down.

D. Effect of age

The *Elderly* group revealed little differences compared to the young one (Figure 2). Considering the kinematics, only in ARR, the *Elderly* group showed lower accuracy characterized by larger variability among subjects (end point error - *Young*: 3.31 [0.60], *Elderly*: 4.62 [1.33], p-value = 0.02). The median value for this PI was also outside the normality range. Figure 3 shows the kinematic of ARR in the horizontal plane (parallel to the table) of a representative young and an elder participant. It can be observed that *Elderly* showed greater variability of reaching the central and controlateral target compared to *Young*. Both *Young* and *Elderly* had a small variability around the median trajectory when reaching the ipsilateral target. Analyzing MOS, we observed a higher path length ratio in the young group (*Young*: 1.24 [0.06], *Elderly*: 1.15 [0.20], p-value = 0.03), meaning that young people followed a trajectory more curvilinear, although less efficient, where the most efficient trajectory is the linear one [3].

In general, the speed of *Elderly* was reduced. Considering EMG, we noticed a general decrease in mean EMG frequency in the *Elderly* for all motor skills, indicating less power (i.e., higher fatigue) compared to *Young* [43].

E. Effect of stroke

All patients were able to complete the motor skills ARR and hand to mouth without object. Three patients could not perform MOS, whereas one patient had the residual ability to complete only ARR and hand to mouth without object. The success rates confirmed the hypothesis of the increasing difficulty in ARR, HMO, and MOS.

As it can be observed in Figure 3, the kinematic profile of patients was characterized by less accuracy and smoothness. We focused our analysis on the post-hoc comparison between *Patients* and *Elderly* groups to maintain the age match.

	Motor ability	Performance Indicator	Minimum Detactable Change			Normality Range		
			ARR	HMO	MOS	ARR	HMO	MOS
KINEMATICS	Accuracy	End point error [cm]	1.21	2.05	0.89	[1.31 - 4.17]	[0.27 - 3.52]	[1.38 - 3.97]
	Efficacy	Success rate [%]	0.00	0.00	0.00	[100.00 - 100.00]	[100.00 - 100.00]	[100.00 - 100.00]
	Efficiency	Movement time [sec]	0.19	0.42	0.32	[0.88 - 1.96]	[1.31 - 2.32]	[1.13 - 1.97]
		Path length ratio	0.20	0.69	0.20	[1.06 - 1.42]	[0.88 - 1.70]	[1.14 - 1.33]
	Intra-limb coordination	Joint angle correlation	0.03	0.67	0.07	[0.95 - 1.00]	[0.24 -0.77]	[0.83 -0.97]
	Movement amplitude	Maximum reached distance [cm]	4.46	6.72	7.03	[27.89 - 34.42]	[13.44 - 32.66]	[29.70 - 49.77]
		ROM elevation angle [deg]	13.88	7.61	21.02	[19.91 - 35.78]	[14.26 - 28.86]	[41.99 - 80.73]
		ROM elbow flex/ext [deg]	16.05	6.05	19.91	[149.45 - 189.95]	[35.27 - 65.23]	[116.39 - 179.79]
	Planning predictability	Time to peak velocity [%]	6.77	11.53	7.28	[42.48 - 57.37]	[37.36 - 59.22]	[42.10 - 52.61]
	Smoothness	Number of velocity peaks	0.00	0.00	0.00	[1.00 - 1.00]	[1.00 - 1.00]	[1.00 - 1.00]
		Movement arrest period ratio [%]	7.76	6.35	5.16	[42.93 - 62.82]	[40.03 - 58.15]	[43.45 - 58.38]
		Spectral arc length	0.05	0.11	0.06	[- 1.54 – 1.38]	[- 1.68 – 1.36]	[- 1.51 – 1.39]
	Speed	Mean velocity [cm/s]	4.57	6.86	6.35	[15.84 - 29.35]	[6.03 - 24.05]	[16.51 - 36.89]
EMG	Efficiency	Waveform length - Triceps	1.21	0.89	0.79	[0.14 - 2.65]	[0.34 - 2.39]	[0.33 - 2.69]
		Waveform length - Biceps	0.77	1.22	0.73	[0.44 - 1.54]	[0.41 - 2.09]	[0.49 - 2.67]
	Muscular effort	Root mean square - Triceps	0.32	0.36	0.54	[0.05 - 0.67]	[0.02 - 0.59]	[0.12 - 0.65]
		Root mean square - Biceps	0.34	0.35	0.32	[0.06 - 0.64]	[0.19 - 0.65]	[0.14 - 0.49]
	Power	Mean frequency [Hz] - Triceps	11.42	28.60	9.25	[71.54 - 132.12]	[34.75 - 77.72]	[74.38 - 97.90]
		Mean frequency [Hz] - Biceps	11.06	27.26	13.17	[73.99 - 136.35]	[58.35 - 98.42]	[80.54 - 103.38]
	Intra-limb coordination	Co-contraction Index	0.19	0.22	0.24	[0.04 - 0.34]	[0.03 - 0.36]	[0.09 - 0.42]

TABLE II: Optimized set of Performance Indicators to be used for the benchmarking scheme and associated Minimum Detectable Change and normality ranges. ARR = Anterior reaching at rest position height; HMO = Hand to mouth with object; MOS = Move object at shoulder height.

In ARR, HMO, and MOS, smoothness, movement amplitude, and power were significantly different between *Patients* and *Elderly*. In particular, for smoothness, the number of velocity peaks and the spectral arc length were able to differentiate between groups, and results obtained by *Patients* also fell outside the normality ranges. For movement amplitude, the ROM of the elbow flexion/extension joint was statistically different between groups, even if the median results of patients were within the normality ranges.

Finally, the power analysis revealed lower mean frequencies of activation for all muscles, suggesting higher fatigue [43] and reduced power in patients. We also observed differences in efficiency and planning predictability, which are related to the timing of movements. In ARR and MOS, patients were less efficient, as detected by the movement time, while in ARR and HMO, the time required to reach the peak velocity decreased. HMO and MOS also revealed a decreased accuracy of patients, as quantified by the end point error.

With MOS, we observed more differences between the elderly and patients compared to ARR and HMO (Figure 2). The results confirmed the significant differences observed for ARR and HMO, but the difference in the median value of the PIs was more pronounced. The movement was significantly less smooth, as also detected by the movement arrest period ratio. MOS was associated with a reduced shoulder elevation, but the decrease was not significant (*Patients*: 30.36 [19.28], *Elderly*: 59.11 [22.08], p-value = 0.075). It could be hypothesized that patients who could reach the target with complete elbow extension compensated with the trunk.

Considering the correlation between PIs and the ARAT scale, we found an excellent correlation with the spectral arm length for both motor skills (ARR: τ = 0.91, MOS: τ = 0.83) and the EMG mean frequency (ARR: τ = 0.75, MOS: τ = 0.84). The correlation was poor for the movement time (ARR: τ = 0.31, MOS: τ = 0.29), the number of velocity peaks (ARR: τ = 0.33, MOS: τ = 0.29), and the Movement arrest period ratio (ARR: τ = 0.27, MOS: τ = 0.38). For the other PIs, the correlation was good (0.61 $<\tau<$ 0.74).

IV. Discussion

This work presents the experimental validation of the benchmarking scheme for upper limb capabilities developed in our previous work [3]. We performed an instrumented assessment exploring the kinematics and EMG domains. The scheme was validated in terms of repeatability, reproducibility, and clinical meaningfulness. First, we performed a Test-Retest protocol on healthy young subjects to validate repeatability. Then, we reproduced the scheme on healthy yuong participants in a different laboratory with different assessors to investigate reproducibility. Finally, the framework was performed on elderly individuals and post-stroke patients.

Kinematics can quantify smoothness and coordination between different joints within the same limb, essential factors for natural and efficient movement, which cannot be captured by conventional scales. Incorporating kinematics not only provides a comprehensive view of recovery, but also allows for nuanced subject stratification. This approach enables a deeper evaluation of the impact and effects of technology, including robotics, in guiding rehabilitation toward correct movement patterns. Our results are in line with those of a recent systematic review [44]. PIs with sufficient summarized evidence according to their criteria were classified in our analysis with excellent or good repeatability. We obtained an opposite result for the trunk displacement. It could be due to our marker placement, which includes only one marker on C7. Other protocols available in the literature include more markers for the trunk (e.g., sternum, clavicle, T8). We preferred to reduce the number of markers to develop a feasible and easy-to-implement protocol. Moreover, adding more markers to the trunk would involve a bare-chested protocol, which could not be comfortable for patients/elderly subjects who are less comfortable with their bodies. Indeed, the evaluation of trunk movement is useful for detecting possible compensatory strategies often present in neurological patients [45], and we could integrate the protocol in this sense.

This work investigated for the first time the repeatability of EMG PIs. From our results, we can draw the following guidelines. Medial deltoid, brachioradialis, and pronator teres were not considered repeatable. As a result, we suggest that an instrumented evaluation of the upper limb should include the following three couples of antagonist muscles: trapezius and pectoralis, anterior and posterior deltoid, biceps and triceps. We observed the highest repeatability for the mean frequency.

For each PI, we quantified the MDC, useful for assessing the efficacy of rehabilitative interventions at different time points, or computing the required sample size for randomized controlled studies involving kinematics or EMG PIs as primary outcome measures. It has to be underlined that this comparison requires following the same approach for data pre-processing.

Another goal of this work was the identification of a reduced set of PIs with an optimal trade-off between number of variables reduction, while retaining as needed variables to describe performance. We identified 13 kinematics PIs, mutually uncorrelated, repeatable, and reproducible, useful for assessing the spatio-temporal aspects of upper limb movement behavior, and 4 EMG PIs to evaluate the muscular activation in the time and frequency domains. These PIs also showed a good or excellent correlation with the ARAT scale, except for movement time, number of velocity peaks, and movement arrest period ratio. We can therefore suggest the PIs listed in Table II, excluding these last 3 PIs. In this way, each relevant motor ability could be quantified by at least one PI.

The normality ranges were defined among the population of 30 healthy young subjects. Some PIs (e.g., mean velocity, shoulder flexion/extension ROM) are characterized by greater variability, which we could associate with the large natural variation between the subjects. Anyway, our results are comparable with those obtained in two similar studies performed on healthy subjects on reaching and hand to mouth motor skills [12], [13]. Another study [46] investigated the hand to mouth and central reaching motor skills in a larger population of young and elderly healthy subjects. However, their marker protocol was extremely simplified (only 5 markers), and did not consider the axial rotations of the upper and forearm.

As seen in the literature [46], age did not significantly im-

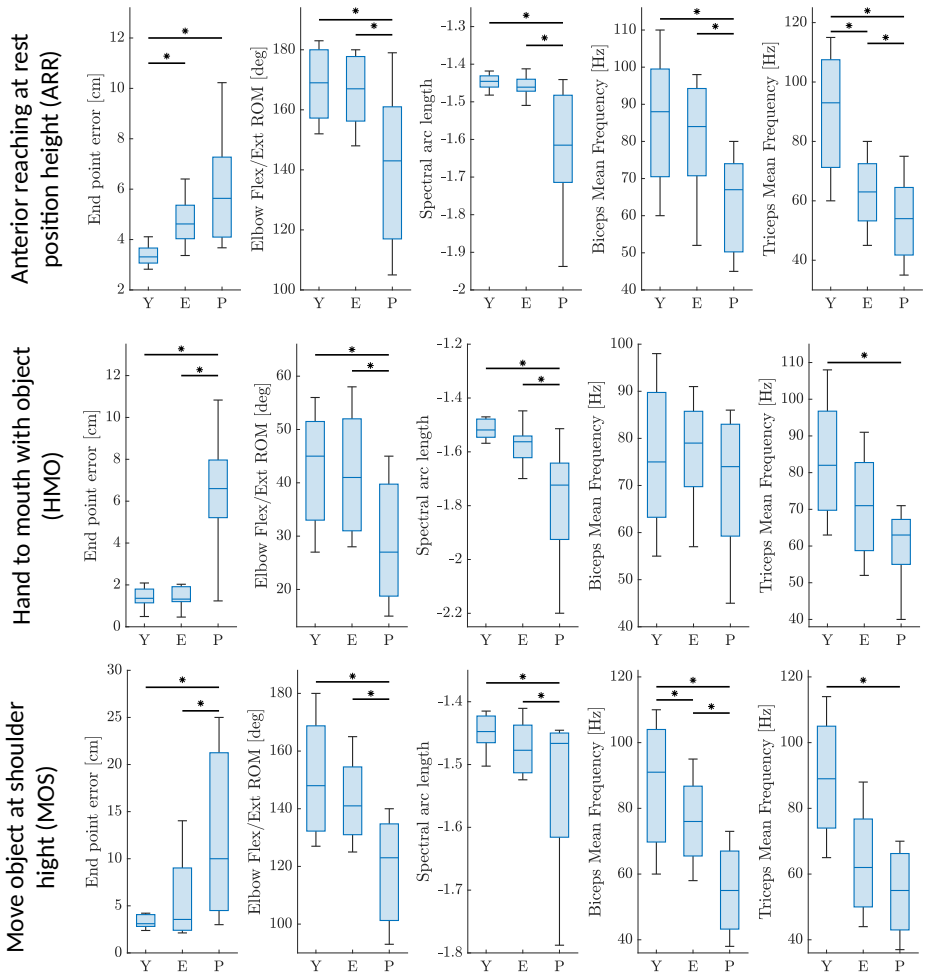

Fig. 2: Kinematics and electromyographic PIs of anterior reaching at rest position height (ARR), hand to mouth with object (HMO), and move object at shoulder height (MOS) across healthy young (H), elderly (E), and patients (P) groups.

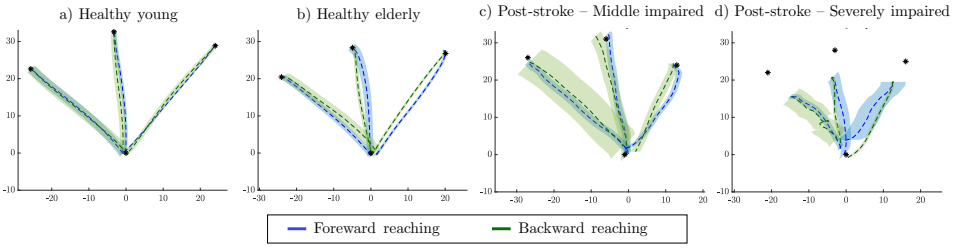

Fig. 3: End-effector kinematics profile of motor skill anterior reaching at rest position height of a healthy young subject, an elderly, a mild impaired post-stroke patient, and a severely-impaired post-stroke patient in the transversal plane. Blue dotted line = Foreward reaching; Green dotted line = Backward reaching.

pact kinematic performance, where we only detected lower accuracy and speed. In contrast, EMG analysis revealed reduced power in the *Elderly* group. This highlights the relevance of EMG measures in clinical evaluation.

Stroke caused differences among groups, demonstrating the effectiveness of the benchmarking scheme in detecting and quantifying different neurological conditions. The motor skills anterior reaching at shoulder height, and hand to mouth without object were the easiest ones, and all patients were able to complete them successfully. Despite this, patients' movements were characterized by reduced smoothness, speed, movement amplitude, power, and intra-limb coordination. The motor skills HMO and move objects at rest position height could be considered as an intermediate level of difficulty. We observed greater differences between healthy and patient groups. In particular, the path length ratio allowed distinguishing patients who can lift the object from the table from patients who dragged it to the target point. This result is in agreement with

the ARAT scale, since dragging the bottle was the strategy selected by compromised patients (i.e., ARAT < 35). Two patients with ARAT equal to 7 and 14 points, respectively, could not lift the bottle to the mouth and, hence, failed to perform HMO. It has to be underlined that these motor skills also require hand functionality. Results between ARR, anterior reaching at shoulder height, and hand to mouth without object are qualitatively comparable. The same consideration could be done for HMO and move object at rest position height. Therefore, in the case of time constraints, we suggest reducing the protocol to only ARR and HMO, which represent a functional task and an ADL, respectively. The MOS was the most difficult motor skill. Only patients with a good residual level of ability (i.e., ARAT > 37) were able to complete it.

Despite the relevance of this work, some limitations can be identified. The assessment has been performed in a controlled laboratory environment, and the results could change transferring the assessment into an ecological environment. Our current setup considers kinematics and EMG sensors that are the gold-standard in biomechanical assessments to properly validate the benchmarking scheme, technology already present in some clinical settings for other uses (e.g., gait analysis). However, the complexity of the setup, the time needed to place the sensors and the post-processing time might limit the integration of the scheme in clinical scenarios, even if a cost/effectiveness analysis is out of the scope of this manuscript. Other sensors might be considered such as inertial measurement units. Recently in the scientific literature, simultaneous measurements obtained wit motion capture system and IMUs have demonstrated the potential use of IMUs in clinical settings to quantify movement quality in stroke patients performing the drinking task ([47]), even if suggested in rehabilitation programs in unsupervised settings not requiring a high level of detail ([48]). However, further studies are needed in this direction, since the effect of sensors positioning and calibration might affect outcome measures reliability ([49]). In this view, we envisage on our side or other research groups on extending an accurate analysis of the proposed framework using simpler and cost-effective sensors. Secondly, our sample sizes were limited, and further research is needed to confirm our results. Another aspect we did not consider was the influence of the dominant side of post-stroke patients. Literature demonstrated that subjects with impairment on the dominant side showed fewer impairments, but this was not translated into better performances in activities of daily living [50]. However, future studies could explore the effect of arm dominance in the results. Finally, we considered only the median scores computed among different motor primitives. Future studies could focus on the analysis of motor primitives. Indeed, their analysis could uncover maladaptation and relevant limitations in movement behavior typical of stroke. Test-retest assessment can be conducted on the elderly group, where patients are more common, with particular attention to possible fatigue, and maybe considering to reduce the protocol to facilitate protocol acceptance. Finally, a further study including longitudinally joint evaluation of clinical scales and the benchmarking scheme might highlight the capability of the benchmarking scheme to detect changes occurring in recovery that current clinical scales fail to identify.

V. Conclusion

In this work, we validated a benchmarking framework for the quantitative assessment of upper limb capacity through kinematics and electromyography measures [3]. We involved young and elderly neurologically-intact participants, as well as post-stroke patients. The scheme was repeatable, reproducible, and clinically meaningful. It is also feasible and easy to implement from the point of view of both the patient and the operator. Considering the vast range of information that can be obtained through a set of simple motor skills, its potential use is extensive, making it a significant tool for assessing upper limb functionality in clinical settings.

References

[1] S. H. Scott, C. R. Lowrey, I. E. Brown, and S. P. Dukelow, "Assessment of neurological impairment and recovery using statistical models of neurologically healthy behavior," *Neurorehabilitation and Neural Repair*, p. 15459683221115413, 2022.

[2] A. Schwarz, M. Bhagubai, S. H. Nies, J. P. Held, P. H. Veltink, J. H. Buurke, and A. R. Luft, "Characterization of stroke-related upper limb motor impairments across various upper limb activities by use of kinematic core set measures," *Journal of NeuroEngineering and Rehabilitation*, vol. 19, no. 1, pp. 1–18, 2022.

[3] V. Longatelli, D. Torricelli, J. Tornero, A. Pedrocchi, F. Molteni, J. L. Pons, and M. Gandolla, "A unified scheme for the benchmarking of upper limb functions in neurological disorders," *Journal of NeuroEngineering and Rehabilitation*, vol. 19, p. 102, Sept. 2022.

[4] O. Lambercy, L. Lünenburger, R. Gassert, and M. Bolliger, "Robots for measurement/clinical assessment," *Neurorehabilitation technology*, pp. 443–456, 2012.

[5] M. Santello, "Synergistic control of hand muscles through common neural input," in *the human hand as an inspiration for robot hand development*, pp. 23–48, Springer, 2014.

[6] A. Schwarz, G. Averta, J. M. Veerbeek, A. R. Luft, J. P. Held, G. Valenza, A. Bicchi, and M. Bianchi, "A functional analysis-based approach to quantify upper limb impairment level in chronic stroke patients: a pilot study," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4198–4204, IEEE, 2019.

[7] T. Platz, C. Pinkowski, F. van Wijck, I.-H. Kim, P. Di Bella, and G. Johnson, "Reliability and validity of arm function assessment with standardized guidelines for the fugl-meyer test, action research arm test and box and block test: a multicentre study," *Clinical rehabilitation*, vol. 19, no. 4, pp. 404–411, 2005.

[8] J. H. Van Der Lee, H. Beckerman, G. J. Lankhorst, L. M. Bouter, *et al.*, "The responsiveness of the action research arm test and the fugl-meyer assessment scale in chronic stroke patients," *Journal of rehabilitation medicine*, vol. 33, no. 3, pp. 110–113, 2001.

[9] F. Garro, M. Chiappalone, S. Buccelli, L. De Michieli, and M. Semprini, "Neuromechanical biomarkers for robotic neurorehabilitation," *Frontiers in Neurorobotics*, vol. 15, p. 742163, 2021.

[10] I. Campanini, C. Disselhorst-Klug, W. Z. Rymer, and R. Merletti, "Surface emg in clinical assessment and neurorehabilitation: barriers limiting its use," *Frontiers in neurology*, p. 934, 2020.

[11] R. Garimella, T. Peeters, K. Beyers, S. Truijen, T. Huysmans, and S. Verwulgen, "Capturing joint angles of the off-site human body," in *2018 IEEE SENSORS*, pp. 1–4, IEEE, 2018.

[12] M. A. Murphy, C. Willén, and K. S. Sunnerhagen, "Kinematic variables quantifying upper-extremity performance after stroke during reaching and drinking from a glass," *Neurorehabilitation and neural repair*, vol. 25, no. 1, pp. 71–80, 2011.

[13] A. de los Reyes-Guzmán, A. Gil-Agudo, B. Peñasco-Martín, M. Solís-Mozos, A. del Ama-Espinosa, and E. Pérez-Rizo, "Kinematic analysis of the daily activity of drinking from a glass in a population with cervical spinal cord injury," *Journal of neuroengineering and rehabilitation*, vol. 7, pp. 1–12, 2010.

[14] F. Grimm, J. Kraugmann, G. Naros, and A. Gharabaghi, "Clinical validation of kinematic assessments of post-stroke upper limb movements with a multi-joint arm exoskeleton," *Journal of NeuroEngineering and Rehabilitation*, vol. 18, no. 1, p. 92, 2021.

[15] A. Merlo, M. Longhi, E. Giannotti, P. Prati, M. Giacobbi, E. Ruscelli, A. Mancini, M. Ottaviani, L. Montanari, and D. Mazzoli, "Upper limb evaluation with robotic exoskeleton. normative values for indices of accuracy, speed and smoothness," *NeuroRehabilitation*, vol. 33, no. 4, pp. 523–530, 2013.

[16] J. M. Kenzie, J. A. Semrau, M. D. Hill, S. H. Scott, and S. P. Dukelow, "A composite robotic-based measure of upper limb proprioception," *Journal of neuroengineering and rehabilitation*, vol. 14, no. 1, pp. 1–12, 2017.

[17] F. Aller, D. Pinto-Fernandez, D. Torricelli, J. L. Pons, and K. Mombaur, "From the state of the art of assessment metrics toward novel concepts for humanoid robot locomotion benchmarking," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 914–920, 2019.

[18] D. Torricelli, C. Rodriguez-Guerrero, J. F. Veneman, S. Crea, K. Briem, B. Lenggenhager, and P. Beckerle, "Benchmarking wearable robots: challenges and recommendations from functional, user experience, and methodological perspectives," *Frontiers in Robotics and AI*, p. 168, 2020.

[19] D. Torricelli and J. L. Pons, "Eurobench: Preparing robots for the real world," in *Wearable Robotics: Challenges and Trends: Proceedings of the 4th International Symposium on Wearable Robotics, WeRob2018, October 16-20, 2018, Pisa, Italy 3*, pp. 375–378, Springer, 2019.

[20] R. M. van Sluijs, D. Rodriguez-Cianca, C. B. Sanz-Morère, S. Massardi, V. Bartenbach, and D. Torricelli, "A method to quantify the reduction of back and hip muscle fatigue of lift-support exoskeletons," *Wearable Technologies*, vol. 4, 2023.

[21] C. Rodrigues-Carvalho, M. Fernández-García, D. Pinto-Fernández, C. Sanz-Morere, F. O. Barroso, S. Borromeo, C. Rodríguez-Sánchez, J. C. Moreno, and A. J. del Ama, "Benchmarking the effects on humanndash;exoskeleton interaction of trajectory, admittance and emg-triggered exoskeleton movement control," *Sensors*, vol. 23, no. 2, 2023.

[22] T. Lencioni, M. Semprini, V. Bandini, J. Jonsdottir, S. Maludrottu, A. Marzegan, S. Scarpetta, C. Vassallo, L. De Michieli, and M. Ferrarin, "Motor control of the lower limbs while walking with the twin exoskeleton operated by twinacta in healthy subjects," *Gait Posture*, vol. 97, pp. 27–28, 2022. Abstracts of the 22nd National Congress of SIAMOC.

[23] M. Goffredo, P. Romano, F. Infarinato, M. Cioeta, M. Franceschini, D. Galafate, R. Iacopini, S. Pournajaf, and M. Ottaviani, "Kinematic analysis of exoskeleton-assisted community ambulation: An observational study in outdoor real-life scenarios," *Sensors*, vol. 22, no. 12, 2022.

[24] H. E. Plesser, "Reproducibility vs. replicability: a brief history of a confused terminology," *Frontiers in neuroinformatics*, vol. 11, p. 76, 2018.

[25] S. A. Julious, "Sample size of 12 per group rule of thumb for a pilot study," *Pharamceutical Statistics*, vol. 4, p. 5, 2005.

[26] G. Wu, F. C. T. van der Helm, H. E. J. D. Veeger, M. Makhsous, P. Van Roy, C. Anglin, J. Nagels, A. R. Karduna, K. McQuade, X. Wang, F. W. Werner, B. Buchholz, and International Society of Biomechanics, "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," *Journal of Biomechanics*, vol. 38, pp. 981–992, May 2005.

[27] G. Rab, K. Petuskey, and A. Bagley, "A method for determination of upper extremity kinematics," *Gait & Posture*, vol. 15, pp. 113–119, Apr. 2002.

[28] H. J. Hermens, B. Freriks, C. Disselhorst-Klug, and G. Rau, "Development of recommendations for semg sensors and sensor placement procedures," *Journal of electromyography and Kinesiology*, vol. 10, no. 5, pp. 361–374, 2000.

[29] C. J. De Luca, "The use of surface electromyography in biomechanics," *Journal of applied biomechanics*, vol. 13, no. 2, pp. 135–163, 1997.

[30] S. Dalla Gasperina, V. Longatelli, F. Braghin, P. Alessandra, and M. Gandolla, "Development and electromyographic validation of a compliant human-robot interaction controller for cooperative and personalized neurorehabilitation," *Frontiers in Neurorobotics*, 2022.

[31] S. Solnik, P. DeVita, P. Rider, B. Long, and T. Hortobágyi, "Teager–kaiser operator improves the accuracy of emg onset detection independent of signal-to-noise ratio," *Acta of bioengineering and biomechanics/Wroclaw University of Technology*, vol. 10, no. 2, p. 65, 2008.

[32] X. Zhang, X. Li, O. W. Samuel, Z. Huang, P. Fang, and G. Li, "Improving the robustness of electromyogram-pattern recognition for prosthetic control by a postprocessing strategy," *Frontiers in Neurorobotics*, vol. 11, p. 51, 2017.

[33] L. Puka, "Kendall's Tau," in *International Encyclopedia of Statistical Science* (M. Lovric, ed.), pp. 713–715, Berlin, Heidelberg: Springer, 2011.

[34] G. Nuyens, W. De Weerdt, P. Ketelaer, H. Feys, L. De Wolf, L. Hantson, A. Nieuwboer, A. Spaepen, and H. Carton, "Inter-rater reliability of the ashworth scale in multiple sclerosis," *Clinical Rehabilitation*, vol. 8, no. 4, pp. 286–292, 1994.

[35] D. V. Cicchetti, "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology," *Psychological Assessment*, vol. 6, pp. 284–290, 1994. Place: US Publisher: American Psychological Association.

[36] R. A. Charter, "A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability," *The Journal of General Psychology*, vol. 130, pp. 290–304, July 2003.

[37] S. K. Rai, J. Yazdany, P. R. Fortin, and J. A. Aviña-Zubieta, "Approaches for estimating minimal clinically important differences in systemic lupus erythematosus," *Arthritis Research & Therapy*, vol. 17, p. 143, June 2015.

[38] J. Tighe, I. McManus, N. G. Dewhurst, L. Chis, and J. Mucklow, "The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP(UK) examinations," *BMC Medical Education*, vol. 10, p. 40, June 2010.

[39] J. J. Deeks, J. P. Higgins, and D. G. Altman, "Analysing Data and Undertaking Meta-Analyses," *Cochrane Handbook for Systematic Reviews of Interventions*, Sept. 2008.

[40] M. Gandolla, A. Antonietti, V. Longatelli, E. Biffi, E. Diella, M. Delle Fave, M. Rossini, F. Molteni, G. D'Angelo, M. Bocciolone, and A. Pedrocchi, "Test-retest reliability of the Performance of Upper Limb (PUL) module for muscular dystrophy patients," *PLoS ONE*, vol. 15, p. e0239064, Sept. 2020.

[41] D. E. Beaton, M. Boers, and G. A. Wells, "Many faces of the minimal clinically important difference (mcid): a literature review and directions for future research," *Current opinion in rheumatology*, vol. 14, no. 2, pp. 109–114, 2002.

[42] M. J. Johnson, J. M. Bland, P. M. Davidson, P. J. Newton, S. G. Oxberry, A. P. Abernethy, and D. C. Currow, "The relationship between two performance scales: New york heart association classification and karnofsky performance status scale," *Journal of pain and symptom management*, vol. 47, no. 3, pp. 652–658, 2014.

[43] T. Schmalz, J. Schändlinger, M. Schuler, J. Bornmann, B. Schirrmeister, A. Kannenberg, and M. Ernst, "Biomechanical and metabolic effectiveness of an industrial exoskeleton for overhead work," *International journal of environmental research and public health*, vol. 16, no. 23, p. 4792, 2019.

[44] A. Schwarz, C. M. Kanzler, O. Lambercy, A. R. Luft, and J. M. Veerbeek, "Systematic review on kinematic assessments of upper limb movements after stroke," *Stroke*, vol. 50, no. 3, pp. 718–727, 2019.

[45] M. Cirstea and M. F. Levin, "Compensatory strategies for reaching in stroke," *Brain*, vol. 123, no. 5, pp. 940–953, 2000.

[46] M. Caimmi, E. Guanziroli, M. Malosio, N. Pedrocchi, F. Vicentini, L. Molinari Tosatti, and F. Molteni, "Normative data for an instrumental assessment of the upper-limb functionality," *BioMed Research International*, vol. 2015, 2015.

[47] T. Unger, R. de Sousa Ribeiro, M. Mokni, T. Weikert, J. Pohl, A. Schwarz, J. P. O. Held, L. Sauerzopf, B. Kühnis, E. Gavagnin, A. R. Luft, R. Gassert, O. Lambercy, C. Awai Easthope, and J. G. Schönhammer, "Upper limb movement quality measures: comparing imus and optical motion capture in stroke patients performing a drinking task," *Front. Digit. Health*, 2024.

[48] S. Cerforglio, P. Capodaglio, P. Rossi, P. Conforti, V. D'Angeli, E. Milani, M. Galli, and V. Cimolin, "Evaluation of upper body and lower limbs kinematics through an imu-based medical system: A comparative study with the optoelectronic system," *Sensors*, 2023.

[49] S. L. Wang, G. Civillico, W. Niswander, and K. L, "Comparison of motion analysis systems in tracking upper body movement of myoelectric bypass prosthesis users," *Sensors*, 2022.

[50] J. E. Harris and J. J. Eng, "Individuals with the dominant hand affected following stroke demonstrate less impairment than those with the nondominant hand affected," *Neurorehabilitation and neural repair*, vol. 20, no. 3, pp. 380–389, 2006.