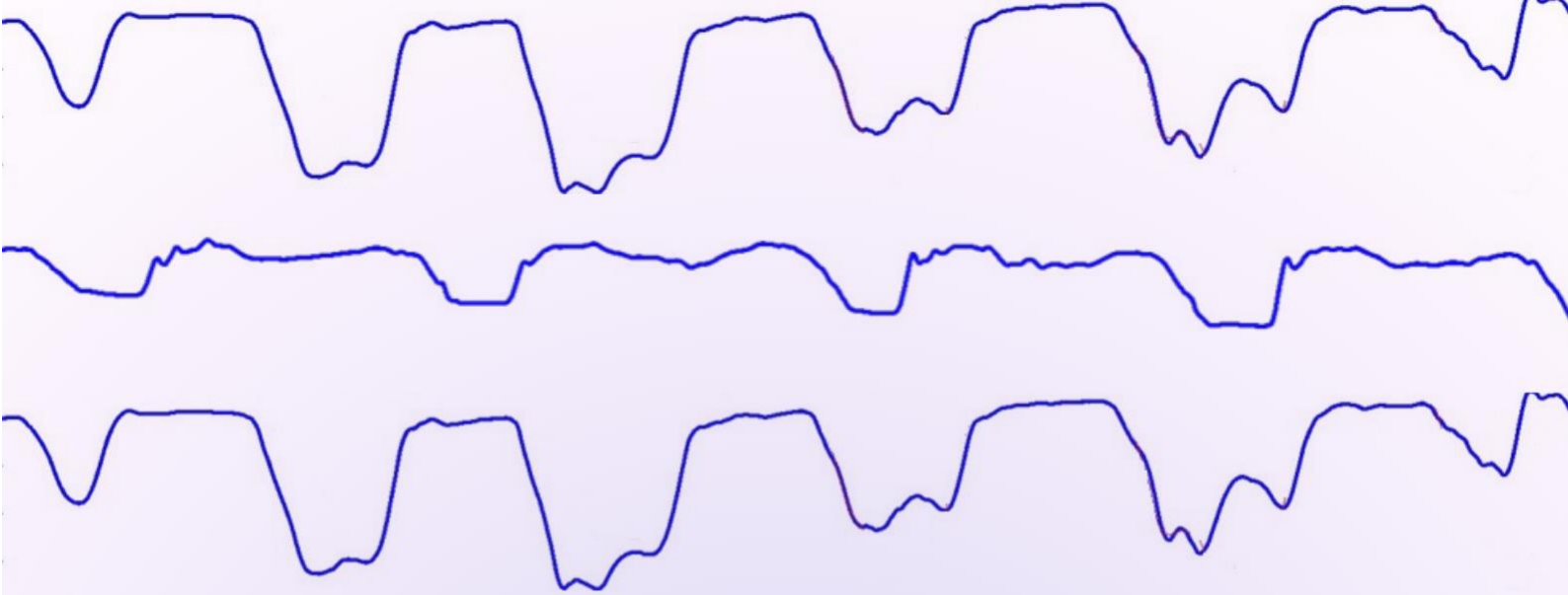
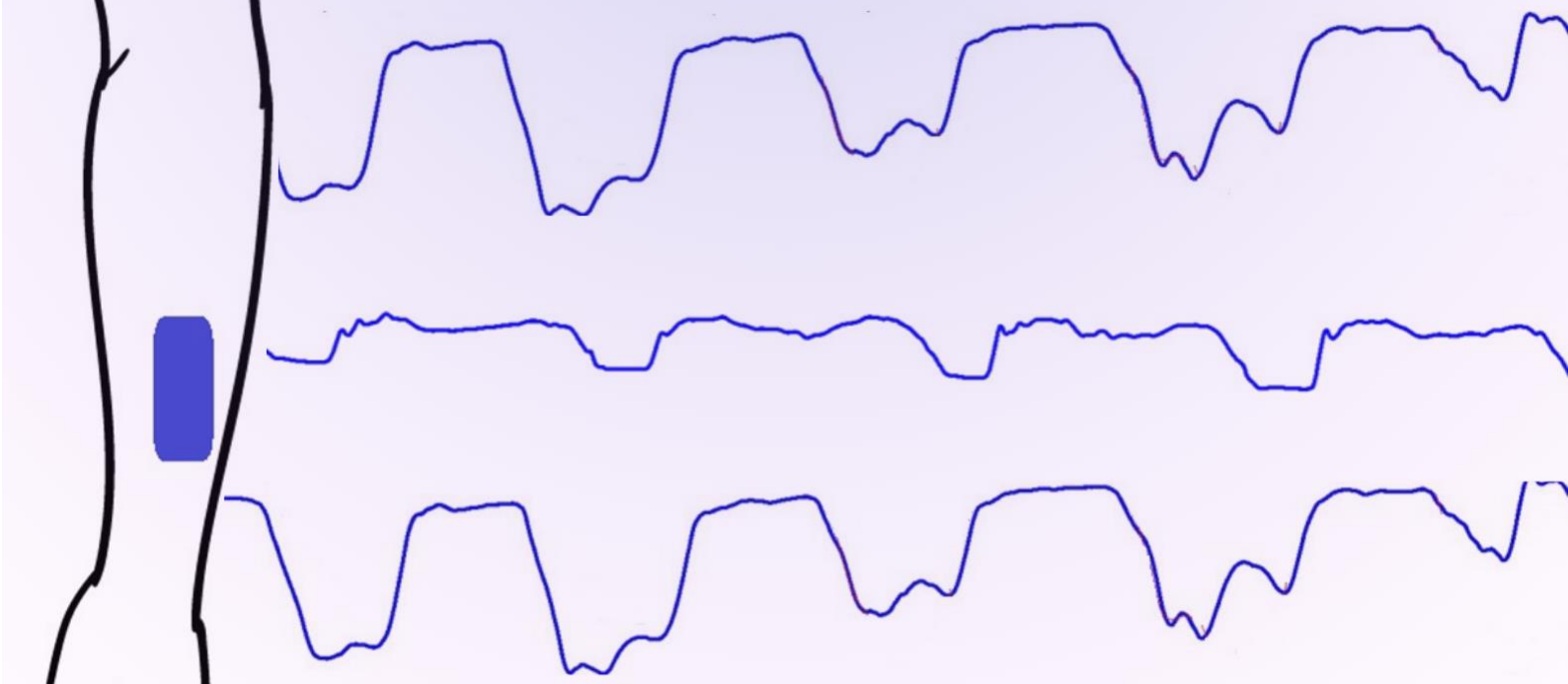


Investigating the Validity and Reliability of Inertial Measurement Units for Post-Stroke Upper Extremity Rehabilitation



Marije Spaaks

MSc. Thesis Technical Medicine

INVESTIGATING THE VALIDITY AND RELIABILITY OF INERTIAL MEASUREMENT UNITS FOR POST-STROKE UPPER EXTREMITY REHABILITATION

- MSc Thesis -

Marije Spaaks

Student number: 4674804

12-08-2024

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University ; Delft University of Technology ; Erasmus University Rotterdam

Master thesis project (TM30004 ; 35 ECTS)

Department of Rehabilitation Medicine, Erasmus MC
05-02-2024 – 26-08-24

Supervisor(s):

Dr. J.B.J. (Hans) Bussmann	Technical supervision
Dr. E. (Erik) Grauwmeijer	Medical supervision
MSc. A.J. (Nienja) Langerak	Daily supervision

Thesis committee members:

Dr. J.B.J. (Hans) Bussmann (Chair)	Erasmus MC
Dr. E. (Erik) Grauwmeijer	Erasmus MC, Rijndam Revalidatie
Dr. A. (Ajay) Seth	TU Delft

This page was intentionally left blank.

Contents

Preface

The completion of this thesis report marks the end of seven years of studying. Eight years ago, I had never even heard of 'Technical Medicine'. Until one day, when I by chance found information on the BSc. Clinical Technology on the website of the University of Twente. A combination of medicine, maths and physics, meaning I did not need to choose between my interests. Later, I found out that the study recently started at the TU Delft as well. There was no time left to visit both universities, so I had to directly apply for the selection procedure on the last day before the deadline. The rest is history.

Now, a few years later and looking back at that moment in high school, I am glad that I took a leap of faith without ever visiting Delft before. The multidisciplinary courses and trips to Leiden, Rotterdam, and even Ho Chi Minh never failed to surprise me. The unique insight that our study gave us into the fascinating world of hospitals and healthcare is unforgettable. Although the exam weeks were intense and the early mornings during all our internships were often tough, it was all worth it. I am grateful for the opportunities I was given and the personal growth I experienced by being forced out of my comfort zone. Now, I am (hopefully) ready for the next chapter.

But first, I would like to express my gratitude to my technical supervisor, Hans, for giving me the opportunity to perform my graduation internship at the Rehabilitation Medicine department of the Erasmus Medical Centre. Thank you for your valuable insights and feedback, and your dedication to the project. Furthermore, I would like to thank my medical supervisor Erik, for providing clinical guidance and showing me Rijndam Rehabilitation and the beautiful, but also emotional moments of post-stroke rehabilitation. Nienja, I am very grateful for your daily guidance and input. Our weekly meetings really helped me to provide structure not only to the project, but also to my thoughts. To Herwin, thank you for your technical input during our frequent meetings with Bert-Jan and Nienja and the feedback on my draft. Bert-Jan, thank you for your work and insights regarding the MATLAB scripts.

In general, I would like to thank all the staff at Rijndam for their hospitality and for providing me an insight into the intriguing world of rehabilitation. Special thanks to Sabine, for taking me under your wing and making me feel part of the team. Not only your clinical expertise, but also the enjoyable lunch walks were valuable to me. In addition, I would like to thank all staff members and PhD's at the Erasmus MC Rehabilitation Department for involving me in your research projects. Samantha, thank you for showing me your interesting research and providing me with the unique experience of interviewing patients within their own vulnerable environments. The motivation and mindset of these patients has been truly inspiring to me.

Last but not least, I want to thank Merel, Jorijn, Emma and Juliette for all 'study' sessions, lunch and coffee breaks. Without your help and mental support, I honestly do not think I would have made it to this point. Our pictionary and pizzanights during exam weeks, but also our rant sessions were really valuable to me and my time in Delft would not have been the same without you. And of course, thanks to my family (especially my parents, Wout, Ali, Nico(line) and Max) for the unconditional support, and always believing in me (and cooking for me) during the final stages of my thesis!

List of abbreviations

Abbreviation	Definition
AC4S	ArmCoach4Stroke
ARAT	Action Research Arm Test
DALYs	Disability adjusted life years
FMA	Fugl-Meyer assessment
ICC	Intraclass correlation coefficient
ICF	International Classification of Functioning, Disability and Health
IMU	Inertial Measurement Unit
IQR	Interquartile range
L	Left
MDC	Minimal detectable change
NIHSS	National Institutes of Health Stroke Scale
R	Right
SD	Standard deviation
SEM	Standard error of measurement
SRRR	Stroke Recovery and Rehabilitation Roundtable
UE	Upper extremity

Summary

Background: Stroke contributes significantly to the global rates of disability and mortality. A widely recognised consequence of a stroke is motor impairment, including upper extremity impairments. Rehabilitation has a significant role in the improvement of upper extremity functioning. However, the field of rehabilitation care is currently confronted with several challenges. Telerehabilitation, which includes wearable motion sensors such as Inertial Measurement Units (IMUs), offers a promising solution to improve rehabilitation care in a home-based environment. Therefore, the ArmCoach4Stroke (AC4S) project, initiated by the Erasmus Medical Centre and Rijndam Rehabilitation Institute, was established to develop and evaluate a wearable system for upper extremity rehabilitation at home for post-stroke patients.

Aim: As part of the AC4S project, the aim of this thesis was to investigate the test-retest reliability and criterion validity of an IMU-sensor based method, compared to an optoelectronic system, for measuring kinematic metrics of the upper extremity in post-stroke patients.

Methods: The data from two protocols were subjected to analysis: a reliability and validity protocol. During both protocols, participants were equipped with IMUs and executed predefined calibration movements, as well as reaching and lifting exercises. For the validity protocol, patients also wore markers of the gold standard marker-based Vicon system. The measurements of the validity protocol were performed in a laboratory setting, whereas the reliability protocol simulated measurements in an ecological environment. The indices of calibration movements and exercises were selected manually. The orientations of the IMU sensor were estimated and joint angle metrics were calculated for the reaching and reach-and-return phase of movements. A statistical analysis of the metrics was performed to investigate test-retest reliability and criterion validity separately. This involved the use of the intraclass correlation coefficients (ICC) and Spearman correlation coefficients, respectively.

Results: Eventually, 18 participants were analysed for the reliability protocol and 7 for the validity protocol. The reliability protocol yielded ICCs indicating both good reliability (values >0.75) and poor reliability (values <0.5) for some metrics. Significant test-retest differences were identified for certain metrics. The Spearman correlation coefficients for the validity protocol indicated strong correlations (values >0.7). However, the p-values were above the level of significance. The results further demonstrated considerable differences in absolute values between IMU and Vicon joint angles across various metrics.

Conclusion: A large variability and poor-moderate reliability were observed for multiple metrics, indicating underlying issues with the measurements or metric calculations. Substantial differences in absolute joint angle values were evident, not only between Vicon and IMU, but also between test-retest sessions or between patients. This indicates suboptimal orientation estimation, angle definitions or other fundamental problems that need further investigation.

1

Introduction

Neurological disorders contribute significantly to global rates of disability and mortality. Among all neurological disorders, stroke accounts for the largest proportion of DALY's (Disability Adjusted Life Years) and deaths (1). Since most patients survive the initial stroke, the long-term effects have the most significant impact on the daily lives of patients and their families. The location and size of the stroke lesion primarily determine the consequences of the stroke (2). A widely recognised functional impairment after stroke is motor impairment, affecting about 80% of the patients. Although spontaneous recovery of function may occur, a substantial number of stroke survivors continues to experience limitations in their upper extremity (UE) function several months after the initial stroke (3, 4).

Rehabilitation has a significant role in improving the upper extremity functioning of patients. Earlier start of rehabilitation can be associated with more improvement of executing activities after stroke (5, 6). Rehabilitation of the upper extremity can be complex, since UE activities require multi-joint coordination and often involve movements of the trunk. Task completion is emphasized, leaving little time for the quality of task performance and evaluation of compensatory movements (7) during rehabilitation (8). Furthermore, rehabilitation facilities are dealing with financial pressure, decreased lengths of stay and staff shortages. Motivational or compliance issues of patients are also a perceived barrier in exercise training. These factors lead to difficulties in providing patients with the necessary intensive practice (5, 9).

An affordable alternative to the provision of supervised outpatient training are home-based rehabilitation methods. Remote rehabilitation strategies are performed in the patient's home and could be self-directed or (remotely) supervised by a therapist, with or without technology. Recent studies describe promising results regarding the effectiveness of home-based rehabilitation interventions. The home-based setting offers contextual learning and directly relevant objectives, which could increase the transfer of practiced skills to real-world activities (10, 11). However, in a systematic review of Fong Mei Toh et al., home-based interventions with self-directed or remote supervision were found to be less effective compared to home-based interventions under direct supervision of a therapist (12). Therapists emphasise the heterogeneity of stroke patients and the necessity of personalising (home-based) rehabilitation care, as well as the importance of maintaining in-person contact. Sufficient communication strategies should be developed for effective interaction with patients and their caregivers (13, 14). Patients themselves report barriers such as fatigue, the absence of assistance during exercises and insufficient knowledge regarding exercise execution (15).

To overcome these barriers and to improve patient adherence to home-based therapy, the possibilities of telerehabilitation are increasingly being investigated. Telerehabilitation may include the use of wearable motion sensors, including accelerometers, gyroscopes and inertial measurement units (IMUs), for purposes of intervention, assessment or monitoring (8). These motion sensors could provide therapists with detailed information regarding upper extremity impairment or activity limitation of patients, to enable more targeted therapy (16). By using these types of Information and Communication Technologies (ICT), rehabilitation services could be provided remotely, without continuous supervision of therapists (17). Hence, the use of telerehabilitation can not only reduce travel time (for both patients and therapists) and

improve care for patients facing reduced mobility or residing in rural areas, but can also increase continuity and personalisation of care by adjusting treatment based on a patient's activity patterns. Consequently, it may enable a single therapist to assist a larger number of patients within the same time frame (18).

Building on the aforementioned potential of telerehabilitation, the ArmCoach4Stroke (AC4S) project was initiated by the Erasmus Medical Centre and Rijndam Rehabilitation Institute. The project aims to develop and evaluate an IMU-based wearable system for upper extremity rehabilitation at home (19), to improve upper extremity exercise of post-stroke patients. A suitable body-fixed hardware configuration as well as analysis software to calculate performance metrics are being investigated. Based on these calculated metrics, personalized and real-time feedback regarding the execution of upper extremity movements will be provided to the patient through an app and a tablet. The information will also be accessible to their therapists.

When developing such new measurement instruments, assessment of its clinimetric properties is crucial. Firstly, it is important to evaluate validity: determining if the developed instrument measures what it intends to measure. Criterion validity, in particular, examines the correspondence of measurements with the current gold standard. This enables the determination of the instrument's validity for its intended clinical purpose. Another essential requirement of measurement instruments is reliability (20). If patients repeatedly perform the same activity under identical conditions, the instrument should yield similar results over time. This property, known as test-retest reliability, is fundamental to the provision of insights into the exercise execution of patients.

As part of this graduation project, data analysis within the AC4S-project was performed, regarding the test-retest reliability and the criterion validity of the IMUs and an optoelectronic measurement system. This analysis aims to answer the following principal research question: What is the test-retest reliability and criterion validity of an IMU-sensor based method, compared to an optoelectronic system, for measuring kinematic metrics of the upper extremity during predefined exercises in post-stroke patients?

2

Background

For a deeper understanding of the concepts of stroke impairments and assessment, validity and reliability, IMUs and the optoelectronic Vicon system, further information regarding these topics is provided in this chapter to serve as a theoretical research background.

2.1 Stroke and upper limb impairments

Stroke is often described as a neurological disorder during which oxygen supply to the brain is disrupted by occlusion (ischaemic stroke) or haemorrhage (haemorrhagic stroke) of blood vessels (Figure 1) (21). The majority of strokes, about 85%, are classified as ischemic strokes, typically resulting from blood clots due to atherosclerosis or an embolism. In the case of a haemorrhagic stroke, stress on the brain tissue (such as hypertension) and internal injury can cause rupture of the blood vessels (22). The impaired perfusion or interruption of blood supply leads to pathological events such as neuroinflammation and oxidative stress, which can eventually result in necrosis of brain cells (23). The extend of neuronal death is a key determinant of morbidity and mortality among stroke patients.

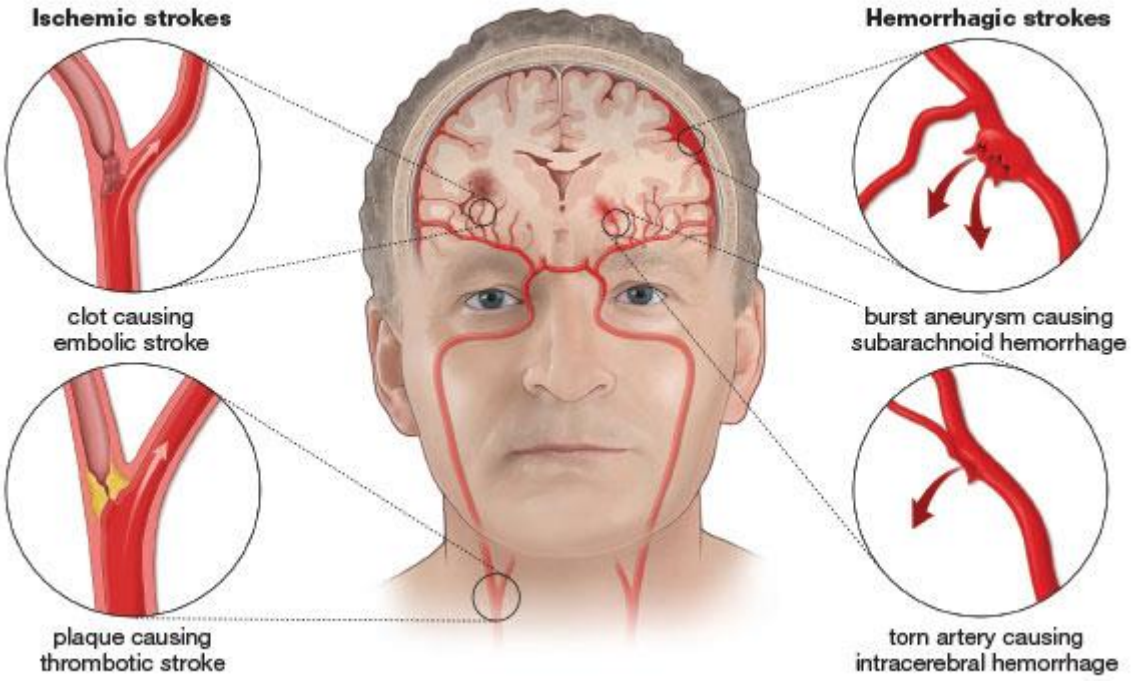


Figure 1: This figure shows the different mechanisms behind the two types of strokes: an ischemic and haemorrhagic stroke. Retrieved from(24).

According to the framework of the International Classification of Functioning, Disability and Health (ICF), a stroke might influence functioning and disability across multiple domains (Figure 2) (25). For example, a stroke might cause abnormalities in body structures, such as the nervous system, or impairments of body functions, resulting in loss of mobility, muscle strength, or muscle tone. These functional impairments can impact a patients' performance at an activity level or on the level of societal participation. Multiple types of impairments can be present simultaneously and the type and severity of these impairments may change over time. Weakness or paralysis of the upper extremities is common after stroke, due to disrupted signal transmission from the brain's motor cortex. Muscle weakness can result in physical disability (26) and may eventually lead to muscle fibrosis, abnormal posturing, pain, spasticity and decreased function.

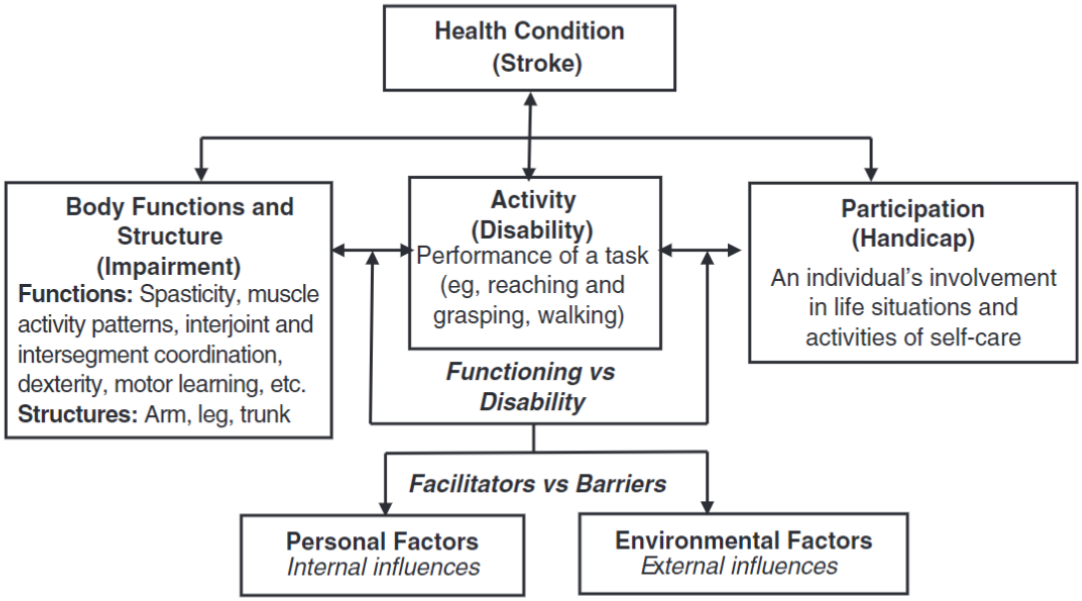


Figure 2: An overview of the International Classification of Functioning Model, showing the effects of stroke on the level of body functions and structure, activity and participation. Retrieved from:(27).

As previously mentioned, earlier start of rehabilitation can be associated with more functional improvements, which could eventually lead to improved performance at an activity or participatory level. The primary aim of recovery therapy after stroke is to induce brain plasticity. However, during the rehabilitation phase, some patients may develop compensatory movements, including the use of trunk movements to assist during arm and hand movements (28). Over time, these compensatory movements can lead to negative functional outcomes, such as learned non-use or misuse of the upper extremity, eventually limiting motor function recovery (26). Research suggests that in mild-to-moderate impaired patients, appropriate interventions may facilitate functional improvement of the upper extremity even in the chronic stage of stroke (29). This recovery of motor functions could be achieved by activity-based therapies, where home-based telerehabilitation shows promising results (30). These types of interventions may not only provide a tool for home-based exercise, but may also help differentiate between compensation and recovery.

2.2 Clinical assessment scales

Nowadays, severity of upper extremity motor impairments is often assessed using clinical assessment scales: the Action Research Arm Test (ARAT) and the Fugl-Meyer Assessment (FMA). For both assessments, an experienced physician is required. The ARAT can be described as a tool to (quickly) assess limitations in upper-extremity activities for patients with

cortical damage (31, 32). Patients perform various actions and are scored on 19 items, with a score of 0 indicating no movement and 3 indicating normal movement. Normal performance corresponds with the maximum score of 57 points across all subscales. The included subscales are: gross motor (9 points), grasp (18 points), grip (12 points) and pinch (18 points) movements.

The FMA is a quantitative measure of motor impairment for stroke patients, suggested for assessment of impairments on a body function and structure level. The ordinal scale has been designed as an objective index to assess recovery of hemiplegic stroke patients (33). The scale includes items measuring movement, coordination and reflex action. The maximum score for the upper extremity is 66 points (33).

Furthermore, the National Institutes of Health Stroke Scale (NIHSS) can be used to describe the severity of a patient's stroke. This numerical scale allows caregivers to measure neurological function and deficits by having patients answer questions and performing physical and mental tasks across 11 categories. Higher scores reflect more severe impairments, with scores <5 described as mild impairments (34). The NIHSS, ARAT and FMA measures are also recommended by the Stroke Recovery and Rehabilitation Roundtable (SRRR) to be included in every stroke recovery trial (35).

2.3 Validity and reliability

While developing a measurement instrument for clinical applications, assessment of its validity and reliability is important.

2.3.1 Validity

Validity can be defined as "the degree to which an instrument truly measures the construct(s) it purports to measure" (20, 36). Several types of validity could be described, among which content and criterion validity seem most relevant for this study. Content validity describes the degree to which the instrument adequately reflects the measured concept, including face validity: if measurements appear as though they are an adequate reflection of the content. Criterion validity focuses on the correspondence of measurements of the instrument of interest with the current gold standard. For continuous measurements using the same units, Bland and Altman limits or intraclass correlation coefficients (ICCs) are often used to assess validity(20). If the gold standard does not use the same units, Spearman (non-parametric) or Pearson's (parametric) correlation coefficients can be used. Measurement instruments should be validated again every time they are applied in a new setting or within a different population. Therefore, the developed AC4S wearables should be validated for use within a stroke rehabilitation setting.

2.3.2 Reliability

Additionally, reliability is an essential aspect of clinical measurements. When repeated measurements are conducted under identical conditions by the same subject, the measurement instrument should yield similar results. Reliability is defined as "the degree to which the measurement is free from measurement error" (37). Since these measurement errors can affect statistical analysis and interpretation of results (38), the reliability should be assessed. The score and error of a measurement can be described according to formula 1:

$$\Upsilon = \eta + \varepsilon \quad (1)$$

Within this formula, Υ represents the observed score, η represents the true score and ε represents the measurement error. Reliability parameters range from values of 0 (unreliable) to 1 (perfect reliability). A parameter of measurement error is the standard error of

measurement (SEM). Variation between measurements could be assessed by the standard deviation (SD) or interquartile range (IQR). Perfect reliability is approached if the measurement errors (SEM) are small compared to the variability within patients(20).

Multiple types of reliability exist, such as internal consistency, inter-rater, intra-rater or test-retest reliability. The types of reliability can be influenced by several sources of variation within measurements: the therapists performing the measurement, the patients undergoing the measurements, the measurement instrument or measurement circumstances (39). The focus of the ArmCoach4Stroke project was on test-retest reliability. To describe the reliability of continuous variables, ICCs are often used (38).

2. 4 Inertial Measurement Units

In this study, the validity and reliability of IMUs within a rehabilitation context were evaluated. The use of wearable sensors, such as IMUs, has been increasingly explored for upper extremity rehabilitation in the past few years (40). IMUs are easily portable and low-cost sensors, which do not need a laboratory setting to perform measurements. The systems can contain an accelerometer, gyroscope and magnetometer. Most commonly used are the accelerometer and gyroscope data, since the magnetometer measurements of the magnetic field can be disturbed by environmental influences (41). Accelerometer and gyroscope measurements combine linear changes in velocity and the rate of rotation or angular velocity in three axes (x, y and z) respectively. Accelerometers can rely on several mechanisms, for example the measurement of the displacement of a mass suspended by springs by measuring a signal proportional to the force acting on the mass (caused by acceleration or gravity). Traditional mechanical gyroscopes consist of spinning discs. While spinning, the orientation of the discs remains the same due to angular momentum, unless an external force is applied. This external force causes a shift ("recession") proportional to the rate of rotation, thus the angular velocity can be determined (42, 43).

Several kinematic parameters can be calculated using IMUs, such as duration, distance, joint angles, velocity or smoothness (44). Velocity and position can be computed by (double) integrating the acceleration with respect to time. However, measurement inaccuracies (such as noise) can be accumulated over time during integration. Such signal drift is often present in gyroscopes. These errors in gyroscope data can lead to linear errors in orientation estimation (45). To avoid drift and track the motion, orientation and position of an object in a three-dimensional space, the accelerometer and gyroscope data can be combined. Therefore, sensor fusion algorithms, such as a Kalman filter or a Madgwick filter can be applied (46).

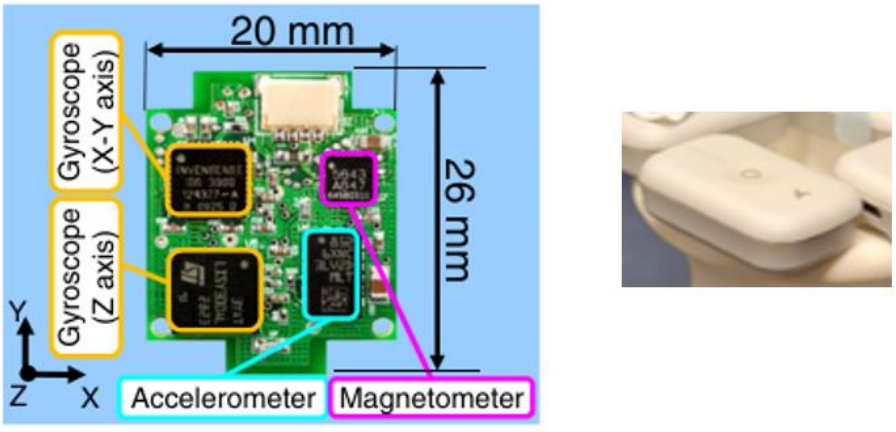


Figure 3: An example of an IMU with an accelerometer, gyroscope and magnetometer (left, from: (46)) and an example of the IMUs as used within the ArmCoach4Stroke project (right).

2.5 IMU data processing and analysis

To accurately measure and analyse IMU data, several steps have to be performed. The methodology of the IMU data analysis of the ArmCoach4Stroke study is based on a study of Bhagubai et al. (47). They conducted a pilot study in which an inertial sensing system to measure UE kinematics was designed and tested. The sensors were aligned to the respective limb segments based on data acquired during predefined static postures or dynamic movements. Measurements of the gravity vector of the accelerometer during these predefined postures represent an axis of the anatomical frame of a limb segment. A rotation matrix with three vectors (x, y and z) defines the orientation of the segment's coordinate frame relative to the sensor's coordinate frame. They also defined a common global frame to be able to relate the orientations of segments to each other. The common vertical axis is defined by measurement of the gravity vector in all sensors during the static pose. The horizontal axis is derived from the measured angular velocity during the dynamic movement. After definition of the global frame, movements can be reconstructed and eventually, joint angles can be calculated. Accurate estimation of onset and termination of movements is complex, since real movements usually show periods of low (but non-zero) velocity and acceleration signals before and after the movements. Therefore, predefined thresholds of 5% or 10% of the peak velocity are sometimes described to indicate start and end of movements (48, 49).

2.6 Vicon motion capture system

Currently, measurements of the upper extremity functions are frequently captured with 3D marker-based, optoelectronic motion capture systems, such as Vicon. System markers are placed on anatomical landmarks and infrared cameras record the marker trajectories based on emitted infrared light (50). Standardized software models, such as the Vicon Plug-In Gait Upper Body model (51), can be used to assign markers to the corresponding anatomical landmarks and calculate kinematic outputs. An example of the Plug-In Gait Upper Body model is shown in Figure 4. Elbow angles are relative angles calculated (using Euler angles, YXZ convention) between the upper arm and forearm segments. Shoulder angles are calculated between the upper arm and thorax. The markers placed on the torso define the thorax. Thorax angles are calculated between the thorax and the laboratory coordinate system.

Optoelectronic systems such as Vicon have been found to be valid and reliable and have high spatial and temporal sensitivity (52). However, they are technically complex, expensive and require a large (controlled) environment. Therefore, these systems are less suitable for in clinic or home-based measurements (53). Despite these limitations, optoelectrical motion capture systems such as Vicon are considered the gold standard movement measurement method in studies investigating other motion capture systems (52, 53). During these studies, metrics such as movement time, spectral ark length or peak velocity of the measurement system of interest (such as IMU's) are then compared to the outcomes of the gold standard method (53).

The following image shows the front view. The left upper body markers are not labeled; attach markers on that side in a similar way to those on the right (with some asymmetry as described above).

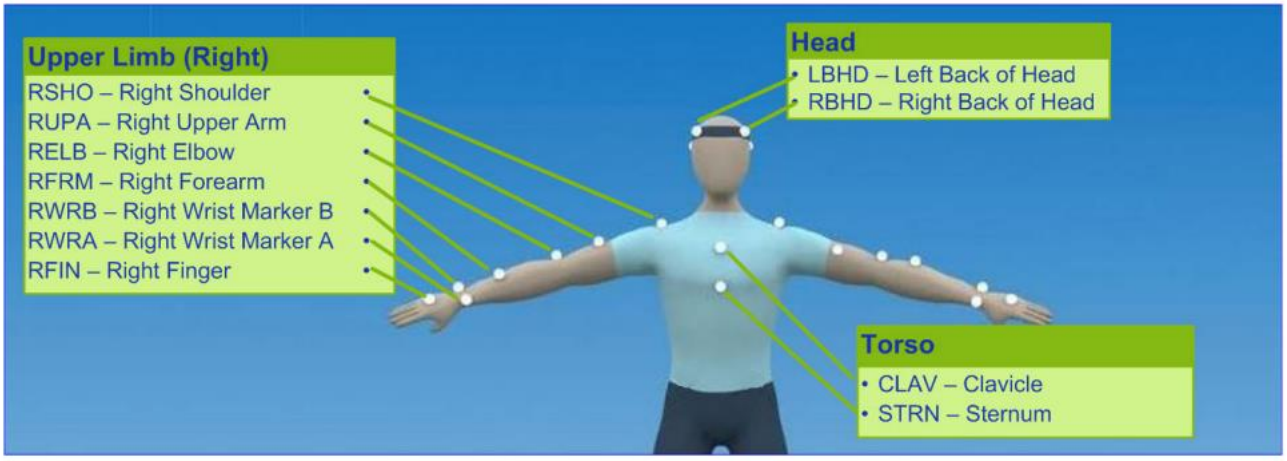


Figure 4: A snapshot of the Plug-In Gait Upper Body model used for Vicon-analysis, showing the placement of markers on upper extremity locations (51).

3

Methods

The following section describes the methods of the two protocols: the reliability protocol and the validity protocol. The primary aim of the reliability protocol was to investigate the test-retest reliability. The primary aim of the validity protocol was to investigate the criterion validity by comparing IMUs to the gold standard. Despite the different objectives, these two protocols show great similarity. For example, the subject inclusion criteria were identical for both protocols. Furthermore, the exercises of the reliability protocol were also included in the validity protocol, with the addition of two other exercises. Both protocols used similar methods for calibration and trial indices selection. For the calculation of kinematic metrics and statistical analysis of both protocols, comparable data processing methods were used. Nevertheless, notable differences between the two protocols exist. The protocols were conducted at different locations. The number and placement of IMU sensors also differed. Additionally, participants of the validity protocol also wore Vicon markers. The validity protocol could also be divided into two parts: Part A and Part B. The following sections will provide a more detailed examination of these aspects.

3.1 Subjects

Adult patients in the subacute (>1 week, <6 months post-stroke) or chronic phase (>6 months post-stroke) following an ischemic or haemorrhagic unilateral stroke and with reduced one-sided upper limb motor function (Brunnstrom stage ≥ 3) were included. The participants had to be able to engage in upper extremity exercise sessions lasting at least 30 minutes (with rest). Furthermore, sufficient trunk control to sit on a chair without side support for 30 minutes was required. Patients with major upper limb complications or comorbidities (such as frozen shoulder or upper limb pain), comprehensive aphasia or cognitive impairments were excluded. The study participants were recruited from the outpatient or inpatient clinics of Roessingh Centre for Rehabilitation in Enschede and Rijndam Rehabilitation Institute in Rotterdam.

3.2 Study Materials

Inertial Measurement Units produced by 2M Engineering (Valkenswaard, The Netherlands) were attached to the body using (medical) tape during both protocols. Also, the optoelectronic Vicon motion capture system (© Vicon Motion Systems Ltd UK) and Vicon Nexus software (Plug-In Gait analysis Upper Body Model pipeline) were used as part of the validity protocol. The sampling frequency of the IMUs was 50 Hz. The Vicon system makes use of eight high-speed infrared cameras, with a sampling frequency of 100 Hz. Furthermore, two commercially available Rollei video cameras were used to record the participants' upper bodies during the exercise performance. Additionally, an object with a cylindrical shape was 3D-printed to be used during some of the exercises.

3.3 Data collection

Data was collected according to the two study protocols: a reliability and validity protocol (Figure 5). Measurements of the validity protocol were performed in a laboratory setting at Roessingh Centre for Rehabilitation in Enschede. Measurements of the reliability protocol mainly took place at Rijndam Rehabilitation Institute in Rotterdam, simulating an ecological environment in which patients had to apply the sensors themselves. Baseline characteristics of participants were collected and clinical assessments (FMA, ARAT) were performed on the first day of the measurements.

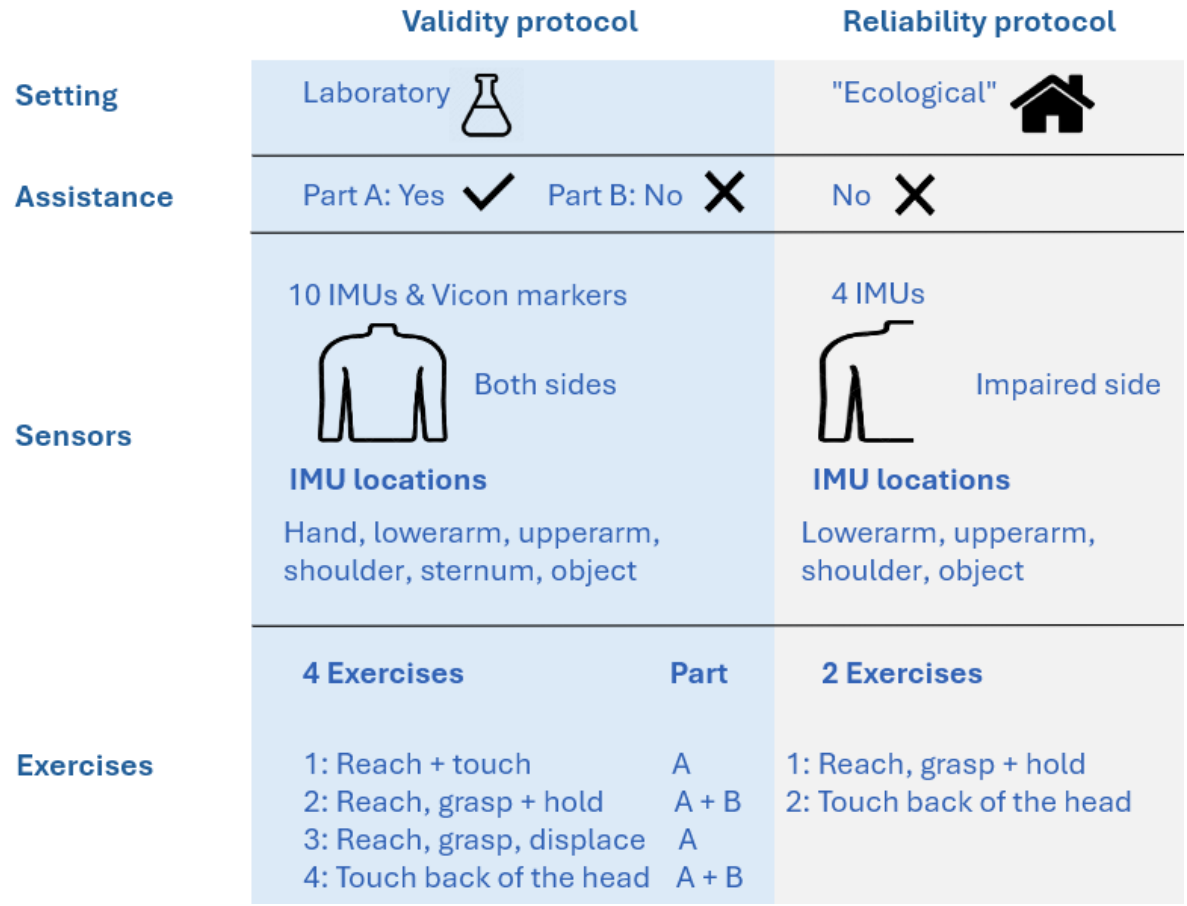


Figure 5: A side-by-side comparison of the validity (Part A + B) and reliability protocol, indicating differences in setting, provided assistance, sensors worn and executed exercises.

For both protocols, IMUs were attached to the upper extremities using medical tape to measure movements during several exercises. The Vicon Nexus software (Plug-in Gait analysis Upper Body Model pipeline) and custom-made MATLAB scripts were used for data analysis.

3.3.1 Reliability protocol

The reliability protocol simulated an ecological setting. Three IMUs were attached to the impaired upper extremity by the patients themselves. One IMU was placed inside the cylindrical object. The exact placement of the IMUs on the upper extremity is described in Table 1 and shown in Figure 6.

Table 1: Within this table, the exact placement of each IMU sensor location of each protocol is described, as well as the side of attachment: on the impaired side, Left (L) + Right (R) or not applicable (N.A.).

Sensor location	**Exact placement**	**Part of protocol**		**Attachment side**	
		Reliability	Validity	Reliability	Validity
Hand	Dorsal side	✗	✓	-	L + R
Lower arm	Near the wrist, dorsal side	✓	✓	Impaired	L + R
Upper arm	Near the elbow, lateral side	✓	✓	Impaired	L + R
Shoulder	Near the acromioclavicular joint and the superior border of the scapula and clavicle	✓	✓	Impaired	L + R
Sternum	At the centre of the chest	✗	✓	N. A.	N. A.
Object	Within the 3D-printed object	✓	✓	N. A.	N. A.

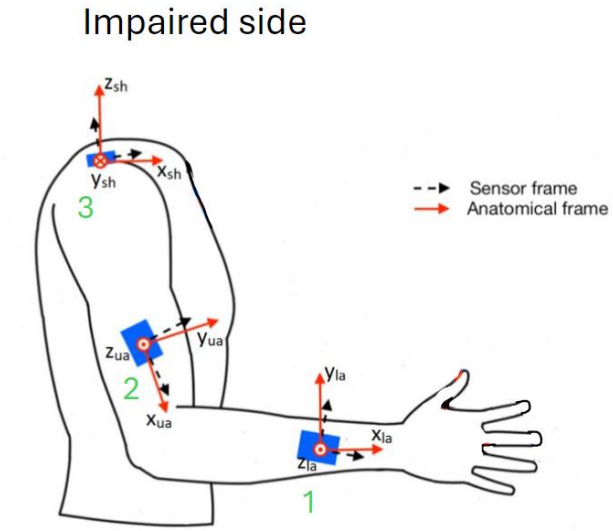


Figure 6: Example of the positioning of IMU-sensors 1-3 (green numbers) for the impaired upper extremity during the reliability protocol, adjusted from the study of Bhagubai et al. (47). The corresponding sensor frames and anatomical frames (X-, Y- and Z-axis) are also shown.

Patients were instructed about the execution of the calibration movements and exercises. They did not receive any assistance or feedback during the application of the IMUs and performance of the calibration exercises. For this protocol, patients began this pre-exercise calibration in a seated position, with both arms on the table (palms facing downward) for five seconds. Then, they performed a supination-pronation movement of the hand, directly followed by a trunk flexion movement (Figure 7), finished by five more seconds in the initial position. After performance of these calibration movements, participants relocated their hand to the designated starting position. Then, they performed two exercises in a randomized order: the 'lift' and 'head' exercise. During the 'lift exercise', patients lifted and held the object, which was positioned at a distance of 45 cm from the starting position. During the 'head exercise', the participants touched the back of their heads. Five trials of each exercise were executed, during

a test and a retest session. There was a 5-10 minute break between the two sessions, during which the sensors were taken off. The sensor were reapplied by the patients before start of the retest session. Detailed descriptions of the executed exercises of both protocols are provided in Table 2. Prior to the commencement of an exercise, the object was shaken by the researcher in order to facilitate the identification of the exercises during the subsequent data processing phase.

Table 2: The names and descriptions of the performed exercises of both protocols are provided, and for each exercise the applicable protocol and exercise number are defined

Exercise name	**Exercise description**	**Part of protocol**	
		Reliability	Validity
Touch exercise	Reach for and briefly touch the object	✗	✓
Lift exercise	Reach for and grasp the object, lift it a few centimetres for 3 seconds, place it back and return to the initial position	✓	✓
Head exercise	Touch the back of the head with the palm of the hand for 1 second and return to the initial position	✓	✓

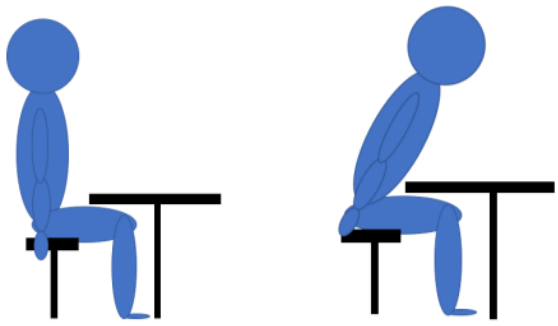


Figure 7: Schematic overview of the beginning and end position of the calibration movement (left) and performed trunk flexion (right). Retrieved from: (54).

3.3.2 Validity protocol

The validity protocol consisted of a part A and part B. During part A, IMU-sensors and Vicon-markers were attached by the researcher. The placement of the IMU-sensors on both arms is described in Table 1 and shown in Figure 8. The placement of the Vicon-markers is shown in Figure 4. After attachment of the sensors, participants were instructed on the performance of the exercises. First, they performed a pre-exercise calibration movement needed for sensor-to-segment calibration, followed by four upper extremity exercises. For each exercise, five trials were executed during a test and a retest session. As a calibration movement, participants had to sit straight with their arms besides their body, perform a trunk flexion and then sit straight again (Figure 7). The researcher could correct the patients to make sure the calibration movements and exercises were performed correctly. In between the test and retest session, the sensors were taken off and participants had a five minute break before reapplication of the sensors by the researcher. Then, calibration movements were performed again, before five repetitions of each arm exercise. Due to the observed difficulty and variability in execution of the third exercise, only three exercises were included for further analysis. The included exercises are described in Table 2.

During part B, the participants needed to apply and remove the IMU-sensors themselves, similar to the reliability protocol. They did not receive any feedback during the performance of the calibration movements. This way, measurements performed in a home-based situation were simulated. Also, participants were asked to perform only two exercises (the lift and head exercises) during this part of the study protocol. Part B will not be included in the data analysis of the validity protocol, but was considered and analysed as part of the reliability protocol. New patient numbers for use in the reliability protocol analysis were created by adding '20' to the patient numbers from validity protocol part B.

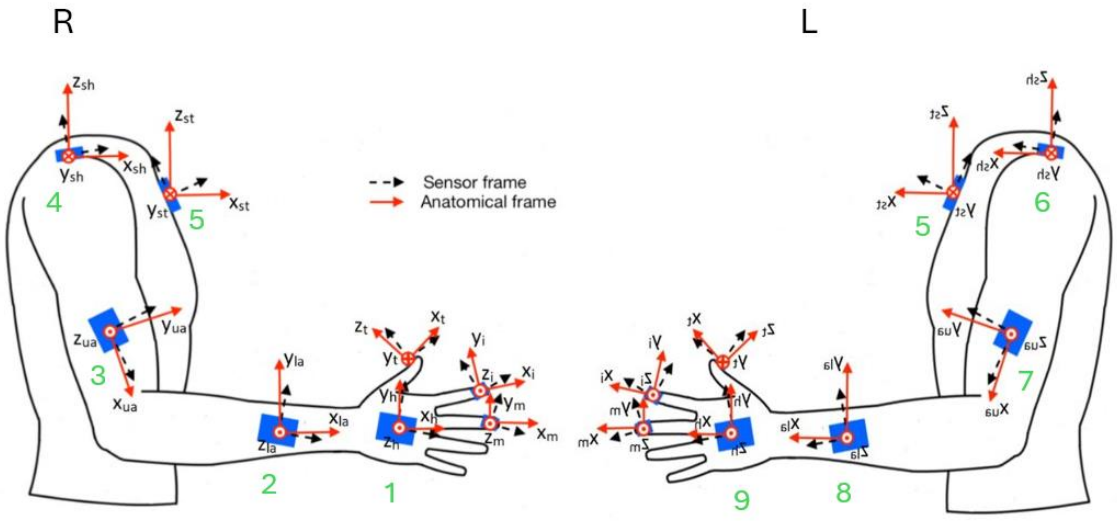


Figure 8: Example of the position of IMU-sensors 1-9 (green numbers) for the left (L) and right (R) arm during the validity protocol, adjusted from the study of Bhagubai et al(47). The corresponding sensor frames and anatomical frames (X-, Y- and Z-axis) are also shown.

3.4 Data (pre)processing

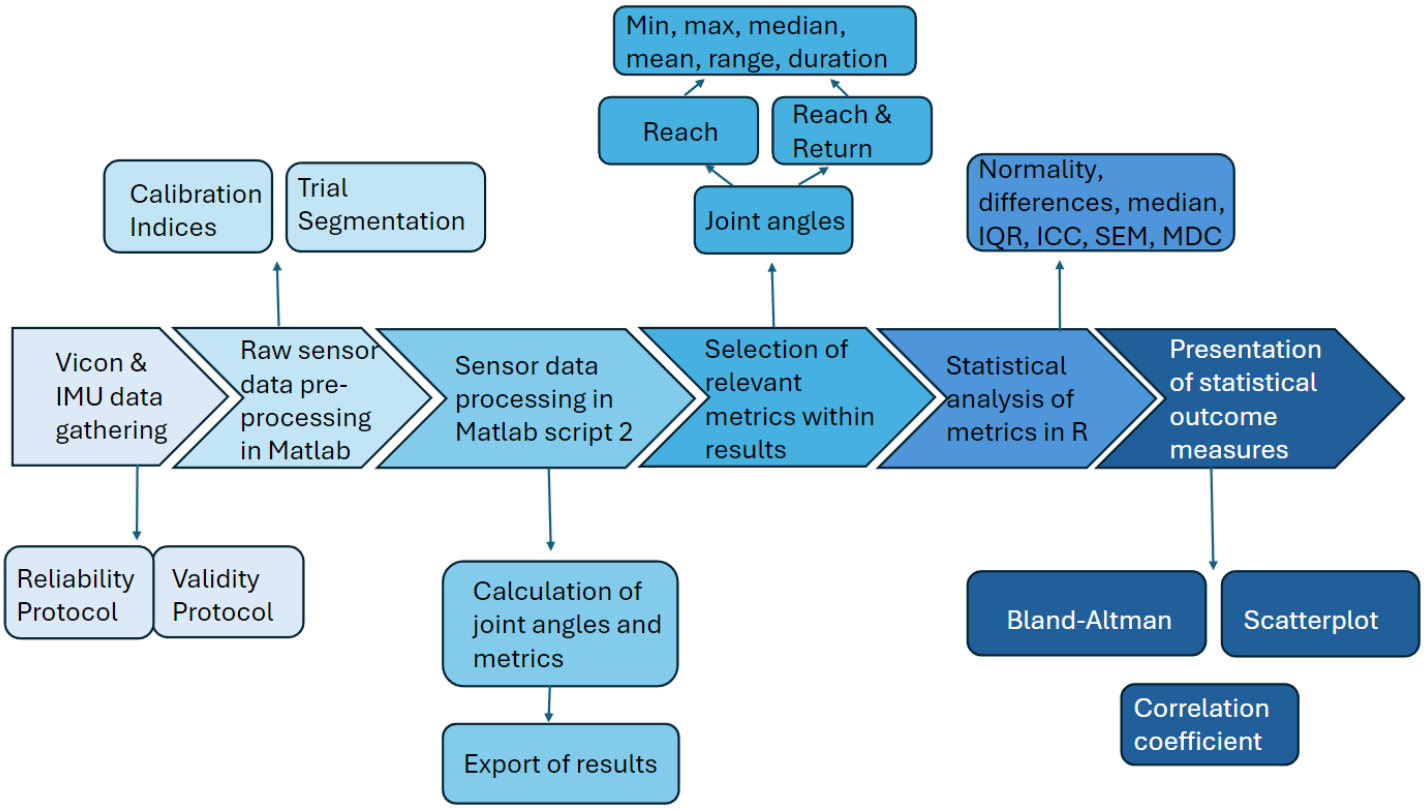


Figure 9: This figure shows a schematic overview of the different stages of data gathering, preprocessing, metric calculation and selection and statistical analysis. The arrows indicate relevant aspects of each stage. Abbreviations: Min = minimum, max = maximum, IQR = interquartile range, ICC = intraclass correlation, SEM = standard error of measurement, MDC = minimal detectable change.

A schematic overview of the data collection, pre(processing) and analysis is shown in Figure 9. The raw IMU data from both protocols was analysed using custom-made Matlab scripts (version 2023B, Mathworks, Inc., USA). Two types of scripts were developed in cooperation with the University of Twente: one preprocessing script for determination of indices and one for calculation of joint angle metrics. The IMU data of the reliability protocol was provided as .parquet files containing all three IMU sensors, whereas the validity protocol IMU data was provided as separate .txt files for each of the ten individual sensors. Thus, the data loading method of the Matlab scripts was adjusted according to the protocol. For preprocessing of the Vicon marker data, the Vicon Nexus software (Plug-in Gait analysis Upper Body Model pipeline) was used. Vicon data was available as .C3D files.

For each patient, IMU data was matched to the corresponding sensor location. Additionally, the Euclidean norm of the accelerometer and gyroscope data was calculated prior to selection of indices. Figures of raw IMU sensor data were inspected and face validity was assessed before proceeding to selection of calibration and trial indices. Video recordings of patients were reviewed to verify the selection of calibration as well as trial indices.

3.4.1 Synchronicity analysis

Unexpected asynchrony between individual IMU-sensors was observed during the first preprocessing steps and face validity checks of sensor data of the reliability protocol. Therefore, multiple synchronicity tests were conducted to investigate the occurrence of asynchrony and identify potential influencing factors. Visual inspection of data of multiple experimental

sessions revealed that signal asynchrony was present, even when the sensors were attached to the object and moved simultaneously. Due to the found inconsistency in asynchrony during patient trials and between sessions, and the absence of any phase within the measurements where the sensors were known to move simultaneously, it was concluded that no asynchrony correction could be applied before further processing of the reliability protocol data. A more detailed description of these synchronicity tests and results can be found in Appendix B.

3.4.2 Calibration indices selection

Based on the results of the synchronicity tests (Appendix B), no IMU synchronization could be performed for the reliability protocol and analysis was continued. To perform the sensor-to-segment calibration and define the global frame, the following indices for every individual IMU had to be identified within the norm of the gyroscope signal of both protocols (Figure 10 & Appendix A):

1. Start of the static phase
2. End of the static phase
3. Start of the calibration movement
4. End of the calibration movement
5. Return to the static pose

Static phases were defined as phases during which the gyroscopic signal was below a predefined threshold (<0.5 ◦/s). During the selection of calibration and trial indices, data had to be inspected for missing values. In the case of missing data within the calibration or trial sections, patients or trials had to be excluded from further analysis.

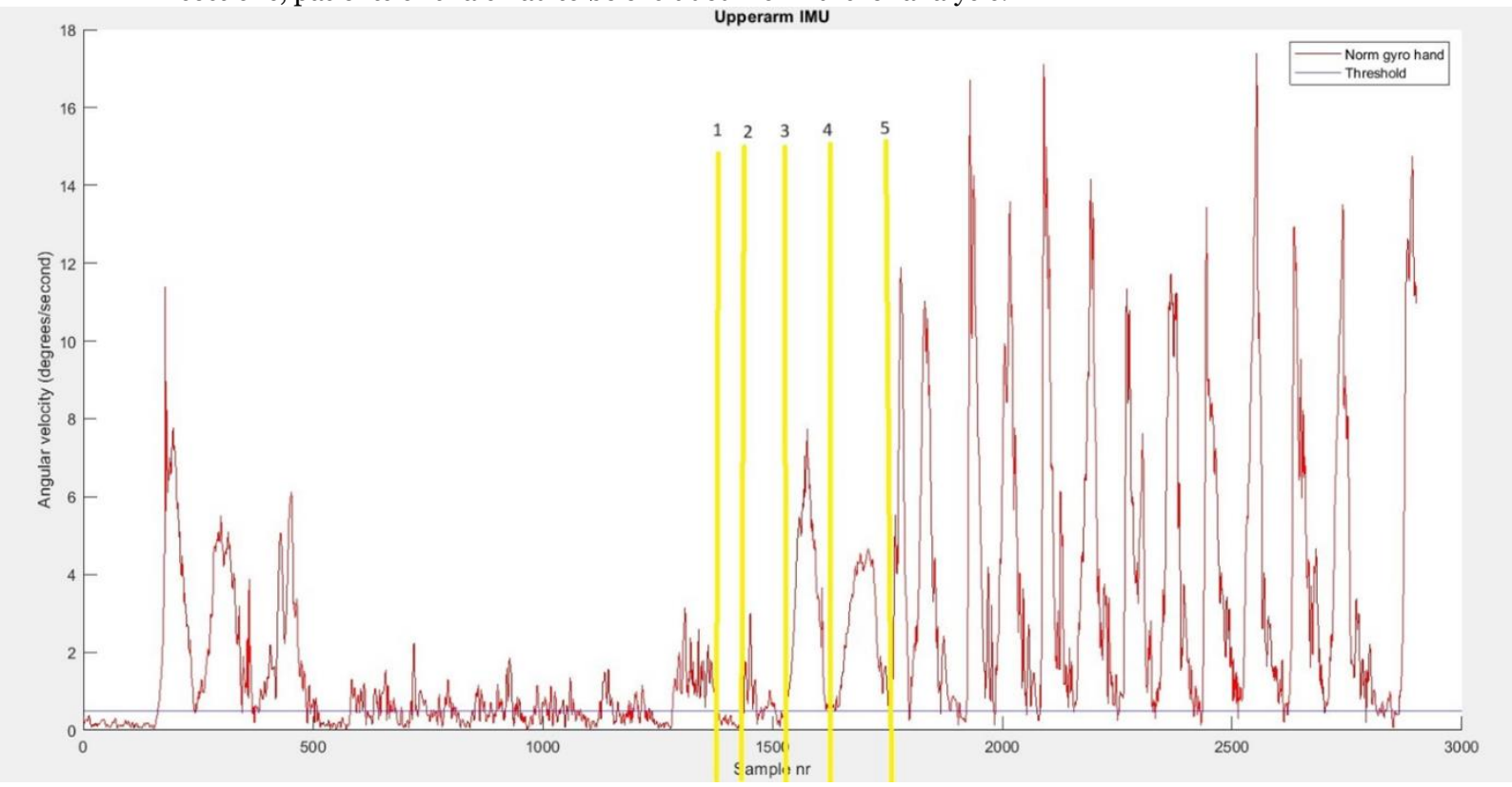


Figure 10: Example of the selection of five calibration indices (numbers 1-5) using Matlab in the norm gyroscope signal for the upperarm IMU during the trunk flexion calibration movement (validity protocol): 1 = start of static phase, 2 = end of static phase, 3 = start of trunk flexion, 4 = end of trunk flexion, 5 = return to initial position.

3.4.3 Trial indices selection

After selection of the calibration indices, also indices of specific time points within the exercise trials had to be manually selected based on sensor morphology. Before selection of these trial indices, a second-order low-pass Butterworth filter with a cutoff frequency of 8 Hz was applied to the norm of the accelerometer data of the lower arm (and IMU object data for the exercises using an object). Then, the following timepoints were selected:

1. Start of the reaching movement
2. End of the reaching movement
3. Return of the hand to the starting position after finishing the movement

Trial segmentation was performed based on IMU accelerometer data for the reliability protocol (Figure 11) and based on Vicon position data for the validity protocol (Figure 12), since Vicon data showed more discernible patterns and five distinctive peaks across the five trials. When applicable, the IMU object accelerometer signal was reviewed to assist in the selection of trial indices. An example of IMU trial segmentation of the head exercise is shown in Appendix A.

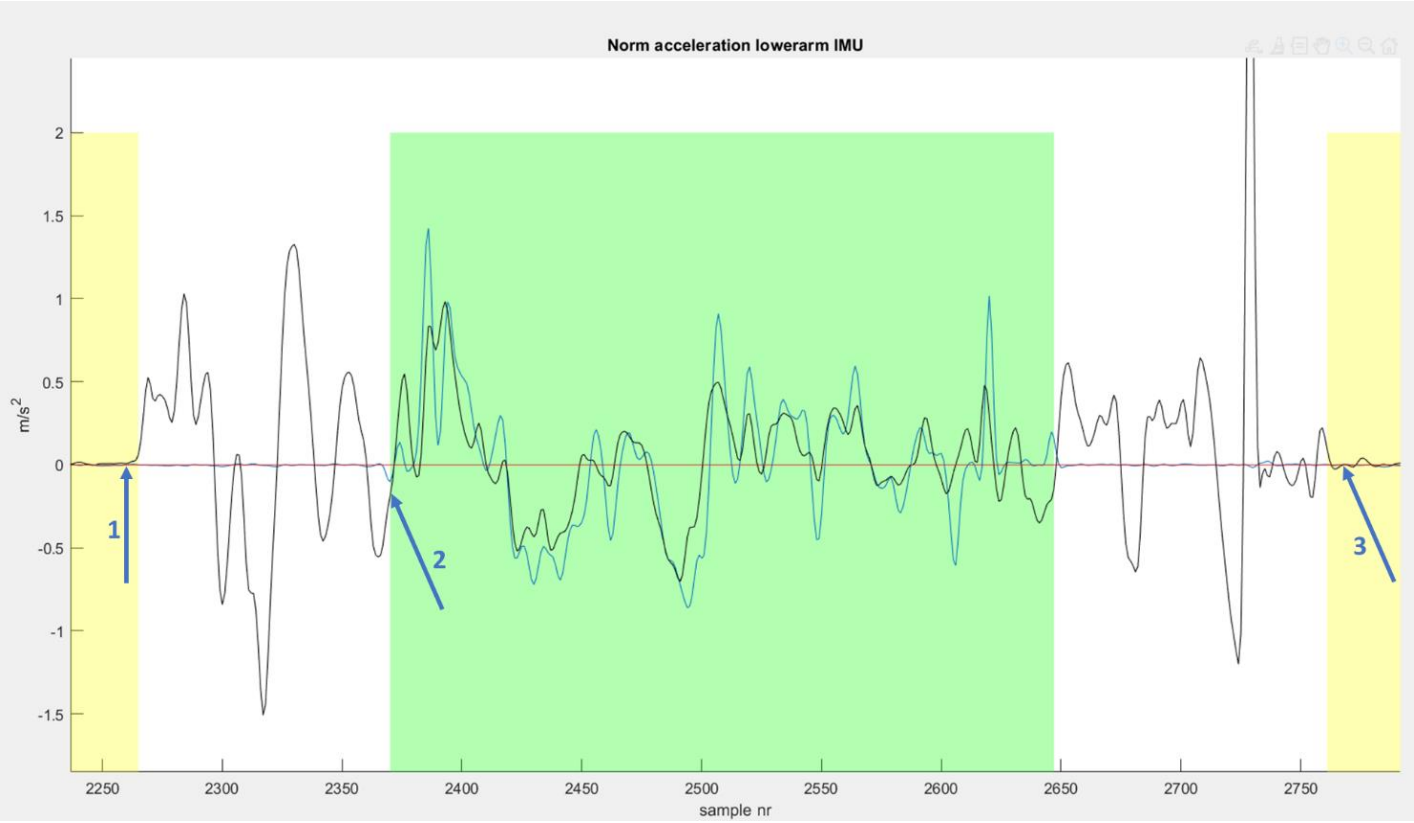


Figure 11: Example of IMU norm accelerometer data during the lift exercise of the reliability protocol. The selection of trial indices using Matlab is shown with numbers 1-3 corresponding to: 1 = start of reach, 2 = end of reach, 3 = return to initial position. The black line corresponds to the lower arm IMU. The green area indicates movement of the object (blue line) after the reaching phase. The yellow area indicates minimal movement.

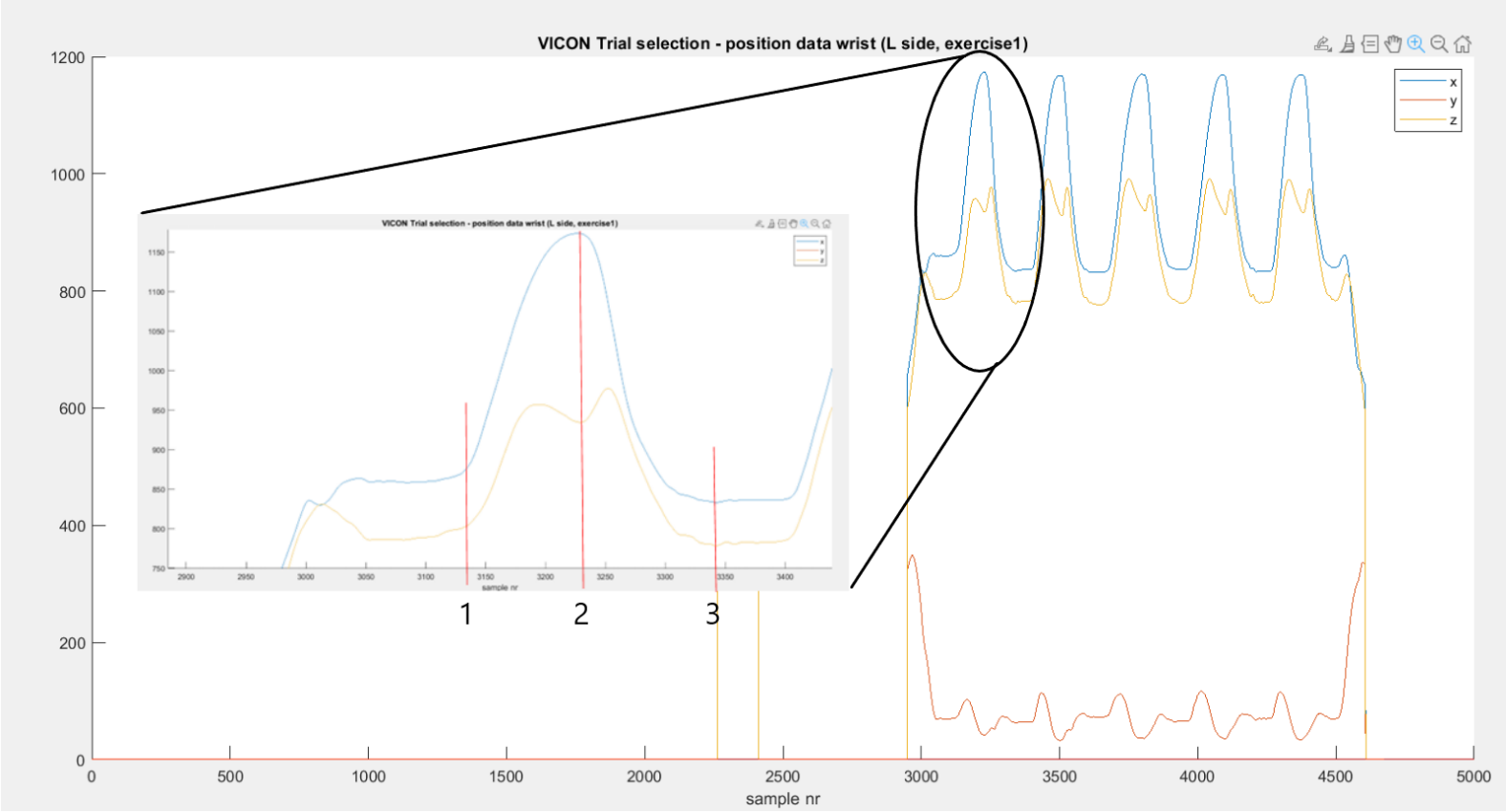


Figure 12: Example of Vicon position data in the x-, y- and z-axis for P02, during the touch exercise of the Test session of part A, validity protocol. Trial indices selection using Matlab is shown with numbers 1-3 corresponding to: 1 = start of reach, 2 = end of reach, 3 = return to initial position.

3.4.4 Metric calculations

After selection of the indices, raw data files of every individual patient were loaded into the main processing Matlab script. The 50 Hz IMU data from the validity protocol was upsampled to 100 Hz and calibration indices were therefore multiplied by two. The accelerometer and gyroscope data of both protocols were scaled to convert raw ADC values to the right units: m/s^2 and degrees/second respectively using the following scale factors (provided by 2M Engineering):

- Accelerometer: $\frac{0.244}{1000} \times 9.81$
- Gyroscope: $\frac{17.50}{1000}$

For the validity protocol data, a 4^{th} order low-pass Butterworth filter with a cutoff frequency of 20 Hz was applied to the Vicon data. Every IMU sensor was synchronised to a nearby Vicon marker based on the lag at which maximum cross-correlation of the accelerations (IMU accelerometer data and second derivative of Vicon position data) occurred. A drift correction for gyroscope data was applied by subtracting the mean of every sensor. Based on a threshold ($\leq 3/180$ rad/s), static phases in gyroscope data were detected during which angular velocity was set to zero (zero-velocity update).

To determine the IMU sensor orientation of both protocols with respect to the corresponding segments ("sensor-to-segment calibration"), the static and dynamic calibration movements

were examined. As mentioned in section 2.5, the measured gravity vector during a static pose of the upper extremity represents the vertical axis of the anatomical frame of the respective extremity. The angular velocity during the dynamic phase represents the horizontal axis of the anatomical frame of the segment. Using the cross-product of the two axes, the third axis was calculated. Furthermore, a common global frame needed to be defined to relate all segment orientations to each other. The common vertical axis is defined using the gravity vector during the static phase. The angular velocity measured during the dynamic phase is used to define the common horizontal axis. Using the cross-product of the two axes, the third axis was calculated. The initial orientation of the sensors within the global frame is expressed using a rotation matrix. A Madgwick filter (Beta = 0.03) was used for orientation estimation in the whole-body frame. The definitions of the IMU measurement axes can be found in Table 3.

Table 3: The definitions of the X-, Y- and Z-axis and the corresponding anatomical body planes in which the movements of that axis occur.

Axis	**Definition**	**Plane**
X-axis	Forward (+) from the trunk	Sagittal plane
Y-axis	Mediolateral to the left (+)	Transverse plane
Z-axis	Vertical, upward (+)	Frontal plane

Segment angles (the shoulder elevation and trunk flexion angles) were calculated relative to reference axes. Joint angles (in degrees) were calculated between the axes of the two adjacent limb segments, relative to the initial position. For example, elbow angles were calculated by relating the upper arm frame to the lower arm frame. The following metrics were calculated for each patient of both protocols: shoulder flexion, shoulder abduction, elbow flexion and trunk flexion. Additionally, shoulder elevation was calculated for participants of the reliability protocol, while shoulder internal rotation was calculated for those in the validity protocol. For the validity protocol, two methods for shoulder angle calculations were investigated: shoulder angles based on the shoulder IMU or angles based on the sternum IMU. No sternum sensor was used in the reliability protocol, thus the shoulder joint angles could only be calculated between the upper arm and shoulder sensor. Moreover, trunk flexion (reliability protocol) was calculated based on movement of the shoulder sensor in the sagittal plane.

Since motion in a plane occurs by rotation around a perpendicular axis(55), movement axes were defined. Clockwise rotation was defined as negative, counterclockwise was defined as positive. The initial positions (corresponding to 0°) and movement axes relevant to these metrics are outlined in Table 4. Orientations were shown in Euler angles using the ZYX convention.

Table 4: The definitions of initial or reference positions (corresponding to 0°) and movement axes are shown for the calculated joint angle metrics of shoulder, elbow and trunk movements.

Joint angle metric	**Initial position (0°)**	**Movement axis**
Shoulder joint flexion/extension	Arm straight besides trunk	In x-z axis, around y-axis
Shoulder joint abduction/adduction	Arm straight besides trunk	In y-z axis, around the x-axis
Shoulder segment elevation	Y-axis	In y-z axis, around the x-axis
Shoulder internal rotation	Arm straight besides trunk	In x-y axis, around the z-axis
Elbow joint flexion	Arm straight besides trunk	In x-z axis, around the y-axis
Trunk segment flexion	Z-axis	In x-z axis, around the y-axis

Based on the trial indices, the joint angles metrics were calculated for different phases of movement: the reaching phase and the reach-and-return phase. The reaching phase was defined between trial indices 1 and 2, the reach-and-return phase between trial indices 1 and 3. For each metric and phase, several statistical outcomes were determined: minimum, maximum, mean and median joint angle, range of joint motion, time to the maximum value and duration of the movement. Currently, only the mean angles and range of joint angles (in degrees) for the two phases were further analysed. A schematic representation of the calculated and included metrics is shown in Figure 13. All output metrics were exported to an Excel file for each patient separately.

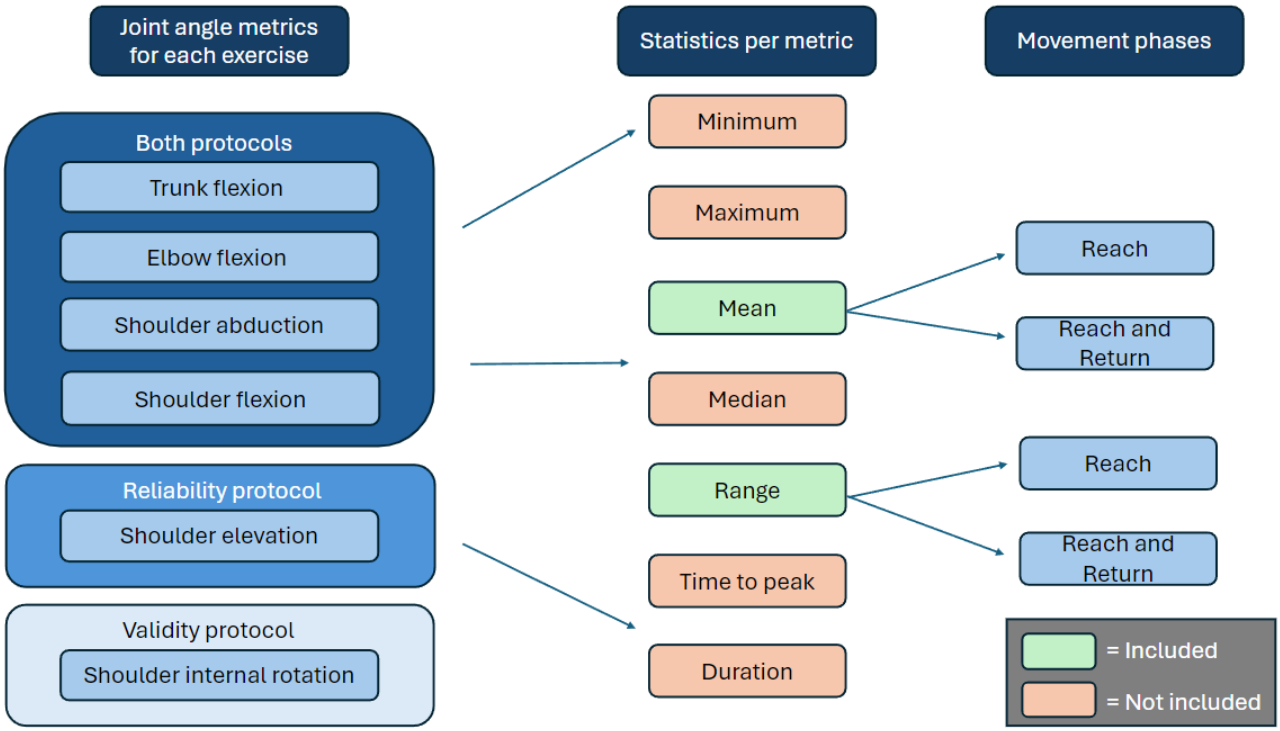


Figure 13: An overview of the calculated metrics of each protocol. For both protocols, only the mean joint angle and joint angle range are included for further analysis. The mean and range were calculated during two phases: the reaching phase and the reach-and-return phase.

3.5 Statistical analysis

Multiple scripts for statistical analysis were created using R and RStudio (R Core Team, version 2023.12.1). Please refer to Appendix C for further details regarding access to the scripts. First, the Excel files with individual patient outcomes were loaded into R. The Excel files of the reliability protocol patients contained a sheet with metrics calculated across the test session and a separate sheet for the retest session. For the validity protocol participants, currently only the test sessions (of part A) were analysed. After the initial data load, the individual patient outcomes were combined into a data frame with metric outcomes of all patients. Then, before conducting calculations to investigate the validity and reliability of IMUs, descriptive statistics were performed, and the normality of the data was assessed. The overall median and interquartile range (IQR) of all metrics for each session (test or retest, IMU and Vicon) and for each exercise (reaching and reach-and-return phase) were calculated. Histograms were created to visually examine the bell-shaped, symmetrical distribution of all individual metric data. Furthermore, the Shapiro-Wilk test was conducted to numerically assess the normal distribution (56). Due to small sample sizes in the validity protocol (57, 58), low Shapiro-Wilk

test p-values (Appendix C), and asymmetric histogram distributions for several reliability protocol metrics, the null hypothesis of normal data distribution was rejected. Consequently, non-parametric methods were chosen for further statistical analysis. Individual patient data frames with raw data and calculated descriptive statistics were combined and exported as Excel documents for further use in the reliability and validity scripts.

3.5.1 Test-Retest reliability

For each patient and metric, the average values across all five trials within the test and retest session were computed. Also, median values (and IQRs) across all patients per metric and session were calculated. The differences and absolute values of the differences between these test and retest trial averages per metric (for each patient) were calculated by subtracting the retest values from the test values. The level of significance of these differences was assessed using the non-parametrical Wilcoxon test for paired data. The level of significance was set at $p \leq 0.05$. Additionally, the overall median test-retest differences and overall median of absolute values of differences (and IQRs) across all patients were calculated.

Furthermore, the ICC was determined to assess the test-retest reliability. A two-way mixed effects ICC-model for absolute agreement was used. Values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively (59). The Standard Error of Measurement (SEM) and Minimal Detectable Change (MDC) were also calculated. The SEM is a measure of the spread of measurement outcomes, while the MDC represents the magnitude of change needed to exceed the variability and the measurement error of the data (60). They were calculated using formula 2 and formula 3:

$$\text{SEM} = \sigma \times \sqrt{(1 - ICC)} \quad (2)$$

$$\text{MDC} = SEM \times 1.96 \times \sqrt{2} \quad (3)$$

3.5.2 Criterion validity

The average values over all trials per metric of the test sessions for both IMU and Vicon data were calculated for each patient. Then, the (absolute values of) differences between these trial averages were calculated. The level of significance of these IMU-Vicon differences was assessed, again using the non-parametrical Wilcoxon test. Furthermore, the overall median (+IQR) absolute differences between IMU and Vicon were determined across all patients.

For each metric, the (trial) average IMU and Vicon patient values were plotted against each other in a Spearman scatter plot, including a regression line. Bland-Altman plots were created using each patient's average of IMU and Vicon combined and the difference between the two averages for each metric. The mean difference (across all patients) between IMU and Vicon and limits of agreement ($1.96 \times \sigma$) were also calculated and plotted to detect outliers.

4

Results

First, the patient characteristics of both protocols are shown. Then, results for the test-retest reliability (reliability protocol) and criterion validity (validity protocol) are presented separately.

4.1 Patient characteristics

Due to computational differences between the right- and left-sided patients, the test-retest reliability could only be evaluated for the remaining 18 right-sided patients, instead of all 32 patients. Out of the 12 recruited patients of the validity protocol, eventually only seven could be included for further analysis. Only four of these patients could be included for the test sessions of the touch exercise and lift exercise, and only three patients were included for the head exercise. Thus, currently only an interim analysis on subgroups of the measured patients was performed. Characteristics of the patients of both protocols measured at both rehabilitation centres (including part B of the validity protocol) are presented in Table 5. Table 6 indicates which patients were included in which exercise. A flowchart of the patient in- and exclusion for the reliability protocol is shown in Figure 14 and in Figure 15 for the validity protocol.

Table 5: Characteristics of patients included for the reliability and validity protocols. Separate and combined characteristics for patients measured at Rijndam (= Centre 1) and Roessingh Rehabilitation (= Centre 2) of the reliability protocol are shown.

Protocol:	**Reliability**	**Reliability**	**Reliability**	**Validity**
Measurement location:	***Centre 1***	***Centre 2***	***Combined***	***Centre 2***
Number of patients	N = 10	N = 8	N = 18	N = 7
Mean age (SD)	61 (8)	58 (11)	60 (9)	57 (14)
Gender (Male/Female)	7/3 (70%)	5/3 (62%)	12/6 (67%)	5/2 (71%)
Type of stroke (Ischemic/hemorrhagic)	6/4 (60%)	6/2 (75%)	12/6 (67%)	6/1 (86%)
Mean time since stroke in days (SD)	198 (423)	99 (131)	154 (323)	138 (140)
Mean NIHSS-score (SD)	3.5 (2)	1 (1)	2 (2)	1 (1)
Mean FMA (SD)	43 (12)	58 (5)	50 (12)	57 (5)
Mean ARAT (SD)	29 (21)	52 (6)	39 (20)	51 (6)
Most affected side (R/L)	10/0 (100%)	8/0 (100%)	18/0 (100%)	5/2 (71%)

Abbreviations: SD = standard deviation, R = right, L = left

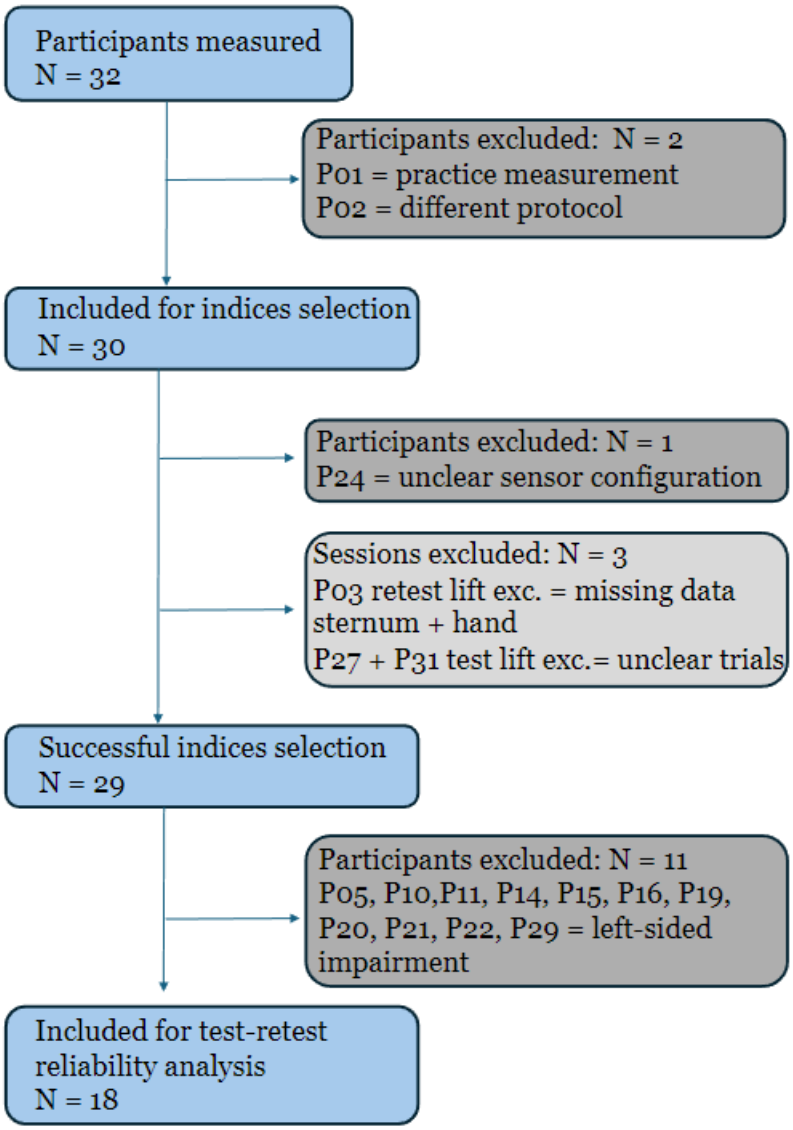


Figure 14: A flowchart of the inclusion and exclusion of the reliability protocol participants, also showing exclusion of a substantial number of participants because of left-sided impairment.

Table 6: An overview of how many and which patients were included for every exercise and session, shown for both the reliability and validity protocol.

Protocol	**Exercise**	**Patients included**
Reliability	Lift exercise – Test & Retest	N = 15; P03, P04, P06, P07, P08, P09, P12, P13, P17, P18, P25, P26, P28, P30, P32
	Head exercise – Test & Retest	N = 18; P03, P04, P06, P07, P08, P09, P12, P13, P17, P18, P23, P25, P26, P27, P28, P30, P31, P32
Validity	Touch exercise – Test	N = 4; P02, P05, P08, P10
	Lift exercise – Test	N = 4; P03, P05, P07, P08
	Head exercise – Test	N = 3; P02, P08, P09

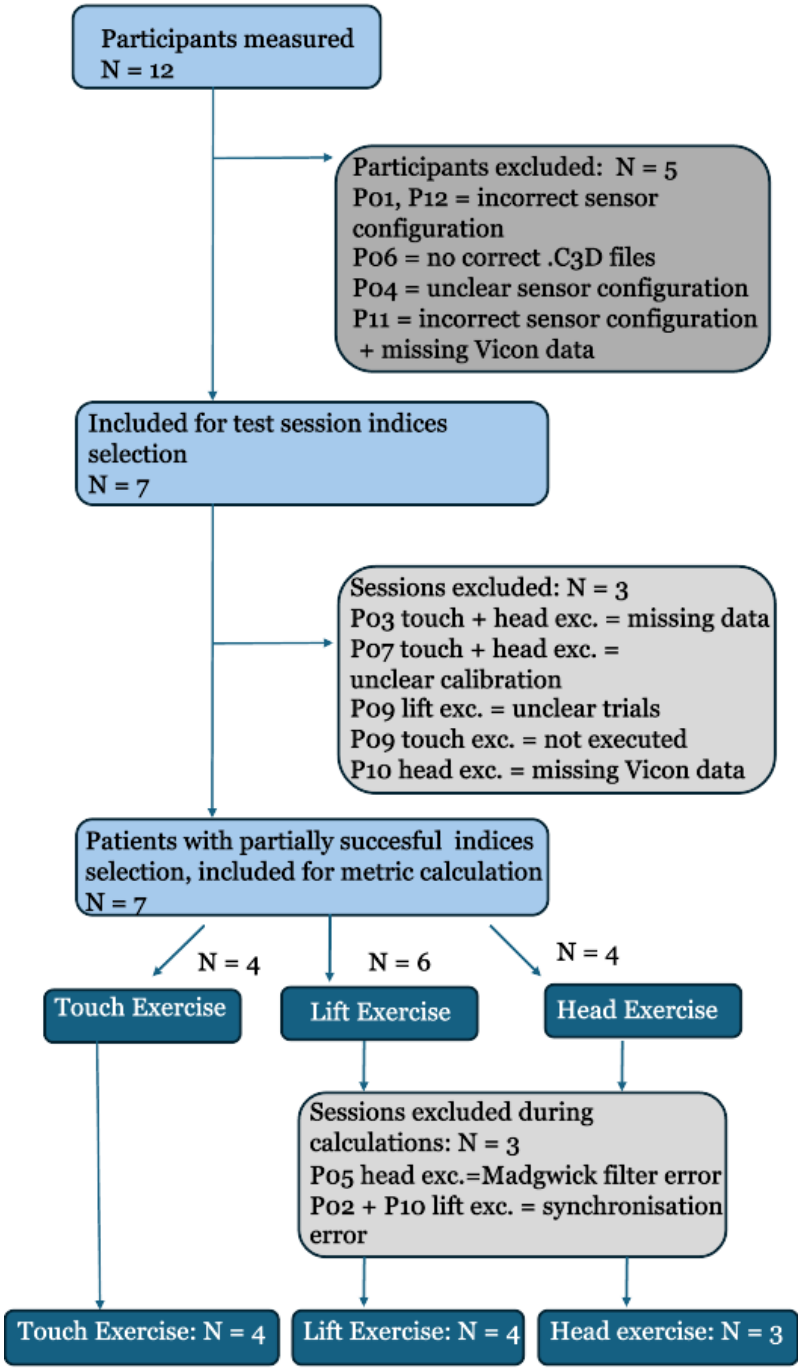


Figure 15: A flowchart of the inclusion and exclusion of the validity protocol participants, showing the resulting numbers of included participants for each exercise (= exc.).

4.2 Test-Retest Reliability

4.2.1 Segment angle trajectories and joint angle visualisation

Two examples of calculated segment angle trajectories in the frontal plane and joint angles for a less (P13) and more impaired patient (P09) during the lift exercise are shown in Figures 16 and 17. The upper arm segment angle trajectory starts at a negative angle, whereas the lower arm segment starts at an angle of about +90°. During the forward reaching movement, the upper arm angle becomes increasingly negative and the lower arm angle becomes increasingly positive (and the vector length slightly decreases). The shoulder segment trajectories tend to fluctuate between +180° and -180°. While reaching for the object, the elbow flexion angle decreases. Cyclic patterns are visible for P13 and less clearly for P09.

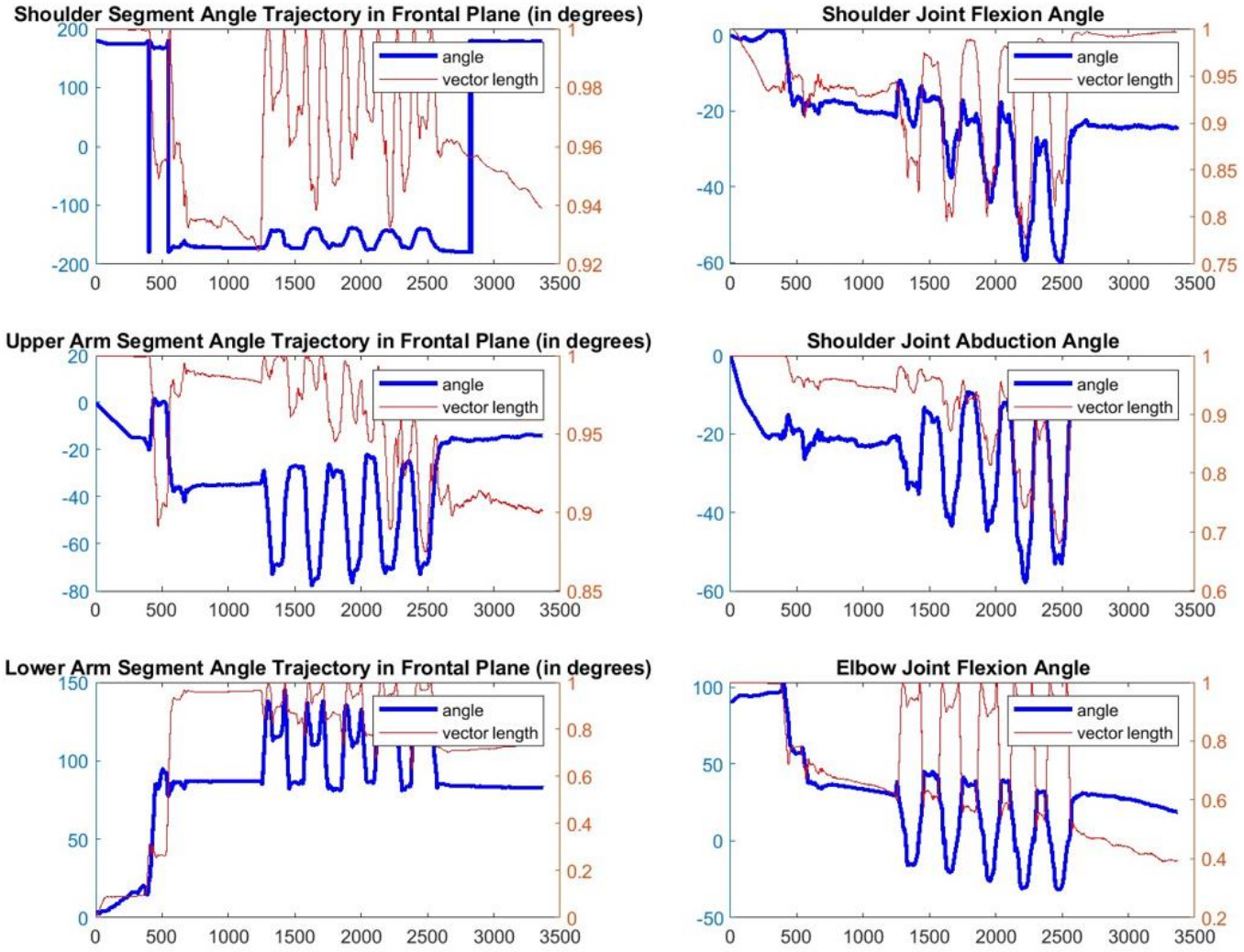


Figure 16: An example of segment angle trajectories in the frontal plane and joint angles for the shoulder flexion and abduction and elbow flexion during the test session of the lift exercise of the reliability protocol, shown for a less impaired patient: P13 (FMA = 60, ARAT = 57, NIHSS = 0). The x-axis shows the sample number, the y-axis the angle in degrees.

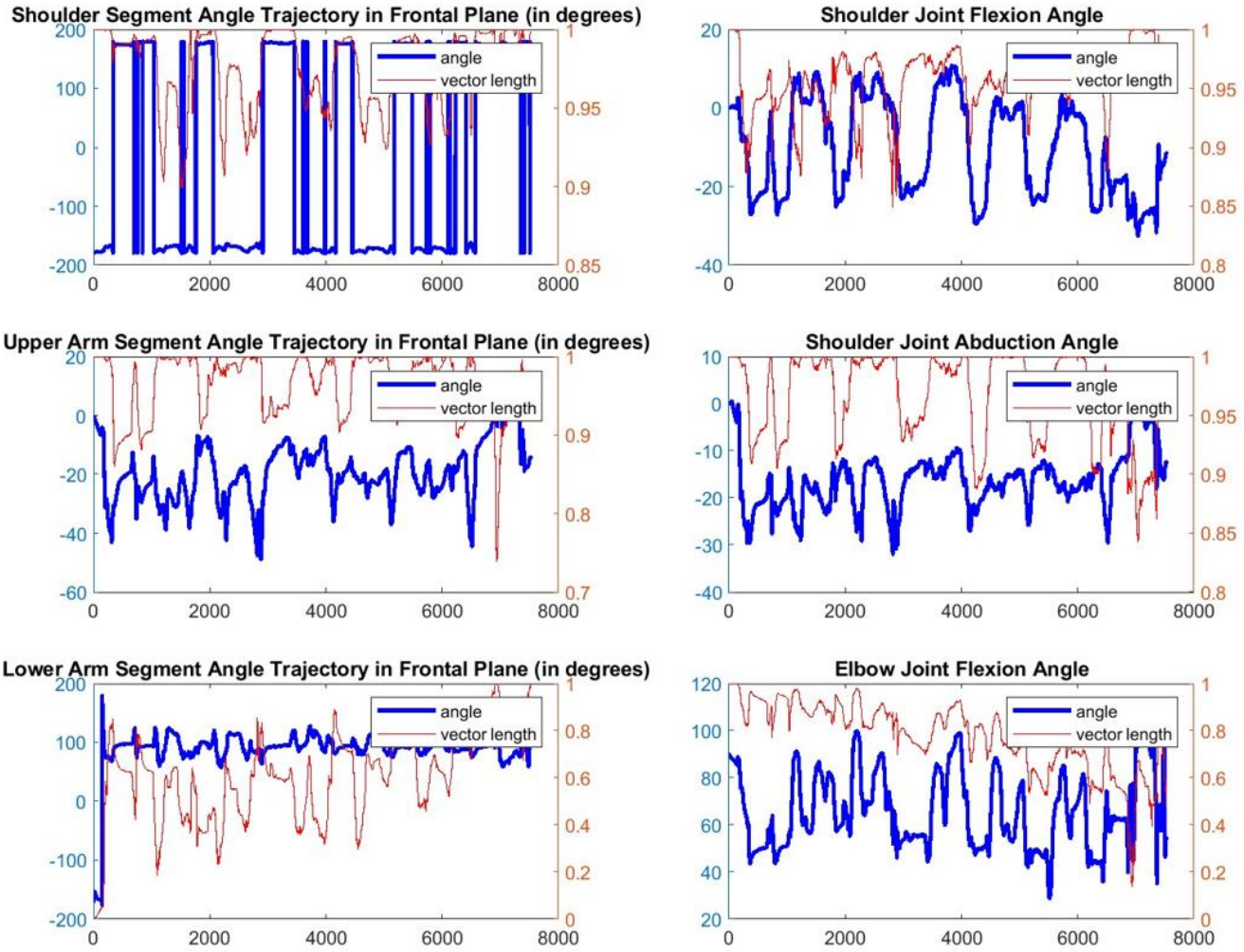


Figure 17: An example of segment angle trajectories in the frontal plane and joint angles for the shoulder flexion and abduction and elbow flexion during the test session of the lift exercise of the reliability protocol, shown for a more impaired patient: P09 (FMA = 35, ARAT = 2, NIHSS = 7). The x-axis shows the sample number, the y-axis the angle in degrees.

4.2.2 Test-Retest Reliability of metrics

Tables 7 and 8 provide an example of the calculated results, shown for the elbow flexion metric. Full tables including all metrics are available in Appendix D. Both the elbow flexion test and retest IQR values demonstrate relatively large ranges (>10°, for the reaching as well as the reach-and-return phases) and median absolute differences. The elbow flexion range during the reaching phase of the head exercise even shows an IQR of 266°. The median absolute differences appear to be higher for the head exercise compared to the lift exercise, for both phases. Values of median absolute test-retest differences for all metrics ranged from below 5◦ (mean shoulder elevation during the reach-and-return phase, head exercise) to 41◦ (elbow flexion range during the reach phase, head exercise).

ICC-values ≥ 0.75 were achieved for elbow flexion (both Range and Mean, during the reaching phase, lift exercise), trunk flexion (Mean, during the reaching phase, head exercise), elbow flexion and shoulder flexion (Range, during the reach-and-return phase, lift exercise), elbow flexion (Mean, during the reach-and-return phase, lift exercise) and shoulder flexion (Mean, during the reach-and-return phase, head exercise). ICC-values ≤0.5 were achieved for multiple metrics: shoulder abduction and trunk flexion (Range, during the reaching phase, lift exercise), shoulder flexion and elevation (Range, reaching phase, head exercise), shoulder flexion (Mean,

during the reaching phase, lift exercise) and elevation (Mean, during the reaching phase, lift and head exercises), shoulder abduction and trunk flexion (Range, during the reach-and-return phase, lift exercise), shoulder flexion (Range, during the reach-and-return phase, head exercise) and shoulder abduction (Mean, during the reach-and-return phase, head exercise). Significant differences ($p \leq 0.05$) were found between the Test and Retest session for shoulder flexion (Range, during the reaching phase, head exercise), elbow and trunk flexion (Mean, during the reaching phase, lift exercise), shoulder flexion and elevation (Range, during the reach-and-return phase, head exercise) and shoulder elevation (Mean, during the reach-and-return phase, head exercise).

Table 7: This table shows the calculated reliability results of the elbow joint angle range and mean elbow joint angle (in degrees) during the reaching phase: Test and Retest median (+ interquartile range, IQR), median absolute test-retest difference (+IQR), Wilcoxon test results, ICC, SEM and MDC

	Reliability protocol – Reaching phase						
Metric	**Test Median (IQR)**	**Retest Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Retest vs Test**	**ICC agree-ment**	**SEM**	**MDC**
Reach, grasp, lift and hold object – Lift exercise							
Elbow Flexion - ***Range***	50.50 (138.81)	52.47 (90.57)	11.74 (34.00)	0.84	0.75	43.73	121.21
Elbow Flexion - ***Mean***	72.30 (108.66)	78.70 (149.98)	14.59 (11.84)	0.04	0.77	37.93	105.14
Reach to the back of the head– Head exercise							
Elbow Flexion - ***Range***	81.15 (167.78)	181.68 (269.16)	40.56 (123.74)	0.06	0.60	79.80	221.19
Elbow Flexion - ***Mean***	78.70 (149.98)	34.64 (49.84)	19.30 (43.14)	0.90	0.50	41.92	116.19

Abbreviations: IQR = interquartile range, ICC = intraclass correlation coefficient, SEM = standard error of measurement, MDC = minimal detectable change

Table 8: This table shows the calculated reliability results of the joint angle range and mean joint angle (in degrees) during the reach-and-return phase: Test and Retest median (+IQR), median absolute test-retest difference (+IQR), Wilcoxon test results, ICC, SEM and MDC

	Reliability protocol – ReachAndReturn phase						
Metric	**Test Median (IQR)**	**Retest Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Retest vs Test**	**ICC agree-ment**	**SEM**	**MDC**
Reach, grasp, lift and hold object – Lift exercise							
Elbow Flexion - ***Range***	64.58 (181.45)	57.08 (105.53)	14.27 (46.19)	0.67	0.84	42.41	117.56
Elbow Flexion - ***Mean***	81.19 (114.46)	86.49 (128.33)	12.64 (13.99)	0.29	0.87	30.03	83.25
Reach to the back of the head– Head exercise							
Elbow Flexion - ***Range***	105.65 (220.30)	195.42 (265.58)	33.93 (51.40)	0.28	0.73	66.01	182.97
Elbow Flexion - ***Mean***	86.49 (128.33)	23.32 (65.82)	24.23 (57.68)	0.57	0.60	38.03	105.40

Abbreviations: IQR = interquartile range, ICC = intraclass correlation coefficient, SEM = standard error of measurement, MDC = minimal detectable change

4.3 Criterion validity

4.3.1 Data inspection

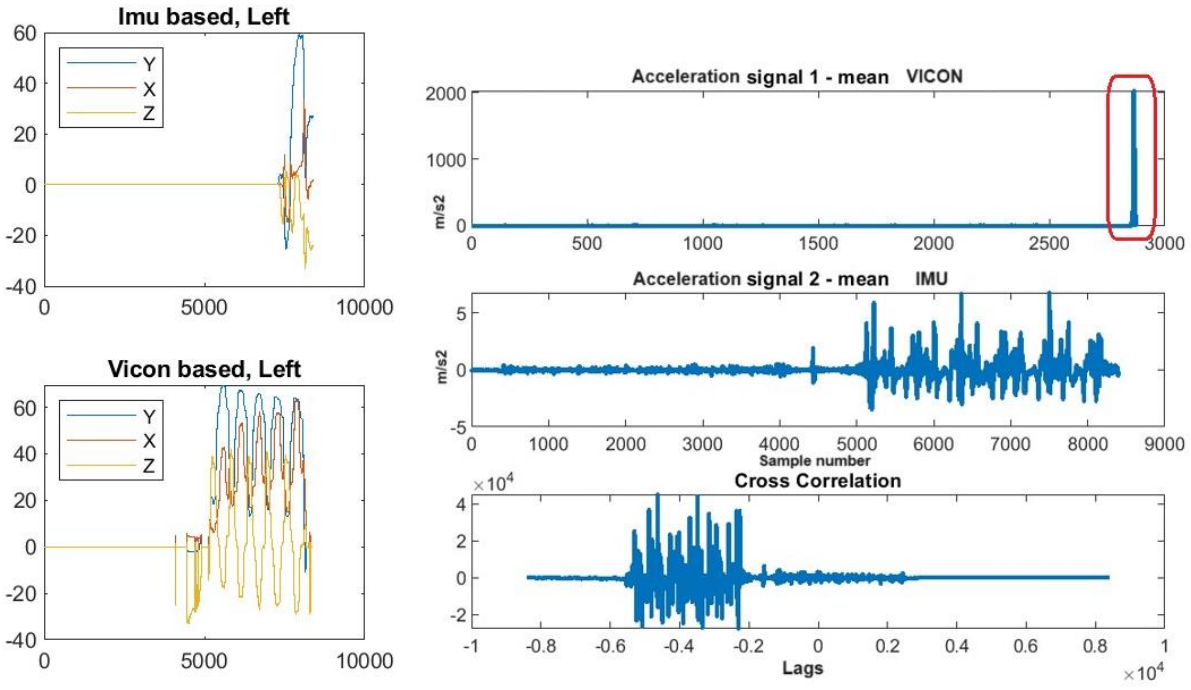


Figure 18: On the left: Example of the lowerarm Euler angles of P02 during the lift exercise (test session), showing asynchrony of IMU and Vicon data. On the right: The corresponding cross-correlation shows an unexpected peak near the end of the calculated Vicon acceleration, probably resulting in wrongly selected lags of maximum cross-correlation.

While assessing face validity of the Euler angles, asynchrony appeared and corrupted the results of some patients. The cross-correlation figures showed deviations in the second derivative of Vicon signals (Figure 18), distorting cross-correlation results and leading to incorrectly updated indices. Two patients were therefore excluded.

Figure 19 shows examples of IMU Sternum Euler angles and Vicon Thorax angles for the included P02 and P09. Movements of the thorax on the impaired side can be evaluated, especially around the Y-axis (indicating trunk flexion), during the execution of the movements. Although the Vicon and IMU data show similar movement patterns, differences in patterns and absolute Euler angle values can be observed.

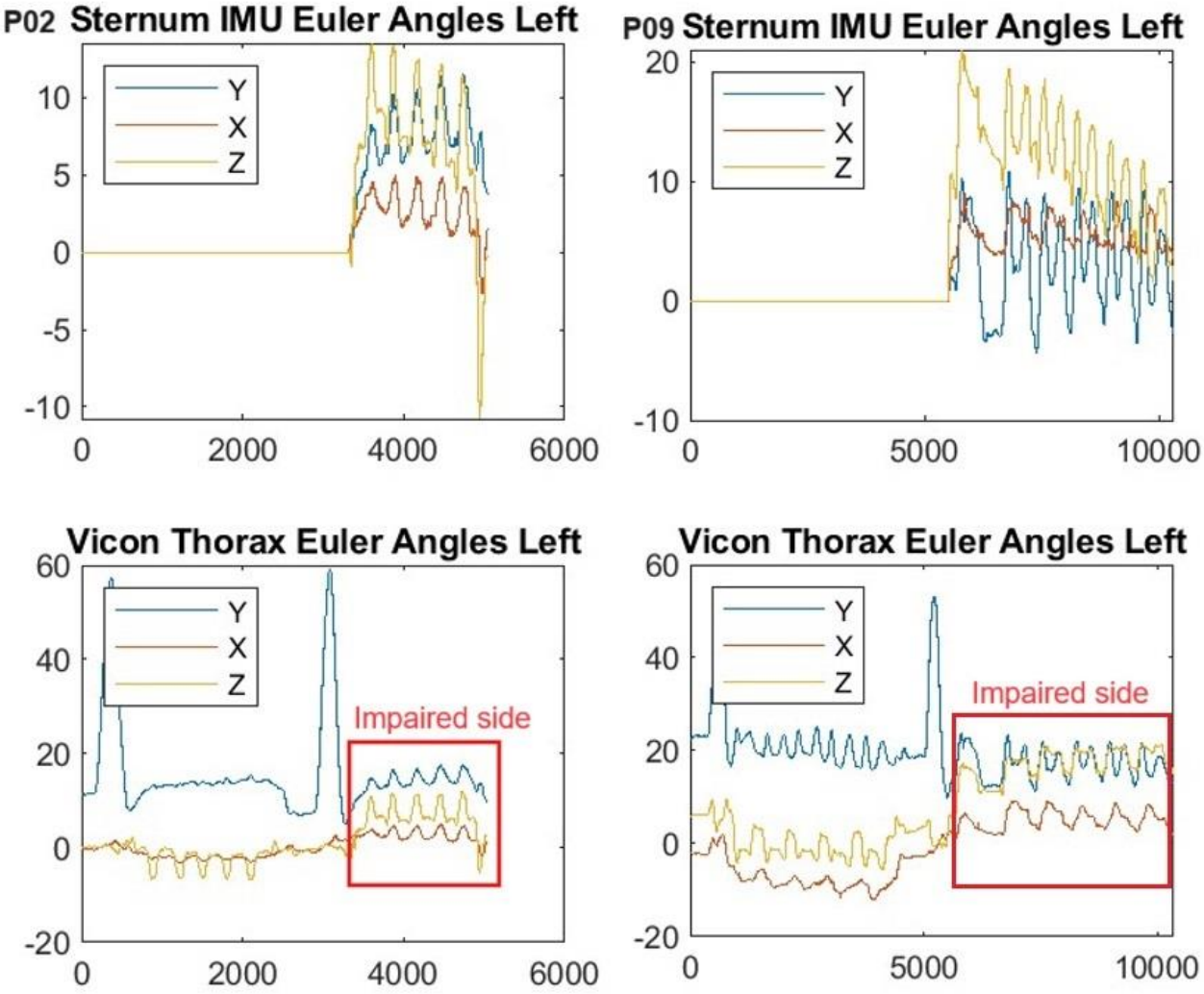


Figure 19: On the left: Synchronised Sternum IMU and Thorax Vicon angles for the less impaired patient P02 (FMA = 60, ARAT = 57, NIHSS = 0) are shown (touch exercise), with a difference in thorax movements for the impaired and unimpaired upper extremity. The red area indicates the movement phase of the impaired extremity. On the right: The same results for the more impaired patient P09 (FMA = 50, ARAT = 46, NIHSS = 2), are shown (lift exercise), with thorax movements during movement phases of both the unimpaired and impaired arm. The y-axis shows the angles in degrees, the x-axis the sample number.

4.3.2 Criterion validity

Eventually, 16 of the calculated metrics were included for analysis. Again, examples of the results calculated for the elbow flexion are shown in Tables 9 and 10. Full tables of results can be found in Appendix E. Wilcoxon test p-values were >0.05 for all metrics, initially suggesting no statistically significant differences between IMU and Vicon results. However, large median absolute differences were observed, especially for the elbow flexion mean and range of the head exercise (during both movement phases). Median absolute Vicon-IMU differences of all metrics ranged from below 1◦ (trunk flexion range during the reaching phase, touch exercise) to 158◦ (mean elbow flexion during the reach-and-return phase, head exercise) (Appendix E).

Although the Spearman correlation coefficient indicated strong positive or negative correlations (rho ≥ 0.7) between IMU and Vicon outcomes for 44 out of 64 of the calculated metrics in total, all p-values were above the level of significance (> 0.05). The head exercise (with only 3 patients included for analysis) frequently yielded p-values of 1, indicating no statistical evidence to support the observed correlations between IMU and Vicon values. Results were likely to have occurred by chance.

Table 9: This table shows the calculated validity results of the joint angle range and mean joint angle (in degrees) during the reaching phase: IMU and Vicon median (+IQR), median absolute IMU-Vicon difference (+IQR), Wilcoxon test results and Spearman correlation coefficient with corresponding p-values.

	Validity protocol – Reaching phase				
Metric	**IMU Median (IQR)**	**Vicon Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Vicon vs IMU, p-value**	**Spearman correlation Vicon vs IMU (p-value)**
Reach to touch object – Touch exercise					
Elbow flexion - ***Range***	46.97 (14.22)	35.37 (3.06)	12.26 (8.41)	0.20	0.2 (0.92)
Elbow flexion - ***Mean***	16.49 (33.03)	73.18 (2.96)	54.06 (27.81)	0.10	0.4 (0.75)
Reach, grasp, lift and hold object - Lift exercise					
Elbow flexion- ***Range***	62.92 (15.39)	44.54 (8.93)	17.12 (10.57)	0.10	0.4 (0.75)
Elbow flexion - ***Mean***	27.20 (18.55)	63.28 (8.25)	41.21 (22.58)	0.10	-0.4 (0.75)
Reach to the back of the head – Head exercise					
Elbow flexion - ***Range***	187.67 (120.92)	59.73 (12.21)	123.73 (110.81)	0.18	0.5 (1)
Elbow flexion - ***Mean***	-16.28 (28.43)	116.19 (1.71)	130.81 (27.55)	0.18	0.5 (1)

Abbreviations: IQR = interquartile range, IMU = Inertial Measurement Unit

Table 10: This table shows the calculated validity results of the joint angle range and mean joint angle (in degrees) during the reach-and-return phase: IMU and Vicon median (+ IQR), median absolute IMU-Vicon difference (+IQR), Wilcoxon test results and Spearman correlation coefficient with corresponding p-values.

	Validity protocol – Reach-and-return phase				
Metric	**IMU Median (IQR)**	**Vicon Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Vicon vs IMU, p-value**	**Spearman correlation Vicon vs IMU (p-value)**
Reach to touch object – Touch exercise					
Elbow flexion - ***Range***	48,12 (15,59)	39.75 (7.36)	13.32 (7.50)	0.36	0.4 (0.75)
Elbow flexion - ***Mean***	19.34 (27.86)	74.30 (3.60)	54.39 (26.26)	0.10	0.4 (0.75)
Reach, grasp, lift and hold object – Lift exercise					
Elbow flexion- ***Range***	80.27 (20.38)	51.14 (4.55)	27.06 (18.04)	0.10	0.2 (0.92)
Elbow flexion - ***Mean***	29.53 (24.68)	62.43 (8.44)	37.08 (24.77)	0.10	-0.2 (0.92)
Reach to the back of the head – Head exercise					
Elbow flexion - ***Range***	192.97 (119.79)	62.28 (12.40)	126.08 (109.70)	0.18	0.5 (1)
Elbow flexion - ***Mean***	-34.66 (9.56)	121.21 (2.40)	157.90 (8.18)	0.18	0.5 (1)

Abbreviations: IQR = interquartile range, IMU = Inertial Measurement Unit

5

Discussion

5.1 Interpretation of results

The aim of this study was to investigate the test-retest reliability and criterion validity of IMUs for calculation of upper extremity metrics in stroke patients. Currently, an interim analysis of results has been conducted with subgroups of the included patients. Therefore, results cannot be generalized to the whole study population. Nevertheless, it can be concluded that the current data processing methods are not yet sufficient. Initially, the study aimed to include distance and velocity metrics. However, inspection of the face validity of these outcomes revealed underlying issues in the data or the metric calculation. Consequently, only joint angle outcomes were deemed sufficient for statistical analysis. The exclusion of some patients due to synchronization or sensor configuration difficulties underscores the challenges encountered even before calculation of the joint angle metrics. Additionally, the large variability and poor-moderate reliability results for multiple metrics indicate underlying issues with the measurements and/or the current joint angle calculation processes. Large differences in absolute joint angle values were observed, not only between Vicon and IMU, but also between test-retest sessions or between patients. This indicates suboptimal orientation estimation, angle definitions or other fundamental problems. A more detailed description of the preliminary results of each protocol is provided in the following sections.

5.1.1 Test-retest reliability

The relatively large IQRs of the joint angles (especially for shoulder elevation and elbow flexion) suggest considerable variations in repeated measurements and between patients, while executing the same movements. A possible explanation for the large range of elbow flexion and shoulder elevation might be the quick transitions from a negative angle (-180°) to a positive angle (+180°). These transitions were observed for multiple patients in the segment angle trajectories of the elbow and shoulder in the frontal planes. They could occur when a patient was operating near the edge of the plane (180° or -180°). Inspection of the Bland-Altman plots also revealed outliers for several metrics. This suggests that there are individual cases with large variations in test-retest measurements. These angle transitions and large ranges in outcomes may be attributed to the current definitions of the axes in all planes and the methods employed for angle calculation. Moreover, the mean and range of trunk flexion appear to be substantially higher for the reliability protocol compared to the validity protocol. An increased trunk flexion following stroke is anticipated, with values observed by Roby-Brami et al. ranging from approximately 5 to 13◦ (61). However, these current trunk flexion results (far above 13◦) are deemed improbable. This discrepancy may be attributed to the use of the shoulder sensor to define trunk flexion.

For six of the calculated metrics, the Wilcoxon test showed statistically significant differences between test and retest sessions. Thus, considerable discrepancies in the calculated joint angles between the two sessions exist. It is noteworthy that these significant differences were found in multiple metrics for multiple exercises, rather than consistently in a single metric. A possible factor influencing the test-retest results could have been the consistency of patient performance across sessions. Merlau et al. addressed this by excluding the first repetition for

each series and each subject, treating these as a test run (62). Within this project, all five repetitions of a session were included for analysis, since the movements were considered to be intuitive and the influence of a test run was assumed to be minimal. Nonetheless, the effects of learning or fatigue cannot be entirely ruled out and might have influenced test-retest reliability. However, examination of the recordings did not reveal clear differences in execution of the test and retest sessions. Therefore, it can be hypothesised that these differences might be attributed to random errors in the measurements or calculations rather than the exercise execution. This hypothesis is also supported by the relatively large SEM (>5°) and MDC (>15°) values for most of the metrics, indicating significant random errors.

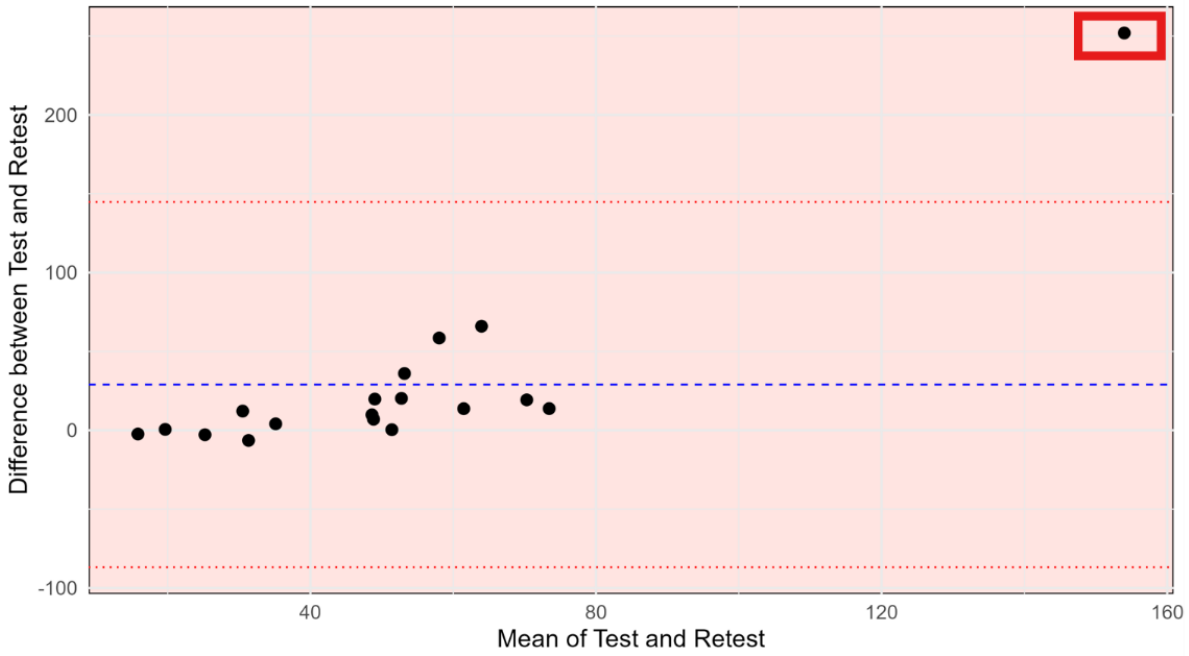


Figure 20: Bland-Altman plot for test-retest comparison of shoulder flexion angle range during the reach-and-return phase of the head exercise, showing an outlier corresponding to P30 marked in red. The test-retest mean is shown on the x-axis, the difference on the y-axis.

Multiple metrics achieved an ICC <0.5, indicating poor reliability. For example, shoulder joint flexion angle range (during the reach-and-return phase, head exercise) yielded both a low ICC (0.025) and Wilcoxon test p-value (0.0017). The Bland-Altman plot reveals an outlier with a very large test-retest difference (Figure 20), corresponding to P30 with Test average of 280° and Retest average of 28°. Repeated calculations without the outlier resulted in an increased ICC (0.38) and Wilcoxon test p-value (0.04), but still indicated low test-retest reliability. Another remarkable example is the mean elbow joint flexion angle (during the reaching phase, lift exercise) that resulted in an ICC value of 0.77 (good reliability), but a Wilcoxon test p-value of 0.04, suggesting significant differences between the test and retest sessions. Inspection of the corresponding scatterplot in Figure 21 shows a patient (P09) with a positive test value and a negative retest value, resulting in large test-retest differences. The underlying explanation for such differences should be further investigated.

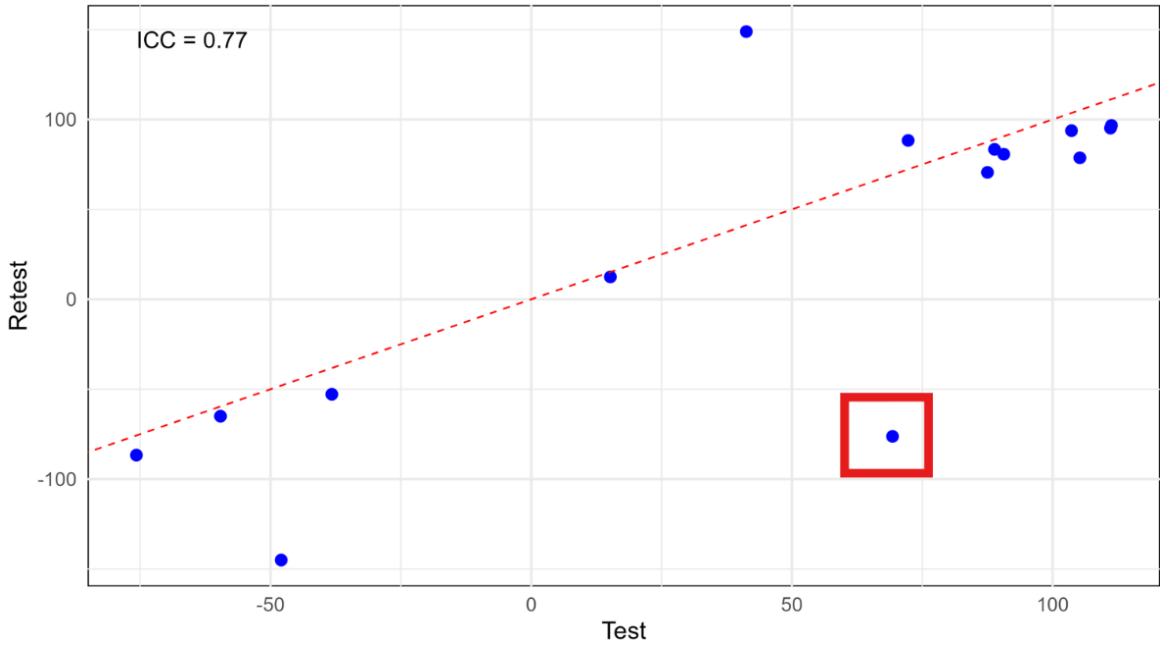


Figure 21: The scatterplot of the mean elbow joint flexion angle during the reaching phase of the lift exercise, complementary to figure X. The datapoint marked in red shows a participant (P09) with a positive test value and negative retest value.

5.1.2 Criterion validity

Although the calculated IMU and Vicon angles exhibit similar patterns, there are notable differences in absolute values, suggesting fundamental differences between the two systems. Other studies have reported similar offsets in absolute values (63). The discrepancy might be due to differences in sensor placement, orientation estimations, or data processing algorithms. Furthermore, the IMUs show signal drift around some of the axes, despite the currently applied gyroscopic drift correction. Overall, while the IMU and Vicon systems may track similar movement patterns, these discrepancies in absolute values and the presence of signal drift highlight the need for further refinement of IMU calibration and data processing techniques to ensure accurate estimation of criterion validity.

Several metrics showed relatively large median absolute differences (in degrees) between Vicon and IMU measurements, especially for the elbow flexion (mean and range). The degree of acceptable measurement variation depends on the clinical application. According to the literature review of McGinley et al., errors between 2° and 5° are often deemed reasonable while measuring joint angles (64). The large median absolute differences between Vicon and IMU data suggest that the discrepancies are well beyond this 5° threshold that would raise concern in clinical contexts. Morrow et al. found an underestimation of large joint angles and overestimation of small joint angles while comparing IMUs to a lab-based motion capture system. They observed the largest error in shoulder elevation data (65). Errors in (proximal) shoulder angles could propagate in (distal) elbow angles (66), highlighting the importance of accurate shoulder measurements. The preliminary results indicate a relatively large median absolute difference between Vicon and IMU measurements for shoulder abduction and rotation, similar for both methods of shoulder angle calculations. Investigation with larger sample sizes is necessary to determine the optimal methodology.

A study of El-Gohary et al. found an excellent agreement with an average correlation coefficient of > 0.97 for all tasks among all subjects, while comparing an inertial and optical motion tracking system(66). Although some metrics of our study also yield a Spearman correlation coefficient close to 1, all corresponding p-values were >0.05. Thus, these results are likely to have occurred by chance. Figure 22 shows two Spearman correlation examples, illustrating the uncertainty of the results. One is an example of a scatterplot for the elbow flexion range of the head exercise with a Spearman rho of 0.5 and p-value of 1. Although moderate correlation is implied, there is no statistical evidence to support this. The other example is the trunk flexion range, which showed a strong correlation (rho =1) during the reaching phase of the lift exercise. Since the p-value was 0.083, no statistically significant correlation could be determined.

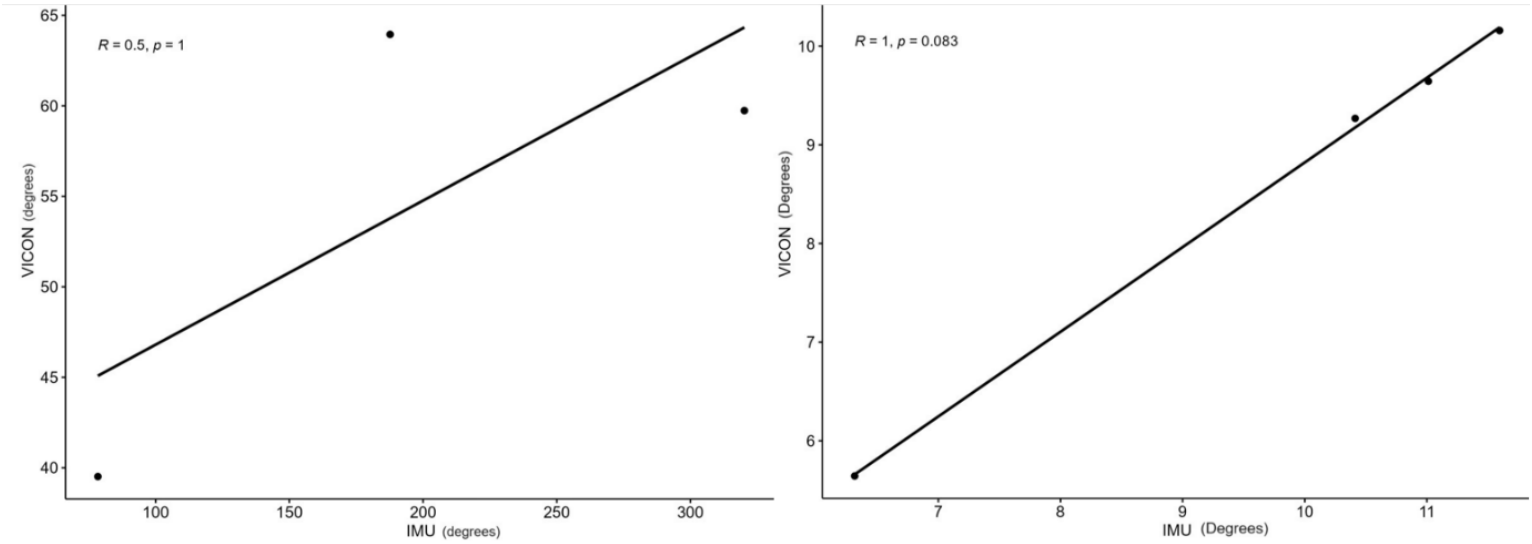


Figure 22: Left: An example of a scatterplot for the elbow flexion range (in degrees) during the reaching phase of the head exercise, with a Spearman rho of 0.5 and p-value of 1. Right: A scatterplot for the trunk flexion range during the reaching phase of the lift exercise with a Spearman rho of 1 and a p-value of 0.083 is shown.

As mentioned in the results section, the current IMU-Vicon synchronization processes appeared to yield inaccurate results for two participants. Although the calibration and trial indices were correct, the updated calibration indices (required for synchronization) were found to be incorrect for some sensors, because of large differences in individual sensor lags. The cross-correlation figures revealed unexpected patterns for the second derivative of the Vicon data in these cases. Probably, the cross-correlation of these patients was performed on erroneous signals and therefore lead to incorrect synchronization and metric calculation. Due to a second derivative needed to calculate acceleration of Vicon position data, errors might be introduced (67), and the results might not be directly comparable to the IMU acceleration data.

5.2 Strengths and limitations

5.2.1 General strengths

A key strength of the ArmCoach4Stroke project is the conduction of measurements in both a laboratory and an ecological setting. These two environments provide complementary insights. The ecological setting has offered valuable initial insights into the usability of the sensors and data, as well as how a diverse range of patients interact with the IMUs. Because of the manual selection of indices, deviating data, missing data or incorrect trials could easily be identified.

Furthermore, the aim of the project is to calculate a large variety of metrics, while current studies often only investigate a subset of metrics. Eventually, not only joint angles, but also distance and velocity profiles will be calculated. That way, validity and reliability can be investigated within a broad range of outcomes and applications.

5.2.2 General limitations

There are several limitations to this study. Firstly, calibration and trial indices were manually selected and therefore possibly subject to human errors. Suboptimal indices selection could have affected sensor to segment calibration and thus result accuracy. Ideally, the use of an algorithm for the automatic detection of calibration movements and trial segmentation is desirable to enhance objectivity and repeatability of results and increase time efficiency. Additionally, trial segmentation was based solely on indices from the wrist Vicon marker or IMU sensor, potentially causing inaccuracies due to timing differences in movement onset between anatomical locations. For example, sternum movement may start later than lower arm movement. These discrepancies are particularly relevant in the reliability protocol, where asynchrony could not be corrected. The precise extent of this underlying sensor asynchrony remained unknown, due to the absence of a phase at which simultaneous movement was known.

A limitation of the use of skin-mounted sensors is the occurrence of soft tissue artefacts. Muscle contractions deform the surrounding skin (45), affecting the measurements and thus joint angle calculations (68). According to the systematic review of Fang et al., measurements near the shoulder joint are the most susceptible to the influence of such skin motion artefacts (69). However, soft tissue artefacts are difficult to prevent and occur in both the IMU and Vicon data. Therefore, validity calculations are not necessarily compromised by these tissue artifacts. To enhance shoulder joint calculations in the reliability protocol, exploring the addition of a sternum sensor could be beneficial. Then, trunk flexion could also be based on movement of the sternum sensor rather than the shoulder IMU. Nonetheless, according to Fiorentino et al., metrics such as the range of motion may be underestimated using sensors attached to the skin, so caution while interpreting the results is still recommended (68).

5.2.3 Limitations: Reliability protocol

Due to the ecological setting of the reliability protocol, small variations in the execution of the exercises were not corrected by the investigators. Some patients experienced difficulties with the execution of calibration movements because of their impairments, potentially affecting the calculated kinematic metrics. Calibration should ideally involve data showing rotation about only one axis, but some patients may have used compensatory movements. Similar issues could occur during trials, where patients might need assistance from the investigator. These variations could have influenced the results, though the exact extent is unknown and will most likely be present during clinical applications as well.

5.2.4 Limitations: Validity protocol

The current design of the Matlab data processing scripts was not able to handle deviations in sensor configurations and resulted in small sample sizes. Larger sample sizes are recommended to achieve valid correlations (70). Also, the Wilcoxon signed-rank test requires a sample size of at least six to accurately accept or reject the null-hypothesis (71). Therefore, current correlation coefficients and Wilcoxon results were deemed insufficient to draw any conclusions. Interpretation of underestimation or overestimation of IMUs compared to Vicon results based on the bias within Bland-Altman plots can be done after inclusion of more patients. Currently, no trends can be evaluated.

Although marker-based motion capture systems are often described as the gold standard for three-dimensional kinematic joint measurements (55, 72, 73), their limitations must be considered. These systems require laboratory conditions and skilled personnel, due to their complex set-up. Therefore, the validity was examined in a laboratory setting, even though the IMUs are being designed for an ecological setting. However, if an instrument exhibits low validity in a laboratory setting, it is probable that its validity in a less controlled environment would also be low. Besides the design for a different setting, it is important to acknowledge the existence of discrepancies in calculations of IMU and Vicon. According to the Vicon Plug-In Gait model (51), the combination of markers on the torso (near C7, T10, clavicle and sternum) defines the 'thorax'. These Vicon thorax angles were compared to the measurements of the IMU on the sternum for the trunk flexion angles. However, it should be noted that the thorax segment and sternum IMU are not identical and were defined differently. This may have resulted in minor discrepancies between the Vicon and IMU trunk flexion results.

5.3 Future recommendations

The interim results indicate the need for further improvement of the currently used sensors and data processing methods. Orientation estimations and joint angle definitions should be thoroughly reviewed and verified. Possibly, the initial positions and reference axes should be redefined and gyroscopic drift corrections could be improved. Optimal settings of lowpass filters and the use of the Kalman filter instead of the currently used Madgwick filter could be explored. Besides that, the script of the validity protocol should be adjusted to deal with deviations of the standard IMU sensor configuration or missing IMU sensors. Moreover, calculation of the second derivative of Vicon signals and the cross-correlation methods for synchronization should be further investigated and improved to be able to include all measured participants. For example, cross-correlation could be performed on a smaller fragment of Vicon data. Once these improvements are made, the reliability and validity of joint angle calculations should be enhanced.

The next step would be to extend the current methods of metric calculation to allow for accurate calculation of velocity and distance profiles. Other studies, such as Hughes et al. (74), already provide promising results regarding the reliability and validity of IMUs for movement metrics such as peak velocity during different movement phases. Within the ArmCoach4Stroke project, data from healthy subjects who performed a similar laboratory exercise protocol as the validity protocol is also available. This data could provide further insights into the validity and reliability of IMUs, independent of variations caused by impairments or slight deviations of the protocol. An analysis of these datasets in a similar manner to the current analysis methods is therefore recommended, allowing for comparison of validity and reliability results.

Within this research, the hypothesis that IMUs are reliable and yield comparable results as a gold standard marker-based system was tested. However, the ArmCoach4Stroke project alone may not confirm the IMUs' clinical validity. Correlation between changes in IMU-derived metrics and a clinical scale for motor recovery (such as the FMA) could be investigated. Eventually, the wearable system should be validated in clinical practice to assess its impact on therapy efficacy, long-term outcomes, and patient quality of life., such as investigated by Lin et al. Or Chae et al. (75, 76). Ideally, a randomized controlled trial should be conducted, comparing IMU-assisted therapy with standard rehabilitation. Furthermore, therapist perspectives and the organizational context should be carefully considered.

6

Conclusion

The purpose of this thesis was to investigate the test-retest reliability and criterion validity of IMUs within the context of upper extremity rehabilitation of stroke patients. Due to computational difficulties, unforeseen project delays and exclusion of participants, only a preliminary investigation of a subset of the data could be performed.

Within the reliability protocol, large IQRs were observed for patients with a right-sided impairment, especially for the shoulder elevation and elbow flexion angles. These findings suggest variations in measurements and exercise executions. The large variability and poor-moderate reliability indicate the existence of underlying issues within the measurements, orientation estimations or current calculation processes.

Because of the small remaining sample sizes of the validity protocol, no definitive conclusions regarding criterion validity could be drawn. Although Vicon and IMU signal patterns were similar, interim results indicate differences in absolute values of both sensors and large variations between subjects. Inclusion of more subjects is essential to accurately assess validity. Thus, data processing methods of both protocols should be thoroughly reviewed to allow for further improvement and interpretation of results.

7

References

1. Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Neurology. 2017;16(11):877-97.
2. Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. Lancet (London, England). 2011;377(9778):1693-702.
3. Feigin VL, Stark BA, Johnson CO, Roth GA, Bisignano C, Abady GG, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet Neurology. 2021;20(10):795-820.
4. Langhorne P, Coupar F, Pollock A. Motor recovery after stroke: a systematic review. The Lancet Neurology. 2009;8(8):741-54.
5. Kwakkel Senior Researcher G. Impact of intensity of practice after stroke: Issues for consideration. Disability and Rehabilitation. 2006;28(13-14):823-30.
6. Langhorne P, Wagenaar R, Partridge C. Physiotherapy after stroke: more is better? Physiotherapy research international : the journal for researchers and clinicians in physical therapy. 1996;1(2):75-88.
7. Bernhardt J, Hayward KS, Kwakkel G, Ward NS, Wolf SL, Borschmann K, et al. Agreed definitions and a shared vision for new standards in stroke recovery research: The Stroke Recovery and Rehabilitation Roundtable taskforce. International Journal of Stroke. 2017;12(5):444-50.
8. Zheng H, Black ND, Harris ND. Position-sensing technologies for movement analysis in stroke rehabilitation. Medical and Biological Engineering and Computing. 2005;43(4):413-20.
9. Oyake K, Suzuki M, Otaka Y, Tanaka S. Motivational Strategies for Stroke Rehabilitation: A Descriptive Cross-Sectional Study. 2020;11.
10. Cunningham P, Turton AJ, Van Wijck F, Van Vliet P. Task-specific reach-to-grasp training after stroke: development and description of a home-based intervention. Clinical rehabilitation. 2016;30(8):731-40.
11. Dobkin BH. A Rehabilitation-Internet-of-Things in the Home to Augment Motor Skills and Exercise Training. Neurorehabil Neural Repair. 2017;31(3):217-27.
12. Toh SFM, Chia PF, Fong KNK. Effectiveness of home-based upper limb rehabilitation in stroke survivors: A systematic review and meta-analysis. Front Neurol. 2022;13:964196.
13. Velez M, Lugo-Agudelo LH, Patiño Lugo DF, Glenton C, Posada AM, Mesa Franco LF, et al. Factors that influence the provision of home-based rehabilitation services for people needing rehabilitation: a qualitative evidence synthesis. The Cochrane database of systematic reviews. 2023;2(2):Cd014823.
14. van der Veen DJ, Döpp CME, Siemonsma PC, Nijhuis-van der Sanden MWG, de Swart BJM, Steultjens EM. Factors influencing the implementation of Home-Based Stroke Rehabilitation: Professionals' perspective. PLoS One. 2019;14(7):e0220226.
15. Langerak AJ, Regterschot GRH, Selles RW, Meskers CGM, Evers M, Ribbers GM, et al. Requirements for home-based upper extremity rehabilitation using wearable motion sensors for stroke patients: a user-centred approach. Disabil rehabil, Assist. 2023;technol..1-13.
16. Kim GJ, Parnandi A, Eva S, Schambra H. The use of wearable sensors to assess and treat the upper extremity after stroke: a scoping review. Disabil Rehabil. 2022;44(20):6119-38.
17. Zheng H, Davies R, Zhou H, Hammerton J, Mawson S, Ware PM, et al. SMART project: Application of emerging information and communication technology to home-based

rehabilitation for stroke patients. International Journal on Disability and Human Development. 2006;5.
18. Brennan DM, Mawson S, Brownsell S. Telerehabilitation: enabling the remote delivery of healthcare, rehabilitation, and self management. Stud Health Technol Inform. 2009;145:231-48.
19. MC E. Arm coach for stroke 2024 [Available from: https://www.erasmusmc.nl/en/research/projects/arm-coach-for-stroke#b9598a65-812f-4f9b-847a-17c545d39ca5.
20. De Vet H, Terwee, C., Mokkink, L., Knol, D. Measurement in Medicine: Practical Guides to Biostatistics and Epidemiology: Cambridge University Press; 2011.
21. Lo EH, Dalkara T, Moskowitz MA. Mechanisms, challenges and opportunities in stroke. Nature Reviews Neuroscience. 2003;4(5):399-414.
22. Kuriakose D, Xiao Z. Pathophysiology and Treatment of Stroke: Present Status and Future Perspectives. International journal of molecular sciences. 2020;21(20).
23. Salaudeen MA, Bello N, Danraka RN, Ammani ML. Understanding the Pathophysiology of Ischemic Stroke: The Basis of Current Therapies and Opportunity for New Ones. Biomolecules. 2024;14(3).
24. Foundation SA. Stroke Facts & Statistics 2024 [Available from: https://www.strokeinfo.org/stroke-facts-statistics/.
25. Üstün T, WHO Family of International Classifications (WHO-FIC). The ICF: An Overview.
26. Raghavan P. Upper Limb Motor Impairment After Stroke. Physical medicine and rehabilitation clinics of North America. 2015;26(4):599-610.
27. Levin MF, Kleim JA, Wolf SL. What Do Motor "Recovery" and "Compensation" Mean in Patients Following Stroke? Neurorehabilitation and Neural Repair. 2008;23(4):313-9.
28. Cirstea MC, Levin MF. Compensatory strategies for reaching in stroke. Brain. 2000;123(5):940-53.
29. Michaelsen SM, Dannenbaum R, Levin MF. Task-specific training with trunk restraint on arm recovery in stroke: randomized control trial. Stroke. 2006;37(1):186-92.
30. Shehjar F, Maktabi B, Rahman ZA, Bahader GA, James AW, Naqvi A, et al. Stroke: Molecular mechanisms and therapies: Update on recent developments. Neurochemistry International. 2023;162:105458.
31. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. International Journal of Rehabilitation Research. 1981;4(4):483-92.
32. Lang CE, Wagner JM, Dromerick AW, Edwards DF. Measurement of Upper-Extremity Function Early After Stroke: Properties of the Action Research Arm Test. Archives of Physical Medicine and Rehabilitation. 2006;87(12):1605-10.
33. Gladstone DJ, Danells CJ, Black SE. The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties. Neurorehabil Neural Repair. 2002;16(3):232-40.
34. Know Stroke NIoNDaS. NIH Stroke Scale [Available from: https://www.ninds.nih.gov/sites/default/files/2024-05/KnowStroke_NIHStrokeScale_May2024_508c.pdf.
35. Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, et al. Standardized Measurement of Sensorimotor Recovery in Stroke Trials: Consensus-Based Core Recommendations from the Stroke Recovery and Rehabilitation Roundtable. Neurorehabil Neural Repair. 2017;31(9):784-92.
36. COSMIN. COSMIN definitions of domains, measurement properties and aspects of measurement properties [Available from: https://cosmin.nl/wp-content/uploads/COSMIN-definitions-domains-measurement-properties.pdf.
37. Mokkink LB, Knol DL, Zekveld AA, Goverts ST, Kramer SE. Factor structure and reliability of the Dutch version of seven scales of the Communication Profile for the Hearing Impaired (CPHI). Journal of speech, language, and hearing research : JSLHR. 2009;52(2):454-64.
38. Intraclass correlations: Uses in assessing rater reliability, (1979).

39. De Vet H, Terwee, C., Mokkink, L., Knol, D. Measurement in Medicine: A practical guide. Cambridge University Press. 2011.
40. Spaaks MJ. "What are the current methods and applications of wearable motion sensors during (home-based) upper extremity rehabilitation for stroke patients?". Scoping Literature Review - Technical Medicine Thesis. 2024.

41. Ahmad N, Ghazilla RABR, Khairi NM, Kasi V, editors. Reviews on Various Inertial Measurement Unit (IMU) Sensor Applications. IEEE Workshop on Signal Processing Systems; 2013.
42. Ribeiro NF, Santos CP, editors. Inertial measurement units: A brief state of the art on gait analysis. 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG); 2017 16-18 Feb. 2017.
43. Brain M, Bowie., D. How the Gyroscope works 2023 [Available from: https://science.howstuffworks.com/gyroscope.htm.
44. Martino Cinnera A, Picerno P, Bisirri A, Koch G, Morone G, Vannozzi G. Upper limb assessment with inertial measurement units according to the international classification of functioning in stroke: a systematic review and correlation meta-analysis. Topics in Stroke Rehabilitation. 2024;31(1):66-85.
45. Rong S. Quantifications and Characteristics of Dynamic Soft Tissue Artifacts Captured by Wearable Inertial Measurement Unit Sensors: The University of British Columbia; 2020.

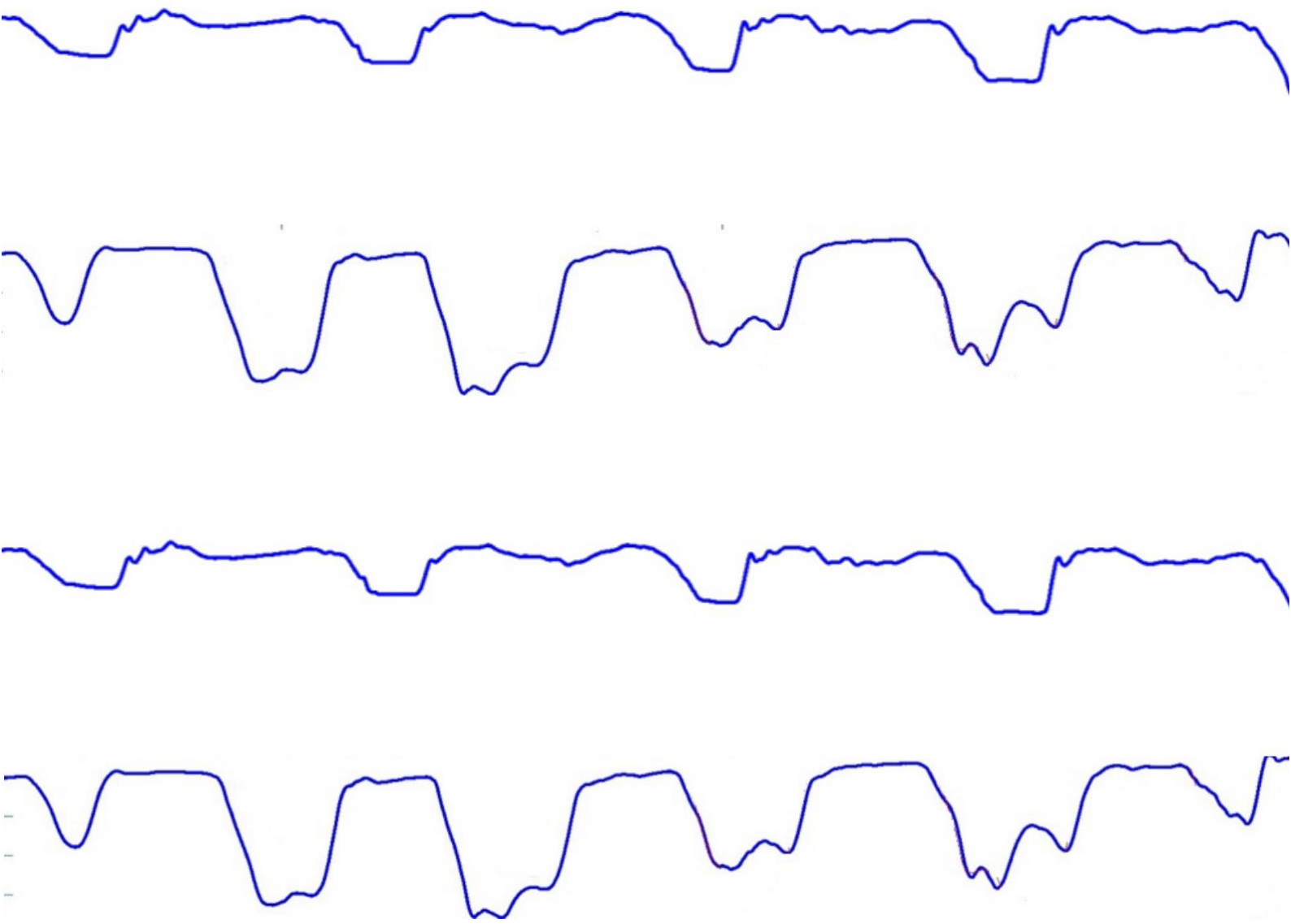
46. Sessa S, Zecca M, Lin Z, Bartolomeo L, Ishii H, Takanishi A. A Methodology for the Performance Evaluation of Inertial Measurement Units. Journal of Intelligent & Robotic Systems. 2013;71(2):143-57.
47. Bhagubai MMC, Wolterink G, Schwarz A, Held JPO, Van Beijnum BF, Veltink PH. Quantifying Pathological Synergies in the Upper Extremity of Stroke Subjects With the Use of Inertial Measurement Units: A Pilot Study. IEEE J Transl Eng Health Med. 2021;9:2100211.
48. Iuppariello L, D'Addio G, Lanzillo B, Balbi P, Andreozzi E, Improta G, et al. A novel approach to estimate the upper limb reaching movement in three-dimensional space. Informatics in Medicine Unlocked. 2019;15:100155.
49. Saes M, Mohamed Refai MI, van Kordelaar J, Scheltinga BL, van Beijnum BF, Bussmann JBJ, et al. Smoothness metric during reach-to-grasp after stroke: part 2. longitudinal association with motor impairment. J Neuroeng Rehabil. 2021;18(1):144.
50. Ozkaya G, Jung HR, Jeong IS, Choi MR, Shin MY, Lin X, et al. Three-dimensional motion capture data during repetitive overarm throwing practice. Scientific Data. 2018;5(1):180272.
51. NEXUS V. Plug-In Gait Reference Guide 2023 [Available from: https://help.vicon.com/download/attachments/11378719/Plug-in%20Gait%20Reference%20Guide.pdf.
52. Hughes CML, Louie A, Sun S, Gordon-Murer C, Belay GJ, Baye M, et al. Development of a Post-stroke Upper Limb Rehabilitation Wearable Sensor for Use in Sub-Saharan Africa: A Pilot Validation Study. Frontiers in bioengineering and biotechnology. 2019;7:322.
53. Cai L, Ma Y, Xiong S, Zhang Y. Validity and Reliability of Upper Limb Functional Assessment Using the Microsoft Kinect V2 Sensor. Applied Bionics and Biomechanics. 2019;2019:1-14.
54. Poldervaart P. Accuracy of metrics measured with IMU

sensors during a reach-to-point movement in

healthy adults: University of Twente; 2022.
55. Das K, de Paula Oliveira T, Newell J. Comparison of markerless and marker-based motion capture systems using 95% functional limits of agreement in a linear mixed-effects modelling framework. Scientific Reports. 2023;13(1):22880.
56. Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Annals of cardiac anaesthesia. 2019;22(1):67-72.
57. Krithikadatta J. Normal distribution. Journal of conservative dentistry : JCD. 2014;17(1):96-7.

58. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism. 2012;10(2):486-9.
59. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of chiropractic medicine. 2016;15(2):155-63.
60. Ries JD, Echternach JL, Nof L, Gagnon Blodgett M. Test-Retest Reliability and Minimal Detectable Change Scores for the Timed "Up & Go" Test, the Six-Minute Walk Test, and Gait Speed in People With Alzheimer Disease. Physical Therapy. 2009;89(6):569-79.
61. Roby-Brami A, Feydy A, Combeaud M, Biryukova EV, Bussel B, Levin MF. Motor compensation and recovery for reaching in stroke patients. Acta neurologica Scandinavica. 2003;107(5):369-81.
62. Merlau B, Cormier C, Alaux A, Morin M, Montané E, Amarantini D, et al. Assessing Spatiotemporal and Quality Alterations in Paretic Upper Limb Movements after Stroke in Routine Care: Proposal and Validation of a Protocol Using IMUs versus MoCap. Sensors (Basel). 2023;23(17).
63. Al-Amri M, Nicholas K, Button K, Sparkes V, Sheeran L, Davies JL. Inertial Measurement Units for Clinical Movement Analysis: Reliability and Concurrent Validity. Sensors (Basel). 2018;18(3).
64. McGinley JL, Baker R, Wolfe R, Morris ME. The reliability of three-dimensional kinematic gait measurements: a systematic review. Gait & posture. 2009;29(3):360-9.
65. Morrow MMB, Lowndes B, Fortune E, Kaufman KR, Hallbeck MS. Validation of Inertial Measurement Units for Upper Body Kinematics. Journal of applied biomechanics. 2017;33(3):227-32.
66. El-Gohary M, McNames J. Shoulder and Elbow Joint Angle Tracking With Inertial Sensors. IEEE Transactions on Biomedical Engineering. 2012;59(9):2635-41.
67. Skogstad S, Nymoen K, Quay Y, Jensenius A. OSC Implementation and Evaluation of the Xsens MVN suit. Proceedings of New Interfaces for Music Expression. 2011.
68. Fiorentino NM, Atkins PR, Kutschke MJ, Goebel JM, Foreman KB, Anderson AE. Soft tissue artifact causes significant errors in the calculation of joint angles and range of motion at the hip. Gait & posture. 2017;55:184-90.
69. Fang Z, Woodford S, Senanayake D, Ackland D. Conversion of Upper-Limb Inertial Measurement Unit Data to Joint Angles: A Systematic Review. 2023;23(14):6535.
70. Bujang MA. An elaboration on sample size determination for correlations based on effect sizes and confidence interval width: a guide for researchers. Restorative dentistry & endodontics. 2024;49(2):e21.
71. Dwivedi AK, Mallawaarachchi I, Alvarado LA. Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method. Statistics in medicine. 2017;36(14):2187-205.
72. Oliva Lozano J, Martín Fuentes I, Muyor J. Validity and Reliability of a New Inertial Device for Monitoring Range of Motion at the Pelvis during Sexual Intercourse. International Journal of Environmental Research and Public Health. 2020;17.
73. Brice SM, Phillips EJ, Millett EL, Hunter A, Philippa B. Comparing inertial measurement units and marker-based biomechanical models during dynamic rotation of the torso. European journal of sport science. 2020;20(6):767-75.
74. Hughes CML, Tran B, Modan A, Zhang X. Accuracy and Validity of a Single Inertial Measurement Unit-Based System to Determine Upper Limb Kinematics for Medically Underserved Populations. Frontiers in bioengineering and biotechnology. 2022;10:918617.
75. Lin LF, Lin YJ, Lin ZH, Chuang LY, Hsu WC, Lin YH. Feasibility and efficacy of wearable devices for upper limb rehabilitation in patients with chronic stroke: a randomized controlled pilot study. Eur J Phys Rehabil Med. 2018;54(3):388-96.
76. Chae SH, Kim Y, Lee KS, Park HS. Development and Clinical Evaluation of a Web-Based Upper Limb Home Rehabilitation System Using a Smartwatch and Machine Learning Model for Chronic Stroke Survivors: Prospective Comparative Study. JMIR Mhealth Uhealth. 2020;8(7):e17216.

Appendices



Appendix A: Indices selection

The following two figures further clarify the trial indices selection of the reliability protocol (Figure A.1) and phases present in the IMU signal of the validity protocol (Figure A.2).

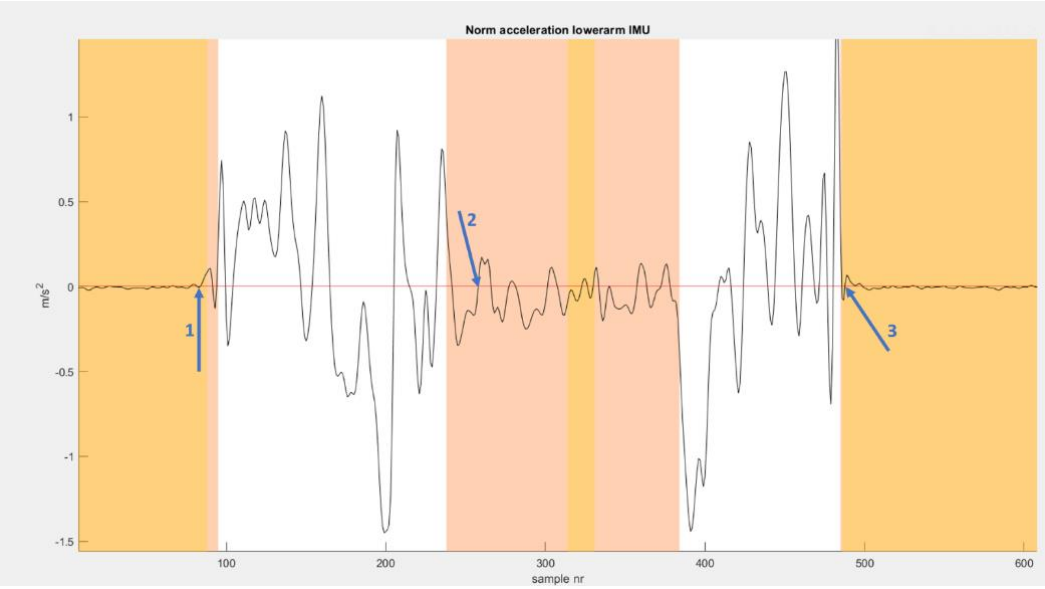


Figure A.1: Example of trial indices selection of the head exercise in the acceleration signal of the reliability protocol. 1 = start of reaching movement, 2 = end of reaching movement, 3 = return to initial position. Yellow areas indicate periods of minimal movement.

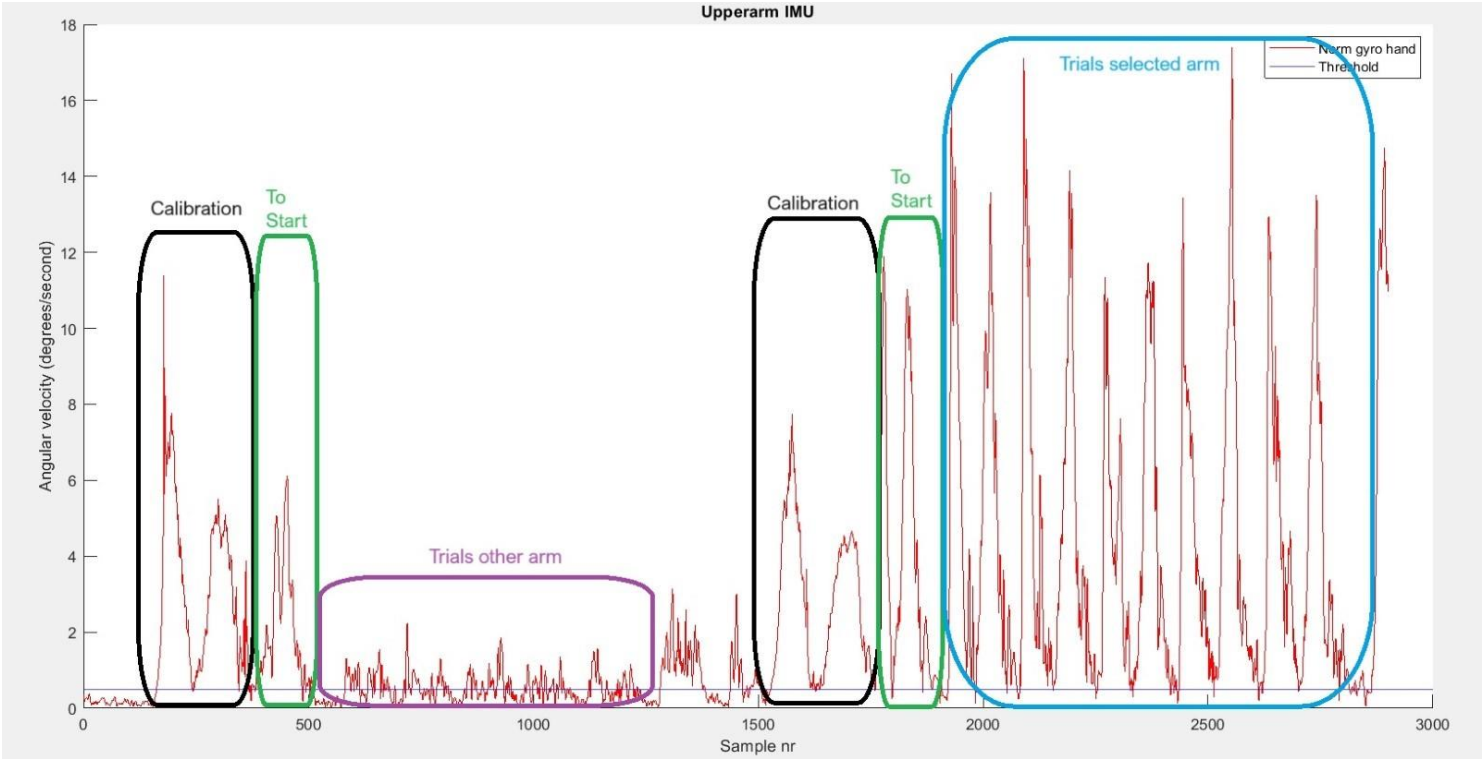


Figure A.2: Example of an upper arm signal of the validity protocol, with several movement phases: calibration movement (black), movement to start position (green), trials of opposite extremity (purple), trials of the currently selected extremity (blue).

Appendix B: Synchronicity analysis

Since unexpected asynchrony between individual IMU-sensors was observed during the first preprocessing steps and face validity checks of sensor data of the reliability protocol, multiple synchronicity tests were conducted to investigate the occurrence of asynchrony and identify potential influencing factors. These experiments and their results are briefly described here.

Experimental methods:

Tests with data of own experiments

Firstly, four IMUs were attached to the object and lifted five times, three times in a row. In between these three sessions, the IMUs were switched off to also assess the influence of a restart on possible asynchrony. The data was analysed using Matlab's "xcorr" function to investigate the normalized cross-correlation between two sensors. By assessing the resulting figures, the maximum "spike" is identified, indicating the lag indices at which maximum correlation of sensor data occurs. Perfect synchronization would show a maximum spike at a lag of zero.

Furthermore, the influence of the sequence of sensor activation on the asynchrony was assessed. The IMUs were activated and attached to the object using two different sequences of activation. Then, the IMU data was plotted to investigate if the chronological sequence of the asynchronous signals matched the sequence of activation.

The consistency of asynchrony was investigated, by comparison of the cross-correlation lags at maximum correlation in two different phases. In the "object phase", four IMUs were attached to the object and shaken. In the "movement phase", sensors were attached to the lower arm, upper arm, and shoulder of the investigator, who performed pronation, supination, and trunk flexion movements. In a separate session, the consistency of lags was investigated by three repetitions of a movement of the object with the four IMU sensors attached to it. Every repetition was selected as a separate phase (so three times an "object phase"), and for every phase the lags of sensor combinations were calculated, to investigate consistency without anatomical influences.

Tests with patient data

The IMUs cross-correlation of patient data were assessed. Lags of the whole data segment of trial execution (between trial indices 1 and 3) and of only the movement phase of the trial were described for P05 and P07 (head exercise) of the reliability protocol. The movement phase was defined as the part of the trial at which all sensors were expected to move (between trial indices 1 and 2).

Experimental results:

Tests with laboratory data

Visual inspection of data of multiple experimental sessions revealed that signal asynchrony was present, even when the sensors were attached to the object and moved simultaneously. The lag of maximum cross-correlation between the IMU signals varied across sessions where the sensors were switched off in between (Figure B.1). Lags appeared to vary from zero to ten samples. The order in which the signals were turned on did not appear to have a direct relationship with the occurrence of asynchrony. The lag resulting from cross-correlation at multiple phases during a single measurement appeared to be quite consistent, with an estimated difference in lags of only one sample (Figure B.2).

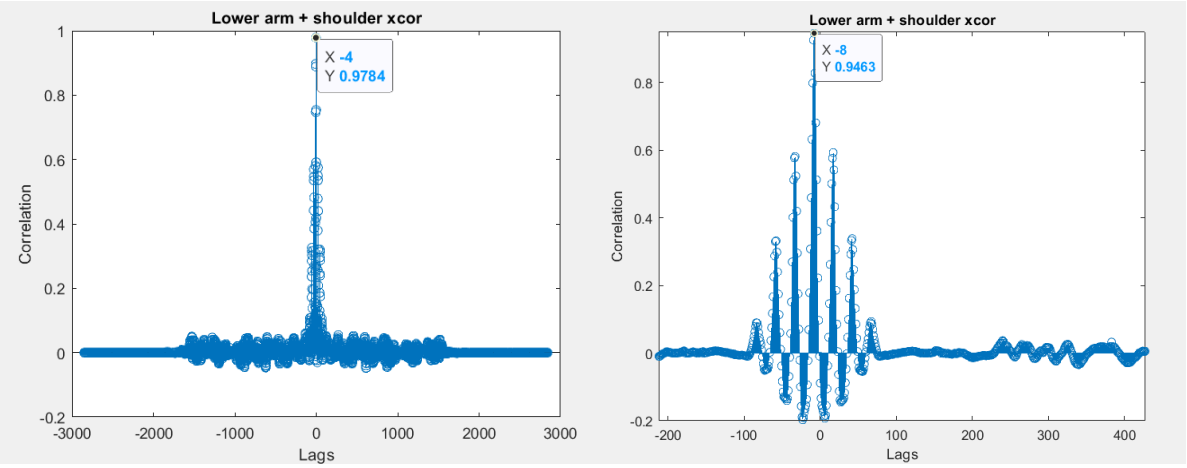


Figure B.1: Examples of cross-correlation for the two IMUs "lower arm" and "shoulder" during session 1 (left image) and session 3 (right image). The lag at which maximum cross-correlation occurred is visible as the X-coordinate: -4 and -8 respectively.

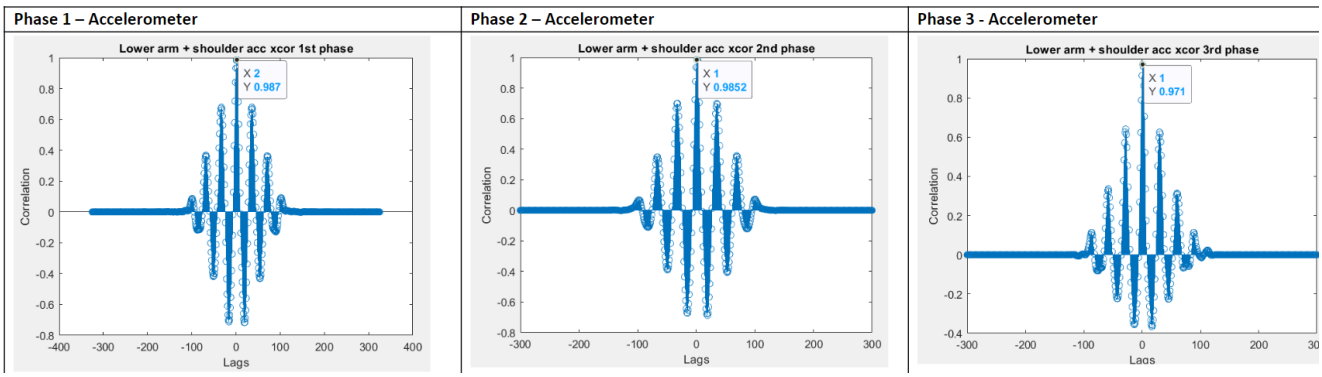


Figure B.2: Examples of the cross-correlation during three different phases of a single measurement. The lag at which maximum cross-correlation occurred is visible as the X-coordinate: 2, 1 and 1 respectively.

Tests with patient data

For both P05 and P07 of the reliability protocol, the lags at which maximum cross-correlation occurred were inconsistent throughout the entire trial or movement phase of the trial (Figure B.3). Due to this inconsistency in asynchrony during patient trials and between sessions, and the absence of any phase within the measurements where the sensors were known to move simultaneously, it was concluded that no asynchrony correction could be applied before further processing of the reliability protocol data.

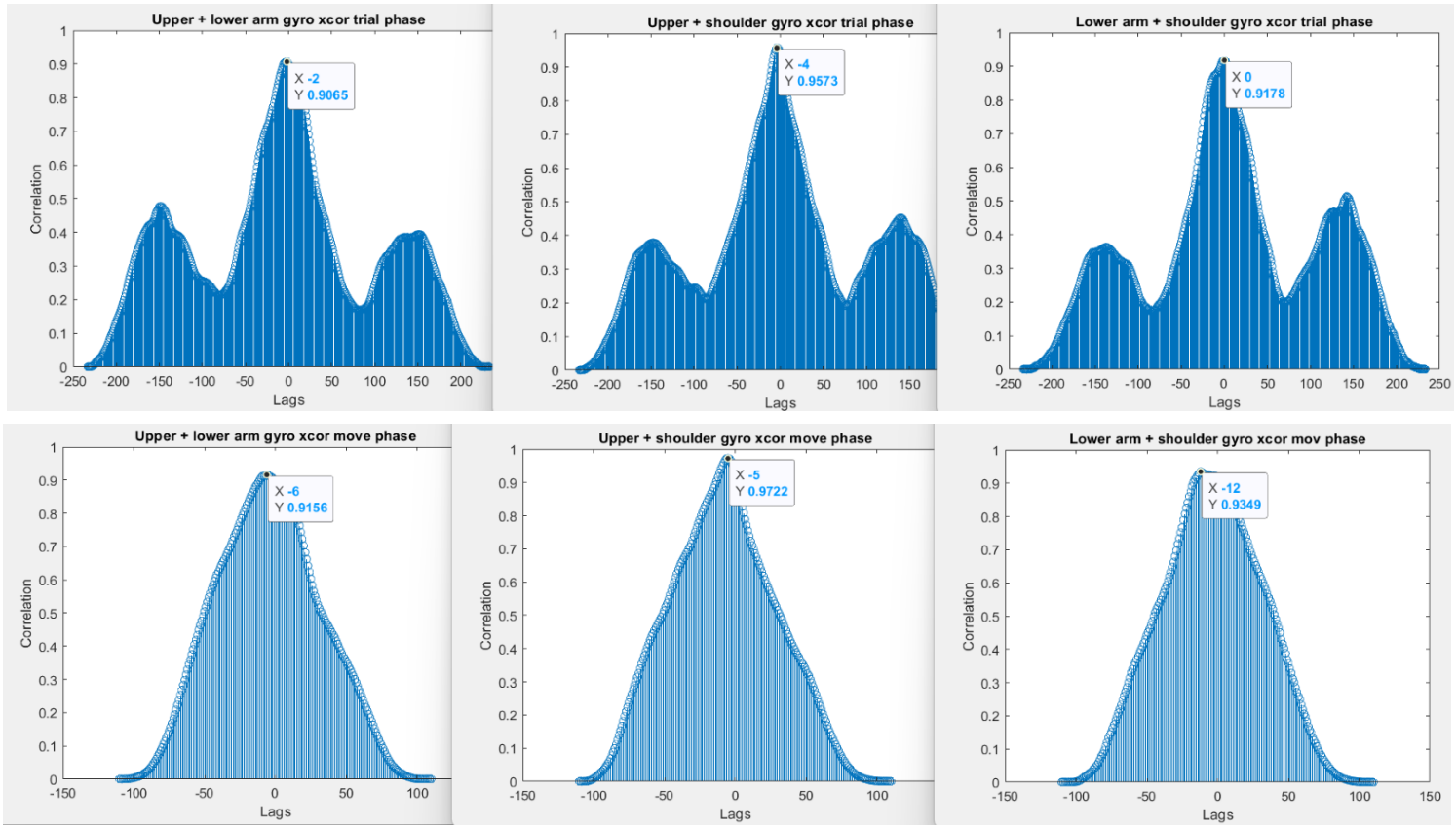


Figure B.3: The lags at which maximum cross-correlation between different IMU sensors occurred for the trial phase (above) or the movement phase (below) of data of P05 (head exercise) were not consistent. The lags are shown as the X-coordinates for multiple sensor combinations, the normalized cross-correlation is shown as the Y-coordinate.

Appendix C: Shapiro-Wilk normality test (reliability)

All the statistical scripts were uploaded to the GitHub repository and can be accessed via the following (private) link, which is restricted to authorised users only: https://github.com/mspaaks/AC4S . In case permission is required, please contact Marije Spaaks (m.spaaks@erasmusmc.nl).

The following tables show the results of the Shapiro-Wilk normality test for the reliability protocol. The hypothesis of normality is rejected if the p-value is ≤0.05.

*Table C.1: Results of the Shapiro-Wilk test W-statistic assessing normality of the lift and head exercise, **joint angle range**, for the reaching and reach-and-return phase, both the Test and Retest session. The corresponding p-values are also shown.*

Metric (range)	Session	Lift Exercise		Head Exercise	
		W	p-value	W	p-value
ReachReturn – Elbow Flexion	Test	0,75	0,00	0,82	0,00
ReachReturn – Shoulder Abduction	Test	0,93	0,32	0,85	0,01
ReachReturn – Shoulder Flexion	Test	0,96	0,67	0,65	0,00
ReachReturn – Shoulder Elevation	Test	0,77	0,00	0,66	0,00
ReachReturn – Trunk Flexion	Test	0,82	0,01	0,90	0,06
Reach – Elbow Flexion	Test	0,77	0,00	0,83	0,00
Reach – Shoulder Abduction	Test	0,94	0,39	0,90	0,06
Reach – Shoulder Flexion	Test	0,94	0,43	0,69	0,00
Reach – Shoulder Elevation	Test	0,83	0,01	0,61	0,00
Reach – Trunk Flexion	Test	0,81	0,00	0,87	0,02
ReachReturn – Elbow Flexion	Retest	0,74	0,00	0,80	0,00
ReachReturn – Shoulder Abduction	Retest	0,95	0,48	0,86	0,01
ReachReturn – Shoulder Flexion	Retest	0,97	0,86	0,96	0,70
ReachReturn – Shoulder Elevation	Retest	0,71	0,00	0,76	0,00
ReachReturn – Trunk Flexion	Retest	0,91	0,14	0,76	0,00
Reach – Elbow Flexion	Retest	0,73	0,00	0,80	0,00
Reach – Shoulder Abduction	Retest	0,90	0,09	0,92	0,12
Reach – Shoulder Flexion	Retest	0,94	0,38	0,94	0,34
Reach – Shoulder Elevation	Retest	0,78	0,00	0,76	0,00
Reach – Trunk Flexion	Retest	0,84	0,01	0,75	0,00

Table C.2: Results of the Shapiro-Wilk test W-statistic assessing normality of the lift and head exercise, ***mean joint angle****, for the reaching and reach-and-return phase, Test and Retest session.*

Metric (mean)	Session	Lift exercise		Head exercise	
		W	p-value	W	p-value
ReachReturn – Elbow Flexion	Test	0,84	0,01	0,89	0,04
ReachReturn – Shoulder Abduction	Test	0,96	0,67	0,99	1,00
ReachReturn – Shoulder Flexion	Test	0,96	0,72	0,93	0,19
ReachReturn – Shoulder Elevation	Test	0,75	0,00	0,68	0,00
ReachReturn – Trunk Flexion	Test	0,93	0,28	0,91	0,10
Reach – Elbow Flexion	Test	0,83	0,01	0,85	0,01
Reach – Shoulder Abduction	Test	0,94	0,40	0,98	0,94
Reach – Shoulder Flexion	Test	0,96	0,75	0,97	0,72
Reach – Shoulder Elevation	Test	0,67	0,00	0,63	0,00
Reach – Trunk Flexion	Test	0,98	0,93	0,96	0,52
ReachReturn – Elbow Flexion	Retest	0,83	0,01	0,97	0,87
ReachReturn – Shoulder Abduction	Retest	0,96	0,74	0,98	0,95
ReachReturn – Shoulder Flexion	Retest	0,94	0,36	0,98	0,91
ReachReturn – Shoulder Elevation	Retest	0,74	0,00	0,74	0,00
ReachReturn – Trunk Flexion	Retest	0,85	0,02	0,97	0,87
Reach – Elbow Flexion	Retest	0,86	0,03	0,98	0,96
Reach – Shoulder Abduction	Retest	0,88	0,04	0,96	0,59
Reach – Shoulder Flexion	Retest	0,94	0,40	0,98	0,96
Reach – Shoulder Elevation	Retest	0,71	0,00	0,71	0,00
Reach – Trunk Flexion	Retest	0,90	0,08	0,98	0,89

Appendix D: Test-Retest reliability

Table D.1: This table shows the calculated reliability results of the joint angle range (in degrees) during the reaching phase: Test and retest median (+ IQR), median absolute test-retest difference (+IQR), Wilcoxon test results, ICC, SEM, MDC.

	Reliability protocol – Range (Reaching phase)						
Metric	**Test Median (IQR)**	**Retest Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Retest vs Test**	**ICC agree-ment**	**SEM**	**MDC**
Reach, grasp, lift and hold object – Lift exercise							
Elbow Flexion	50.50 (138.81)	52.47 (90.57)	11.74 (34.00)	0.84	0.75	43.73	121.21
Shoulder Abduction	12.59 (11.46)	13.40 (6.79)	4.93 (5.82)	0.67	0.39	5.43	15.06
Shoulder Flexion	25.65 (8.06)	19.75 (9.80)	6.05 (4.24)	0.35	0.70	5.39	14.94
Shoulder Elevation	76.56 (200.27)	24.96 (281.80)	12.88 (100.85)	0.84	0.70	72.67	201.44
Trunk Flexion	77.93 (45.74)	54.19 (24.10)	30.19 (50.44)	0.11	0.06	48.09	133.29
Reach to the back of the head– Head exercise							
Elbow Flexion	81.15 (167.78)	181.68 (269.16)	40.56 (123.74)	0.06	0.60	79.80	221.19
Shoulder abduction	39.83 (24.05)	30.46 (31.91)	7.66 (13.93)	0.16	0.59	13.55	37.55
Shoulder Flexion	45.92 (39.84)	28.82 (20.53)	12.74 (21.76)	0.02	0.04	35.10	97.28
Shoulder Elevation	27.72 (70.92)	66.26 (230.79)	11.44 (111.80)	0.06	0.49	96.53	267.58
Trunk Flexion	49.22 (66.45)	62.32 (47.87)	14.72 (34.12)	0.04	0.52	45.05	124.88

Abbreviations: IQR = interquartile range, ICC = intraclass correlation coefficient, SEM = standard error of measurement, MDC = minimal detectable change

Table D.2: This table shows the calculated reliability results of the mean joint angle (in degrees) during the reaching phase: Test and retest median (+IQR), median absolute test-retest difference (+IQR), Wilcoxon test results, ICC, SEM, MDC.

	Reliability protocol – Mean (Reaching phase)						
Metric	**Test Median (IQR)**	**Retest Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Retest vs Test**	**ICC agree-ment**	**SEM**	**MDC**
Reach, grasp, lift and hold object – Lift exercise							
Elbow Flexion	72.30 (108.66)	78.70 (149.98)	14.59 (11.84)	0.04	0.77	37.93	105.14
Shoulder abduction	-10.27 (13.50)	-16.29 (24.89)	7.46 (7.19)	0.13	0.60	8.15	22.59
Shoulder Flexion	9.17 (28.45)	7.05 (26.22)	14.35 (10.69)	0.06	0.44	15.77	43.70
Shoulder Elevation	-158.80 (82.49)	-161.61 (107.07)	9.21 (34.89)	0.71	0.49	76.80	212.87
Trunk Flexion	-18.23 (75.28)	-61.97 (81.11)	19.14 (38.43)	0.05	0.51	38.27	106.08
Reach to the back of the head– Head exercise							
Elbow Flexion	78.70 (149.98)	34.64 (49.84)	19.30 (43.14)	0.90	0.50	41.92	116.19
Shoulder abduction	-16.29 (24.89)	-17.65 (14.94)	11.83 (21.65)	0.54	0.51	16.98	47.07
Shoulder Flexion	7.05 (26.22)	9.17 (28.91)	10.47 (12.34)	0.86	0.74	13.05	36.18
Shoulder Elevation	-161.61 (107.07)	-151.73 (19.70)	6.79 (31.17)	0.08	0.42	38.02	105.38
Trunk Flexion	-61.97 (81.11)	-10.89 (32.27)	19.20 (14.78)	0.83	0.75	17.22	47.74

Abbreviations: IQR = interquartile range, ICC = intraclass correlation coefficient, SEM = standard error of measurement, MDC = minimal detectable change

Table D.3: This table shows the calculated reliability results of the joint angle range (in degrees) during the reach-and-return phase: Test and retest median (+IQR), median absolute test-retest difference (+IQR), Wilcoxon test results, ICC, SEM, MDC.

	Reliability protocol – Range (Reach-and-return phase)						
Metric	**Test Median (IQR)**	**Retest Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Retest vs Test**	**ICC agree-ment**	**SEM**	**MDC**
Reach, grasp, lift and hold object – Lift exercise							
Elbow Flexion	64.58 (181.45)	57.08 (105.53)	14.27 (46.19)	0.67	0.84	42.41	117.56
Shoulder abductio n	20.93 (12.09)	19.45 (11.30)	4.46 (9.27)	0.93	0.31	7.62	21.13
Shoulder Flexion	32.57 (16.78)	30.98 (14.52)	2.90 (5.27)	0.32	0.78	5.89	16.33
Shoulder Elevatio n	78.09 (329.68)	25.41 (346.47)	5.00 (24.50)	0.44	0.74	82.45	228.55
Trunk Flexion	79.53 (42.03)	67.49 (24.22)	31.10 (64.25)	0.11	0.35	45.25	125.44
Reach to the back of the head– Head exercise							
Elbow Flexion	105.65 (220.30)	195.42 (265.58)	33.93 (51.40)	0.28	0.73	66.01	182.97
Shoulder abductio n	45.87 (25.27)	34.27 (35.31)	13.49 (22.81)	0.21	0.53	19.99	55.40
Shoulder Flexion	56.20 (41.01)	34.88 (16.72)	12.91 (15.44)	0.00	0.03	43.75	121.26
Shoulder Elevatio n	32.43 (83.90)	69.12 (304.45)	9.48 (107.83)	0.05	0.61	88.43	245.11
Trunk Flexion	64.79 (62.04)	77.71 (57.78)	12.65 (24.00)	0.10	0.57	46.19	128.04

Abbreviations: IQR = interquartile range, ICC = intraclass correlation coefficient, SEM = standard error of measurement, MDC = minimal detectable change

Table D.4: This table shows the calculated reliability results of the mean joint angle (in degrees) during the reach-and-return phase: Test and retest median (+IQR), median absolute test-retest difference (+IQR), Wilcoxon test results, ICC, SEM, MDC.

	Reliability protocol – Mean (Reach-and-return phase)						
Metric	**Test Median (IQR)**	**Retest Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Retest vs Test**	**ICC agree-ment**	**SEM**	**MDC**
Reach, grasp, lift and hold object – Lift exercise							
Elbow Flexion	81.19 (114.46)	86.49 (128.33)	12.64 (13.99)	0.29	0.87	30.03	83.25
Shoulder abduction	-9.74 (13.35)	-15.92 (23.52)	6.65 (13.92)	0.32	0.54	9.66	26.78
Shoulder Flexion	15.12 (35.40)	11.09 (37.42)	12.51 (13.10)	0.07	0.59	14.76	40.92
Shoulder Elevation	-155.41 (82.53)	-159.93 (90.54)	8.16 (64.44)	0.75	0.52	58.63	162.52
Trunk Flexion	-29.37 (78.31)	-71.13 (68.39)	16.84 (25.33)	0.11	0.54	35.29	97.81
Reach to the back of the head– Head exercise							
Elbow Flexion	86.49 (128.33)	23.32 (65.82)	24.23 (57.68)	0.57	0.60	38.03	105.40
Shoulder abduction	-15.91 (23.16)	-18.62 (14.57)	14.59 (20.74)	0.76	0.42	20.41	56.58
Shoulder Flexion	11.09 (37.42)	15.22 (25.37)	8.65 (11.25)	0.57	0.75	13.82	38.30
Shoulder Elevation	-159.93 (90.54)	-154.24 (22.97)	4.73 (14.17)	0.02	0.62	18.74	51.94
Trunk Flexion	-71.13 (68.39)	-8.03 (35.31)	20.61 (12.43)	0.97	0.72	16.01	44.39

Abbreviations: IQR = interquartile range, ICC = intraclass correlation coefficient, SEM = standard error of measurement, MDC = minimal detectable change

Appendix E: Criterion validity

Table E.1: This table shows the calculated validity results of the joint angle range (in degrees) during the reaching phase: IMU and Vicon median (+IQR), median absolute IMU-Vicon difference (+IQR), Wilcoxon test results and Spearman correlation (+ p-value). Both methods for shoulder angle calculations are shown.

	Validity protocol – Range (Reaching phase)				
Metric	**IMU Median (IQR)**	**Vicon Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Vicon vs IMU, p-value**	**Spearman correlation Vicon vs IMU (p-value)**
Reach to touch object – Touch exercise					
Trunk flexion	4.35 (1.24)	4.07 (1.33)	0.59 (0.35)	0.86	0.4 (0.75)
Elbow flexion	46.97 (14.22)	35.37 (3.06)	12.26 (8.41)	0.20	0.2 (0.92)
ShoulderUsing Sternum - flexion	45.15 (8.75)	37.64 (10.18)	5.67 (2.31)	0.10	1 (0.083)
ShoulderUsing Sternum - abduction	10.86 (3.82)	11.89 (1.99)	3.77 (1.65)	0.58	-0.2 (0.92)
ShoulderUsing Sternum- rotation	10.02 (3.51)	28.89 (1.85)	16.81 (5.78)	0.10	0 (1)
ShoulderUsing Shoulder- flexion	37.81 (5.97)	37.64 (10.18)	2.44 (0.80)	0.58	1 (0.083)
ShoulderUsing Shoulder - abduction	10.52 (4.50)	11.89 (1.99)	5.85 (1.43)	1	0.4 (0.75)
ShoulderUsing Shoulder - rotation	18.40 (2.78)	28.89 (1.85)	10.49 (4.34)	0.10	-0.8 (0.33)
Reach, grasp, lift and hold object - Lift exercise					
Trunk flexion	10.71 (1.77)	9.46 (1.41)	1.26 (0.36)	0.10	1 (0.083)
Elbow flexion	62.92 (15.39)	44.54 (8.93)	17.12 (10.57)	0.10	0.4 (0.75)
ShoulderUsing Sternum - flexion	48.08 (8.67)	40.99 (15.20)	3.75 (3.92)	0.10	0.8 (0.33)
ShoulderUsing Sternum - abduction	16.79 (5.60)	9.51 (7.71)	10.33 (6.43)	0.86	0.6 (0.42)
ShoulderUsing Sternum- rotation	19.06 (13.33)	39.05 (19.66)	20.00 (6.34)	0.10	1 (0.083)

ShoulderUsing Shoulder-flexion	43.73 (5.47)	40.99 (15.20)	6.73 (4.75)	0.86	0.4 (0.75)
ShoulderUsing Shoulder - abduction	14.32 (5.32)	9.51 (7.71)	7.73 (7.18)	0.86	0.6 (0.42)
ShoulderUsing Shoulder - rotation	27.41 (13.54)	39.05 (19.66)	13.86 (3.90)	1	1 (0.083)
Reach to the back of the head – Head exercise					
Trunk flexion	9.23 (4.28)	2.41 (0.22)	6.81 (4.49)	0.42	-1 (0.33)
Elbow flexion	187.67 (120.92)	59.73 (12.21)	123.73 (110.81)	0.18	0.5 (1)
ShoulderUsing Sternum - flexion	102.98 (23.22)	25.33 (14.72)	59.50 (23.00)	0.18	0.5 (1)
ShoulderUsing Sternum - abduction	50.95 (8.47)	66.22 (17.29)	27.67 (15.72)	0.18	0.5 (1)
ShoulderUsing Sternum-rotation	76.17 (38.46)	100.35 (15.90)	27.62 (23.19)	0.42	0.5 (1)
ShoulderUsing Shoulder-flexion	59.59 (6.77)	25.33 (14.72)	25.63 (17.18)	0.18	-0.5 (1)
ShoulderUsing Shoulder - abduction	35.60 (14.56)	66.22 (17.29)	41.74 (8.29)	0.18	1 (0.33)
ShoulderUsing Shoulder - rotation	25.89 (12.35)	100.35 (15.90)	56.43 (12.56)	0.18	0.5 (1)

Abbreviations: IQR = interquartile range, IMU = Inertial Measurement Unit

Table E.2: This table shows the calculated validity results of the mean joint angle (in degrees) during the reaching phase: IMU and Vicon median (+ interquartile range, IQR), median absolute IMU-Vicon difference (+IQR), Wilcoxon test results and Spearman correlation (+ p-value). Both methods for shoulder angle calculations are shown.

	Validity protocol – Mean (Reaching phase)				
Metric	**IMU Median (IQR)**	**Vicon Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Vicon vs IMU, p-value**	**Spearman correlation Vicon vs IMU (p-value)**
Reach to touch object – Touch exercise					
Trunk flexion	4.05 (3.58)	12.78 (4.11)	7.84 (4.19)	0.10	0.4 (0.75)
Elbow flexion	16.49 (33.03)	73.18 (2.96)	54.06 (27.81)	0.10	0.4 (0.75)
ShoulderUsing Sternum - flexion	41.86 (11.54)	43.04 (2.93)	4.90 (5.89)	1	0.8 (0.33)
ShoulderUsing Sternum - abduction	-13.074 (21.40)	23.49 (0.88)	35.95 (20.28)	0.10	-0.4 (0.75)
ShoulderUsing Sternum- rotation	-0.69 (5.46)	18.03 (1.84)	19.93 (8.51)	0.10	-0.8 (0.33)
ShoulderUsing Shoulder- flexion	30.54 (18.09)	43.04 (2.93)	11.36 (13.15)	0.20	0.8 (0.33)
ShoulderUsing Shoulder - abduction	-8.21 (17.58)	23.49 (0.88)	31.08 (16.46)	0.10	-0.4 (0.75)
ShoulderUsing Shoulder - rotation	-12.78 (11.38)	18.03 (1.84)	30.80 (13.22)	0.10	-1 (0.083)
Reach, grasp, lift and hold object - Lift exercise					
Trunk flexion	5.82 (7.03)	17.52 (4.79)	18.63 (14.81)	0.10	0.4 (0.75)
Elbow flexion	27.20 (18.55)	63.28 (8.25)	41.21 (22.58)	0.10	-0.4 (0.75)
ShoulderUsing Sternum - flexion	46.49 (4.93)	47.57 (12.05)	8.08 (6.70)	0.20	0.8 (0.33)
ShoulderUsing Sternum - abduction	-5.47 (8.26)	16.62 (9.53)	25.66 (8.83)	0.10	0.8 (0.33)
ShoulderUsing Sternum- rotation	-6.88 (1.48)	6.88 (14.72)	13.12 (8.44)	0.20	0 (1)

ShoulderUsing Shoulder-flexion	34.38 (15.32)	47.57 (12.05)	20.43 (8.59)	0.10	0.8 (0.33)
ShoulderUsing Shoulder - abduction	-6.26 (10.66)	16.62 (9.53)	19.91 (5.73)	0.10	0.8 (0.33)
ShoulderUsing Shoulder - rotation	-15.87 (7.74)	6.88 (14.72)	19.72 (9.45)	0.20	-0.4 (0.75)
Reach to the back of the head – Head exercise					
Trunk flexion	-2.45 (4.06)	0.97 (2.36)	11.47 (3.80)	0.18	0.5 (1)
Elbow flexion	-16.28 (28.43)	116.19 (1.71)	130.81 (27.55)	0.18	0.5 (1)
ShoulderUsing Sternum - flexion	59.93 (7.20)	42.67 (3.24)	17.25 (3.97)	0.18	1 (0.33)
ShoulderUsing Sternum - abduction	5.87 (54.31)	55.63 (10.09)	31.53 (50.60)	0.42	-0.5 (1)
ShoulderUsing Sternum-rotation	0.37 (18.84)	-24.81 (12.95)	5.39 (15.79)	0.18	0.5 (1)
ShoulderUsing Shoulder-flexion	51.27 (5.99)	42.67 (3.24)	10.80 (6.25)	0.42	-0.5 (1)
ShoulderUsing Shoulder - abduction	3.39 (40.96)	55.63 (10.09)	34.01 (41.93)	0.18	-0.5 (1)
ShoulderUsing Shoulder - rotation	-15.87 (12.56)	-24.81 (12.95)	2.27 (3.72)	0.18	1 (0.33)

Abbreviations: IQR = interquartile range, IMU = Inertial Measurement Unit

Table E.3: This table shows the calculated validity results of the joint angle range (in degrees) during the reach-and-return phase: IMU and Vicon median (+ interquartile range, IQR), median absolute IMU-Vicon difference (+IQR), Wilcoxon test results and Spearman correlation (+ p-value). Both methods for shoulder angle calculations are shown.

	Validity protocol – Range (Reach-and-return phase)				
Metric	**IMU Median (IQR)**	**Vicon Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Vicon vs IMU, p-value**	**Spearman correlation Vicon vs IMU (p-value)**
Reach to touch object – Touch exercise					
Trunk flexion	5,05 (1,26)	4.62 (2.25)	0.71 (0.44)	0.20	1 (0.083)
Elbow flexion	48,12 (15,59)	39.75 (7.36)	13.32 (7.50)	0.36	0.4 (0.75)
ShoulderUsing Sternum - flexion	45,86 (6.99)	37.79 (10.00)	6.99 (1.93)	0.10	1 (0.083)
ShoulderUsing Sternum - abduction	16,58 (3.14)	14.40 (2.02)	3.72 (1.19)	0.58	0.4 (0.75)
ShoulderUsing Sternum- rotation	11.40 (3.57)	31.91 (4.31)	19.90 (2.46)	0.10	0.6 (0.42)
ShoulderUsing Shoulder- flexion	38.93 (5.28)	37.79 (10.00)	2.76 (1.78)	1	0.8 (0.33)
ShoulderUsing Shoulder - abduction	16.57 (4.01)	14.40 (2.02)	4.73 (1.19)	0.58	0.4 (0.75)
ShoulderUsing Shoulder - rotation	19.16 (1.88)	31.91 (4.31)	12.75 (6.19)	0.10	-1 (0.083)
Reach, grasp, lift and hold object - Lift exercise					
Trunk flexion	11.96 (2.95)	10.82 (2.29)	1.14 (0.66)	0.10	1 (0.083)
Elbow flexion	80.27 (20.38)	51.14 (4.55)	27.06 (18.04)	0.10	0.2 (0.92)
ShoulderUsing Sternum - flexion	48.83 (9.13)	41.75 (12.98)	5.40 (2.18)	0.10	0.8 (0.33)
ShoulderUsing Sternum - abduction	20.96 (5.12)	14.76 (6.87)	9.63 (8.38)	0.58	0.8 (0.33)
ShoulderUsing Sternum- rotation	20.87 (11.69)	42.71 (19.01)	22.21 (6.95)	0.10	0.8 (0.33)

ShoulderUsing Shoulder-flexion	45.01 (3.73)	41.75 (12.98)	5.82 (4.35)	0.86	0.8 (0.33)
ShoulderUsing Shoulder - abduction	20.98 (4.24)	14.76 (6.87)	9.03 (6.55)	0.86	0.8 (0.33)
ShoulderUsing Shoulder - rotation	28.32 (12.53)	42.71 (19.01)	15.85 (5.02)	0.1	1 (0.083)
Reach to the back of the head – Head exercise					
Trunk flexion	9.62 (4.53)	3.16 (0.14)	6.46 (4.67)	0.18	-1 (0.33)
Elbow flexion	192.97 (119.79)	62.28 (12.40)	126.08 (109.70)	0.18	0.5 (1)
ShoulderUsing Sternum - flexion	113.22 (45.61)	25.73 (13.83)	62.79 (47.09)	0.18	-0.5 (1)
ShoulderUsing Sternum - abduction	51.08 (9.09)	71.05 (16.34)	29.59 (16.37)	0.18	0.5 (1)
ShoulderUsing Sternum-rotation	89.65 (46.76)	101.47 (16.37)	21.30 (13.83)	0.79	0.5 (1)
ShoulderUsing Shoulder-flexion	64.77 (9.71)	25.73 (13.83)	26.55 (17.29)	0.18	-0.5 (1)
ShoulderUsing Shoulder - abduction	39.78 (13.33)	71.05 (16.34)	43.72 (9.24)	0.18	1 (0.33)
ShoulderUsing Shoulder - rotation	38.17 (14.07)	101.47 (16.37)	54.56 (10.89)	0.18	0.5 (1)

Abbreviations: IQR = interquartile range, IMU = Inertial Measurement Unit

Table E.4: This table shows the calculated validity results of the mean joint angle (in degrees) during the reach-and-return phase: IMU and Vicon median (+ interquartile range, IQR), median absolute IMU-Vicon difference (+IQR), Wilcoxon test results and Spearman correlation (+ p-value). Both methods for shoulder angle calculations are shown.

	Validity protocol – Mean (Reach-and-return phase)				
Metric	**IMU Median (IQR)**	**Vicon Median (IQR)**	**Median Absolute Difference (IQR)**	**Wilcoxon test Vicon vs IMU, p-value**	**Spearman correlation Vicon vs IMU (p-value)**
Reach to touch object – Touch exercise					
Trunk flexion	4.20 (3.66)	12.43 (4.80)	7.83 (4.22)	0.10	0.4 (0.75)
Elbow flexion	19.34 (27.86)	74.30 (3.60)	54.39 (26.26)	0.10	0.4 (0.75)
ShoulderUsing Sternum - flexion	38.68 (11.53)	41.46 (1.93)	5.90 (4.67)	1	0.4 (0.75)
ShoulderUsing Sternum - abduction	-13.85 (23.52)	24.24 (2.09)	36.60 (22.63)	0.10	0.8 (0.33)
ShoulderUsing Sternum-rotation	-0.89 (5.99)	22.10 (4.58)	24.05 (6.68)	0.10	-0.4 (0.75)
ShoulderUsing Shoulder-flexion	28.79 (18.05)	41.46 (1.93)	12.07 (13.45)	0.20	0.4 (0.75)
ShoulderUsing Shoulder - abduction	-9.45 (19.40)	24.24 (2.09)	32.20 (18.51)	0.10	-0.2 (0.92)
ShoulderUsing Shoulder - rotation	-12.31 (11.46)	22.10 (4.58)	34.41 (11.08)	0.10	-0.8 (0.33)
Reach, grasp, lift and hold object - Lift exercise					
Trunk flexion	5.04 (7.28)	16.85 (4.75)	18.62 (14.84)	0.10	0.2 (0.92)
Elbow flexion	29.53 (24.68)	62.43 (8.44)	37.08 (24.77)	0.10	-0.2 (0.92)
ShoulderUsing Sternum - flexion	46.99 (4.81)	50.22 (14.38)	8.10 (6.42)	0.20	0.4 (0.75)
ShoulderUsing Sternum - abduction	-5.34 (10.17)	16.47 (9.72)	25.09 (9.10)	0.10	0.8 (0.33)
ShoulderUsing Sternum-rotation	-7.01 (1.29)	4.41 (18.30)	10.80 (12.64)	0.36	0.6 (0.42)

ShoulderUsing Shoulder-flexion	36.67 (13.15)	50.22 (14.38)	19.92 (9.46)	0.10	0.8 (0.33)
ShoulderUsing Shoulder - abduction	-6.11 (12.33)	16.47 (9.72)	20.30 (4.13)	0.10	0.8 (0.33)
ShoulderUsing Shoulder - rotation	-16.74 (8.36)	4.41 (18.30)	18.16 (10.13)	0.10	0 (1)
Reach to the back of the head – Head exercise					
Trunk flexion	-3.47 (4.39)	10.12 (2.37)	11.71 (4.10)	0.18	0.5 (1)
Elbow flexion	-34.66 (9.56)	121.21 (2.40)	157.90 (8.18)	0.18	0.5 (1)
ShoulderUsing Sternum - flexion	64.29 (9.92)	43.05 (3.98)	21.24 (5.94)	0.18	1 (0.33)
ShoulderUsing Sternum - abduction	8.46 (57.84)	63.22 (11.57)	34.53 (56.39)	0.18	0.5 (1)
ShoulderUsing Sternum-rotation	3.98 (8.44)	-26.03 (18.37)	31.96 (17.02)	0.18	0.5 (1)
ShoulderUsing Shoulder-flexion	56.84 (9.40)	43.05 (3.98)	13.79 (13.38)	0.18	-1 (0.33)
ShoulderUsing Shoulder - abduction	5.42 (44.38)	63.22 (11.57)	37.56 (42.93)	0.18	0.5 (1)
ShoulderUsing Shoulder - rotation	-13.81 (10.93)	-26.03 (18.37)	12.23 (7.08)	0.42	1 (0.33)

Abbreviations: IQR = interquartile range, IMU = Inertial Measurement Unit