DISSERTATION

Toward clinical application of RGB-depth camera-based instrumental motion analysis

Meilensteine auf dem Weg zur klinischen Anwendung der instrumentellen Bewegungsanalyse mit RGB-Tiefenkameras

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

von
Hanna Marie Röhling

Erstbetreuung: PD Dr. med. Tanja Schmitz-Hübsch

Datum der Promotion: 29.11.2024

Contents

List of tables

List of figures

List of abbreviations

CIS clinically isolated syndrome

EDSS Expanded Disability Status Scale

MS multiple sclerosis

MSFC Multiple Sclerosis Functional Composite

PD Parkinson's disease

POCO postural control

POCO-DUAL dual-task postural control

P1QC project 1–quality control

P2NORM project 2–normative data

SAS standing up and sitting down

SCSW short comfortable speed walk

SIP stepping in place

SLW short line walk

SMSW short maximum speed walk

SOPs standard operating procedures

Abstract

In many neurological diseases, evaluating motor impairments is essential for making diagnoses, assessing disease severity, monitoring symptoms, and measuring treatment effects. Accordingly, the further development of tools for the accurate assessment of motor function is a critical concern in current research. Compared with prevailing clinical scales, instrumental motion analysis technologies promise a more objective, detailed, and efficient assessment of motor impairment independent from clinical raters. However, these technologies have yet to be established as standard tools in research settings and clinical practice.

In the work presented here, we focused on the use of an RGB-depth camera-based system for instrumental motion analysis in persons with multiple sclerosis. This system extracts spatiotemporal motor parameters from recordings of short, standardized movement tasks. In two projects, we have addressed constructive next steps to improve the applicability and interpretability of this technology. In the first project, we developed a quality control process that enables users to review existing data and flag poor-quality recordings. This allowed us to systematically identify and quantify technical and performance issues in a large dataset of 4692 recordings from individuals with multiple sclerosis and healthy controls. How these findings can contribute to the development of fully automated quality control was then illustrated using a common performance-related issue. In the second project, we determined normative reference values for 43 spatiotemporal parameters from 133 healthy adults between 20 and 60. We reported descriptive statistics for these parameters and provided an approach to model relationships between parameters and confounding anthropometric and demographic factors. In addition, we presented a user-friendly z-score-based representation of data points in the context of the normative data.

The quality control results showed that a systematic cleaning of the respective data is necessary and can be ensured with the help of the developed process. Some of the findings can be directly integrated into quality assurance processes. The obtained normative data help to interpret newly collected datasets and identify altered movement patterns of diseased individuals at the individual or group level. Overall, the results enhance utility and promote the application of the system in further research use. At the same time, they lay a relevant foundation for future broader application in clinical practice.

Zusammenfassung

Die Erfassung motorischer Beeinträchtigungen ist bei vielen neurologischen Erkrankungen von entscheidender Bedeutung für die Diagnosestellung, die Beurteilung der Krankheitsschwere, die Überwachung von Symptomen und die Messung von Behandlungseffekten. Entsprechend ist die Weiterentwicklung von Methoden zur präzisen Bewertung der motorischen Funktion ein zentrales Anliegen aktueller Forschung. Im Vergleich zu gängigen klinischen Skalen versprechen Technologien zur instrumentellen Bewegungsanalyse eine objektivere, detailliertere und effizientere Beurteilung motorischer Beeinträchtigungen, die unabhängig von klinischen Untersuchern erfolgt. Diese Technologien müssen jedoch erst als Standardinstrumente in der Forschung und klinischen Praxis etabliert werden.

In der hier vorgestellten Arbeit haben wir uns mit der Anwendung eines RGB-Tiefenkamera Systems zur instrumentellen Bewegungsanalyse bei Personen mit Multipler Sklerose befasst. Dieses System ermittelt räumlich-zeitliche motorische Parameter aus Aufnahmen von kurzen, standardisierten Bewegungsaufgaben. In zwei Projekten haben wir konstruktive nächste Schritte zur Verbesserung der Anwendbarkeit und Interpretierbarkeit dieser Technologie umgesetzt. Im ersten Projekt haben wir einen Qualitätskontrollprozess entwickelt, der es Nutzern ermöglicht, vorhandene Daten zu sichten und Aufnahmen schlechter Qualität zu kennzeichnen. So konnten wir technische und ausführungsbedingte Probleme in einem großen Datensatz von 4692 Aufnahmen von Personen mit Multipler Sklerose und gesunden Kontrollen systematisch identifizieren und quantifizieren. Am Beispiel eines häufigen auftretenden ausführungsbedingten Problems wurde veranschaulicht, wie diese Erkenntnisse zur Entwicklung einer vollautomatischen Qualitätskontrolle beitragen können. Im zweiten Projekt haben wir normative Referenzwerte für 43 räumlich-zeitliche Parameter von 133 gesunden Erwachsenen zwischen 20 und 60 Jahren ermittelt. Wir haben deskriptive Statistiken für diese Parameter bereitgestellt und einen Ansatz für die Modellierung der Zusammenhänge mit anthropometrischen und demografischen Störfaktoren erarbeitet. Außerdem haben wir eine nutzerfreundliche z-Score-basierte Darstellung von Datenpunkten im Kontext der normativen Daten vorgestellt.

Die Ergebnisse der Qualitätskontrolle haben gezeigt, dass eine systematische Säuberung der jeweiligen Datensätze notwendig ist und mithilfe des entwickelten Prozesses

sichergestellt werden kann. Die gewonnenen Erkenntnisse können zum Teil direkt in Qualitätssicherungsprozesse integriert werden. Die normativen Daten helfen bei der Interpretation von neu erhobenen Datensätzen und bei der Identifikation veränderter Bewegungsmuster erkrankter Personen auf Individual- oder Gruppenniveau. Insgesamt steigern die Ergebnisse die Nutzbarkeit und fördern die Anwendung des Systems in der weiteren Forschung. Zugleich legen sie eine relevante Grundlage für eine breitere zukünftige Nutzung in der klinischen Praxis.

1 Introduction

1.1 Motivation

The high prevalence and increasing burden of neurological disorders pose a challenge for healthcare systems worldwide [1]. The clear mandate for specialized care for affected individuals and the concurrent shortage of neurologists in many (aging) societies inevitably lead to a predicament. Leveraging efficient technologies that support healthcare professionals and caregivers is one way to soften the impact of this increase in the need for neurological care.

Motor impairments of various manifestations are among the hallmark symptoms of neurological diseases, such as Parkinson's disease (PD), stroke, or multiple sclerosis (MS). The accumulation and progression of motor symptoms often serve as an indicator of overall disease progression. Observer-based clinical scales, such as the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale [2] or the Expanded Disability Status Scale (EDSS) [3] in MS, are the most widely used tools for assessing disease severity in clinical and research settings. These scales rely substantially on clinicians' assessment of the motor behavior of patients or study subjects. However, these conventional methods often do not adequately reflect disease progression clinically, e.g., due to limited sensitivity to change [4, 5] or low inter-rater reliability [6].

The use of instrumental motion analysis systems as an alternative has yet to be established in this context. Respective systems bear the opportunity to objectively assess human movement in great detail as well as streamline and automate the monitoring of motor symptoms, potentially remotely. In increasingly complex treatment landscapes, such technologies further anticipate simplifying individual clinical decision-making. As such, motion analysis technologies can improve care and democratize access to care, especially as physicians' time becomes an increasingly valuable commodity. In research, such systems can be used to improve the understanding of motor function and to measure intervention effects more sensitively and reliably. Remotely applicable systems also have the advantage of being deployable in decentralized trials.

This thesis addresses aspects of advancing the applicability of an RGB-depth camera-based instrumental motion analysis system. The use and potential of such systems are

introduced below using MS as a case example. However, they could be considered analogously for other neurological diseases.

1.2 Multiple sclerosis

MS is an immune-mediated neurological disease that causes demyelination and degeneration of axons in the central nervous system. In 2020, an estimated 2.8 million people worldwide were affected by MS (incidence of 35.9 per 100,000), amounting to a considerable increase in the last years and decades [7, 8]. The mean age of onset is 32, with women affected about twice as often as men [7]. There are four main MS phenotypes generally considered: Manifestations of MS symptoms before a confirmed MS diagnosis are referred to as clinically isolated syndrome (CIS); after confirmed diagnosis, the phenotype characterized by acute inflammatory relapses and subsequent remission is called relapsing-remitting MS; for some, the relapsing-remitting disease course develops into a progressive disease course, characterized by gradual decline rather than relapses, which is referred to as secondary progressive MS; lastly, approximately 15% are affected by a progressive course of the disease without remissions, which is referred to as primary progressive MS [8]. It has to be noted that these diagnostic distinctions have recently been challenged by the concept of *progress independent of relapse*, short *PIRA* [9]. Kappos et al. found that disability accumulation in persons diagnosed with relapsing-remitting MS was primarily not associated with relapse events. They argue that MS should not be dichotomized into the subtypes described above but should be conceptualized as a continuum that generally follows a progressive disease course that may be associated with varying degrees of additional disability due to relapses.

In addition to non-motor impairments, including visual, cognitive, and sensory dysfunction, people with MS are usually affected by a decline in motor function, which can significantly impact their quality of life [10, 11]. Motor impairments can manifest very differently in each individual and encompass a wide range of motor symptoms, such as gait disturbances, spasticity, paresis, ataxia, tremor, fatigue, and fine motor disturbances [12].

Treatment options are multifaceted and include disease-modifying therapies (primarily licensed for the relapsing-remitting type), treatment of acute relapses, as well as symptomatic treatment and rehabilitation. As knowledge about the disease increases,

treatment options become more complex [13]. There is widespread agreement that early, specialized, and personalized care can improve overall prognosis and delay or prevent long-term disability [13, 14].

1.2.1 Conventional assessment of motor function in persons with multiple sclerosis

The clinician-administered EDSS is most frequently used for measuring disease severity in persons with MS [15]. The EDSS total score (often termed EDSS step) is derived from eight ordinally scaled subscores considering individual functional brain systems and the maximum gait distance. In this context, the pyramidal, cerebellar, and maximal gait distances are subscales associated with motor function. The EDSS has been widely criticized regarding objectivity, reliability, and sensitivity to (subtle) change as well as its ordinal non-linear scale characteristics (see reviews from Meyer-Moock et al. [15] and Inojosa et al. [16] and references therein). Motor assessments via timed functional tests such as the ones included in the Multiple Sclerosis Functional Composite (MSFC) [17], that is, the Timed 25-Foot Walk and 9-Hole Peg Test, have the advantage of being on a continuous scale and expectedly more reliable [18]. Whether the MSFC outperforms the EDSS regarding its sensitivity to change remains debatable [16, 19]. Instrumental motion analysis methods, which have gained increasing importance in recent years, promise a more precise and in-depth analysis of human movements.

1.3 Instrumental motion analysis technologies

Systems for the instrumental analysis of human movement in healthcare settings include portable inertial measurement units, pressure plates, force plates, instrumented walkways, and camera systems. These technologies promise accurate descriptions of motor impairments, surpassing human observational assessments in detail and objectivity. Sufficiently easy-to-use systems further free up time and cost resources required for clinician-administered scales. As such, the outcomes of these systems are hypothetically better suited as the prevailing clinical scales to support diagnostic and treatment decisions, monitor symptoms and disease progression over time, or even serve as clinical trial endpoints.

For motion capture technologies, continuous and task-based point measurements are two approaches to be distinguished, which may differ in their objectives. In turn, measures of motor performance, i.e., how people move in their familiar environment, and measures of motor capacity, i.e., how people can move in a standardized environment, have been distinguished (see [20]). This distinction coincides with a subdivision between remote and lab-based measurements, with remote measurements mostly involving continuous measurements of motor performance. In contrast, lab-based measurements mainly involve monitoring motor capacity with specific point assessments of standardized motor tasks (similar to clinical neurological rating procedures). Essential differences between the two types of assessment have been described. For example, a comparison of the results of different measurement approaches (in-lab gait vs. community ambulation) for individuals with MS is presented in Shema-Shiratzky et al. [21]. Of note, remote data acquisition has also been explored for task-based instrumental motion assessment in persons with MS, e.g., via the smartphone applications dreaMS [22] or Floodlight Proof-of-Concept [23].

Arguably, the most commonly assessed movement pattern for persons with MS is gait in both continuous/remote and task-based/in-lab settings. Examples include the evaluation of whether daily step counts, measured by a continuously worn accelerometer over four weeks, can augment disability assessment [24]. Another study by Zanotto et al. [25] examined gait variability within groups with comparable EDSS scores in a task-based scenario. They measured parameters such as gait speed, step/stride length, and time spent in either double or single support during gait cycles with an instrumented walkway. Beyond gait, instrumented motion analysis systems have been used to assess other spatiotemporal parameters in persons with MS, such as body sway behavior in balance/postural control tasks or kinematics of hand-to-mouth movements [26, 27].

To date, instrumental motion analysis systems rarely entered healthcare settings in research or clinical practice. General limitations of instrumental motion analysis systems as of today include that there is a plethora of options, configurations, and settings proposed for each of the many system categories, and there is no clear consensus on best practices and relevant outcomes, resulting in low comparability among existing evidence [28, 29, 30]. In addition, the interpretability of the outcomes is often insufficient in the absence of valid cut-offs with respect to normal motor function or minimally important

change for individuals with MS. The various devices are mostly not interchangeable, and often, a tradeoff between system applicability, clinical applicability, and accuracy arises. For instance, simple activity trackers such as Fitbit (Fitbit Inc, San Francisco, CA, USA) are comparatively easy to use. However, Feehan et al. found that they do not provide accurate motion parameters–apart from mere step counts–and should thus not inform healthcare decisions [31]. On the other hand, camera-based systems often have a limited sensor range. They can thus only record motor tasks such as gait within a limited space. Not least, there are regulatory requirements and hurdles regarding the application of instrumental motion analysis, as described in [32]. For the market approval and subsequent clinical application of medical devices, a corresponding formal approval by a regulatory body (e.g., according to the requirements of the Medical Device Regulation in the European Union or of the Food and Drug Administration in the USA) is necessary. To this end, the corresponding requirements must first be met, which are derived from European and German law for the German market. The legal framework includes, for example, the European Regulation (EU) 2017/745 and the German Medizinprodukterecht-Durchführungsgesetz. The requirements include that the product meets basic safety and performance requirements, that a quality management system is established, and that it is suitable for the intended use. National and international standards (e.g., ISO 13485 regarding quality management systems for medical devices) can be used as evidence of compliance. These requirements converge in the clinical evaluation process, where clinical data is used to demonstrate the product's clinical benefit, performance, and safety. In research, instrumental motion analysis devices aim to provide reliable, meaningful, and generalizable outcomes in mostly group-based evaluations. As in the clinical context, reliability, accuracy, and suitability are essential. However, depending on the jurisdiction and context of use, specific device certification is not always necessary in the research context. Regulatory drug trials represent an example of more stringent formal requirements. Viceconti et al. illustrate this by describing the steps toward the qualification of digital mobility outcomes in people with PD as designated biomarkers [33].

Despite these limitations and roadblocks, experts rightly call for broader application [34, 35] and look for ways to overcome the implementation gap. In addition to the advantages over rater-based clinical scales, some systems already provide decent usability by

optimizing their interfaces to be user-friendly and by providing some guidance on how to interpret results. Moreover, with technological progress, these systems are becoming not only more accurate but also more cost-effective. Efforts are underway to validate and standardize the usage of various systems systematically [36, 37, 38, 39, 40]. Consistent dissemination of these advances in scientific publications, as well as accessible demos and training opportunities, can help motivate the scientific and clinical communities to adopt and ultimately benefit from these newer technologies more widely.

1.3.1 RGB-depth camera-based motion analysis systems

This work deals with a subcategory of instrumental motion analysis using camera systems, explicitly 3D or RGB-depth cameras. The gold standard for 3D camera-based technologies in terms of accurate tracking of body landmarks are marker-based multi-camera systems such as Vicon (Vicon Motion Systems, Oxford, UK) [36, 41] or Qualisys motion capture systems (Qualisys, Gothenburg, Sweden) [42, 43]. For persons with MS, these were used to study gait patterns and the risk of falling [44, 45, 46]. However, these systems are resource-intensive as they are comparatively expensive, difficult to operate, and only deployable in environments specifically suited for them. Furthermore, they are marker-based, which, in addition to the lab environment, might irritate recorded subjects, diminish ecological validity, and constitute a potential error source (e.g., incorrect marker placement). Arguably preferable for large-scale application are thus less accurate but portable, low-cost, easy-to-use, and markerless single-sensor technologies.

Plenty of consumer RGB-depth cameras on the market can be used to record human movement in 3D. They include various generations of Intel RealSense cameras (Intel Corporation, Santa Clara, CA, USA) and modern smartphones. Most prominently used in the medical research context, however, is the Microsoft Kinect v2 (Microsoft, Redmond, WA, USA). This camera has been previously validated regarding its reliability, accuracy [36, 47], acceptability, and usability [48]. Regarding potential clinically relevant outcomes in MS, it has been used to assess postural control [47] and gait [49, 50] and as a tool in the context of rehabilitation [51]. Of course, usage of the Kinect v2 is not limited to MS but extends to research into other diseases such as PD, ataxia, or stroke [49, 52, 53].

The system employed for this thesis was Motognosis Labs (Motognosis GmbH, Berlin, Germany), which likewise relies on the Kinect v2 as a sensor. It has proven easy to use for

operators in prior application in different conditions [36, 38, 42, 52, 54, 55, 56, 57, 58, 59]. This system allows the recording of a series of short, standardized motor tasks in standard clinical situations. From these recordings, a collection of spatiotemporal parameters can be extracted that objectively quantify various aspects of the recorded movements.

1.4 Research aims

Motognosis Labs has been used for several years in our and other clinical research groups. It has proven to be feasible in research use, has earlier been shown to be accurate and reliable [36], and has been tested in previous pilot and proof-of-concept studies for its ability to represent divergent motion patterns based on spatiotemporal outcome parameters extracted from Kinect v2 and its predecessor [52, 54, 55, 56, 57]. The now cumulatively large–and growing–body of data allows for more comprehensive analyses and paves the way for wider use in research and prospectively clinical settings. In two projects, we have addressed crucial prerequisites in the transition toward broader application.

Project 1–quality control (P1QC): To conduct high-quality research, an understanding of what constitutes accurate, clean, and meaningful data within the capabilities of the applied technology is essential. In our case, however, there were no standardized procedures for systematic quality control in place that we could have used to verify our data. The search for relevant literature revealed that quality aspects in this area of research are generally understudied and largely unreported. In P1QC, we thus developed, employed, and evaluated a post hoc quality control approach to help identify and address quality issues efficiently and pave the way for standardized post-processing and data-cleaning.

Project 2–normative data (P2NORM): If instrumental motion analysis measures are to be used to assess disease manifestation or severity, comprehensive reference data from a healthy population are an essential prerequisite for interpretability. For this technology, it was not defined what constitutes “healthy” motor behavior in terms of the outcome parameters. In P2NORM, we aimed to create a robust normative database to facilitate the interpretation of spatiotemporal parameter values.

The methods and results presented in this thesis for P1QC and P2NORM are mainly based on the original publications by Röhling et al [60] and Röhling et al. [61].

2 Methods

First, the mutual methods and materials for P1QC and P2NORM are described, relating to the technical configuration of the instrumental motion analysis system and the motor tasks performed. Then, the different studies and the participants pooled from them are presented. Finally, the project-specific methods are addressed.

2.1 Instrumental motion analysis via Motognosis Labs

All data used for P1QC and P2NORM were recorded using the Kinect v2 sensor (Microsoft, Redmond, WA) in combination with Motognosis Labs recording software (version 1.1, 1.4, 2.0, or 2.1; Motognosis GmbH, Berlin, Germany). In addition to RGB (color) recordings, this sensor uses time-of-flight infrared technology to measure depth in a 1.5 to 4.5m distance from the camera (see Figure 1 A) at 30 frames per second and 512x424 pixels. Microsoft's associated software development kit was used to extract three-dimensional time series for 25 anatomical landmarks (see Figure 1 B) from the depth information with a decision forest-based body tracking approach [62]. While facing

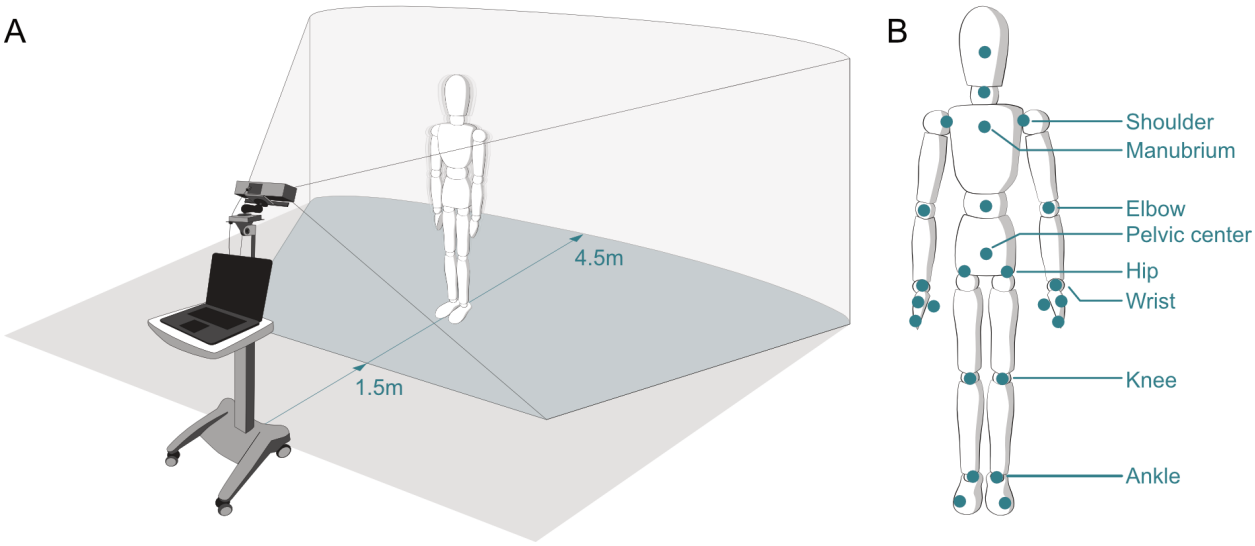


Figure 1: Set-up of Motognosis Labs and illustration of Microsoft Kinect v2 anatomical landmarks. A: Set-up of motion analysis system Motognosis Labs using Kinect v2 with sensor range as indicated. B: Dots indicate the positions of 25 anatomical landmarks tracked via the Microsoft Kinect software development kit. The labeled landmarks correspond to those used to extract spatiotemporal parameters for this work. The illustrations were adapted and modified from [52] Figure 1 as well as [63] Figure 4 and were provided courtesy of Motognosis GmbH.

the camera, participants performed short, standardized movement tasks according to

written standard operating procedures (SOPs). They wore everyday clothes and shoes. Protocols for included movement tasks differed between studies from which participants were pooled, i.e., not every motor task was performed in every study (for more information, see [60, 61]). The tasks analyzed in P2NORM were limited to short comfortable speed walk (SCSW), short maximum speed walk (SMSW), short line walk (SLW), stepping in place (SIP), standing up and sitting down (SAS), and postural control (POCO), which are described in Table 1. For P2NORM, custom (Motognosis GmbH) algorithms were used to extract spatiotemporal parameters quantifying relevant motor aspects of these tasks. Respectively utilized landmark movements and extracted parameters are described in Table 1. For P1QC, dual-task postural control (POCO-DUAL) measurements, which correspond to POCO measurements and simultaneous performance of a cognitive task (Serial 3's subtraction), were additionally included in the analyses. For P1QC, no spatiotemporal parameters were extracted; however, task and movement signal descriptions from Table 1 apply likewise. According to the respective SOPs, SCSW, SMSW, SLW, and SAS were performed three times in a row, whereas the other tasks were performed once.

2.2 Participants

For P1QC and P2NORM, data from different studies were retrospectively pooled. The studies are briefly presented below, focusing on their respective use of Motognosis Labs. The composition of pooled participants per project is listed in Table 2.

AMBOS The study investigated whether arm ergometry can improve gait function in persons with MS [65]. In addition to the primary clinical endpoint (6-Minute Walk Test) and secondary outcome measures, including stopwatch-based mobility measures, Motognosis Labs was exploratively used to measure intervention effects. The study was conducted at Universitätsklinikum Eppendorf, Hamburg, Germany (ClinicalTrials.gov Identifier: NCT03147105).

ASD In this study, Motognosis Labs was used to explore differences in motor patterns between adults with autism spectrum disorder without intellectual impairment and healthy controls [54]. The study was conducted at the Department of Psychiatry and

Table 1: Descriptions of short standardized movement tasks, relevant landmark signals, and respective spatiotemporal outcome parameters. Modified after Röhling et al. [61] Table 2 and Otte et al. [58] Table 2.

Description	Relevant movement signals from Kinect landmarks	Spatiotemporal parameters
Short comfortable speed walk (SCSW)		
The participant stands just outside the sensor range and walks toward the sensor at a comfortable speed in response to an auditory cue	Mean speed derived from pelvic center movement in the direction of walking	Gait speed [m/s]
	Mean step length, mean step width, and mean step duration over all (left and right) detected steps derived from left and right ankle movement in the direction of walking	Step length [cm] Step width [cm] Step duration [s]
	Mean gait cadence extrapolated from detected steps and recording length	Gait cadence [steps/min]
	Mean angular arm swing amplitude (averaged over left and right averages) and absolute symmetry angle [64] (between left and right mean angular arm swing amplitude) derived from left and right wrist movement relative to manubrium movement in anterior-posterior direction	Arm angular amplitude [°] Arm symmetry angle [n.u.]
Short maximum speed walk (SMSW)		
The participant stands just outside the sensor range and walks toward the sensor at maximum speed in response to an auditory cue	Mean speed derived from pelvic center movement in the direction of walking	Gait speed [m/s]
Short line walk (SLW)		
The participant stands just outside the sensor range and, in response to an auditory cue, walks toward the sensor in tandem gait, i.e., walks on an imaginary line with the heels touching the toes at each step	Mean and coefficient of variation of progression speed derived from pelvic center movement in the direction of walking	Progression speed [°/s] Relative progression variability [%]
	Standard deviation and speed of upper body sway starting from pelvic center	Roll sway variability [°] Roll sway speed [°/s]
	Line walk cadence derived from recording length and left and right ankle movement relative to respective hip movement	Line walk cadence [steps/min]
	Standard deviation and speed of arm movement angle (averaged over left and right) derived from elbow movement relative to respective shoulder movement in 3D	Arm variability [°] Arm speed [°/s]
Stepping in place (SIP)		
The participant walks on the spot at a comfortable pace for 40s	Mean knee amplitude, mean step duration, and mean stance duration (averaged over left and right averages) derived from knee movement in anterior-posterior direction	Knee amplitude [m] Step duration [s] Stance duration [s]
	Mean stepping cadence extrapolated from detected steps and recording length	Stepping cadence [steps/min]
	Absolute symmetry angle [64] (between left and right mean knee amplitude)	Knee symmetry angle [n.u.]
	Mean coefficient of variation of left and right “stride times” measured as time between knee amplitude peaks (i.e., slightly adapted from [52])	Arrhythmicity [%]
Standing up and sitting down (SAS)		
The participant sits on an armless chair, arms hanging to the side, stands up after an auditory cue, and sits down again after a second auditory cue	Speed of manubrium movement in vertical and anterior-posterior direction	Transition time (up) [s] Transition time (down) [s]
	Range of manubrium movement in anterior-posterior direction	AP deflection range (up) [m] AP deflection range (down) [m]
Postural control (POCO)		
The participant stands in front of the sensor with closed feet and open eyes for 20s; after an auditory cue, the subject closes their eyes and remains in this position for another 20s	Angular range and mean speed of the body sway vector between mean ankle position and pelvic center during eyes closed and eyes open measurement conditions in pitch, roll, and 3D direction	Pitch/Roll/3D sway range (open eyes) [°] Pitch/Roll/3D sway speed (open eyes) [°/s] Pitch/Roll/3D sway range (closed eyes) [°] Pitch/Roll/3D sway speed (closed eyes) [°/s]
	Romberg ratio of sway range and sway speed in pitch, roll, and 3D direction (i.e., value for closed eyes condition divided by respective value for open eyes condition)	RR of pitch/roll/3D sway range [n.u.] RR of pitch/roll/3D sway speed [n.u.]

Abbreviations: AP: anterior-posterior; RR: Romberg ratio.

Psychotherapy at Charité–Universitätsmedizin Berlin (EA1/392/16).

Chiba In this study, motor patterns were recorded from healthy Japanese adults and populations with ataxia (for comparisons to data from healthy German adults, see [58]). The study was conducted at the Division of Rehabilitation Medicine and Department of Neurology, Graduate School of Medicine, Chiba University, Chiba, Japan.

CIS This prospective observational study included recently diagnosed persons with MS or persons with CIS to investigate disease course and prognosis markers. Among MRI examinations [66], optical coherence tomography [67], and others, Motognosis Labs was used to assess motor function in this context. The study was conducted at Charité–Universitätsmedizin Berlin (EA1/182/10).

OPRIMS This ongoing observational study investigates whether and how the progression of primary progressive MS can be best assessed and measured. Amongst various other procedures, such as myelin oligodendrocyte glycoprotein antibody testing [68] and brain perfusion assessment [69], Motognosis Labs is used to measure motor aspects exploratively. The study is conducted at Universitätsklinikum Eppendorf, Hamburg, Germany (PV3961).

Valkinect This ongoing observational study includes subjects with MS, PD, and healthy controls. It aims to examine whether instrumental motion analysis using a Kinect camera is a suitable (i.e., accurate, reliable, valid, and responsive) examination method for motor function impairments in neurological diseases. Drebinger et al. investigated whether changes in spatiotemporal Motognosis Labs parameters due to motor exertion are suitable as fatigue markers for persons with MS [59]. The study is conducted at Charité–Universitätsmedizin Berlin (EA1/339/16).

VIMS This observational study included participants with relapsing-remitting MS, primary progressive MS, and secondary progressive MS, as well as healthy subjects between 20 and 69. It aimed to explore the use of optical coherence tomography to measure changes in the retina that reflect the disease progression in these subjects [67].

In this context, Motognosis Labs was used exploratively to measure concurrent motor capacity. The study was conducted at Charité–Universitätsmedizin Berlin (EA1/163/12).

WALKIMS-DA This study examined Kinect measurements along with day-to-day activity recorded by a physical activity monitor in people with MS and healthy controls. Krüger et al. report comparisons between measures of the physical activity monitor and a physical activity questionnaire in this context [70]. The study was conducted at Charité–Universitätsmedizin Berlin (EA1/321/14).

Table 2: Demographic and anthropometric information subdivided by projects and further subdivided into studies from which data were pooled. Sample size information is provided as the number of participants (percentage of female participants; percentage of missing information, if any); Age, height, weight, and BMI information is provided as mean (standard deviation; percentage of missing information, if any); Expanded Disability Status Scale (EDSS) score for persons with multiple sclerosis is provided as median (range; percentage of missing information, if any). Adapted from Röhling et al. [60] Table S1 in multimedia appendix 1 and Röhling et al. [61] Table 1.

Study	Population	Sample size	Age [years]	Height [cm]	Weight [kg]	BMI [kg/m^2]	EDSS
Röhling et al. [60] for project 1–quality control (P1QC)							
All	HC	162 (51.2; 1.2)	38.3 (12.8; 1.2)	172.0 (9.6; 3.7)	70.4 (14.6; 8.0)	23.8 (3.9; 8.0)	/
	PwMS	187 (51.9)	45.3 (10.8)	174.1 (8.8; 1.6)	75.0 (14.6; 8.0)	24.7 (4.3; 8.0)	3.0 (0.0-6.5; 2.7)
Ambos	PwMS	26 (46.2)	52.5 (5.4)	176.0 (9.5)	77.5 (20.1)	24.8 (5.3)	5.0 (4.0-6.5)
ASD	HC	43 (51.2)	33.1 (8.5)	174.0 (9.6)	73.0 (16.2)	24.0 (4.5)	/
Chiba	HC	30 (36.7; 6.7)	30.9 (7.0; 6.7)	167.1 (8.6; 6.7)	60.7 (9.2; 6.7)	21.7 (2.1; 6.7)	/
CIS	PwMS	41 (65.9)	37.7 (9.6)	174.8 (8.4)	70.0 (11.8; 22.0)	23.0 (3.7; 22.0)	1.5 (0.0-5.5)
Oprims	PwMS	25 (36.0)	52.2 (6.7)	175.0 (9.5; 4.0)	79.5 (11.5; 4.0)	26.0 (3.8; 4.0)	3.0 (0.0-5.5)
Valkinect	HC	22 (50.0)	48.4 (16.9)	172.2 (10.3)	70.8 (11.0)	23.8 (2.6)	/
	PwMS	7 (28.6)	47.4 (10.9)	175.7 (9.7)	82.0 (16.6)	26.7 (6.3)	2.0 (1.0-4.5)
VIMS	HC	36 (52.8)	34.7 (11.0)	173.4 (10.8; 5.6)	72.2 (16.4; 25.0)	24.5 (4.7; 25.0)	/
	PwMS	57 (50.9)	41.0 (11.8)	172.5 (8.5; 3.5)	71.5 (12.1; 8.8)	23.9 (3.2; 8.8)	2.0 (0.0-6.0; 3.5)
WALKIMS-DA	HC	31 (64.5)	49.1 (8.7)	171.8 (7.3; 6.5)	73.8 (13.9; 6.5)	24.9 (3.9; 6.5)	/
	PwMS	31 (58.1)	50.8 (5.1)	173.3 (8.4)	79.1 (15.1)	26.3 (4.7)	4.0 (1.5-6.0; 9.7)
Röhling et al. [61] for project 2–normative data (P2NORM)							
All	HC	133 (56.4)	36.8 (10.4)	172.9 (9.3)	71.8 (13.9)	23.9 (3.8)	/
ASD	HC	41 (51.2)	33.9 (8.0)	174.2 (9.7)	73.9 (16.1)	24.2 (4.4)	/
Valkinect	HC	35 (51.4)	44.7 (11.2)	172.6 (8.5)	70.9 (11.2)	23.8 (3.0)	/
VIMS	HC	57 (63.2)	34.1 (9.0)	172.2 (9.7)	70.9 (13.7)	23.8 (3.8)	/
Further considerations for project 2–normative data (P2NORM)							
Valkinect	PwMS	19 (42.1)	50.4 (9.4)	175.6 (8.7)	78.4 (15.1)	25.5 (5.1)	/

Abbreviations: /: not applicable; EDSS: Expanded Disability Status Scale; HC: healthy controls; pwMS: persons with multiple sclerosis.

2.3 Project-specific data processing and analysis

2.3.1 Project 1–quality control (P1QC)

Quality control pipeline development For P1QC, a post hoc quality control pipeline for Motognosis Labs was developed and tested [60]. The pipeline comprised a graphical

user interface, which allowed users to browse through visualizations regarding individual Motognosis Labs recordings. The visualizations consisted of summarized depth images (i.e., depth images averaged pixel-wise over time, as well as vertically and horizontally) and selected relevant movement signals from Kinect landmarks (cf. Table 1). Within the graphical user interface, users could view the visualizations, access respective depth videos and operator comments from the time of recording, and document their quality assessment ("keep," "undecided," "discard") based on predefined quality criteria. Quality criteria were established using a representative subsample of data and included known task-specific performance-related issues (e.g., feet open instead of closed during POCO), technical problems (e.g., step detection issues for gait tasks), as well as "Other," with the possibility to leave a free text comment. Figure 2 shows the rating interface, including a respective exemplary visualization for SIP. The interface and visualizations were produced using Python (3.7.3) as well as packages tkinter (8.6) and matplotlib (3.1.0). For a complete list of predefined criteria and more details on the pipeline development, please refer to [60].

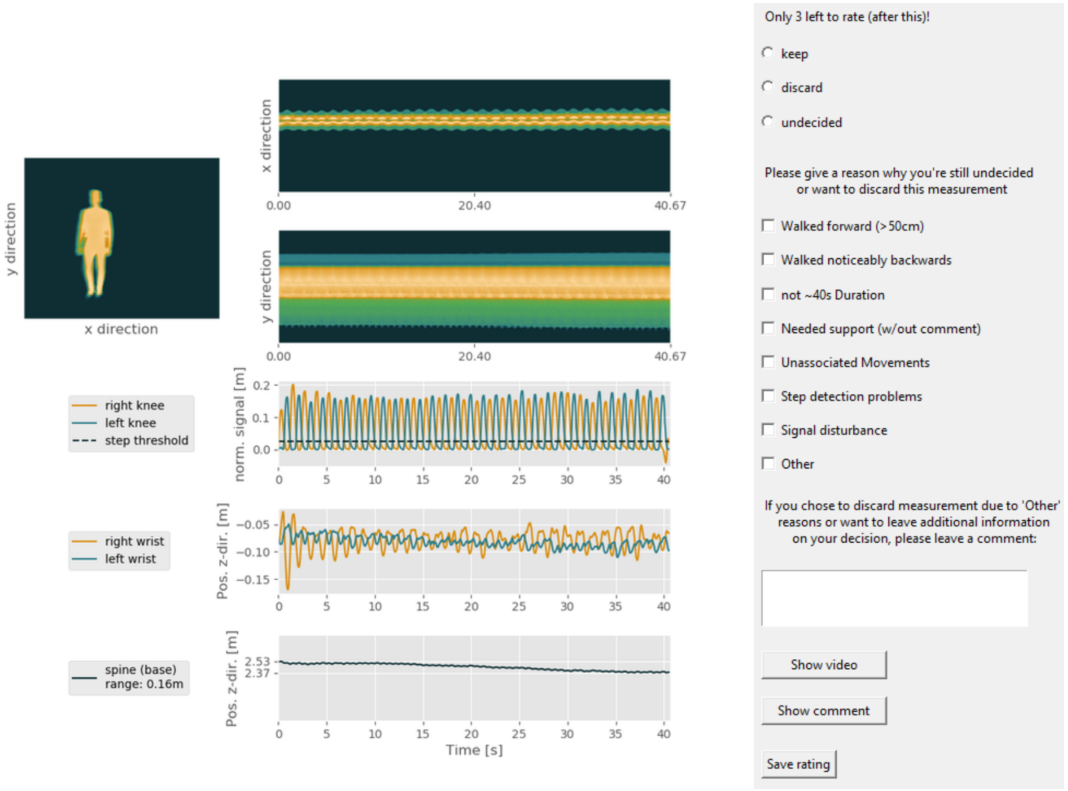


Figure 2: Visualization of a stepping in place (SIP) recording embedded in the graphical user interface developed for quality ratings in project 1–quality control (P1QC). Modified from Figure 1 in Röhling et al. [60].

Quality rating analyses Recordings of 4692 recordings (SCSW, SMSW, SLW, SIP, SAS, POCO, and POCO-DUAL) from 162 healthy controls and 187 persons with MS were quality controlled with the developed pipeline. They were distributed among eight trained raters, such that two people examined each recording. The respective frequencies of "keep," "discard," and "undecided" decisions and the corresponding rater concordance were evaluated for each motor task. Furthermore, we evaluated which categories were selected how often and which were the most frequent quality problems of the recordings per different motor tasks. Disease-associated differences in performance-related quality concerns were assessed by comparing selection frequencies of rating categories between healthy controls and persons with MS in a subset of data. Finally, we examined whether the approach to quality control is user-friendly by assessing the median rating duration per recording and interviewing the participating raters [60].

Subsequent considerations regarding automation Considerations of how the results from [60] can be used to automate quality control processes were outside the scope of the paper itself but were addressed for this work. The performance-related quality issue of subjects standing with open feet instead of closed feet during POCO was used as an example to illustrate this. The median distance of ankle landmarks for each POCO recording included in the dataset from Röhling et al. [60] was computed. This distance was dichotomized using different thresholds and compared to the "ground truth" of whether feet were actually open (i.e., one or more raters marked open feet as a quality issue for a recording). With this information, a receiver operating characteristic curve was generated, and sensitivity (i.e., true positive rate) and specificity (i.e., true negative rate) were computed for sensible thresholds between 8 and 14cm. These steps were carried out with Python (3.9.7) using packages matplotlib (3.4.3), scikit-learn (1.2.0), numpy (1.20.3), and pandas (1.3.4).

2.3.2 Project 2–normative data (P2NORM)

For P2NORM, data from 133 healthy adults were analyzed. We computed descriptive statistics, including mean, standard deviation, coefficient of variation, as well as 25th and 75th percentiles for 43 spatiotemporal parameters (Table 1). Included data passed the

quality control mechanisms developed in P1QC and described in [60]. Spatiotemporal parameters for SCSW, SMSW, SLW, and SAS were averaged over repetitions.

To investigate whether a relevant association between confounders (age, height, weight, and sex) and the spatiotemporal parameters can be robustly modeled, we chose a predictive modeling approach. Multiple linear regression models with independent variables age, height, weight, sex, and study were fitted to each spatiotemporal parameter individually as a dependent variable, with sex being dummy-coded and study effect-coded. This was embedded in a repeated (100 times) five-fold-cross-validation approach. At each repetition and cross-validation fold, predictions regarding parameter values were made on the respective test set (excluding the regression coefficients for study affiliation). The R^2-values between predicted and true values for the test set were averaged over each fold and repetition (mean R^2_{test}) and used to indicate whether the model showed a generalizable association. If the mean R^2_{test} was larger than 0.1, the model was assumed sensible and fitted on the entire data. Normalization using the regression coefficients then corresponds to calculating the normative residuals (ϵ) for subject i:

$$\epsilon_i = x_i - (\beta_0 + \beta_{Age} Age_i + \beta_{Sex} Sex_i + \beta_{Height} Height_i + \beta_{Weight} Weight_i), \tag{1}$$

with x_i being the respective spatiotemporal parameter value, β_0 being the intercept, $\beta_{Age/Sex/Weight/Height}$ being the regression coefficients for respective confounders, and Age_i, Sex_i (female: 0; male: 1), $Height_i$, $Weight_i$ the respective anthropometric and demographic characteristics of subject i.

We selected a z-score-based approach to visualize values across the different parameter scales in a readily interpretable manner. The calculation of the z-scores was based on the raw or, where appropriate, normalized (i.e., normative residuals) parameter values and was achieved as follows:

$$\begin{aligned} z_i = z_{raw,i} &= \frac{x_i - \bar{x}}{s} \text{ if the mean } R^2_{test} \leq 0.1, \\ z_i = z_{res,i} &= \frac{\epsilon_i}{s_\epsilon} \text{ if the mean } R^2_{test} > 0.1, \end{aligned} \tag{2}$$

with x_i being the respective spatiotemporal parameter value, $\bar{x}$ and s being that parameter's mean and standard deviation in the normative sample, and s_ϵ being the normative residuals' standard deviation.

In [61], recordings from a participant with MS from the Valkinect study were used to exemplify the z-score normalization and visualizations. For analyses and visualizations in [61] Python (3.7.3) and packages pandas (1.3.5), numpy (1.21.6), statsmodels (0.13.2), seaborn (0.11.2), matplotlib (3.1.0), scipy (1.7.3), and scikit-learn (0.21.2) were used. Here, the data from 19 persons with MS (cf. Table 2 under *Further considerations for project 2–normative data (P2NORM)*) from the Valkinect study were used to propose how this could be illustrated for a group rather than an individual. This boxplot-based visualization was produced using Python (3.9.7) and packages pandas (1.3.4), numpy (1.20.3), and matplotlib (3.4.3).

3 Results

The following sections present the main results from P1QC and P2NORM. They thus encompass results presented in the respective publications (P1QC: [60]; P2NORM: [61]) and additional project-specific considerations made in the course of this thesis.

3.1 Project 1–quality control (P1QC)

For this project, the tool developed to gather quality information for further analyses, i.e., the quality control pipeline, is a result in and of itself. We designed visualizations such that severe protocol deviations were directly recognizable from summarized depth images. Task-specific signals were illustrated to identify less evident quality issues. The pipeline and the GUI allowed users to make quick and informed decisions regarding the quality of Motognosis Labs recordings and to document them appropriately. The feasibility of the approach was substantiated by the median evaluation time of 6.3 seconds per recording and the positive verbal feedback of the participating raters.

Depending on the motor task, rater concordance was between 71.5% and 92.3%. Unanimous “keep” decisions were made for 39.6% to 85.1% of recordings, whereas unanimous “discard” decisions were between 5.0% and 26.3% (Table 3).

Table 3: Percentage distribution of the different possible combinations of quality rating decisions (“keep,” “discard,” or “undecided” ratings from two independent raters) as well as overall rater concordance. Adapted from Figure 2 in Röhling et al. [60].

	SCSW	**SMSW**	**SLW**	**SIP**	**SAS**	**POCO**	**POCO-DUAL**
Rater concordance [%]	92.3	79.5	74.6	85.6	90.4	71.5	72.7
“keep” & “keep” decisions [%]	85.1	73.3	60.5	70.8	62.9	50.3	39.6
“discard” & “discard” decisions [%]	6.5	5.0	9.4	13.1	26.3	13.0	25.3
“undecided” & “undecided” decisions [%]	0.7	1.2	4.7	1.7	1.2	8.2	7.8
“keep” & “undecided” decisions [%]	5.5	8.2	17.9	10.3	5.9	16.4	14.7
“keep” & “discard” decisions [%]	0.8	10.5	2.8	2.7	1.0	3.1	4.9
“discard” & “undecided” decisions [%]	1.4	1.9	4.7	1.4	2.7	9.0	7.8

Abbreviations: POCO: postural control; POCO-DUAL: dual-task postural control; SAS: standing up and sitting down; SCSW: short comfortable speed walk; SIP: stepping in place; SLW: short line walk; SMSW: short maximum speed walk.

We grouped quality concerns into technical issues and performance-related issues. The most prevalent technical issues for this dataset comprised signal disturbances, i.e., excessive noise in the background, floor, or clothing (5.8% to 31.9% of recordings depending on the motor task) as well as incorrect step detection for SCSW, SMSW,

and SLW (10.9% to 19.7%). The most prevalent performance issues for this dataset comprised open feet position during POCO and POCO-DUAL (18.4% and 29.4%), unassociated movements during POCO-DUAL (21.2%), and incorrect initial arm positioning for SAS (26.8%).

Due to a lack of further information, such as systematic operator comments, disease dependency of performance-related issues could only be estimated based on between-group differences in selected rating criteria frequencies. These between-group differences were assessed for recordings from studies Valkinect, VIMS, and WALKIMS-DA only (cf. Figure 3 in Röhling et al. [60]). We chose this subset to diminish the effects of study affiliation as a confounder, as these were the studies from which both people with MS and healthy controls were included. The most frequently occurring performance-related problems (see above) showed some absolute group differences, but in each case, healthy controls' recordings were likewise affected substantially. In this subset, open feet during POCO were observed in 16.5% (healthy controls) versus 25.0% (people with MS) of recordings; open feet during POCO-DUAL were observed in 31.0% (healthy controls) versus 36.9% (people with MS) of recordings; unassociated movements during POCO-DUAL were observed in 29.3% (healthy controls) versus 21.5% (people with MS) of recordings; and incorrect arm positioning during SAS was observed in 21.1% (healthy controls) versus 25.4% (people with MS) of recordings. Thus, a statement about disease dependency remains inconclusive; instructional errors appear to have a more significant impact here. However, we observed between-group differences in less prevalent performance-related quality problems. Healthy controls barely exhibited the need to support themselves with a walking aid, the wall, or a chair during SCSW, SMSW, SLW, SIP, SAS, POCO, and POCO-DUAL (0.0% to 1.7%). In contrast, we found this in persons with MS to a comparatively higher degree (0.6% to 6.0%). Similarly, we observed sidesteps for both POCO and SLW in 1.1% and 11.5% of cases for individuals with MS and in 0.0% and only 1.9% of cases for healthy controls. For a more detailed account of all predefined quality criteria and distinctions between studies and cohort affiliation, please refer to the original publication and the associated supplementary material [60].

How rating results from [60] can prospectively serve as "ground truth" for automated quality control algorithms was exemplified using the issue of open feet positioning during POCO. In Figure 3, using a receiver operator characteristic, the median distance of the

ankles of recordings used in [61] was contrasted with the fact that one or both raters flagged the feet positioning of the corresponding subject as being open. In this example,

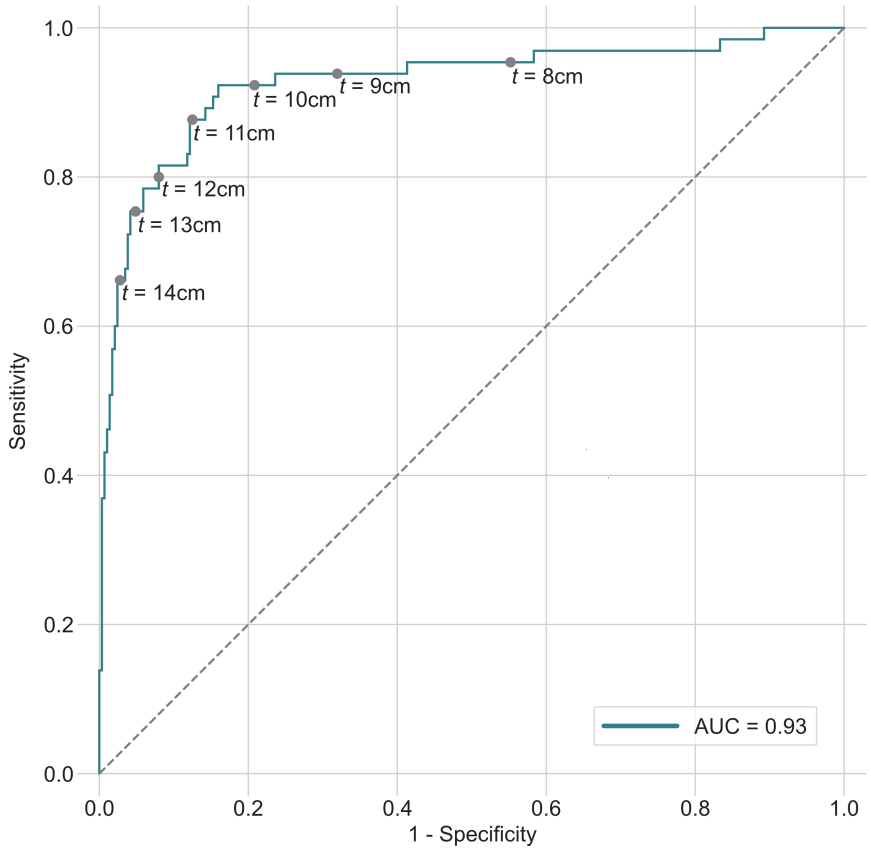


Figure 3: Receiver operator characteristic curve illustrating the discriminatory power of the median ankle landmark distance as an indicator for open feet positioning during postural control (POCO). Abbreviations: AUC: area under the (receiver operator characteristic) curve; t: threshold for median ankle distance. [Own representation: Hanna Marie Röhling].

the discriminatory power of the median ankle distance corresponded to an area under the curve of 0.93. Based on this receiver operator characteristic, different suitable cut-offs for automated assessment of this performance-related issue were conceivable. Table 4 lists the sensitivity and specificity of some sensible whole-numbered thresholds from $t =$ 8cm to $t =$ 14cm. Which threshold is most suitable depends on the context of use. If the automatic quality control would only trigger a short reminder in the recording software during POCO to ask the subjects to close their feet, it could be argued that this is better done too often than rarely. Thus, a smaller threshold with a higher sensitivity would be suitable, such that the vast majority of people with this kind of performance issue get a reminder, but also some people whose feet are already closed. In opposite scenarios, where false negatives hurt more than false positives, a higher threshold with a higher specificity is more suitable.

Table 4: Sensitivity and specificity for different thresholds of the median ankle landmark distance serving as a binary classifier for predicting the performance-related quality issue of open feet during postural control (POCO). [Own representation: Hanna Marie Röhling].

	Threshold						
	8cm	**9cm**	**10cm**	**11cm**	**12cm**	**13cm**	**14cm**
Sensitivity	0.95	0.94	0.92	0.88	0.80	0.75	0.66
Specificity	0.45	0.68	0.79	0.88	0.92	0.95	0.97

3.2 Project 2–normative data (P2NORM)

We excluded data of insufficient quality using the quality control pipeline developed in P1QC. The supplementary material for [61] provides respective “discard” rates and reasoning. The main results from P2NORM were computed on the cleaned data and comprised descriptive statistics for 43 spatiotemporal parameters from 133 healthy adults (Table 5). Table 5 further contains mean R^2_{test} values that indicate whether a normalization regarding confounding factors age, sex, height, and weight is appropriate in this sample. This value exceeded the chosen threshold of 0.1 for parameters SCSW step length and step width. Respective regression coefficients for normalization and the standard deviations for normative residuals of these parameters are provided in Table 6. An example of a visualization of z-scores (incorporating normalization of SCSW step length and SCSW step width; cf. equations 1 and 2 and Table 6) for an individual with MS was given in Röhling et al. [61]. This example was extended for this work to visualize an exemplary group instead of an individual’s measurements (Figure 4). The boxplots show the distribution of spatiotemporal parameters for a group of persons with MS from the Valkinect cohort and, by using z-scores, contextualize it regarding the normative data. This way, shared movement patterns across multiple dimensions, such as higher overall sway ranges and sway speed for POCO but lower overall gait and progression speed for SCSW, SMSW, and SLW compared to healthy controls, are easily recognized in one comprehensive display.

Table 5: Descriptive statistics of normative data for 43 spatiotemporal parameters. Adapted from Table 3 in Röhling et al. [61].

Spatiotemporal parameter	Mean	SD	CoV	Q1	Q3	Mean R^2_{test}
Short comfortable speed walk (SCSW); n=126						
Gait speed [m/s]	1.16	0.17	0.15	1.06	1.28	−0.05
Step length [cm]	69.35	7.69	0.11	64.85	74.38	0.14
Step width [cm]	10.19	2.72	0.27	8.24	11.77	0.13
Step duration [s]	0.52	0.05	0.10	0.48	0.56	−0.03
Gait cadence [steps/min]	112.07	10.55	0.09	103.97	120.04	−0.04
Arm angular amplitude [°]	26.48	10.81	0.41	18.19	32.45	−0.14
Arm symmetry angle [n.u.]	0.23	0.16	0.72	0.11	0.30	−0.10
Short maximum speed walk (SMSW); n=90						
Gait speed [m/s]	1.66	0.18	0.11	1.53	1.77	−0.08
Short line walk (SLW); n=128						
Progression speed [m/s]	0.35	0.10	0.28	0.29	0.39	−0.07
Relative progression variability [%]	0.33	0.08	0.24	0.27	0.38	−0.09
Roll sway variability [°]	1.80	0.76	0.42	1.21	2.16	−0.12
Roll sway speed [°/s]	5.58	1.90	0.34	4.37	6.47	−0.14
Line walk cadence [steps/min]	71.78	16.35	0.23	60.50	81.39	−0.07
Arm variability [°]	5.32	3.27	0.62	2.94	6.47	−0.12
Arm speed [°/s]	18.20	7.54	0.41	13.24	20.74	−0.06
Stepping in place (SIP); n=121						
Knee amplitude [m]	0.18	0.06	0.31	0.15	0.23	−0.15
Step duration [s]	0.83	0.11	0.13	0.74	0.88	−0.11
Stance duration [s]	0.39	0.15	0.38	0.28	0.47	0.01
Stepping cadence [steps/min]	98.29	16.13	0.16	87.07	111.00	−0.06
Knee symmetry angle [n.u.]	0.06	0.05	0.87	0.02	0.09	−0.14
Arrhythmicity [%]	6.00	0.84	0.14	5.37	6.50	−0.08
Standing up and sitting down (SAS); n=90						
Transition time (up) [s]	1.53	0.19	0.12	1.39	1.63	−0.16
Transition time (down) [s]	1.66	0.22	0.13	1.48	1.80	−0.09
AP deflection range (up) [m]	0.37	0.07	0.19	0.32	0.41	−0.11
AP deflection range (down) [m]	0.40	0.08	0.20	0.34	0.46	−0.11
Postural control (POCO); n=113						
Pitch sway range (open eyes) [°]	0.91	0.43	0.48	0.59	1.15	−0.13
Roll sway range (open eyes) [°]	0.89	0.37	0.42	0.64	1.09	−0.11
3D sway range (open eyes) [°]	0.92	0.41	0.44	0.67	1.10	−0.14
Pitch sway speed (open eyes) [°/s]	0.14	0.05	0.39	0.10	0.16	−0.10
Roll sway speed (open eyes) [°/s]	0.15	0.06	0.37	0.10	0.18	−0.11
3D sway speed (open eyes) [°/s]	0.22	0.07	0.34	0.17	0.26	−0.10
Pitch sway range (closed eyes) [°]	1.12	0.50	0.45	0.74	1.40	−0.16
Roll sway range (closed eyes) [°]	1.03	0.43	0.41	0.73	1.27	−0.10
3D sway range (closed eyes) [°]	1.09	0.50	0.46	0.71	1.41	−0.15
Pitch sway speed (closed eyes) [°/s]	0.18	0.06	0.36	0.14	0.22	−0.08
Roll sway speed (closed eyes) [°/s]	0.20	0.08	0.39	0.14	0.25	−0.07
3D sway speed (closed eyes) [°/s]	0.30	0.10	0.33	0.21	0.34	−0.06
RR of pitch sway range [n.u.]	1.42	0.75	0.53	0.94	1.75	−0.13
RR of roll sway range [n.u.]	1.29	0.62	0.49	0.84	1.64	−0.13
RR of 3D sway range [n.u.]	1.33	0.69	0.52	0.87	1.65	−0.14
RR of pitch sway speed [n.u.]	1.46	0.62	0.43	1.03	1.77	−0.14
RR of roll sway speed [n.u.]	1.43	0.62	0.43	0.90	1.89	−0.15
RR of 3D sway speed [n.u.]	1.41	0.51	0.36	0.99	1.64	−0.13

Abbreviations: AP: anterior-posterior; CoV: coefficient of variation; Mean R^2_{test}: mean R^2 value for test sets over all folds and repetitions of cross-validation; Q1: 25th percentile; Q3: 75th percentile; RR: Romberg ratio; SD: standard deviation

Table 6: Multiple linear regression model coefficients and standard deviation for normative residuals (s_ϵ) for two spatiotemporal short comfortable speed walk (SCSW) parameters for normalization regarding age, sex, height, and weight. Coefficients are provided as β-value (p-value; 95% confidence interval). Adapted from Röhling et al. [61] Table 4.

	SCSW step length	**SCSW step width**
β_0	-2.176 (0.882; [-31.187, 26.835])	6.164 (0.274; [-4.953, 17.281])
β_{Age}	-0.073 (0.234; [-0.193, 0.048])	0.038 (0.107; [-0.008, 0.084])
β_{Sex}	-0.913 (0.567; [-4.062, 2.236])	1.301 ($<$0.05; [0.094, 2.507])
β_{Height}	0.510 ($<$0.001; [0.331, 0.689])	-0.020 (0.561; [-0.089, 0.048])
β_{Weight}	-0.181 ($<$0.001; [-0.281, -0.080])	0.076 ($<$0.001; [0.038, 0.115])
s_ϵ	6.005	2.301

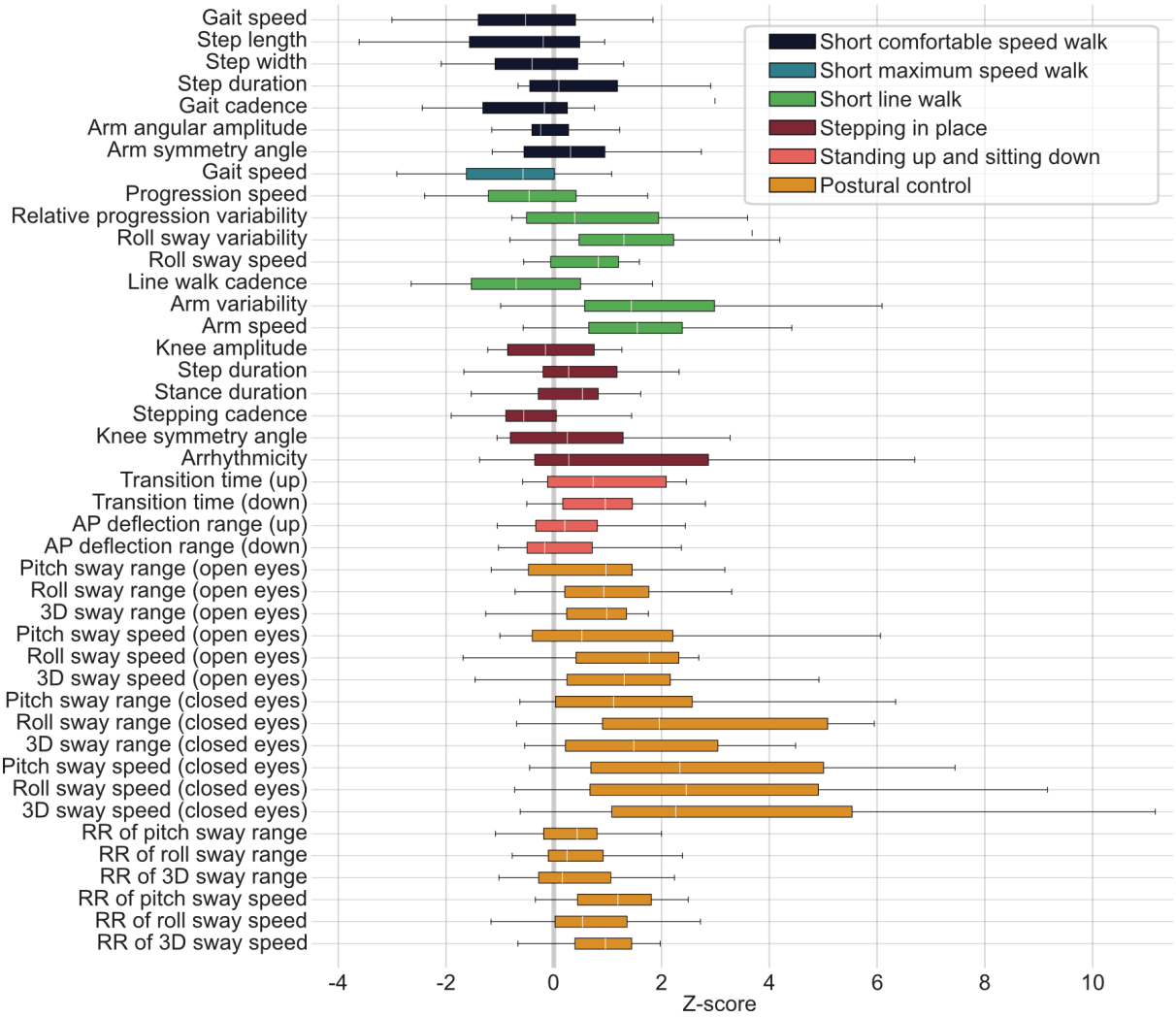


Figure 4: Boxplots showing the distribution of 43 z-score normalized spatiotemporal parameters for a group of 19 persons with multiple sclerosis in the context of normative data. Abbreviations: AP: anterior-posterior; RR: Romberg ratio. [Own representation: Hanna Marie Röhling].

4 Discussion

The starting point for this thesis was the previous development and increasing use of an RGB-depth camera-based motion analysis system, yielding a comparatively large database of recordings. For these recordings, feasible and efficient quality control procedures and extended data from healthy subjects for interpretation were needed. This work addressed these appropriate next steps toward improved applicability of this system in two projects with respective publications (P1QC: [60], P2NORM: [61]).

4.1 Project 1–quality control (P1QC)

In Röhling et al. [60], we presented a post hoc QC pipeline with which we identified the nature and frequency of data quality issues for the system and movement protocol in use. The pipeline proved feasible for users to rate large amounts of Motognosis Labs recordings efficiently. A significant proportion of the tested data was subject to quality problems, which we have divided into categories of technical and performance-related issues. A key finding was the frequency of specific performance-related problems (e.g., open feet positioning during POCO and POCO-DUAL) that cannot be attributed to disease-related problems in the execution of the motor task but to non-compliance with SOPs and inadequate operator control. With 71.5% to 92.3%, rater concordance ranged from generally acceptable to good.

For the application of the system and the field of instrumental motion analysis, this implies that the human factor is still present in supposedly objective technologies. Accordingly, there is a need for iterative checking and improvement of motion analysis technologies and the entire application workflow, including operator training and SOPs definitions. Respective results pave the way for improved automation of data selection for analysis.

An example of post hoc automated assessment for one observed performance-related quality concern–feet positioning during POCO–was introduced in this thesis. Here, a straightforward thresholding approach on the median ankle landmark distance already showed good accuracy regarding the detection of open feet. However, it cannot be assumed that all observed qualitative issues can easily be modeled using the landmark signals alone. Once developed and tested, automated checks could be included in the post hoc analyses or the recording software. Arguably, complete quality control

automation for Motognosis Labs is not required as long as it is not widely rolled out and is still in the research stage. The more manual approach presented in [60] has the advantage of promoting a better understanding of the system and fosters the agency of scientific users. Furthermore, previously unseen quality issues could arise when testing the system in new contexts, such as in a different environment or with a diseased cohort with specific motor limitations. Manual screening of recordings, as opposed to automated quality screening, could facilitate identifying such unknown quality issues. However, a higher degree of automation is undoubtedly desirable for prospective clinical usage.

The rater concordance shows that not only recording quality but also the rating thereof is affected by human factors, which is arguably a limitation of this work. We did expect some inherent ambiguity in the rating process, so we included the “undecided” rating category. One prominent example of this ambiguity is that qualitative problems do not affect every spatiotemporal parameter equally. For example, poorly recognized steps in SCSW are a reason not to trust respective step length estimations, but they do not necessarily indicate poor gait speed estimations. In this sense, we expect higher concordance with more informative and nuanced rater instructions. Limitations of this work further include limited transferability to other instrumental motion analysis systems regarding the employed visualizations. These can perhaps be adopted for other camera-based methods but must be adapted for other technological categories. An example of how such visualizations can be used for inertial measurement units was illustrated by Kroneberg et al. [71]. They used simple visualizations of step time series to quality control the correct detection of turns during gait. The transferability of the observed quality criteria and respective frequencies might also be limited. Technical quality aspects should be partially transferable to other RGB-depth camera systems but not necessarily to other technology families. Performance-related quality concerns should be more transferable when the same or similar motor tasks are employed. However, other performance-related issues might emerge in differently affected clinical populations or during at-home use of similar systems as more heterogeneous performance-related and environment-related deviations are likely.

P1QC directly impacts the work with the system in the research context and will impact future clinical use. Systematic post hoc quality control leads to cleaner and more sustainable databases and, thus, to more robust and valid research results. In addition,

detected quality problems can prompt the adaptation of the custom algorithms that generate the spatiotemporal parameters. In cases where this is possible, qualitative problems can be proactively addressed and not only detected. Furthermore, the quality control approach can lead to a better understanding of the data as the clinical and scientific users gain more data literacy and agency instead of working with "black box" technology. Clear and detailed reports concerning data quality inform experimental planning (e.g., number of motor task repetitions or avoidance of known noise sources) and thus make for more efficient use of respective systems. Lastly, we argue that recording quality and data selection for analysis should be monitored and reported more consistently in the entire field. This might foster further discussion and cooperation regarding necessary quality control mechanisms and better standardization between manufacturers, researchers, and clinical users.

4.2 Project 2–normative data (P2NORM)

In [61], we provided normative data from 133 healthy adults for 43 spatiotemporal parameters measured for assessments SCSW, SMSW, SLW, SIP, SAS, and POCO using a Kinect v2 and the Motognosis Labs system. We further proposed an approach for robust modeling of associations with anthropometric and demographic factors (age, sex, height, and weight). Using this approach, we provided normalization formulas for two SCSW parameters. Lastly, we exemplified how to effectively use z-scores to visualize spatiotemporal parameter values in relation to normative data.

To model associations with age, sex, height, and weight, we opted for a predictive multivariate cross-validation approach rather than report mere bivariate statistics such as correlations. We intended to present an approach that–as far as possible with this limited data set–models generalizable relationships. In our case, this implies that there are indeed significant simple bivariate correlations with more than the two given spatiotemporal parameters in our dataset (see Röhling et al. supplemental material [61]). However, these associations did not lead to a robust model in the context of our approach and our data. It should be explicitly noted here that the associations were modeled for healthy controls. The corresponding regression models for diseased populations may differ considerably.

We provided an in-depth comparison regarding normative values for spatiotemporal

parameters from the literature in [61]. In summary, few large normative datasets for instrumental motion analysis outcomes are mentioned in the literature. In the specific Kinect v2 literature, only the work of Latorre et al. [49] regarding spatiotemporal gait parameters stood out. Direct comparisons of our specific parameter values with values from other instrumented motion analysis systems–especially systems not based on RGB-depth cameras–are often inconclusive because the measurement systems, motor tasks, measurement protocols, and evaluation algorithms differ considerably. This underlines the need to establish normative data anew for each technology applied–even if it can be assumed that normative ranges are similar between systems if applied under comparable conditions or simultaneously [36, 72].

Despite efforts to obtain a diverse sample, our data set was limited to a group of comparatively young and fit people in urban Germany. Therefore, transferability to, for example, people over 60 should not be assumed. This makes this normative dataset unsuitable for neurological diseases such as PD, as the age of onset is usually later in life for those affected. The literature also reports cultural and socioeconomic differences in physical activity patterns [58, 73]. Thus, diversifying the data by expanding the age range or including relevant socioeconomic factors could lead to new insights in the context of the proposed regression models. Other limitations include using the z-score transformation, which is less interpretable in the case of non-normally distributed spatiotemporal parameters such as the knee symmetry angle for SIP or arm variability for SLW (cf. Figure 2 in Röhling et al. [61]). In this case, other transformations may be preferable (e.g., percentiles). However, the use in this thesis and in Röhling et al. [61] was exemplary and can be easily adapted. In addition, we found statistically significant differences for 25 spatiotemporal parameters regarding study affiliation (one-way ANOVA; $p < 0.05$), which we reported in the supplementary material of [61] but did not analyze in detail. Such biases are essential to understand measurement variability better and should be further explored in future studies.

This reference data gives current and future users of this and similar systems an estimate of what spatiotemporal parameter values to expect from a healthy adult cohort. For clinical groups of interest, the normative data and z-scores can thus be used to hypothesize cutoffs for pathology and define abnormal movement patterns, as illustrated using the example of persons with MS from the Valkinect cohort. Importantly, this can also

be applied at the individual level and may contribute to the monitoring of individual symptom progression and definitions of relapse onset or remission.In analyses between different systems, clinics, or laboratories, normative data can help reveal systematic differences that may indicate, for example, deviations in respective SOPs. Ultimately, this work improves the interpretability and further analyzability of outcome parameters for researchers and potential clinical users.

4.3 Outlook

Next steps in building evidence for this technology are studies of the intraindividual variability of spatiotemporal parameters to quantify the naturally occurring measurement noise in healthy controls and individuals with MS. Another sensible step is to test the clinical applicability of the spatiotemporal parameters in comparison to established scores or rating systems such as the EDSS or the MSFC. This includes comparing their respective reliability and sensitivity to change as well as investigating whether and which spatiotemporal parameters can be used as study endpoints for demonstrating intervention effects. In this context, the study of the relationship between measured abnormal movement patterns and patients' perceived functional limitation or improvement is of paramount importance. In addition, for clinical use of the parameters in MS, it is essential to investigate whether they have predictive power for the onset of relapses, events of disease progression, or treatment success. At this point, it should be reiterated that the considerations in the context of this work are made for MS but are analogously transferable to other diseases.

A curse and blessing of instrumental motion analysis technologies is their progressive evolution. Human motion capture continues to become more accurate as a result of new camera generations with higher temporal and spatial resolution and more powerful pose estimation techniques. Furthermore, an increasing number of modern smartphones are equipped with a depth sensor. On the other hand, systems that have reached a certain level of evidence and clinical suitability may become technically obsolete if their algorithms and signal processing cannot be transferred to these advancing technologies. This already applies here; the production of the Kinect v2, which still enjoys scientific popularity, was discontinued in 2017. Only systems that are reasonably robust to such changes can persist in the long term. Thus, for the system used in this work,

transferability is tested for the new Microsoft sensor generation called Azure Kinect [42]. In our research group, this sensor is currently being explored for self-measurements by patients in their homes as part of the HOPE-MS study (German Clinical Trials Register ID: DRKS00027042).

4.4 Conclusion

RGB-depth camera-based instrumental motion analysis promises comprehensive, objective, and reliable assessment of motor impairments in patients and study subjects. Respective systems might augment or, to some extent, replace ubiquitous rater-based scales such as the EDSS as study endpoints and for disease monitoring in clinical care. Thus, they hold the potential to improve the quality of intervention studies and the description of disease progression, which can ultimately contribute to better treatment and prognosis for patients. With this work, we improved the applicability of an RGB-depth camera-based motion analysis system by developing a systematic quality control approach and providing normative reference values. These necessary next steps toward more widespread and efficient use of this technology have since been successfully implemented in data analysis and interpretation. Overall, the results of this work facilitate the handling and interpretation of data from this and similar technologies in both research and future clinical settings.

References

[1] GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the global burden of disease study 2016. *The Lancet Neurology*, 18(5):459–480, May 2019.

[2] Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, Poewe W, Sampaio C, Stern MB, Dodel R, Dubois B, Holloway R, Jankovic J, Kulisevsky J, Lang AE, Lees A, Leurgans S, LeWitt PA, Nyenhuis D, Olanow CW, Rascol O, Schrag A, Teresi JA, van Hilten JJ, and LaPelle N; Movement Disorder Society UPDRS Revision Task Force. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders*, 23(15):2129–2170, November 2008.

[3] Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, 33(11):1444–1452, November 1983.

[4] Sharrack B, Hughes RA, Soudain S, and Dunn G. The psychometric properties of clinical rating scales used in multiple sclerosis. *Brain*, 122(1):141–159, January 1999.

[5] Regnault A, Boroojerdi B, Meunier J, Bani M, Morel T, and Cano S. Does the MDS-UPDRS provide the precision to assess progression in early Parkinson's disease? learnings from the Parkinson's progression marker initiative cohort. *Journal of Neurology*, 266(8):1927–1936, May 2019.

[6] Noseworthy JH, Vandervoort MK, Wong CJ, and Ebers GC. Interrater variability with the expanded disability status scale (EDSS) and functional systems (FS) in a multiple sclerosis clinical trial. *Neurology*, 40(6):971–975, June 1990.

[7] Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, Robertson N, La Rocca N, Uitdehaag B, van der Mei I, Wallin M, Helme A, Angood Napier C, Rijke N, and Baneke P. Rising prevalence of multiple sclerosis worldwide: Insights from the atlas of MS, third edition. *Multiple Sclerosis Journal*, 26(14):1816–1821, December 2020.

[8] Engelhard J, Oleske DM, Schmitting S, Wells KE, Talapala S, and Barbato LM. Multiple sclerosis by phenotype in germany. *Multiple Sclerosis and Related Disorders*, 57:103326, January 2022.

[9] Kappos L, Wolinsky JS, Giovannoni G, Arnold DL, Wang Q, Bernasconi C, Model F, Koendgen H, Manfrini M, Belachew S, and Hauser SL. Contribution of relapse-independent progression vs relapse-associated worsening to overall confirmed disability accumulation in typical relapsing multiple sclerosis in a pooled analysis of 2 randomized clinical trials. *JAMA Neurology*, 77(9):1132–1140, September 2020.

[10] Sutliff MH. Contribution of impaired mobility to patient burden in multiple sclerosis. *Current Medical Research and Opinion*, 26(1):109–119, January 2010.

[11] Barin L, Salmen A, Disanto G, Babačić H, Calabrese P, Chan A, Kamm CP, Kesselring J, Kuhle J, Gobbi C, Pot C, Puhan MA, and von Wyl V; Swiss Multiple Sclerosis Registry (SMSR). The disease burden of multiple sclerosis from the individual and population perspective: Which symptoms matter most? *Multiple Sclerosis and Related Disorders*, 25:112–121, October 2018.

[12] Schmidt RM, Hoffmann F, Faiss JH, Köhler W, and Zettl U. *Multiple Sklerose*. Urban & Fischer in Elsevier, 8 edition, 2021.

[13] Hauser SL and Cree BAC. Treatment of multiple sclerosis: A review. *The American Journal of Medicine*, 133(12):1380–1390.e2, December 2020.

[14] Kalincik T, Manouchehrinia A, Sobisek L, Jokubaitis V, Spelman T, Horakova D, Havrdova E, Trojano M, Izquierdo G, Lugaresi A, Girard M, Prat A, Duquette P, Grammond P, Sola P, Hupperts R, Grand'Maison F, Pucci E, Boz C, Alroughani R, Van Pesch V, Lechner-Scott J, Terzi M, Bergamaschi R, Iuliano G, Granella F, Spitaleri D, Shaygannejad V, Oreja-Guevara C, Slee M, Ampapa R, Verheul F, McCombe P, Olascoaga J, Amato MP, Vucic S, Hodgkinson S, Ramo-Tello C, Flechter S, Cristiano E, Rozsa C, Moore F, Luis Sanchez-Menoyo J, Laura Saladino M, Barnett M, Hillert J, and Butzkueven H; MSBase Study Group. Towards personalized therapy for multiple sclerosis: prediction of individual treatment response. *Brain*, 140(9):2426–2443, August 2017.

[15] Meyer-Moock S, Feng YS, Maeurer M, Dippel FW, and Kohlmann T. Systematic literature review and validity evaluation of the expanded disability status scale (EDSS) and the multiple sclerosis functional composite (MSFC) in patients with multiple sclerosis. *BMC Neurology*, 14:58, March 2014.

[16] Inojosa H, Schriefer D, and Ziemssen T. Clinical outcome measures in multiple sclerosis: A review. *Autoimmunity Reviews*, 19(5):102512, May 2020.

[17] Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, Petkau J, Syndulko K, Weinshenker BG, Antel JP, Confavreux C, Ellison GW, Lublin F, Miller AE, Rao SM, Reingold S, Thompson A, and Willoughby E. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain*, 122(5):871–882, May 1999.

[18] Solari A, Radice D, Manneschi L, Motti L, and Montanari E. The multiple sclerosis functional composite: different practice effects in the three test components. *Journal of the Neurological Sciences*, 228(1):71–74, January 2005.

[19] D'Souza M, Kappos L, and Czaplinski A. Reconsidering clinical outcomes in multiple sclerosis: Relapses, impairment, disability and beyond. *Journal of the Neurological Sciences*, 274(1-2):76–79, November 2008.

[20] Maetzler W, Rochester L, Bhidayasiri R, Espay AJ, Sánchez-Ferro A, and Uem JMT. Modernizing daily function assessment in Parkinson's disease using capacity, perception, and performance measures. *Movement Disorders*, 36(1):76–82, January 2021.

[21] Shema-Shiratzky S, Hillel I, Mirelman A, Regev K, Hsieh KL, Karni A, Devos H, Sosnoff JJ, and Hausdorff JM. A wearable sensor identifies alterations in community ambulation in multiple sclerosis: contributors to real-world gait quality and physical activity. *Journal of Neurology*, 267(7):1912–1921, July 2020.

[22] Woelfle T, Pless S, Reyes O, Wiencierz A, Feinstein A, Calabrese P, Gugleta K, Kappos L, Lorscheider J, and Naegelin Y. Reliability and acceptance of dreaMS, a software application for people with multiple sclerosis: a feasibility study. *Journal of Neurology*, 270(1):262–271, January 2023.

[23] Montalban X, Graves J, Midaglia L, Mulero P, Julian L, Baker M, Schadrack J, Gossens C, Ganzetti M, Scotland A, Lipsmeier F, van Beek J, Bernasconi C, Belachew S, Lindemann M, and Hauser SL. A smartphone sensor-based digital outcome assessment of multiple sclerosis. *Multiple Sclerosis Journal*, 28(4):654–664, April 2022.

[24] Block VJ, Lizée A, Crabtree-Hartman E, Bevan CJ, Graves JS, Bove R, Green AJ, Nourbakhsh B, Tremblay M, Gourraud PA, Ng MY, Pletcher MJ, Olgin JE, Marcus GM, Allen DD, Cree BA, and Gelfand JM. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. *Journal of Neurology*, 264(2):316–326, February 2017.

[25] Zanotto T, Sosnoff JJ, Ofori E, Golan D, Zarif M, Bumstead B, Buhse M, Kaczmarek O, Wilken J, Muratori L, Covey TJ, and Gudesblatt M. Variability of objective gait measures across the expanded disability status scale in people living with multiple sclerosis: A cross-sectional retrospective analysis. *Multiple Sclerosis and Related Disorders*, 59:103645, March 2022.

[26] Carpinella I, Anastasi D, Gervasoni E, Di Giovanni R, Tacchino A, Brichetto G, Confalonieri P, Rovaris M, Solaro C, Ferrarin M, and Cattaneo D. Balance impairments in people with early-stage multiple sclerosis: Boosting the integration of instrumented assessment in clinical practice. *Sensors (Basel)*, 22(23):9558, December 2022.

[27] Coghe G, Corona F, Pilloni G, Porta M, Frau J, Lorefice L, Fenu G, Cocco E, and Pau M. Is there any relationship between upper and lower limb impairments in people with multiple sclerosis? a kinematic quantitative analysis. *Multiple Sclerosis International*, 2019:1–6, October 2019.

[28] Espay AJ, Bonato P, Nahab FB, Maetzler W, Dean JM, Klucken J, Eskofier BM, Merola A, Horak F, Lang AE, Reilmann R, Giuffrida J, Nieuwboer A, Horne M, Little MA, Litvan I, Simuni T, Dorsey ER, Burack MA, Kubota K, Kamondi A, Godinho C, Daneault JF, Mitsi G, Krinke L, Hausdorff JM, Bloem BR, and Papapetropoulos S; Movement Disorders Society Task Force on Technology. Technology in Parkinson's disease: Challenges and opportunities. *Movement Disorders*, 31(9):1272–82, September 2016.

[29] Grimm B and Bolink S. Evaluating physical function and activity in the elderly patient using wearable motion sensors. *EFORT Open Reviews*, 1(5):112–120, March 2017.

[30] Odin P, Chaudhuri KR, Volkmann J, Antonini A, Storch A, Dietrichs E, Pirtošek Z, Henriksen Z, Horne M, Devos D, and Bergquist F. Viewpoint and practical recommendations from a movement disorder specialist panel on objective measurement in the clinical management of Parkinson's disease. *NPJ Parkinson's Disease*, 4:14, May 2018.

[31] Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, Hamilton CB, and Li LC. Accuracy of fitbit devices: Systematic review and narrative syntheses of quantitative data. *JMIR mHealth and uHealth*, 6(8):e10527, August 2018.

[32] Seeck A. Johner Institut - In 7 Schritten zum Medizinprodukt. *https://www.johner-institut.de/blog/regulatory-affairs/schritte-zum-medizinprodukt*, May 2023. Accessed on August 25, 2023.

[33] Viceconti M, Hernandez Penna S, Dartee W, Mazzà C, Caulfield B, Becker C, Maetzler W, Garcia-Aymerich J, Davico G, and Rochester L. Toward a regulatory qualification of real-world mobility performance biomarkers in Parkinson's patients using digital mobility outcomes. *Sensors (Basel)*, 20(20):5920, October 2020.

[34] Espay AJ, Hausdorff JM, Sánchez-Ferro Á, Klucken J, Merola A, Bonato P, Paul SS, Horak FB, Vizcarra JA, Mestre TA, Reilmann R, Nieuwboer A, Dorsey ER, Rochester L, Bloem BR, and Maetzler W; Movement Disorder Society Task Force on Technology. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. *Movement Disorders*, 34(5):657–663, March 2019.

[35] Frechette ML, Meyer BM, Tulipani LJ, Gurchiek RD, McGinnis RS, and Sosnoff JJ. Next steps in wearable technology and community ambulation in multiple sclerosis. *Current Neurology and Neuroscience Reports*, 19(10):80, September 2019.

[36] Otte K, Kayser B, Mansow-Model S, Verrel J, Paul F, Brandt AU, and Schmitz-Hübsch T. Accuracy and reliability of the Kinect version 2 for clinical measurement of motor function. *PLoS ONE*, 11(11):e0166532, November 2016.

[37] Kressig RW and Beauchet O; "European GAITRite Network Group". Guidelines for clinical applications of spatio-temporal gait analysis in older adults. *Aging Clinical and Experimental Research*, 18(2):174–176, April 2006.

[38] Steinert A, Sattler I, Otte K, Röhling H, Mansow-Model S, and Müller-Werdan U. Using new camera-based technologies for gait analysis in older adults in comparison to the established GAITRite system. *Sensors (Basel)*, 20(1):125, January 2020.

[39] McDonough AL, Batavia M, Chen FC, Kwon S, and Ziai J. The validity and reliability of the GAITRite system's measurements: A preliminary evaluation. *Archives of Physical Medicine and Rehabilitation*, 82(3):419–425, March 2001.

[40] Morris R, Stuart S, McBarron G, Fino PC, Mancini M, and Curtze C. Validity of mobility lab (version 2) for gait assessment in young adults, older adults and Parkinson's disease. *Physiological Measurement*, 40(9):095003, September 2019.

[41] Albert JA, Owolabi V, Gebel A, Brahms CM, Granacher U, and Arnrich B. Evaluation of the pose tracking performance of the Azure Kinect and Kinect v2 for gait analysis in comparison with a gold standard: A pilot study. *Sensors (Basel)*, 20(18):5104, September 2020.

[42] Bertram J, Krüger T, Röhling HM, Jelusic A, Mansow-Model S, Schniepp R, Wuehr M, and Otte K. Accuracy and repeatability of the Microsoft Azure Kinect for clinical measurement of motor function. *PLoS ONE*, 18(1):e0279697, January 2023.

[43] do Carmo Vilas-Boas M, Pereira Choupina HM, Rocha AP, Fernandes JM, and Silva Cunha JP. Full-body motion assessment: Concurrent validation of two body tracking depth sensors versus a gold standard system during gait. *Journal of Biomechanics*, 87:189–196, April 2019.

[44] Allali G, Laidet M, Herrmann FR, Armand S, Elsworth-Edelsten C, Assal F, and Lalive PH. Gait variability in multiple sclerosis: a better falls predictor than EDSS in patients with low disability. *Journal of Neural Transmission (Vienna)*, 123(4):447–450, April 2016.

[45] Kelleher KJ, Spence W, Solomonidis S, and Apatsidis D. The characterisation of gait patterns of people with multiple sclerosis. *Disability and Rehabilitation*, 32(15):1242–1250, February 2010.

[46] Liparoti M, Della Corte M, Rucco R, Sorrentino P, Sparaco M, Capuano R, Minino R, Lavorgna L, Agosti V, Sorrentino G, and Bonavita S. Gait abnormalities in minimally disabled people with multiple sclerosis: A 3D-motion analysis study. *Multiple Sclerosis and Related Disorders*, 29:100–107, April 2019.

[47] Clark RA, Pua YH, Oliveira CC, Bower KJ, Thilarajah S, McGaw R, Hasanki K, and Mentiplay BF. Reliability and concurrent validity of the Microsoft xbox one Kinect for assessment of standing balance and postural control. *Gait & Posture*, 42(2):210–213, July 2015.

[48] Morrison C, D'Souza M, Huckvale K, Dorn JF, Burggraaff J, Kamm CP, Steinheimer SM, Kontschieder P, Criminisi A, Uitdehaag B, Dahlke F, Kappos L, and Sellen A. Usability and acceptability of ASSESS MS: Assessment of motor dysfunction in multiple sclerosis using depth-sensing computer vision. *JMIR Human Factors*, 2(1):e11, June 2015.

[49] Latorre J, Colomer C, Alcañiz M, and Llorens R. Gait analysis with the Kinect v2: normative study with healthy individuals and comprehensive study of its sensitivity, validity, and reliability in individuals with stroke. *Journal of NeuroEngineering and Rehabilitation*, 16(1):97, July 2019.

[50] Gholami F, Trojan DA, Kovecses J, Haddad WM, and Gholami B. A Microsoft Kinect-based point-of-care gait assessment framework for multiple sclerosis patients. *IEEE Journal of Biomedical and Health Informatics*, 21(5):1376–1385, September 2017.

[51] Saladino ML, Gualtieri C, Scaffa M, Lopatin MF, Kohler E, Bruna P, Blaya P, Testa C, López G, Reyna M, Piedrabuena R, Mercante S, Barboza A, and Cáceres FJ. Neuro rehabilitation effectiveness based on virtual reality and tele rehabilitation in people with multiple sclerosis in Argentina: Reavitelem study. *Multiple Sclerosis and Related Disorders*, 70:104499, February 2023.

[52] Otte K, Ellermeyer T, Vater TS, Voigt M, Kroneberg D, Rasche L, Krüger T, Röhling HM, Kayser B, Mansow-Model S, Klostermann F, Brandt AU, Paul F, Lipp A, and Schmitz-Hübsch T. Instrumental assessment of stepping in place captures clinically relevant motor symptoms of Parkinson's disease. *Sensors (Basel)*, 20(19):5465, September 2020.

[53] Honda T, Mitoma H, Yoshida H, Bando K, Terashi H, Taguchi T, Miyata Y, Kumada S, Hanakawa T, Aizawa H, Yano S, Kondo T, Mizusawa H, Manto M, and Kakei S.

Assessment and rating of motor cerebellar ataxias with the Kinect v2 depth sensor: Extending our appraisal. *Frontiers in Neurology*, 11:179, March 2020.

[54] Cho AB, Otte K, Baskow I, Ehlen F, Maslahati T, Mansow-Model S, Schmitz-Hübsch T, Behnia B, and Roepke S. Motor signature of autism spectrum disorder in adults without intellectual impairment. *Scientific Reports*, 12(1):7670, May 2022.

[55] Behrens JR, Mertens S, Krüger T, Grobelny A, Otte K, Mansow-Model S, Gusho E, Paul F, Brandt AU, and Schmitz-Hübsch T. Validity of visual perceptive computing for static posturography in patients with multiple sclerosis. *Multiple Sclerosis Journal*, 22(12):1596–1606, October 2016.

[56] Behrens J, Pfüller C, Mansow-Model S, Otte K, Paul F, and Brandt AU. Using perceptive computing in multiple sclerosis - the short maximum speed walk test. *Journal of NeuroEngineering and Rehabilitation*, 11:89, May 2014.

[57] Grobelny A, Behrens JR, Mertens S, Otte K, Mansow-Model S, Krüger T, Gusho E, Bellmann-Strobl J, Paul F, Brandt AU, and Schmitz-Hübsch T. Maximum walking speed in multiple sclerosis assessed with visual perceptive computing. *PLoS ONE*, 12(12):e0189281, December 2017.

[58] Otte K, Ellermeyer T, Suzuki M, Röhling HM, Kuroiwa R, Cooper G, Mansow-Model S, Mori M, Zimmermann H, Brandt AU, Paul F, Hirano S, Kuwabara S, and Schmitz-Hübsch T. Cultural bias in motor function patterns: Potential relevance for predictive, preventive, and personalized medicine. *EPMA Journal*, 12(1):91–101, March 2021.

[59] Drebinger D, Rasche L, Kroneberg D, Althoff P, Bellmann-Strobl J, Weygandt M, Paul F, Brandt AU, and Schmitz-Hübsch T. Association between fatigue and motor exertion in patients with multiple sclerosis—a prospective study. *Frontiers in Neurology*, 11:208, April 2020.

[60] Röhling HM, Althoff P, Arsenova R, Drebinger D, Gigengack N, Chorschew A, Kroneberg D, Rönnefarth M, Ellermeyer T, Rosenkranz SC, Heesen C, Behnia B, Hirano S, Kuwabara S, Paul F, Brandt AU, and Schmitz-Hübsch T. Proposal for post hoc quality control in instrumented motion analysis using markerless motion capture: Development and usability study. *JMIR Human Factors*, 9(2):e26825, April 2022.

[61] Röhling HM, Otte K, Rekers S, Finke C, Rust R, Dorsch EM, Behnia B, Paul F, and Schmitz-Hübsch T. RGB-depth camera-based assessment of motor capacity: Normative data for six standardized motor tasks. *International Journal of Environmental Research and Public Health*, 19(24):16989, December 2022.

[62] Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, Kipman A, and Blake A. Real-time human pose recognition in parts from single depth images. In

Conference on Computer Vision and Pattern Recognition (CVPR) 2011. IEEE, June 2011.

[63] Karen Otte. *Using quantitative measures of stepping in place performance to assess motor symptoms of Parkinson's Disease*. Dissertation, June 2023.

[64] Alves SA, Ehrig RM, Raffalt PC, Bender A, Duda GN, and Agres AN. Quantifying asymmetry in gait: The weighted universal symmetry index to evaluate 3D ground reaction forces. *Frontiers in Bioengineering and Biotechnology*, 8:579511, October 2020.

[65] Heinrich I, Rosenthal F, Patra S, Schulz KH, Welsch GH, Vettorazzi E, Rosenkranz SC, Stellmann JP, Ramien C, Pöttgen J, Gold SM, and Heesen C. Arm ergometry to improve mobility in progressive multiple sclerosis (AMBOS)—results of a pilot randomized controlled trial. *Frontiers in Neurology*, 12:644533, July 2021.

[66] Cooper G, Chien C, Zimmermann H, Bellmann-Strobl J, Ruprecht K, Kuchling J, Asseyer S, Brandt AU, Scheel M, Finke C, and Paul F. Longitudinal analysis of t1w/t2w ratio in patients with multiple sclerosis from first clinical presentation. *Multiple Sclerosis Journal*, 27(14):2180–2190, December 2021.

[67] Motamedi S, Yadav SK, Kenney RC, Lin TY, Kauer-Bonin J, Zimmermann HG, Galetta SL, Balcer LJ, Paul F, and Brandt AU. Prior optic neuritis detection on peripapillary ring scans using deep learning. *Annals of Clinical and Translational Neurology*, 9(11):1682–1691, November 2022.

[68] Jarius S, Ruprecht K, Stellmann JP, Huss A, Ayzenberg I, Willing A, Trebst C, Pawlitzki M, Abdelhak A, Grüter T, Leypoldt F, Haas J, Kleiter I, Tumani H, Fechner K, Reindl M, Paul F, and Wildemann B. MOG-IgG in primary and secondary chronic progressive multiple sclerosis: a multicenter study of 200 patients and review of the literature. *Journal of Neuroinflammation*, 15(1):88, March 2018.

[69] Testud B, Delacour C, El Ahmadi AA, Brun G, Girard N, Duhamel G, Heesen C, Häußler V, Thaler C, Has Silemek AC, and Stellmann JP. Brain grey matter perfusion in primary progressive multiple sclerosis: Mild decrease over years and regional associations with cognition and hand function. *European Journal of Neurology*, 29(6):1741–1752, June 2022.

[70] Krüger T, Behrens JR, Grobelny A, Otte K, Mansow-Model S, Kayser B, Bellmann-Strobl J, Brandt AU, Paul F, and Schmitz-Hübsch T. Subjective and objective assessment of physical activity in multiple sclerosis and their relation to health-related quality of life. *BMC Neurology*, 17(1):10, January 2017.

[71] Kroneberg D, Elshehabi M, Meyer AC, Otte K, Doss S, Paul F, Nussbaum S, Berg D, Kühn AA, Maetzler W, and Schmitz-Hübsch T. Less is more – estimation of the number of strides required to assess gait variability in spatially confined settings. *Frontiers in Aging Neuroscience*, 10:435, January 2019.

[72] Schmitz-Hübsch T, Brandt AU, Pfueller C, Zange L, Seidel A, Kühn AA, Paul F, Minnerop M, and Doss S. Accuracy and repeatability of two methods of gait analysis – GaitRite™ und mobility lab™ – in subjects with cerebellar ataxia. *Gait & Posture*, 48:194–201, July 2016.

[73] Welmer AK, Kåreholt I, Rydwik E, Angleman S, and Wang HX. Education-related differences in physical performance after age 60: a cross-sectional study assessing variation by age, gender and occupation. *BMC Public Health*, 13:641, July 2013.

Statutory declaration

“I, Hanna Marie Röhling, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic *Toward clinical application of RGB-depth camera-based instrumental motion analysis (Meilensteine auf dem Weg zur klinischen Anwendung der instrumentellen Bewegungsanalyse mit RGB-Tiefenkameras)*, independently and without the support of third parties, and that I used no other sources and aids than those stated.
All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons (cf. declaration of contribution). I have correctly marked all texts or parts of texts that were generated in collaboration with other persons.

My contributions to any publications to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; http://www.icmje.org) on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another faculty.

The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me.”

Date	Hanna Marie Röhling

Declaration of author's contributions

I, Hanna Marie Röhling, have made the following contributions to the publications listed below and to further considerations in the context of this thesis:

Publication 1

Röhling HM, Althoff P, Arsenova R, Drebinger D, Gigengack N, Chorschew A, Kroneberg D, Rönnefarth M, Ellermeyer T, Rosenkranz SC, Heesen C, Behnia B, Hirano S, Kuwabara S, Paul F, Brandt AU, and Schmitz-Hübsch T. Proposal for post hoc quality control in instrumented motion analysis using markerless motion capture: Development and usability study. *JMIR Human Factors*, 9(2):e26825, April 2022.

Tanja Schmitz-Hübsch and I conceptualized the study and the corresponding research objectives and developed the methodology.
I implemented the prototypes and finalized versions of the rating software and interface. This included generating the visualizations used for quality control in the rating software. I furthermore contributed to the Motognosis Labs system in my time as an employee at Motognosis GmbH.
Regarding instrumental motion analysis data selection and curation, I pooled the dataset subjected to quality control from existing recordings from multiple studies, which included reviewing for completeness and plausibility, preprocessing, and maintenance efforts.
A first collaborative piloting of the quality control approach (with a subset of the co-authors) was organized and led by me. In this group, we worked out the main rating criteria and suggestions for improving the visualizations, the software, and, in general, the quality control approach. I subsequently incorporated corresponding updates into the visualizations, the rating software, and the rating process. I then organized the central research process for this publication, i.e., the quality control of 4692 recordings. This included the distribution of tasks to different raters (co-authors), the provision of the developed rating software, the preparation of the ratings, the training of the raters, and the consolidation of the rating results. In addition, I participated in the quality control ratings myself.
I carried out the subsequent processing of the rating results and the creation and implementation of the analysis protocol. I formally analyzed the results in consultation with Tanja Schmitz-Hübsch. I created all the figures and tables for the publication. For this thesis, I modified Figure 1 of the publication (see Figure 2). I further adapted Table S1 in multimedia appendix 1 (see Table 2) and information presented in Figure 2 (see Table 3) of the publication for this thesis.
I prepared the first draft of the manuscript in collaboration with Tanja Schmitz-Hübsch and integrated suggestions for improvement in the revision (through feedback from

co-authors and in the publication process by peer reviewers). I likewise carried out formal steps regarding the submission of the manuscript to the journal with support from Tanja Schmitz-Hübsch.

Publication 2

Röhling HM, Otte K, Rekers S, Finke C, Rust R, Dorsch EM, Behnia B, Paul F, and Schmitz-Hübsch T. RGB-depth camera-based assessment of motor capacity: Normative data for six standardized motor tasks. International Journal of Environmental Research and Public Health, 19(24):16989, December 2022.

Tanja Schmitz-Hübsch and I conceptualized the study and the corresponding research objectives.
Regarding instrumental motion analysis data selection and curation, I pooled the normative dataset from existing recordings from multiple studies, which included reviewing for completeness and plausibility, preprocessing, and maintenance efforts. In addition to pooling existing data, we collected new data in the scope of the Valkinect study. To recruit these subjects, I made virtual and physical recruitment postings, set up appointments, and did prescreening interviews on the phone. Furthermore, I collected the motion analysis data from these newly recruited subjects using Motognosis Labs.
I developed an initial approach to the statistical data analysis, which I then finalized with the help of Tanja Schmitz-Hübsch and with one statistical consultation appointment (Institute of Biometry and Clinical Epidemiology at Charité – Universitätsmedizin Berlin).
I performed all data analysis steps and implemented respective scripts. This included quality control of Motognosis Labs data, extraction of spatiotemporal parameters, merging of data, and statistical analysis. I also contributed to the Motognosis Labs system in my time as an employee at Motognosis GmbH. I performed the formal interpretation of the results in consultation with Tanja Schmit-Hübsch. For the publication, I produced all figures and tables except for Figure 1, which was provided courtesy of Motognosis GmbH. For Tables 1, 2, 5, and 6 in this thesis, I modified Tables 2, 1, 3, and 4 of the publication.
I prepared the first draft of the manuscript in collaboration with Tanja Schmitz-Hübsch and integrated suggestions for improvement in the revision (through feedback from co-authors and in the publication process by peer reviewers). I likewise carried out formal steps regarding the submission of the manuscript to the journal with support from Tanja Schmitz-Hübsch.

Further considerations in the context of this thesis

In addition to the work presented in the publications, I performed further analyses for this thesis in the context of P1QC and P2NORM.
For P1QC, this corresponded to the exemplary development of an automated quality control approach for the detection of open feet during POCO. I developed and implemented the corresponding approach, computed the respective results (Table 4), generated the visualization (Figure 3), and analyzed the results independently.
For further considerations regarding P2NORM, I visualized and interpreted the usability of z-scores for a group of persons with MS (Figure 4) rather than an individual (as presented in publication 1, i.e., Röhling et al. [61]). I furthermore added the subsection *further considerations for project 2–normative data (P2NORM)* in Table 2, which describes subject characteristics for said visualization.

Signature, date and stamp of the first supervisor

Signature of the doctoral candidate

Printing copy of publication 1

Röhling HM, Althoff P, Arsenova R, Drebinger D, Gigengack N, Chorschew A, Kroneberg D, Rönnefarth M, Ellermeyer T, Rosenkranz SC, Heesen C, Behnia B, Hirano S, Kuwabara S, Paul F, Brandt AU, and Schmitz-Hübsch T. Proposal for post hoc quality control in instrumented motion analysis using markerless motion capture: Development and usability study. *JMIR Human Factors*, 9(2):e26825, April 2022.

Original Paper

Proposal for Post Hoc Quality Control in Instrumented Motion Analysis Using Markerless Motion Capture: Development and Usability Study

Hanna Marie Röhling[1,2,3,4], MSc; Patrik Althoff[1,2,3]; Radina Arsenova[1,2,3,5], MD; Daniel Drebinger[1,2,3]; Norman Gigengack[1,2,3]; Anna Chorschew[1,2,3]; Daniel Kroneberg[6], MD; Maria Rönnefarth[6,7], MD; Tobias Ellermeyer[1,2,3,8]; Sina Cathérine Rosenkranz[9,10], MD; Christoph Heesen[9,10], MD; Behnoush Behnia[11], MD; Shigeki Hirano[12], MD, PhD; Satoshi Kuwabara[12], MD, PhD; Friedemann Paul[1,2,3,6,13], MD; Alexander Ulrich Brandt[13,14], MD; Tanja Schmitz-Hübsch[1,2,3,13], MD

[1]Experimental and Clinical Research Center, a cooperation between the Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association and the Charité - Universitätsmedizin Berlin, Berlin, Germany

[2]Experimental and Clinical Research Center, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

[3]Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

[4]Motognosis GmbH, Berlin, Germany

[5]Department of Pediatrics, St Joseph Krankenhaus Berlin-Tempelhof, Berlin, Germany

[6]Department of Neurology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

[7]Clinical Study Center, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

[8]Department of Neurology, Vivantes Auguste-Viktoria-Klinikum, Berlin, Germany

[9]Institute of Neuroimmunology and Multiple Sclerosis, Center for Molecular Neurobiology Hamburg (ZMNH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[10]Department of Neurology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[11]Department of Psychiatry, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

[12]Department of Neurology, Graduate School of Medicine, Chiba University, Chiba, Japan

[13]NeuroCure Clinical Research Center, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

[14]Department of Neurology, University of California, Irvine, CA, United States

Corresponding Author:
Hanna Marie Röhling, MSc
Experimental and Clinical Research Center, a cooperation between the Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association and the Charité - Universitätsmedizin Berlin
Lindenberger Weg 80
Berlin, 13125
Germany
Phone: 49 30 450539718
Fax: 49 30 450 539915
Email: hanna-marie.roehling@charite.de

Abstract

Background: Instrumented assessment of motor symptoms has emerged as a promising extension to the clinical assessment of several movement disorders. The use of mobile and inexpensive technologies such as some markerless motion capture technologies is especially promising for large-scale application but has not transitioned into clinical routine to date. A crucial step on this path is to implement standardized, clinically applicable tools that identify and control for quality concerns.

Objective: The main goal of this study comprises the development of a systematic quality control (QC) procedure for data collected with markerless motion capture technology and its experimental implementation to identify specific quality concerns and thereby rate the usability of recordings.

Methods: We developed a post hoc QC pipeline that was evaluated using a large set of short motor task recordings of healthy controls (2010 recordings from 162 subjects) and people with multiple sclerosis (2682 recordings from 187 subjects). For each of these recordings, 2 raters independently applied the pipeline. They provided overall usability decisions and identified technical and performance-related quality concerns, which yielded respective proportions of their occurrence as a main result.

Results: The approach developed here has proven user-friendly and applicable on a large scale. Raters' decisions on recording usability were concordant in 71.5%-92.3% of cases, depending on the motor task. Furthermore, 39.6%-85.1% of recordings were concordantly rated as being of satisfactory quality whereas in 5.0%-26.3%, both raters agreed to discard the recording.

Conclusions: We present a QC pipeline that seems feasible and useful for instant quality screening in the clinical setting. Results confirm the need of QC despite using standard test setups, testing protocols, and operator training for the employed system and by extension, for other task-based motor assessment technologies. Results of the QC process can be used to clean existing data sets, optimize quality assurance measures, as well as foster the development of automated QC approaches and therefore improve the overall reliability of kinematic data sets.

(JMIR Hum Factors 2022;9(2):e26825) doi: 10.2196/26825

KEYWORDS

instrumented motion analysis; markerless motion capture; visual perceptive computing; quality control; quality reporting; gait analysis

Introduction

With technology rapidly advancing, instrumented motion analysis (IMA) has emerged as an auspicious tool to augment clinical decision-making in persons with motor impairments [1-5]. Applications range from complex gait laboratory equipment to consumer grade health apps, which quantify what a person can do in a standardized setting (motor capacity) or what a person does in everyday life (motor performance) [6]. Regarding motor capacity, marker-based optoelectronic motion analysis systems serve as the gold standard for other technologies [7,8] and are, for instance, successfully used in treatment planning for children with cerebral palsy [9]. However, their high cost and complexity of analysis comprise significant disadvantages for clinical use. Thus, technologies that are portable, affordable, and easy to use are more promising for large-scale application. Respective devices developed for clinical use include pressure-sensitive walkways, inertial sensors ("wearables"), and markerless motion capture systems based on consumer depth cameras [2,10]. In the following, the term IMA will be used for this more versatile subcategory of motion analysis systems.

Despite favorable properties, IMA has not been successfully integrated into wide clinical routine yet [11,12]. Although regulatory requirements for medical products address safety and accuracy within the context of use (eg, for application in specific diseases) [13-15], successful implementation of IMA further depends on acceptance from patients and clinicians. Thus, technical usability, interpretability of outcomes, and quantifiable clinical benefits play a major role in this development. Standardized and efficient quality control (QC) procedures, not only during initial development but also during advancement and application of a system, could facilitate this technological maturation process. We found such QC aspects to be largely understudied and underreported.

QC can be applied at three levels: preventive, ad hoc, and post hoc. Preventive QC is applied before data acquisition. Manufacturers or developing groups generate initial results on data quality and publish them in proof-of-concept studies, including small samples of healthy subjects and target groups for clinical application [7,8,16,17]. Such studies can identify major pitfalls and elaborate on correct usage of these systems. For technology that is already in use with a substantial number of researchers or clinicians, expert consensus can further yield guidelines to improve preventive QC [18]. Ad hoc QC is pertained during measurements. Depending on the system, operators can decide to discard, reinstruct, and rerecord upon observing deviations from standard operating procedures (SOPs) or receiving error messages. Lastly, post hoc QC is employed at the data analysis stage. One option in this context is univariate or multivariate outlier analysis based on the kinematic parameters [19-21]. However, these approaches are highly data-dependent, inept to uncover systematic errors or "false normal" parameter values, and do not provide information regarding underlying causes of data deviation. Additional post hoc QC measures constitute postprocessing tools and successive recalculation of kinematic parameters [22,23] as well as plausibility checks based on raw data [24-26]. To date, such processes have only been performed on comparatively small data sets.

In this study, we used data acquired with the emerging Motognosis Labs system (Motognosis GmbH) that extracts kinematic parameters from depth camera recordings. In recent years, this system was extensively used in a research context at our site and our cooperating sites [24-29] with a standardized protocol for short motor tasks specifically designed to assess motor capacities of people with multiple sclerosis (MS) [7,30]. Regarding preventive QC, previously established SOPs for system operators and patient instructions were used for all data analyzed herein. With respect to ad hoc QC, the software provides visual feedback regarding general subject positioning

in the volume of acquisition and real-time tracking of the whole body as well as individual body parts. Regarding post hoc QC, we found previously employed approaches to be either insufficient, incomplete, or not feasible to reliably examine large amounts of data [19-21,24-26]. Likewise, review of IMA literature did not yield any standards or generalizable concepts. Thus, we propose an approach for systematic post hoc QC, enabling clinical users to prevent, detect, and eliminate data of inferior quality.

For the quality concerns considered here, we distinguish technical and performance issues. Technical issues comprise system-specific malfunctioning of hardware and software as well as artifacts specific to the recording technique, such as signal interference due to subjects' clothing or the recording environment in the case of depth sensing technology. Performance issues can be considered less technology-specific and can be attributed either to the operator (eg, by providing faulty instructions) or to noncompliance of the recorded subject. If the latter is unrelated to the disease, it should lead to trial exclusion; however, impairment-related inability can be considered a feature of interest.

The main objectives of this study were to (1) build a post hoc QC pipeline that is efficient, user-friendly, and adaptable, enabling clinical users to make standardized and robust decisions concerning usability of individual recordings; (2) perform QC for a large number of recordings acquired at different study sites and thus investigate the types and frequencies of quality issues; and (3) analyze the feasibility of the approach.

Methods

Data Set

Our study was based on recordings of short, structured motor tasks captured with the Motognosis Labs system. This system relies on a consumer depth camera (Microsoft KinectV2, Microsoft Corporation) and visual perceptive computing. More precisely, the software development kit associated with the camera allows for the markerless tracking of 3D time series from 25 artificial anatomical landmarks for subjects located at 1.5 to 4.5 m from the camera. Custom Motognosis Labs algorithms employ these time series to extract kinematic parameters to quantify various aspects of motor capacity.

Data were pooled from 8 monocentric studies at 3 study sites that used software versions 1.1, 1.4, 2.0, or 2.1 as part of their protocols. These studies will be referred to using the following identifiers: ASD, CIS, Valkinect, VIMS, and WALKIMS-DA (conducted at Charité – Universitätsmedizin Berlin, Berlin, Germany); Ambos and Oprims (conducted at Universitätsklinikum Eppendorf, Hamburg, Germany); and Chiba (conducted at Chiba University, Chiba, Japan). These studies were approved by the respective institutional review boards and all subjects provided written informed consent. The data set comprised recordings from 187 persons with MS and 162 healthy controls. VIMS, Valkinect, and WALKIMS-DA included both groups, whereas the other studies contributed subjects from 1 group only. Descriptive statistics include information on gender, age, anthropometry, and disease severity in case of people with MS, as measured by the Expanded Disability Status Scale [31] (Table 1 and study-specific information in Table S1 in Multimedia Appendix 1).

All subjects performed the Perceptive Assessment in Multiple Sclerosis (PASS-MS) protocol or parts of it between December 2014 and April 2019. PASS-MS consists of 10 structured motor tasks: Postural Control (POCO), Postural Control with Dual Task (POCO-DUAL), Stepping in Place (SIP), Stand Up and Sit Down (SAS), Short Line Walk (SLW), Short Comfortable Speed Walk (SCSW), Short Maximum Speed Walk (SMSW), Pronator Drift Test, Finger-Nose Test, and Finger Tapping. The latter 3 tasks were excluded from this study, as evaluation algorithms were still in an explorative stage at the time, yielding premature claims regarding data quality. A description of the remaining tasks except POCO-DUAL can be found in Otte et al [7,30]. POCO-DUAL equates to POCO with the addition of a cognitive task (Serial 3's subtraction). System operators had received in-depth training on how to use Motognosis Labs according to written SOPs. System SOPs included specifications of the setup, subject instructions, and rejection guidelines for recordings affected by performance and technical issues. According to the protocol, SAS, SLW, SCSW, and SMSW are recorded thrice consecutively, whereas POCO, POCO-DUAL, and SIP are recorded once. Deviations from SOPs occurred when single tasks or task repetitions were omitted, or operators decided to produce additional recordings (all of which should prompt an operator comment that is stored along with raw data of each recording). Such deviations explain incongruencies in the numbers of recordings per task (Table 1 and study-specific information in Table S2 in Multimedia Appendix 1), as all available recordings were included in this post hoc QC initiative.

Table 1. Demographic information about study subjects with missing data indicated as percentages and number of recordings per Perceptive Assessment in Multiple Sclerosis task subdivided by disease status.

Subject characteristics	All	HC[a]	PwMS[b]
Demographics			
N (% female; % —[c])	349 (51.6; 0.6)	162 (51.2; 1.2)	187 (51.9; 0)
Age (years), mean (SD; % —)	42.0 (12.2; 0.6)	38.3 (12.8; 1.2)	45.3 (10.8; 0)
Height (cm), mean (SD; % —)	173.1 (9.2; 2.6)	172.0 (9.6; 3.7)	174.1 (8.8; 1.6)
Weight (kg), mean (SD; % —)	72.9 (14.8; 8.0)	70.4 (14.6; 8.0)	75.0 (14.6; 8.0)
BMI (kg/m^2), mean (SD; % —)	24.3 (4.1; 8.0)	23.8 (3.9; 8.0)	24.7 (4.3; 8.0)
EDSS[d] median (range; % —)	N/A[e]	N/A	3.0 (0.0-6.5; 2.7)
# of recordings per PASS-MS[f] task			
All	4692	2010	2682
POCO[g]	354	165	189
POCO-DUAL[h]	245	88	157
SCSW[i]	1043	489	554
SMSW[j]	907	361	546
SLW[k]	957	428	529
SIP[l]	291	131	160
SAS[m]	895	348	547

[a]HC: healthy controls.
[b]PwMS: people with multiple sclerosis.
[c]—: not available.
[d]EDSS: Expanded Disability Status Scale.
[e]N/A: not applicable.
[f]PASS-MS: Perceptive Assessment in Multiple Sclerosis.
[g]POCO: Postural Control.
[h]POCO-DUAL: Postural Control with Dual Task.
[i]SCSW: Short Comfortable Speed Walk.
[j]SMSW: Short Maximum Speed Walk.
[k]SLW: Short Line Walk.
[l]SIP: Stepping in Place.
[m]SAS: Stand Up and Sit Down.

QC Pipeline Development

The QC pipeline development comprised 2 key components. First, we implemented informative visualizations enabling raters to classify the quality of raw data from PASS-MS recordings and hence implicitly assess the reliability of associated kinematic parameters. Second, we developed an efficient rating strategy for large numbers of recordings.

For the creation of informative visualizations, videos from raw depth streams were generated to enable review of each recorded task. The depth information was further used to produce a condensed representation of each recording in the form of 3 images that are hereafter referred to as motion profiles. They comprise images of depth data averaged over time, over the vertical direction, and over the horizontal direction. As PASS-MS tasks are short and highly standardized, we assumed that major protocol deviations and technical issues would be easily identifiable from motion profiles. To allow for the detection of more subtle quality issues, we also illustrated characteristic signals that are used to calculate kinematic parameters with Motognosis Labs. Visualizations were generated using Python (version 3.7.3) and the matplotlib package (version 3.1.0). A stratified random sample from 15 people with MS and 14 healthy controls was used to test and update visualizations and determine the main rating criteria per task.

We then built a graphical user interface (GUI), which includes a rating window containing visualizations, an overall usability decision checkbox (keep, discard, undecided), and task-specific multiselect checkboxes containing the main rating criteria. Furthermore, on-demand viewers for depth videos and operator comments were integrated. The GUI was programmed in Python (version 3.7.3) using the tkinter package (version 8.6). We

prepared detailed rating manuals as well as oral instructions (~45 minutes) to familiarize raters with the GUI. The entire data set (see Table 1) was subjected to ratings, such that each recording was investigated by 2 independent raters. In this step, 8 raters evaluated a total of 4692 recordings from 162 healthy controls and 187 people with MS. Raters comprised medical students, clinician scientists or researchers in other professions, and trained neurologists, all from Charité, Berlin. Among them, 6 raters had operated Motognosis Labs before, whereas 2 were new to the system. Moreover, 2 raters had been actively involved in the development of the QC pipeline, whereas 6 were new to any systematic QC of the data. After in-depth instructions, ratings were conducted individually by the raters at a self-selected speed.

Statistical Analysis

Statistical analyses included the extraction of frequencies for overall usability decisions, rater concordance and discordance, and selected rating criteria. The former 2 were illustrated as confusion matrices. Furthermore, the median rating duration per recording was extracted from the GUI log files. Figures were produced with Python (version 3.7.3) using the matplotlib package (version 3.1.0).

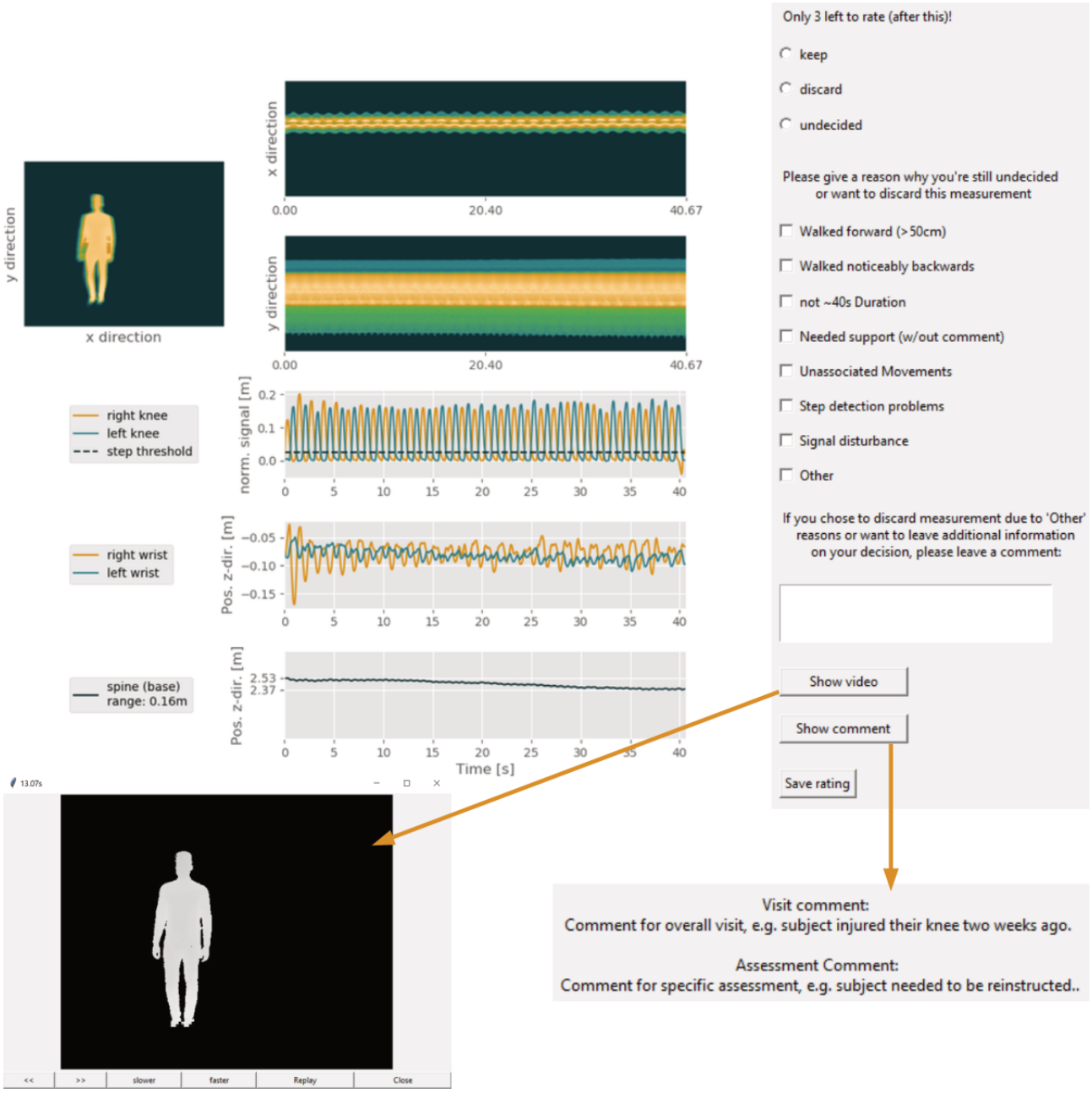
Results

QC Pipeline Usage and Feasibility

After generating visualizations, the implemented GUI can be opened to progressively rate motor task recordings. Intermediate results can be saved in an underlying Excel file, such that raters can flexibly organize their workload. An example of the rating window including respective visualizations, checkboxes, and buttons is shown in Figure 1.

Oral feedback from raters upon completion confirmed that the GUI and the QC pipeline behind it were easy to use and effective. The median rating duration per recording amounted to 6.3 seconds.

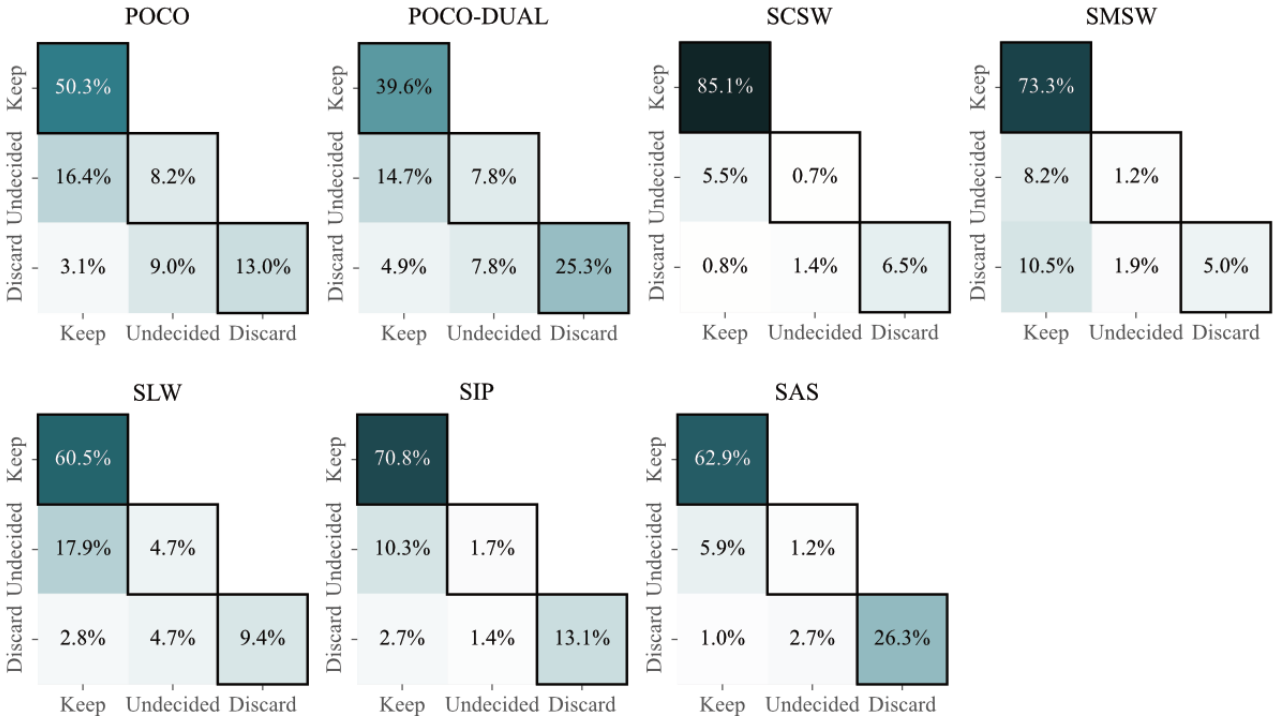
Figure 1. Rating window screenshots for an exemplary Stepping in Place recording. Upper left: motion profiles generated by summation of frontally recorded depth data over time, along horizontal and vertical directions and signal curves characteristic of the task (here: knee amplitudes, arm sway, and overall subject positioning over time). Upper right: checkboxes for usability decisions and main criteria including an option for free-text comments. Lower left: on-demand depth video viewer. Lower right: on-demand operator comment viewer.



Rater Concordance and Usability of Recordings

Concerning keep, discard, or undecided decisions, raters concurred on more than 70% of recordings for each task (POCO: 71.5%, POCO-DUAL: 72.7%, SCSW: 92.3%, SMSW: 79.5%, SLW: 74.6%, SIP: 85.6%, and SAS: 90.4%) (Figure 2). Consequently, we observed discordance for up to 28.5% of recordings, which points to task-specific difficulties in using the rating criteria. However, such discordance was mostly due to 1 rater's undecided decision. Instances of strictly opposing usability, meaning that 1 rater voted keep and the other discard, were uncommon (between 0.8% and 4.9%), except for SMSW (10.5%).

Figure 2. Synopsis of usability decisions by 2 raters per recording per Perceptive Assessment in Multiple Sclerosis task. Rater agreement on usability decisions keep, discard, and undecided are framed. POCO: Postural Control; POCO-DUAL: Postural Control with Dual Task; SAS: Stand Up and Sit Down; SCSW: Short Comfortable Speed Walk; SIP: Stepping in Place; SLW: short line walk; SMSW: Short Maximum Speed Walk.



A task-wise visualization of rater decisions regarding usability of recordings is depicted in Figure 2. Unobjectionable usability, defined as a unanimous keep decision, was obtained for 85.1% of SCSW, more than 70% of SMSW and SIP (73.3% and 70.8%, respectively), more than 60% for SAS and SLW (62.9% and 60.5%, respectively) and less than or close to half for POCO and POCO-DUAL recordings (50.3% and 39.6%, respectively). The highest rates for unanimous discard decisions were observed for SAS (26.3%), followed by POCO-DUAL (25.3%), and POCO and SIP (13.0% and 13.1%, respectively). The respective rates were low for gait tasks including SLW, SCSW, and SMSW (9.4%, 6.5%, and 5.0%, respectively). Rater concordance as well as proportions of unanimous keep and discard decisions subdivided for all studies can be found in Table S3 in Multimedia Appendix 1.

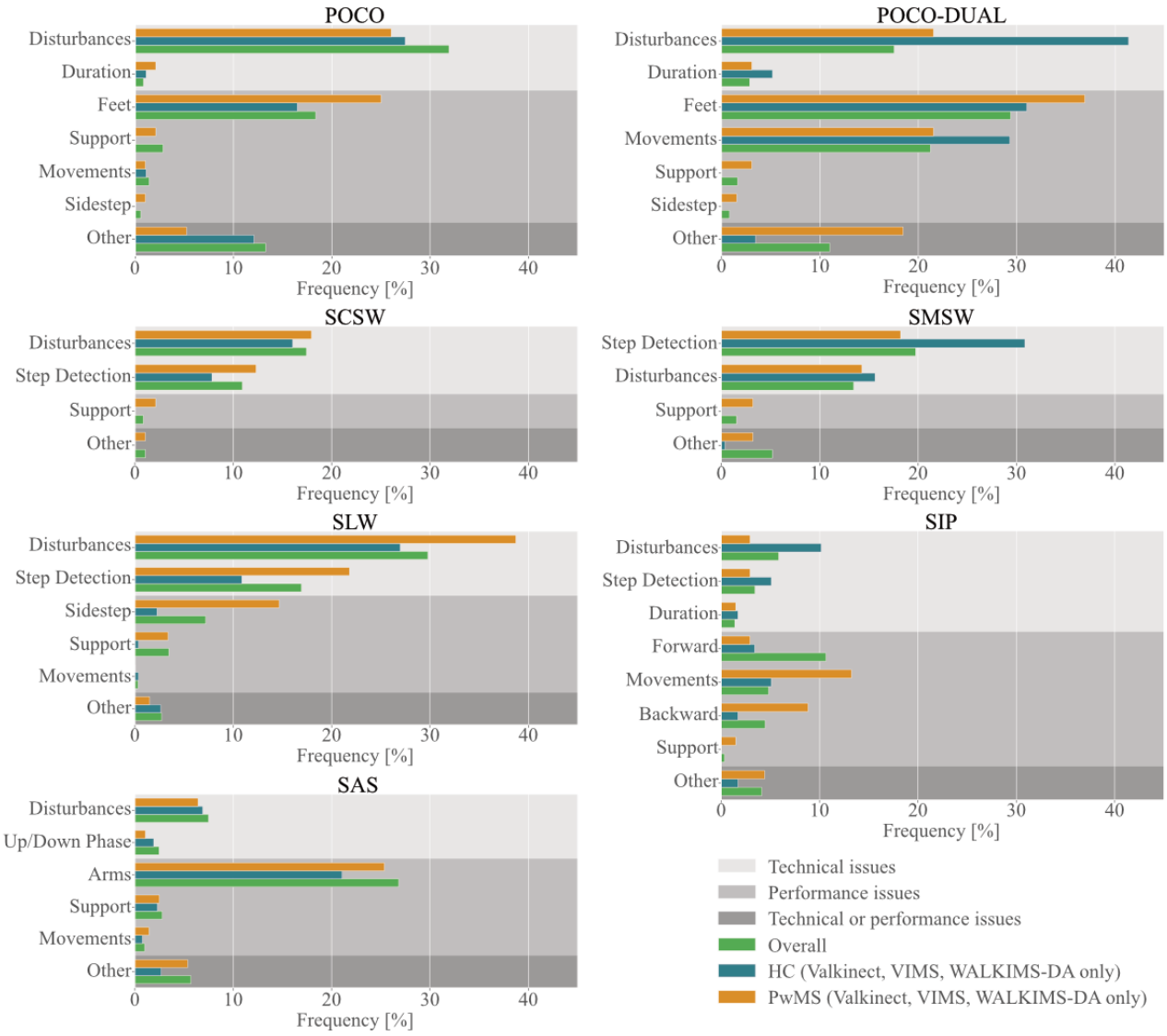
Main Quality Concerns

The main rating criteria compiled during QC pipeline development are listed below, with the respective tasks indicated in parentheses.

- Disturbances, technical issue: Signal disturbances including noisy background, floor, and technical issues with tracking clothing (all tasks)
- Duration, technical issue: Recording duration substantially deviating from 40 seconds, namely a deviation of more than 1 second (POCO, POCO-DUAL, and SIP)
- Step Detection, technical issue: Incorrect Step Detection (SCSW, SMSW, SIP, and SLW)
- Up/Down Phase, technical issue: Incomplete or incorrectly detected standing-up or sitting-down phase (SAS)
- Arms, performance issue: Arms not hanging loosely down at the beginning of the recording (SAS)
- Backward, performance issue: Subject walking backward by more than 50 cm or exhibiting a deliberate backward correction (SIP)
- Feet, performance issue: Deviation from closed feet position, namely if the feet are in an open or a V-shaped position (POCO and POCO-DUAL)
- Forward, performance issue: Subject moving forward by more than 50 cm (SIP)
- Movements, performance issue: Task-unassociated movements such as scratching or gesturing (POCO, POCO-DUAL, SLW, SIP, and SAS)
- Sidestep, performance issue: 1 or multiple sidesteps (POCO, POCO-DUAL, and SLW)
- Support, performance issue: Subject needing support from a walking stick, walls, rollator, or the like (all tasks)
- Other, technical or performance issue: Other/unlisted criterion (all tasks)

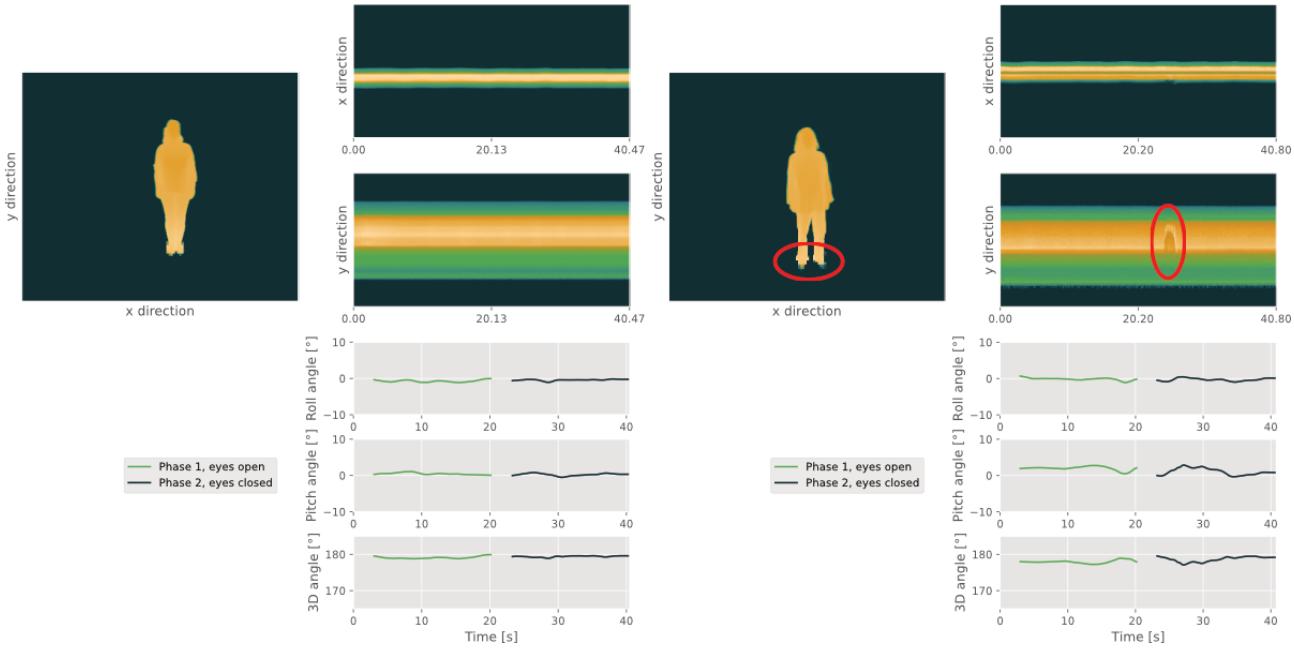
Respective selection frequencies (multiple selections were possible) are illustrated in Figure 3. Possible disease-associated differences in data quality can be estimated from the 3 studies featuring healthy controls and people with MS, namely VIMS, Valkinect, and WALKIMS-DA.

Figure 3. Selection frequencies of technical and performance-related rating criteria for all subjects as well as split by group for the 3 studies featuring healthy controls and people with multiple sclerosis. HC: healthy controls; POCO: Postural Control; POCO-DUAL: Postural Control with Dual Task; PwMS: people with multiple sclerosis; SAS: Stand Up and Sit Down; SCSW: Short Comfortable Speed Walk; SIP: Stepping in Place; SLW: Short Line Walk; SMSW: Short Maximum Speed Walk.



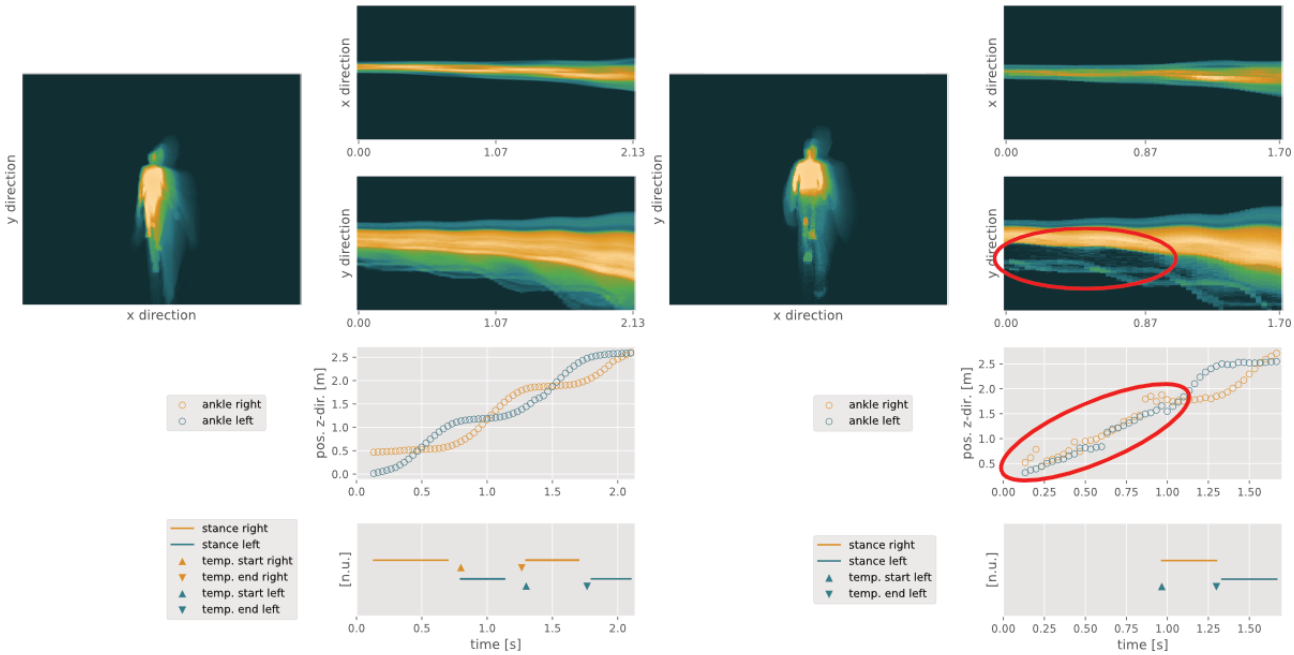
The most prevalent quality concerns comprised Feet, Disturbances, and Other for POCO and additionally Movements for POCO-DUAL. An example of a POCO recording that was discarded due to incorrect Feet positioning as well as unassociated Movements, namely the most frequent performance-associated quality concerns, can be found in Figure 4. For POCO-DUAL, supposedly task-unassociated movements were tagged with Movements and Other by the raters. However, these hand and arm movements often seemed to result from cognitive efforts made during mental arithmetic. In this case, no clear distinction between task-associated and task-unassociated movements can be made. Regarding technical quality concerns, raters' comments suggested that recordings tagged with Disturbances or Other most often exhibited noisy or corrupt leg, feet, or floor signals.

Figure 4. Left: quality control pipeline visualization screenshot of a high-quality Postural Control recording. Right: quality control pipeline visualization screenshot of a Postural Control recording featuring 2 frequently observed performance-related quality concerns, incorrect Feet positioning (according to standard operating procedures, the forefoot and heel should be closed) and unassociated hand Movements around second 22.



Prevalent quality concerns for gait tasks were Disturbances and Step Detection in SLW and— less frequently—SCSW and SMSW. A cross-dependency between the 2 criteria was often observed when unsuitable clothing led to noisy signals (noted as Disturbances by the raters), which in turn leads to issues concerning Step Detection. An example of this issue for an SCSW recording is depicted in Figure 5. Other Disturbances related to floor reflections were not associated with Step Detection issues as often.

Figure 5. Left: quality control pipeline visualization screenshot of a high-quality short comfortable speed walk recording. Right: quality control pipeline visualization screenshot of a Short Comfortable Speed Walk recording featuring a frequently observed technical quality concern, unsuitable clothing causing Disturbances and thus Step Detection issues. Abbreviation temp. represents temporal and indicates the detected stance phases used for temporal rather than spatial parameters.



Excessive forward locomotion (Forward) was the most frequent quality concern for SIP recordings. However, from our experience, the chosen threshold of 50 cm forward motion is rather conservative and distances up to 80-100 cm might be tolerable.

The most prominent problem for SAS was incorrect arm positioning (Arms) at the beginning of a recording. Such incorrect arm positioning was not easily discernible from the motion profile alone and raters usually consulted the provided depth videos to confirm this specific quality concern. Furthermore, a mistake in signal plot generation for

SAS—affecting 3.8% of SAS plots—led to an overestimation of recordings affected by the Up/Down Phase criterion. Figure 3 provides raw ratings, and the represented numbers hence reflect this overestimation.

Disparities between people with MS and healthy controls for performance-related quality aspects were apparent for the generally less often observed Support (all tasks) and Sidestep (POCO, POCO-DUAL, and SLW) issues. This can be interpreted as a disease-related difficulty or the inability to follow task instructions. Results regarding incorrect Feet positioning during POCO and POCO-DUAL did not allow for the interpretation of this criterion as a mainly disease-related one. This criterion as well as Forward and Backward motion during SIP and the incorrect starting position of the Arms during SAS were present in both groups, though slightly more frequent in people with MS. Frequencies of the observed quality criteria further subdivided for all studies can be found in Table S4 in Multimedia Appendix 1.

Discussion

This study presents a post hoc QC pipeline for clinical users of an IMA system. Its core consists of an interface, which enables an intuitive usability decision for individual recordings based on an extendable set of quality criteria. The pipeline proved highly feasible for users—including raters less acquainted with the IMA system itself—and yielded acceptable rater concordance. Its application in a large set of recordings from healthy controls and people with MS demonstrated the utility and necessity of post hoc QC to ensure reliable data and avoid misinterpretation of IMA results. It further identified points for improvement in preventive and ad hoc QC. To our knowledge, this is the first study to systematically investigate QC aspects and propose a clinically applicable QC pipeline for visual perceptive computing.

In the following, we will discuss 2 main aspects of our results. First, the rater concordance, which indicates the feasibility and limitations of our QC approach, and second, the usability decisions themselves, which indicate the quality and limitations of our data.

Rater concordance between 71.5% to 92.3% was generally acceptable. Only for SMSW, strictly opposed keep/discard decisions occurred to a relevant extent (10.5%). This was mostly caused by 1 rater's discard decisions because no full gait cycle was captured. Due to the limited recording range of the depth camera, this is a frequent observation for SMSW and cannot be directly attributed to technical or performance issues. Generally, discordance may reflect ambiguity regarding rating criteria, difficulties in the evaluation of individual cases, or rater oversight. Probably only 1 rater, most likely the operator of the system, will apply post hoc QC in future clinical applications. Thus, possible reasons for rater discordance should be carefully addressed in further development of the QC pipeline, for instance, by specifying the rating criteria, as well as conducting more targeted rater trainings. However, as with other clinical judgments, QC decisions will remain informed, but ultimately intuitive decisions.

Usability decisions were interpreted as follows. Recordings receiving a unanimous keep or discard decision from the corresponding 2 raters were regarded as having assessable and satisfactory or unsatisfactory quality, respectively. Remaining recordings with discordant or undecided usability decisions were classified as needing further investigation, thus being less assessable and with potentially objectionable quality. The proportion of unanimous keep decisions varied substantially between tasks (39.6%-85.1%). In this respect, the SCSW task had the most favorable results with the highest rater concordance (92.3%) and the highest proportion of keep decisions among all tasks. At the other end of the spectrum were POCO and POCO-DUAL with rather moderate rater concordance (71.5% and 72.7%, respectively) and comparatively less unanimous keep decisions (50.3% and 39.6%, respectively). This partial ambiguity supports our inclusion of undecided as an option to avoid forced decisions as well as free text comments to enable marking of unexpected quality concerns.

Regarding technical quality issues, the short walk tasks SCSW, SMSW, and SLW suffered the most from unfavorable properties of clothing that hampered infrared light reflection [32]. POCO and POCO-DUAL often exhibited noisy and cutoff feet signals, attributable to a limited differentiation of feet and ground leading to unstable landmark estimations, as reported earlier [7]. Countermeasures include general recommendations toward subjects' clothing and flooring at the measurement site.

We expected performance-related quality concerns to be associated with physical limitations and thus the disease status to some extent. This seemed to apply to rating criteria Sidestep and Support. However, the more commonly observed performance-related issues (eg, Feet and Movements for POCO and POCO-DUAL, Arms for SAS, and Forward for SIP) occurred in healthy subjects as well. This implies that mistakes in task instruction or ad hoc QC occurred to a relevant degree, despite detailed SOPs and operator training. Even higher proportions of performance-related issues may be expected with wider clinical use or in unsupervised telemedical applications. Thus, further IMA development should aim to implement technical measures for automated real-time detection of performance issues and respective response plans (eg, reinstruction and repetition). Performance-related quality concerns may specifically apply to the assessment of motor capacity in a lab setting or in task-based assessments as opposed to the recently proposed IMA systems for continuous assessment of motor performance [4,5,15].

In the literature, we found generally sparse reporting of QC aspects for IMA. This includes reporting of unobjectionable data quality, which we assume to be unlikely. As an indicator of technical IMA system performance, some authors reported exclusion of IMA recordings due to seemingly blatant technical failures, with rates ranging from a few corrupted examples to recordings of 48.8% of the participants [21,22,33,34]. Unfortunately, respective proportions could not be provided for our data set, as we did not track recordings discarded ad hoc. Regarding data exclusion in postprocessing, outlier detection was the most frequent approach. For univariate outlier detection on normative gait and balance parameters in children, exclusion rates of 2.5% and 6% were reported [20,35]. A multivariate

outlier detection approach on kinematic gait data with successive expert evaluation identified erroneous Step Detection in 3.4% of the subjects [21], whereas a custom post hoc QC procedure applied on SMSW data obtained using Motognosis Labs led to exclusion of 6.7% of the recordings [24]. We consider the QC approach presented here to be rather conservative when compared to outlier detection. It is highly possible that significant quality concerns identified at the raw data level would not be detected by outlier analysis at the kinematic parameter level. For example, failure to stand with closed feet during POCO most likely results in reduced postural sway, which would be mistaken for higher postural stability in the respective subject at the kinematic outcome level.

Lastly, reporting of manual postprocessing, for example, using the GAITRite footfall labeling tool, is often limited to whether it was employed at all [22,36], and respective proportions are only seldom addressed [37].

Beyond IMA, the need for QC has been recognized for other technical procedures. In the context of MS research, magnetic resonance imaging and optical coherence tomography serve as examples for which recommendations have been made regarding standardized protocols, QC, and harmonious reporting thereof [38-42]. Therefore, we propose standardized reporting of IMA results to include information regarding the following: (1) number of recording failures during data acquisition; (2) type and amount of applied postprocessing, both technical and manual; (3) fraction of recordings undergoing QC; (4) fraction of recordings ultimately excluded from analysis (mention of respective causes would be highly valuable for future users)

Limitations of this study may include the decision to have each recording viewed by 2 out of 8 available raters; this limits formal interrater reliability analyses and does not assess individual rater bias. However, we did not aim to establish interrater reliability but focused on obtaining generalizable estimates of rater concordance and determining the feasibility of the approach with a reasonably diverse set of raters. Further, other possible factors influencing usability of the recordings were not specifically analyzed. These include effect of the study site, population, system operators, as well as subjects' age, height, and weight. However, we consider QC results generalizable to and representative of routine applications because of the large size and heterogeneity of our sample. Differences in hardware were not tracked in this study (Kinect 2 sensors and laptops). Likewise, differences in software versions were disregarded because they were considered not substantial. However, recommendations regarding hardware and software may prospectively play a role in preventive QC in large-scale applications.

Regarding transferability, the visualizations employed here were specific to Motognosis Labs. However, appropriate visualizations have been implemented for other IMA systems as well. Examples include footprint depictions from pressure-sensitive walkways or acceleration illustrations from inertial sensors. Thus, we expect the general QC approach presented in this study to be transferable to other IMA systems. As for the observed quality concerns, technical issues are mostly or partially transferable to other depth camera– or visual sensor–based systems, respectively. The performance issues observed here are even more generalizable and thus highly informative for all researchers and clinicians using lab- or task-based IMA. The results of this study clearly support the need for QC of IMA data to ensure objectivity and enhance acceptance by clinical users and regulators alike. As a first step, this approach can advance consensus on the QC standards of different IMA systems and ultimately improve data quality.

Acknowledgments

We thank NeuroCure Clinical Research Center (NCRC), funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2049 – 390688087 and Charité-BIH Clinical Study Center. Parts of this work were funded by a restricted research grant from Roche. DK is a participant in the BIH-Charité Clinician Scientist Program funded by the Charité –Universitätsmedizin Berlin and the Berlin Institute of Health. SCR was supported by the Clinician-Scientist Fellowship from the Stifterverband für die Deutsche Wissenschaft and the Hertie Network of Excellence in Clinical Neuroscience of the Gemeinnützige Hertie-Stiftung.

Conflicts of Interest

AUB is a shareholder and HMR is a paid part-time employee at Motognosis GmbH. The other authors declare no conflict of interest.

Multimedia Appendix 1

Further information on data set, usability decisions, and rating criteria statistics.
[DOCX File , 110 KB-Multimedia Appendix 1]

References

1. Petraglia F, Scarcella L, Pedrazzi G, Brancato L, Puers R, Costantino C. Inertial sensors versus standard systems in gait analysis: a systematic review and meta-analysis. Eur J Phys Rehabil Med 2019 Apr;55(2):265-280 [FREE Full text] [doi: 10.23736/S1973-9087.18.05306-6] [Medline: 30311493]

2. Muro-de-la-Herran A, Garcia-Zapirain B, Mendez-Zorrilla A. Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. Sensors (Basel) 2014 Feb;14(2):3362-3394 [FREE Full text] [doi: 10.3390/s140203362] [Medline: 24556672]
3. Pradhan C, Wuehr M, Akrami F, Neuhaeusser M, Huth S, Brandt T, et al. Automated classification of neurological disorders of gait using spatio-temporal gait parameters. J Electromyogr Kinesiol 2015 Apr;25(2):413-422. [doi: 10.1016/j.jelekin.2015.01.004] [Medline: 25725811]
4. Frechette ML, Meyer BM, Tulipani LJ, Gurchiek RD, McGinnis RS, Sosnoff JJ. Next steps in wearable technology and community ambulation in multiple sclerosis. Curr Neurol Neurosci Rep 2019 Sep;19:80. [doi: 10.1007/s11910-019-0997-9] [Medline: 31485896]
5. Alexander S, Peryer G, Gray E, Barkhof F, Chataway J. Wearable technologies to measure clinical outcomes in multiple sclerosis: a scoping review. Mult Scler 2021 Oct;27(11):1643-1656 [FREE Full text] [doi: 10.1177/1352458520946005] [Medline: 32749928]
6. Maetzler W, Rochester L, Bhidayasiri R, Espay AJ, Sánchez-Ferro A, van Uem JMT. Modernizing daily function assessment in Parkinson's disease using capacity, perception, and performance measures. Mov Disord 2021 Jan;36(1):76-82. [doi: 10.1002/mds.28377] [Medline: 33191498]
7. Otte K, Kayser B, Mansow-Model S, Verrel J, Paul F, Brandt AU, et al. Accuracy and reliability of the kinect version 2 for clinical measurement of motor function. PLoS One 2016;11(11):e0166532 [FREE Full text] [doi: 10.1371/journal.pone.0166532] [Medline: 27861541]
8. Webster KE, Wittwer JE, Feller JA. Validity of the GAITRite® walkway system for the measurement of averaged and individual step parameters of gait. Gait Posture 2005 Dec;22(4):317-321. [doi: 10.1016/j.gaitpost.2004.10.005] [Medline: 16274913]
9. Wren TAL, Tucker CA, Rethlefsen SA, Gorton GEIII, Õunpuu S. Clinical efficacy of instrumented gait analysis: systematic review 2020 update. Gait Posture 2020 Jul;80:274-279. [doi: 10.1016/j.gaitpost.2020.05.031] [Medline: 32563727]
10. Morrison C, D'Souza M, Huckvale K, Dorn JF, Burggraaff J, Kamm CP, et al. Usability and acceptability of ASSESS MS: assessment of motor dysfunction in multiple sclerosis using depth-sensing computer vision. JMIR Hum Factors 2015 Jan;2(1):e11 [FREE Full text] [doi: 10.2196/humanfactors.4129] [Medline: 27025782]
11. Espay AJ, Hausdorff JM, Sánchez-Ferro Á, Klucken J, Merola A, Bonato P, Movement Disorder Society Task Force on Technology. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. Mov Disord 2019 May;34(5):657-663 [FREE Full text] [doi: 10.1002/mds.27671] [Medline: 30901495]
12. Brichetto G, Pedullà L, Podda J, Tacchino A. Beyond center-based testing: understanding and improving functioning with wearable technology in MS. Mult Scler 2019 Sep;25(10):1402-1411. [doi: 10.1177/1352458519857075] [Medline: 31502913]
13. Maetzler W, Klucken J, Horne M. A clinical view on the development of technology-based tools in managing Parkinson's disease. Mov Disord 2016 Sep;31(9):1263-1271. [doi: 10.1002/mds.26673] [Medline: 27273651]
14. Stamate C, Magoulas GD, Kueppers S, Nomikou E, Daskalopoulos I, Jha A, et al. The cloudUPDRS app: a medical device for the clinical assessment of Parkinson's Disease. Pervasive Mob Comput 2018 Jan;43:146-166. [doi: 10.1016/j.pmcj.2017.12.005]
15. Viceconti M, Hernandez Penna S, Dartee W, Mazzà C, Caulfield B, Becker C, et al. Toward a regulatory qualification of real-world mobility performance biomarkers in Parkinson's patients using digital mobility outcomes. Sensors (Basel) 2020 Oct;20(20):5920 [FREE Full text] [doi: 10.3390/s20205920] [Medline: 33092143]
16. Steinert A, Sattler I, Otte K, Röhling H, Mansow-Model S, Müller-Werdan U. Using new camera-based technologies for gait analysis in older adults in comparison to the established GAITRite system. Sensors (Basel) 2019 Dec;20(1):125 [FREE Full text] [doi: 10.3390/s20010125] [Medline: 31878177]
17. Mancini M, Salarian A, Carlson-Kuhta P, Zampieri C, King L, Chiari L, et al. ISway: a sensitive, valid and reliable measure of postural control. J Neuroeng Rehabil 2012 Aug;9:59 [FREE Full text] [doi: 10.1186/1743-0003-9-59] [Medline: 22913719]
18. Kressig RW, Beauchet O, European GAITRite® Network Group. Guidelines for clinical applications of spatio-temporal gait analysis in older adults. Aging Clin Exp Res 2006 Apr;18:174-176. [doi: 10.1007/BF03327437] [Medline: 16702791]
19. Hinton DC, Cheng Y, Paquette C. Everyday multitasking habits: university students seamlessly text and walk on a split-belt treadmill. Gait Posture 2018 Jan;59:168-173. [doi: 10.1016/j.gaitpost.2017.10.011] [Medline: 29032000]
20. Guffey K, Regier M, Mancinelli C, Pergami P. Gait parameters associated with balance in healthy 2- to 4-year-old children. Gait Posture 2016 Jan;43:165-169 [FREE Full text] [doi: 10.1016/j.gaitpost.2015.09.017] [Medline: 26439183]
21. Sunderland KM, Beaton D, Fraser J, Kwan D, McLaughlin PM, Montero-Odasso M, ONDRI Investigators, et al. The utility of multivariate outlier detection techniques for data quality evaluation in large studies: an application within the ONDRI project. BMC Med Res Methodol 2019 May;19:102 [FREE Full text] [doi: 10.1186/s12874-019-0737-5] [Medline: 31092212]
22. Martindale CF, Roth N, Gasner H, List J, Regensburger M, Eskofier BM, et al. Technical validation of an automated mobile gait analysis system for hereditary spastic paraplegia patients. IEEE J Biomed Health Inform 2020 May;24(5):1490-1499. [doi: 10.1109/JBHI.2019.2937574] [Medline: 31449035]

23. Rampp A, Barth J, Schülein S, Gaßmann KG, Klucken J, Eskofier BM. Inertial sensor-based stride parameter calculation from gait sequences in geriatric patients. IEEE Trans Biomed Eng 2015 Apr;62(4):1089-1097. [doi: 10.1109/TBME.2014.2368211] [Medline: 25389237]
24. Grobelny A, Behrens JR, Mertens S, Otte K, Mansow-Model S, Krüger T, et al. Maximum walking speed in multiple sclerosis assessed with visual perceptive computing. PLoS One 2017 Dec;12(12):e0189281 [FREE Full text] [doi: 10.1371/journal.pone.0189281] [Medline: 29244874]
25. Kroneberg D, Elshehabi M, Meyer A, Otte K, Doss S, Paul F, et al. Less is more–estimation of the number of strides required to assess gait variability in spatially confined settings. Front Aging Neurosci 2018 Jan;10:435 [FREE Full text] [doi: 10.3389/fnagi.2018.00435] [Medline: 30719002]
26. Behrens J, Pfüller C, Mansow-Model S, Otte K, Paul F, Brandt AU. Using perceptive computing in multiple sclerosis-the Short Maximum Speed Walk test. J Neuroeng Rehabil 2014 May;11:89 [FREE Full text] [doi: 10.1186/1743-0003-11-89] [Medline: 24886525]
27. Behrens JR, Mertens S, Krüger T, Grobelny A, Otte K, Mansow-Model S, et al. Validity of visual perceptive computing for static posturography in patients with multiple sclerosis. Mult Scler 2016 Oct;22(12):1596-1606. [doi: 10.1177/1352458515625807] [Medline: 26814201]
28. Drebinger D, Rasche L, Kroneberg D, Althoff P, Bellmann-Strobl J, Weygandt M, et al. Association between fatigue and motor exertion in patients with multiple sclerosis—a prospective study. Front Neurol 2020 Apr;11:208 [FREE Full text] [doi: 10.3389/fneur.2020.00208] [Medline: 32351439]
29. Veauthier C, Ryczewski J, Mansow-Model S, Otte K, Kayser B, Glos M, et al. Contactless recording of sleep apnea and periodic leg movements by nocturnal 3-D-video and subsequent visual perceptive computing. Sci Rep 2019 Nov;9:16812 [FREE Full text] [doi: 10.1038/s41598-019-53050-3] [Medline: 31727918]
30. Otte K, Ellermeyer T, Suzuki M, Röhling HM, Kuroiwa R, Cooper G, et al. Cultural bias in motor function patterns: potential relevance for predictive, preventive, and personalized medicine. EPMA J 2021 Mar;12(1):91-101 [FREE Full text] [doi: 10.1007/s13167-021-00236-3] [Medline: 33782636]
31. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983 Nov;33(11):1444-1452. [doi: 10.1212/WNL.33.11.1444] [Medline: 6685237]
32. Lachat E, Macher H, Landes T, Grussenmeyer P. Assessment and calibration of a RGB-D Camera (Kinect v2 Sensor) towards a potential use for close-range 3D modeling. Remote Sens 2015 Oct;7(10):13070-13097. [doi: 10.3390/rs71013070]
33. Ramsperger R, Meckler S, Heger T, van Uem J, Hucker S, Braatz U, SENSE-PARK study team. Continuous leg dyskinesia assessment in Parkinson's disease—clinical validity and ecological effect. Parkinsonism Relat Disord 2016 May;26:41-46. [doi: 10.1016/j.parkreldis.2016.02.007] [Medline: 26952699]
34. Larsson J, Ekvall Hansson E, Miller M. Increased double support variability in elderly female fallers with vestibular asymmetry. Gait Posture 2015 Mar;41(3):820-824. [doi: 10.1016/j.gaitpost.2015.02.019] [Medline: 25800649]
35. Dusing SC, Thorpe DE. A normative sample of temporal and spatial gait parameters in children using the GAITRite® electronic walkway. Gait Posture 2007 Jan;25(1):135-139. [doi: 10.1016/j.gaitpost.2006.06.003] [Medline: 16875823]
36. Schmitz-Hübsch T, Brandt AU, Pfueller C, Zange L, Seidel A, Kühn AA, et al. Accuracy and repeatability of two methods of gait analysis – GaitRite™ und Mobility Lab™ – in subjects with cerebellar ataxia. Gait Posture 2016 Jul;48:194-201. [doi: 10.1016/j.gaitpost.2016.05.014] [Medline: 27289221]
37. Wong JS, Jasani H, Poon V, Inness EL, McIlroy WE, Mansfield A. Inter- and intra-rater reliability of the GAITRite system among individuals with sub-acute stroke. Gait Posture 2014 May;40(1):259-261. [doi: 10.1016/j.gaitpost.2014.02.007] [Medline: 24630463]
38. Tewarie P, Balk L, Costello F, Green A, Martin R, Schippling S, et al. The OSCAR-IB consensus criteria for retinal OCT quality assessment. PLoS One 2012 Apr;7(4):e34823 [FREE Full text] [doi: 10.1371/journal.pone.0034823] [Medline: 22536333]
39. Cruz-Herranz A, Balk LJ, Oberwahrenbrock T, Saidha S, Martinez-Lapiscina EH, Lagreze WA, IMSVISUAL consortium. The APOSTEL recommendations for reporting quantitative optical coherence tomography studies. Neurology 2016 Jun;86(24):2303-2309 [FREE Full text] [doi: 10.1212/WNL.0000000000002774] [Medline: 27225223]
40. Traboulsee A, Simon JH, Stone L, Fisher E, Jones DE, Malhotra A, et al. Revised recommendations of the consortium of MS Centers Task Force for a standardized MRI protocol and clinical guidelines for the diagnosis and follow-up of multiple sclerosis. AJNR Am J Neuroradiol 2016 Mar;37(3):394-401 [FREE Full text] [doi: 10.3174/ajnr.A4539] [Medline: 26564433]
41. Motamedi S, Gawlik K, Ayadi N, Zimmermann HG, Asseyer S, Bereuter C, et al. Normative data and minimally detectable change for inner retinal layer thicknesses using a semi-automated OCT image segmentation pipeline. Front Neurol 2019 Nov;10:1117 [FREE Full text] [doi: 10.3389/fneur.2019.01117] [Medline: 31824393]
42. Chien C, Juenger V, Scheel M, Brandt AU, Paul F. Considerations for mean upper cervical cord area implementation in a longitudinal MRI setting: methods, interrater reliability, and MRI quality control. AJNR Am J Neuroradiol 2020 Feb;41(2):343-350. [doi: 10.3174/ajnr.A6394] [Medline: 31974079]

Abbreviations

GUI: graphical user interface
IMA: instrumented motion analysis
MS: multiple sclerosis
PASS-MS: Perceptive Assessment in Multiple Sclerosis (ie, name of short motor assessment battery recorded with Motognosis Labs)
POCO: Postural Control
POCO-DUAL: Postural Control with Dual Task
QC: quality control
SAS: Stand Up and Sit Down
SCSW: Short Comfortable Speed Walk
SIP: Stepping in Place
SLW: Short Line Walk
SMSW: Short Maximum Speed Walk
SOP: standard operating procedure

Edited by A Kushniruk; submitted 29.12.20; peer-reviewed by C Mazza, E Ravera; comments to author 16.03.21; revised version received 02.05.21; accepted 07.12.21; published 01.04.22

Please cite as:
Röhling HM, Althoff P, Arsenova R, Drebinger D, Gigengack N, Chorschew A, Kroneberg D, Rönnefarth M, Ellermeyer T, Rosenkranz SC, Heesen C, Behnia B, Hirano S, Kuwabara S, Paul F, Brandt AU, Schmitz-Hübsch T
Proposal for Post Hoc Quality Control in Instrumented Motion Analysis Using Markerless Motion Capture: Development and Usability Study
JMIR Hum Factors 2022;9(2):e26825
URL: https://humanfactors.jmir.org/2022/2/e26825
doi: 10.2196/26825
PMID:

Printing copy publication 2

Röhling HM, Otte K, Rekers S, Finke C, Rust R, Dorsch EM, Behnia B, Paul F, and Schmitz-Hübsch T. RGB-depth camera-based assessment of motor capacity: Normative data for six standardized motor tasks. *International Journal of Environmental Research and Public Health*, 19(24):16989, December 2022.

International Journal of
Environmental Research and Public Health

MDPI

Article

RGB-Depth Camera-Based Assessment of Motor Capacity: Normative Data for Six Standardized Motor Tasks

Hanna Marie Röhling [1,2,3,4,*], Karen Otte [1,2,3,4], Sophia Rekers [5,6], Carsten Finke [5,6], Rebekka Rust [1,2,3,7], Eva-Maria Dorsch [1,2,3,5], Behnoush Behnia [8], Friedemann Paul [1,2,3,5,7] and Tanja Schmitz-Hübsch [1,2,3,7,*]

1 Experimental and Clinical Research Center, a Cooperation between the Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association and the Charité—Universitätsmedizin Berlin, 13125 Berlin, Germany
2 Experimental and Clinical Research Center, Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, 13125 Berlin, Germany
3 Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany
4 Motognosis GmbH, 10119 Berlin, Germany
5 Department of Neurology, Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, 10117 Berlin, Germany
6 Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, 10117 Berlin, Germany
7 NeuroCure Clinical Research Center, Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, 10117 Berlin, Germany
8 Department of Psychiatry and Psychotherapy, Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, 12203 Berlin, Germany
* Correspondence: hanna-marie.roehling@charite.de (H.M.R.); tanja.schmitz-huebsch@charite.de (T.S.-H.)

check for updates

Citation: Röhling, H.M.; Otte, K.; Rekers, S.; Finke, C.; Rust, R.; Dorsch, E.-M.; Behnia, B.; Paul, F.; Schmitz-Hübsch, T. RGB-Depth Camera-Based Assessment of Motor Capacity: Normative Data for Six Standardized Motor Tasks. *Int. J. Environ. Res. Public Health* **2022**, *19*, 16989. https://doi.org/10.3390/ijerph192416989

Academic Editor: Monika Błaszczyszyn

Received: 11 October 2022
Accepted: 13 December 2022
Published: 17 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: Background: Instrumental motion analysis constitutes a promising development in the assessment of motor function in clinical populations affected by movement disorders. To foster implementation and facilitate interpretation of respective outcomes, we aimed to establish normative data of healthy subjects for a markerless RGB-Depth camera-based motion analysis system and to illustrate their use. Methods: We recorded 133 healthy adults (56% female) aged 20 to 60 years with an RGB-Depth camera-based motion analysis system. Forty-three spatiotemporal parameters were extracted from six short, standardized motor tasks—including three gait tasks, stepping in place, standing-up and sitting down, and a postural control task. Associations with confounding factors, height, weight, age, and sex were modelled using a predictive linear regression approach. A z-score normalization approach was provided to improve usability of the data. Results: We reported descriptive statistics for each spatiotemporal parameter (mean, standard deviation, coefficient of variation, quartiles). Robust confounding associations emerged for step length and step width in comfortable speed gait only. Accessible normative data usage was lastly exemplified with recordings from one randomly selected individual with multiple sclerosis. Conclusion: We provided normative data for an RGB depth camera-based motion analysis system covering broad aspects of motor capacity.

Keywords: instrumental motion analysis; normative data; RGB-Depth camera; Microsoft Kinect v2; gait analysis; tandem gait; postural control; stepping in place; standing up and sitting down

1. Introduction

Identification and monitoring of motor impairments are key elements in the management of diseases impacting motor function. The instrumental task-based assessment of motor capacity provides an alternative to observation and assessment by clinical experts and analog standardized tests, such as the Timed Up and Go Test [1] or Timed 25-Foot Walk [2]. Due to anticipated time and cost efficiency as well as outcome objectivity, instrumental motion analysis has drawn increasing attention in recent years. Telemedical use of such technologies from patients' homes can further protect vulnerable groups and provide relief for overburdened healthcare systems.

Interpretability of outcomes from instrumental motion analysis is, however, still limited. Heterogeneous usage of different technologies, movement protocols, extracted parameters and respective algorithms rarely allows for robust between-system compatibility. Reliable normative data for the system in use thus comprises a crucial prerequisite for the interpretation of outcomes. Respective values can help to define meaningful thresholds for assumed pathology, expected variability, as well as dependencies on anthropometric and demographic features. Normative datasets can further aid in improving harmonization by revealing systematic biases between system outputs.

Large normative datasets of gait parameters ranging from gait speed to arm swing asymmetry continue to be of high interest to the scientific community and are comparatively prevalent [3–7]. Few larger databases exist for other motor tasks, such as tandem gait and postural control [8–10]. However, most reported parameter values from healthy controls stem from small case-control or proof-of-concept studies and comprise around 30 healthy subjects or less [11–21]. For this study, a motion analysis system based on the use of a single RGB-Depth camera (Microsoft Kinect v2) was employed, which has been evaluated for accuracy and reliability [11,22] and has been used in various clinical populations [5,23,24]. Previous related works from our groups include outcome parameters from healthy controls [23,25–29], but these do not represent a robust normative database by themselves due to likewise limited sample sizes, restriction to single motor tasks, or general study design.

In this study, we thus aimed to provide elaborate normative values for spatiotemporal parameters of an RGB-Depth camera-based motion analysis system for six different motor tasks. We further assessed associations of the parameters with confounding demographic and anthropometric factors and illustrated usage of the normative data.

2. Materials and Methods

2.1. Participants

One-hundred-thirty-three participants were pooled from control groups of two multiple sclerosis studies (acronyms: Valkinect, VIMS) and one autism spectrum disorder study (acronym ASD) at Charité—Universitätsmedizin Berlin, Berlin, Germany. Participants were recruited via social media posts, institutional databases, intranet, and by approaching accompanying persons from respective case cohorts.

Exclusion criteria were psychiatric disorders, chronic neurological diseases, or acute motor impairments. Adapted from norm data specifications for another commercially available system [30,31], we used five persons per sex and age decade as a lower limit in sample size planning. We focused on adult, decidedly non-geriatric, individuals and thus included participants within the ages of 20 and 60, representing an adequate control group for common neuroimmunological conditions. All included participants received instrumental motion analysis and had complete information regarding age, sex, height, and weight (Table 1, visualized in Supplementary Figure S1).

Table 1. Anthropometric and demographic subject characteristics overall and subdivided by study. Abbreviations: SD: standard deviation.

Study	Sample Size (% Female)	Age Mean (SD; Range) [Years]	Height Mean (SD; Range) [cm]	Weight Mean (SD; Range) [kg]	BMI Mean (SD; Range) [kg/m2]
All	133 (56%)	36.83 (10.44; 20–60)	172.89 (9.34; 153–194)	71.80 (13.86; 46–115)	23.94 (3.79; 17.75–34.33)
ASD	41 (51%)	33.88 (7.99; 20–49)	174.17 (9.65; 155–194)	73.85 (16.06; 46–115)	24.24 (4.41; 17.75–33.90)
VIMS	57 (63%)	34.14 (9.06; 20–60)	172.16 (9.66; 153–193)	70.86 (13.69; 47–110)	23.83 (3.79; 18.29–34.33)
Valkinect	35 (51%)	44.69 (11.23; 22–60)	172.60 (8.52; 157–190)	70.91 (11.24; 53–97)	23.76 (3.01; 18.93–32.04)

Data from one randomly selected Valkinect participant with multiple sclerosis was used to exemplify usage of the normative data (male, 53 years, 183 cm, 73 kg).

2.2. Instrumental Motion Analysis

Labs (Motognosis GmbH, Berlin, Germany; versions 1.4.0.2, 1.4.0.3, 2.0.1.0, 2.1.2.0), and a markerless motion analysis system based on a single RGB-Depth consumer camera (Microsoft Kinect v2; Microsoft cooperation, Redmond, WA, USA), were used to record short movement tasks from participants wearing standard clothing and comfortable footwear in a 1.5–4.5 m distance from the sensor (Figure 1). The Kinect v2 sensor was positioned at a height of 1.4 m and tilted in a pitch direction of roughly $-8°$ to $-9°$. Scientific staff operated the system following written standard operating procedures for technical setup and task instruction.

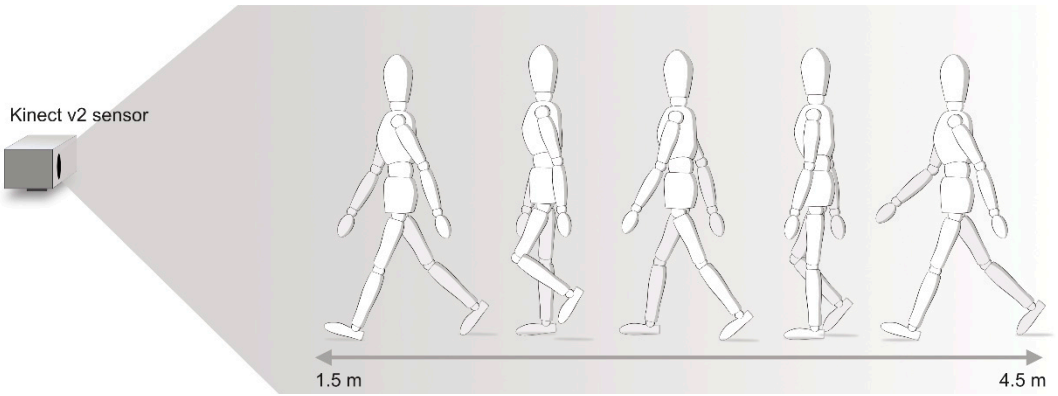


Figure 1. Technical set-up of the Motognosis Labs motion analysis system with a single Microsoft Kinect v2 sensor with exemplary sketch of a gait task. The illustration was included by courtesy of Motognosis GmbH.

We included data for six tasks: short comfortable speed walk (SCSW), short maximum speed walk (SMSW; VIMS and Valkinect only as this task was not included in the ASD measurement protocol), tandem gait referred to as short line walk (SLW), stepping in place (SIP), standing up and sitting down (SAS; VIMS and Valkinect only as this task was not included in the ASD measurement protocol), and postural control (POCO). Within each session SCSW, SMSW, SLW, and SAS were recorded three consecutive times, SIP and POCO were recorded once.

The Microsoft Kinect SDK (version 2.0.14) enables extraction of 25 three-dimensional time series of body landmarks from these recordings, which were used to extract spatiotemporal parameters with custom algorithms. Here, we extracted 43 parameters (Table 2), most of which have been previously introduced [22,23,25–28]. Others were carefully vetted in terms of clinical interest and statistical properties when tested in an independent dataset [22].

Table 2. Motor task descriptions, information on respectively extracted movement signals and spatiotemporal parameters as well as parameter names. Abbreviations: AP: anterior-posterior; n.u.: unitless; RR: Romberg ratio.

Task Description	Movement Signal and Spatiotemporal Parameter Description	Parameter Names
Short comfortable speed walk (SCSW)		
The participant stands just outside the sensor range and walks towards the sensor at comfortable speed in response to an auditory cue	Mean speed derived from pelvic center landmark movement in walk direction	Gait speed [m/s]
	Mean step length, mean step width, and mean step duration over all (left and right) detected steps derived from left and right ankle landmark movement in walk direction	Step length [cm]; Step width [cm]; Step duration [s]
	Mean gait cadence extrapolated from detected steps and recording length	Gait cadence [steps/min]
	Mean angular arm swing amplitude (averaged over left and right averages) and absolute symmetry angle [32] (between left and right mean angular arm swing amplitude) derived from left and right wrist landmarks relative to manubrium landmark movement in anterior-posterior direction	Arm angular amplitude [°]; Arm symmetry angle [n.u.]

Table 2. *Cont.*

Task Description	Movement Signal and Spatiotemporal Parameter Description	Parameter Names
	Short maximum speed walk (SMSW)	
The participant stands just outside the sensor range and walks towards the sensor at maximum speed in response to an auditory cue	Mean speed derived from pelvic center landmark movement in walk direction	Gait speed [m/s]
	Short line walk (SLW)	
The participant stands just outside the sensor area and, in response to an auditory cue, walks towards the sensor in tandem gait, i.e., walks on an imaginary line with the heels touching the toes at each step	Mean and coefficient of variation of progression speed derived from pelvic center landmark movement in walk direction Angular standard deviation and speed of upper body sway starting from pelvic center landmark Line walk cadence derived from recording length and peaks of left and right ankle landmark movement relative to respective hip landmarks Angular standard deviation and speed of arm movement angle (averaged over left and right) derived from elbow landmarks relative to respective shoulder landmarks movement in 3D	Progression speed [°/s]; Relative progression variability [%] Roll sway variability [°]; Roll sway speed [°/s] Line walk cadence [steps/min] Arm variability [°]; Arm speed [°/s]
	Stepping in place (SIP)	
The participant walks on the spot at comfortable pace for 40 s	Mean knee amplitude, mean step duration, and mean stance duration (averaged over left and right averages) derived from knee landmark movement in anterior-posterior direction Mean stepping cadence extrapolated from detected steps and recording length Absolute symmetry angle [32] (between left and right mean knee amplitudes) Mean coefficient of variation of left and right "stride times" measured as time between knee amplitude peaks (i.e., slightly adapted from [23])	Knee amplitude [m]; Step duration [s]; Stance duration [s] Stepping cadence [steps/min] Knee symmetry angle [n.u.] Arrhythmicity [%]
	Standing up and sitting down (SAS)	
The participant sits on an armless chair, arms hanging to the side, stands up after an auditory cue and sits down again after a second auditory cue	Speed of manubrium landmark movement in vertical and anterior-posterior direction Range of manubrium landmark movement in anterior-posterior direction	Transition time (up) [s]; Transition time (down) [s] AP deflection range (up) [m]; AP deflection range (down) [m]
	Postural control (POCO)	
The participant stands with closed feet and open eyes facing the sensor for 20 s; after an auditory cue subject closes eyes and remains in this position for another 20 s	Angular range and mean speed of the body sway vector between mean ankle landmark position and pelvic center landmark during eyes closed and eyes open measurement conditions in pitch, roll, and 3D direction Romberg ratio of sway range and sway speed in pitch, roll, and 3D direction—i.e., value for closed eyes condition divided by respective value for open eyes condition	Pitch/Roll/3D sway range (open eyes) [°]; Pitch/Roll/3D sway speed (open eyes) [°/s]; Pitch/Roll/3D sway range (closed eyes) [°]; Pitch/Roll/3D sway speed (closed eyes) [°/s] RR of pitch/roll/3D sway range [n.u.]; RR of pitch/roll/3D sway speed [n.u.]

2.3. Data Analysis

Recordings with gross performance deviations (e.g., wrong feet position for POCO) or technical errors were identified using a previously described post hoc quality control pipeline [33] and discarded. Statistical analyses were performed, and visualizations were generated using Python 3.7.3 (packages pandas 1.3.5, numpy 1.21.6, statsmodels 0.13.2, seaborn 0.11.2, matplotlib 3.1.0, scipy 1.7.3, scikit-learn 0.21.2). For SCSW, SMSW, SLW and SAS, the extracted spatiotemporal parameters were averaged per participant over all remaining repetitions. Data are presented as group mean, standard deviation, coefficient of variation, and quartiles and distributions are visualized.

To model and address the influence of potential confounders in a generalizable way, we fitted ordinary least squares regression models—*parameter ~ age + sex + height + weight + study*—where sex was dummy-coded (female: 0; male: 1), and the study was used as an effect-coded control variable. Models were first fitted in a repeated (100 times) five-fold cross-validation procedure, using the R^2-value (R^2_{test}) between true and predicted parameter values of respective test sets at each fold, and repetition as a performance indicator. Predicted parameter values were calculated using derived β-values from respective training sets, omitting the study to simulate assessing model performance on external data. For an averaged $R^2_{test} > 0.1$, models were assumed to describe generalizable associations. For these models β-values, corresponding 95% confidence intervals and p-values for the independent variables were extracted after fitting them on the full dataset.

For supplementary information, we extracted bivariate statistics (Pearson's correlation coefficient, independent samples t-test, one-way ANOVA) regarding associations between spatiotemporal parameters and potential confounders.

To better interpret individual behavior in comparison to normative data and across different parameter scales, we applied z-score normalization. The z-score of a value refers to its relative distance to the mean measured in numbers of standard deviations:

$$z_{raw,\,i} = \frac{x_i - \overline{x}}{s}, \tag{1}$$

with x_i being the raw value for a given parameter and subject i, and $\overline{x}$ and s being the raw normative data's sample mean and standard deviation. For spatiotemporal parameters with previously detected confounding associations with age, sex, height and weight, respectively standardized residuals of the linear models were favored over z_{raw}:

$$z_{res,i} = \frac{\varepsilon_i}{s_\varepsilon}, \tag{2}$$

with s_ε being the standard deviation of the normative residuals and

$$\varepsilon_i = x_i - \left(\beta_0 + \beta_{Age}x_{i,Age} + \beta_{Sex}x_{i,Sex} + \beta_{Height}x_{i,Height} + \beta_{Weight}x_{i,Weight}\right). \tag{3}$$

3. Results

3.1. Normative Values

Distributions and statistics presented in this section were produced using only data that passed quality control, performed by one trained researcher. Discard rates (3.2–15.0% depending on task) and reasons for exclusion are provided in supplementary Table S1. The normative values are distributed highly variable (Figure 2): SCSW gait speed, for instance, is approximately normally distributed, while other parameters such as SLW roll sway speed are highly skewed or feature outliers. Respective descriptive statistics are provided in Table 3.

Table 3. Descriptive statistics of spatiotemporal parameters and generalizability estimates regarding linear models describing associations with confounders age, sex, height, and weight. Abbreviations: AP: anterior-posterior; CV: cross-validation; CoV: coefficient of variation; n.u.: unitless; Q1: 25th percentile; Q3: 75th percentile; RR: Romberg ratio; SD: standard deviation.

Spatiotemporal Parameter	Mean	SD	CoV	Q1	Q3	Mean R^2_{test} for Repeated (100×) 5-Fold CV
Short comfortable speed walk (SCSW); n = 126						
Gait speed [m/s]	1.16	0.17	0.15	1.06	1.28	−0.05
Step length [cm]	69.35	7.69	0.11	64.85	74.38	0.14
Step width [cm]	10.19	2.72	0.27	8.24	11.77	0.13
Step duration [s]	0.52	0.05	0.10	0.48	0.56	−0.03
Gait cadence [steps/min]	112.07	10.55	0.09	103.97	120.04	−0.04
Arm angular amplitude [°]	26.48	10.81	0.41	18.19	32.45	−0.14
Arm symmetry angle [n.u.]	0.23	0.16	0.72	0.11	0.30	−0.10
Short maximum speed walk (SMSW); n = 90						
Gait speed [m/s]	1.66	0.18	0.11	1.53	1.77	−0.08
Short line walk (SLW); n = 128						
Progression speed [m/s]	0.35	0.10	0.28	0.29	0.39	−0.07
Relative progression variability [%]	0.33	0.08	0.24	0.27	0.38	−0.09
Roll sway variability [°]	1.80	0.76	0.42	1.21	2.16	−0.12
Roll sway speed [°/s]	5.58	1.90	0.34	4.37	6.47	−0.14
Line walk cadence [steps/min]	71.78	16.35	0.23	60.50	81.39	−0.07
Arm variability [°]	5.32	3.27	0.62	2.94	6.47	−0.12
Arm speed [°/s]	18.20	7.54	0.41	13.24	20.74	−0.06

Table 3. *Cont.*

Spatiotemporal Parameter	Mean	SD	CoV	Q1	Q3	Mean R^2_{test} for Repeated (100×) 5-Fold CV
Stepping in place (SIP); n = 121						
Knee amplitude [m]	0.18	0.06	0.31	0.15	0.23	−0.15
Step duration [s]	0.83	0.11	0.13	0.74	0.88	−0.11
Stance duration [s]	0.39	0.15	0.38	0.28	0.47	0.01
Stepping cadence [steps/min]	98.29	16.13	0.16	87.07	111.00	−0.06
Knee symmetry angle [n.u.]	0.06	0.05	0.87	0.02	0.09	−0.14
Arrhythmicity [%]	6.00	0.84	0.14	5.37	6.50	−0.08
Standing up and sitting down (SAS); n = 90						
Transition time (up) [s]	1.53	0.19	0.12	1.39	1.63	−0.16
Transition time (down) [s]	1.66	0.22	0.13	1.48	1.80	−0.09
AP deflection range (up) [m]	0.37	0.07	0.19	0.32	0.41	−0.11
AP deflection range (down) [m]	0.40	0.08	0.20	0.34	0.46	−0.11
Postural control (POCO); n = 113						
Pitch sway range (open eyes) [°]	0.91	0.43	0.48	0.59	1.15	−0.13
Roll sway range (open eyes) [°]	0.89	0.37	0.42	0.64	1.09	−0.11
3D sway range (open eyes) [°]	0.92	0.41	0.44	0.67	1.10	−0.14
Pitch sway speed (open eyes) [°/s]	0.14	0.05	0.39	0.10	0.16	−0.10
Roll sway speed (open eyes) [°/s]	0.15	0.06	0.37	0.10	0.18	−0.11
3D sway speed (open eyes) [°/s]	0.22	0.07	0.34	0.17	0.26	−0.10
Pitch sway range (closed eyes) [°]	1.12	0.50	0.45	0.74	1.40	−0.16
Roll sway range (closed eyes) [°]	1.03	0.43	0.41	0.73	1.27	−0.10
3D sway range (closed eyes) [°]	1.09	0.50	0.46	0.71	1.41	−0.15
Pitch sway speed (closed eyes) [°/s]	0.18	0.06	0.36	0.14	0.22	−0.08
Roll sway speed (closed eyes) [°/s]	0.20	0.08	0.39	0.14	0.25	−0.07
3D sway speed (closed eyes) [°/s]	0.30	0.10	0.33	0.21	0.34	−0.06
RR of pitch sway range [n.u.]	1.42	0.75	0.53	0.94	1.75	−0.13
RR of roll sway range [n.u.]	1.29	0.62	0.49	0.84	1.64	−0.13
RR of 3D sway range [n.u.]	1.33	0.69	0.52	0.87	1.65	−0.14
RR of pitch sway speed [n.u.]	1.46	0.62	0.43	1.03	1.77	−0.14
RR of roll sway speed [n.u.]	1.43	0.62	0.43	0.90	1.89	−0.15
RR of 3D sway speed [n.u.]	1.41	0.51	0.36	0.99	1.64	−0.13

3.2. Associations with Age, Sex, Height, and Weight

Negative and low mean R^2_{test} values (Table 3) indicated that most fitted models did not generalize well when presented with the new data and modelled confounding associations were not sustainable for this dataset. R^2_{test} values greater than 0.1 were only observed for SCSW step length and step width. For these parameters, linear models fitted using the full dataset showed an association of increased step length in taller, lighter individuals and increased step width in heavier, male individuals (Table 4). We thus suggest normalizing new datapoints for these spatiotemporal parameters using the provided models (Table 4) and (3). Resulting residuals should then be compared to residuals of the normative data using (2). Bivariate statistics regarding associations between spatiotemporal factors age, sex, height, weight, and study are provided in Supplementary Table S2.

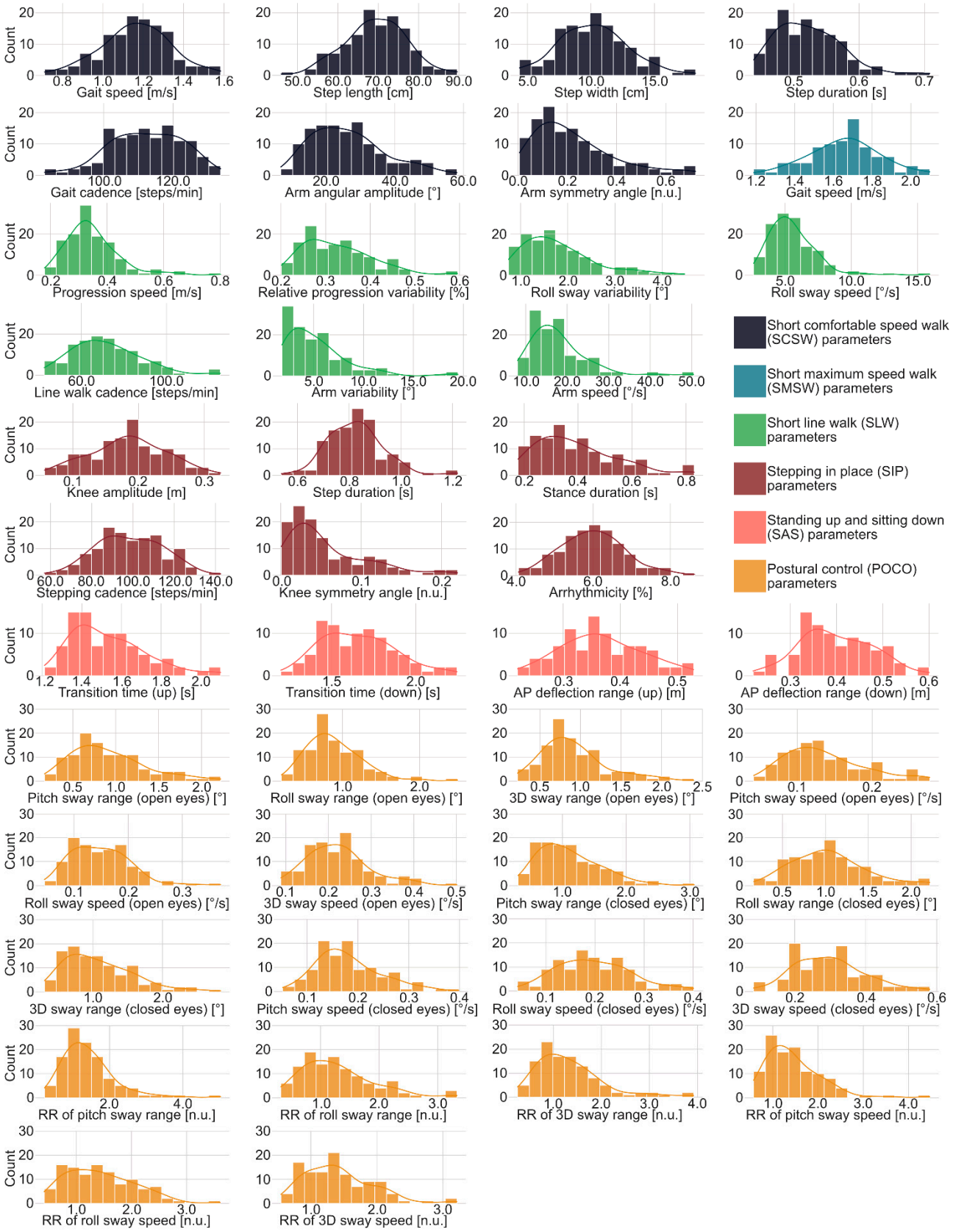


Figure 2. Distributions of raw spatiotemporal parameters color coded by motor task. Abbreviations: AP: anterior-posterior; n.u.: unitless; POCO: postural control; RR: Romberg ratio; SAS: standing up and sitting down; SCSW: short comfortable speed walk; SIP: stepping in place; SLW: short line walk; SMSW: short maximum speed walk.

Table 4. Linear model coefficients describing associations with confounders age, sex, height, and weight. Abbreviations: CI: confidence interval; SCSW: short comfortable speed walk; sϵ: standard deviation of residuals.

Spatiotemporal Parameter	β_0	β_0 *p*-Value; 95% CI	β_{Age}	β_{Age} *p*-Value; 95% CI	β_{Sex}	β_{Sex} *p*-Value; 95% CI	β_{Height}	β_{Height} *p*-Value; 95% CI	β_{Weight}	β_{Weight} *p*-Value; 95% CI	s_ϵ
SCSW step length [cm]	−2.176	0.882; [−31.187, 26.835]	−0.073	0.234; [−0.193, 0.048]	−0.913	0.567; [−4.062, 2.236]	0.510	<0.001; [0.331, 0.689]	−0.181	<0.001; [−0.281, −0.080]	6.005
SCSW step width [cm]	6.164	0.274; [−4.953, 17.281]	0.038	0.107; [−0.008, 0.084]	1.301	<0.05; [0.094, 2.507]	−0.020	0.561; [−0.089, 0.048]	0.076	<0.001; [0.038, 0.115]	2.301

3.3. Usage of Normative Values

Usage of the provided normative data was exemplified for data from a person with multiple sclerosis (Figure 3). The illustration of z-score transformed values allows for straightforward overview of individual patterns, and cross-checking whether findings from the literature apply for the individual at hand. For instance, the person in Figure 3 shows above (healthy) average POCO pitch/roll/3D sway speed with closed eyes and below average SMSW gait speed, which has been likewise found at group level in pilot studies using Motognosis Labs and the Kinect v1 (healthy participant overlap with Valkinect and VIMS: n = 9) [26,27].

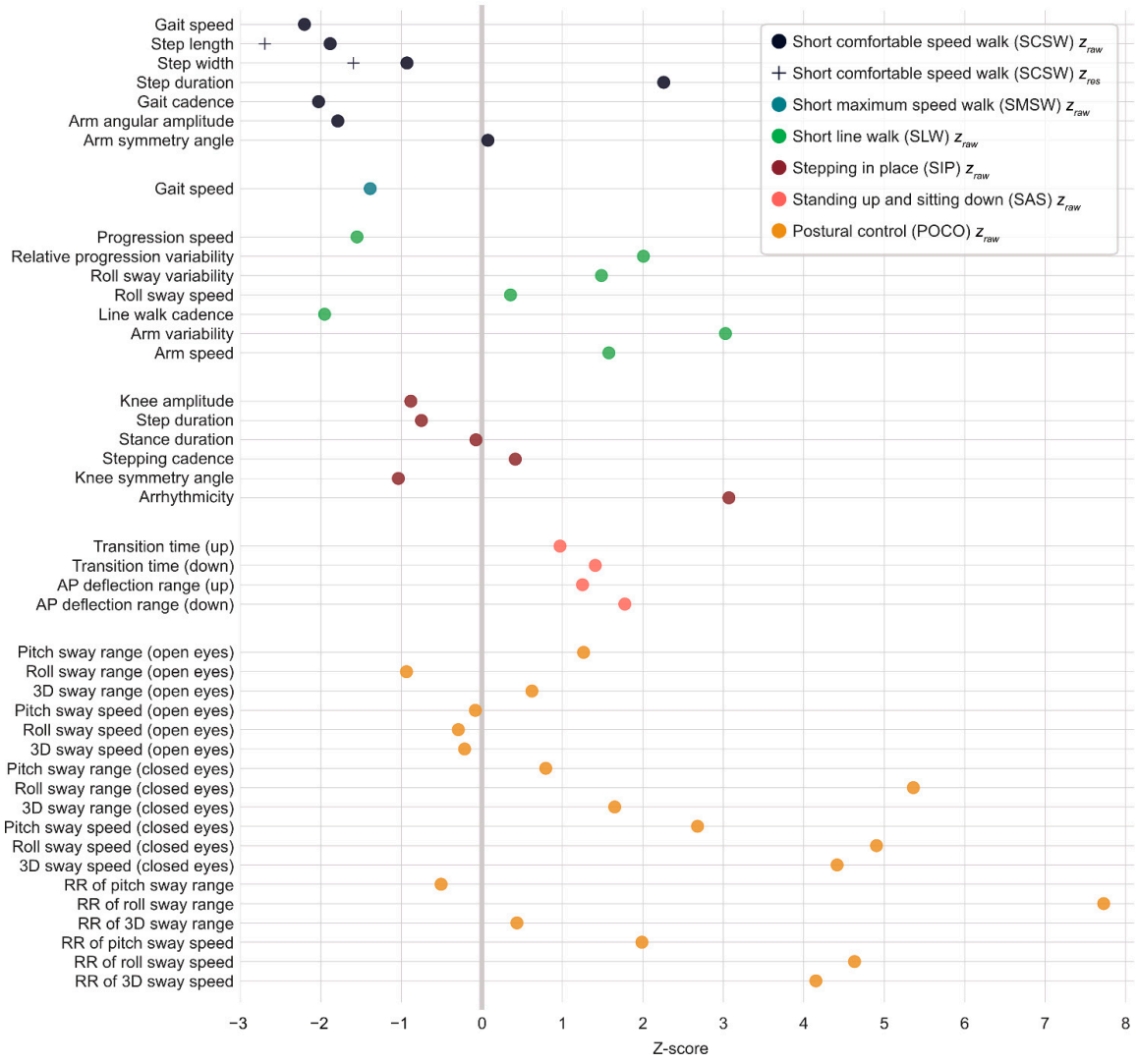


Figure 3. Z-score transformed spatiotemporal parameters from recordings of a randomly picked person with multiple sclerosis from the Valkinect study. Abbreviations: AP: anterior-posterior; RR: Romberg ratio; z_{raw}: z-score for raw values; z_{res}: z-score for residual values.

For SCSW step length and step width, we suggest using z_{res} over z_{raw} in visualizations and further analysis. In Figure 3, both are depicted to illustrate effects of regression-based normalization for identified confounding associations. The patient shows below average SCSW step length and step width when looking at raw data (54.86 cm and 7.65 cm; $z_{raw} = -1.88$ and -0.93). These values deviate further from the healthy mean when controlling for age, sex, height, and weight ($z_{res} = -2.70$ and -1.60).

4. Discussion

In this study, we provided normative data for 43 spatiotemporal parameters of six short motor tasks recorded from 20- to 60-year-old healthy adults using an inexpensive and easy-to-use RGB depth camera system. The necessity for regression-based normalization regarding demographic and anthropometric confounders was found mostly negligible for this sample except for two gait parameters. Further use of the raw and normalized normative parameters was exemplified by means of z-score visualizations.

The presented values will aid interpretation of outcomes for clinical users and researchers employing instrumental motion analysis—especially RGB-Depth camera systems. At group level, the results can serve for hypothesis generation about populations of interest, even without a sufficiently matched control group. Observed patterns may define univariate or multivariate "motor biomarkers" indicating pathology. For longitudinal studies, cross-sectional normative data is arguably less relevant, as subjects provide their own baseline data. However, it can still aid overall interpretation of intraindividual changes. For instance, changes of a similar magnitude might be clinically more meaningful if they exceed certain normative thresholds.

4.1. Short Comfortable and Maximum Speed Walk (SCSW and SMSW)

Gait speed measurement highly depends on start protocol (static or dynamic), path length, and speed instructions. Our means (SCSW: 1.16 m/s; SMSW: 1.66 m/s) are consistent with values for adults under 60 recorded with a stopwatch (4 m gait, static start; SCSW: 1.11–1.21 m/s; SMSW: 1.57–1.88 m/s) [3] and a Kinect v2 study by Latorre et al. (measurements starting at 6 m from the sensor; SCSW: 1.16–1.19 m/s) [5]. However, SMSW gait speed in Motognosis Labs studies with the Kinect v1 (healthy participant overlap with Valkinect and VIMS: n = 9) substantially exceeded our results (1.83 and 1.85 m/s) [27,28], likely because of a more dynamic starting protocol. The limited sensor range of the Kinect v2 only allows for few SMSW gait cycles to be recorded. Thus, parameters other than speed lack robustness [33] and were not reported here.

Latorre et al. further report comparable SCSW cadence (107.43–112.37 steps/min versus our 112.07 steps/min), but divergent step lengths and widths (62–67 cm and 11–12 cm versus our 69.35 cm and 10.19 cm) [5], which may result from a differing set-up and algorithmic step definition.

In line with our findings, mean arm angular amplitudes of 25.0–26.2° were measured during 4 km/h treadmill walking with an ultrasound motion capture system [7]. During 1-min walking at preferred speed in adults younger than 60, Mirelman et al. measured considerably higher arm swing amplitudes (42.0–53.4°) but lower asymmetry (corresponding to an arm symmetry angle of 0.164–0.202) [6]. A comparison with other asymmetry data from the literature was mostly inconclusive because of different metrics, e.g., in [7].

Despite clinically well-established effects, age did not emerge as a relevant confounder for gait parameters—possibly because of our comparatively young cohort. This is consistent with findings that gait speed does not change significantly under the age of 60 [3,4]. We expected to find associations with height, as respective gait parameter normalization approaches have long been proposed and comprise for example scaling as a function of leg length [34] or body height [35]. The extent to which size differences explain sex or weight differences and vice versa cannot be reliably determined using our statistical approach.

4.2. Short Line Walk (SLW)

Test performance of SLW can reveal subtle balance deficits and depends highly on instructions and individual implementation strategies. Conventionally reported SLW parameters include time-to-complete (stopwatch measure) and number of missteps (observational measure) [36], only few studies regarding instrumental motion analysis of SLW are available.

Velázquez-Pérez et al. derived trunk, lumbar and arm ranges of motion using wearables, which are kindred parameters to SLW arm or roll sway variability and speed, but not directly comparable [37]. Grinberg et al. focused on lower limb gait parameters during 3 m tandem gait at self-selected speed and found a substantially higher line walk cadence compared to our results (84.5 steps/min vs. 71.78 steps/min) [21]. Ganz et al., on the other hand, provide rather synoptic parameters derived from a single wearable that describe postural corrections, overall movement, as well as regularity and complexity. They reported composite factors of these parameters to be associated with age, sex, and BMI in older adults [8]. However, time-to-complete was reported not to be associated with age, sex, or BMI [36], which is more in line with our results that yielded no robust model describing associations between spatiotemporal parameters and respective confounders.

4.3. Stepping in Place (SIP)

The observed mean stepping cadence of 98.29 steps/min compared well to the 99 steps/min reported by Garcia et al. Interestingly, they found similar cadences for SIP and SCSW in their samples [14], while SIP stepping cadence was substantially lower than SCSW gait cadence here. Other authors implicitly report higher mean cadence (104.35–112.15 steps/min; extrapolated from reported cycle durations) and lower arrhythmicity (2.63–3.89) [13,15]. Substantially higher arrhythmicity (slightly adapted algorithm) was measured in persons with Parkinson's disease using Motognosis Labs [23], yielding the parameter as a potential gait variability substitute. In [23] further parameters, such as longest stance time, were extracted to assess festination or freezing of gait behavior. This was omitted here, as no such behavior was expected in healthy adults.

4.4. Standing up and Sitting down (SAS)

Although sit-to-stand transitions are widely used in clinical ratings or timed assessments of various disorders (e.g., as part of the Timed Up and Go Test), heterogeneous transition phase definitions, phase segmentation procedures and outcome parameters obstruct direct comparisons to our data.

For instance, Weiss et al. [17,18] reported comparatively short mean durations for Sit-to-Stand (0.5 s and 0.56 s) and Stand-to-Sit transitions (0.7 s and 0.85 s) during performance of the Timed Up and Go Test.

However, they took the extrema of accelerometer-derived anterior-posterior acceleration for phase segmentation, which systematically underestimates these phases, when considering respective formal definitions [19]. Definitions from van Lummel et al. are more consistent with our approach and yielded slightly lower values (1.45 s and 1.47 s) during five times Sit-To-Stand at self-selected speed in young adults [16].

While we found low inter-individual variability and negligible confounding for transition times in our sample, differences in transition times were previously described between age-groups 25 and younger and 70 and older [16,20]. Furthermore, possible cultural bias has been observed for this task [25].

4.5. Postural Control (POCO)

Direct comparison to data from the literature is futile due to major differences in measurement technologies (e.g., force plates, pressure plates, and accelerometers), motor tasks (e.g., reaching, single-legged, and open stance tasks) and outcome measures (e.g., path lengths and displacement of center of pressure) [9–12].

Previously published values from healthy adults using Motognosis Labs with a Microsoft Kinect v1 (healthy participant overlap with Valkinect and VIMS: n = 9) feature slightly lower sway speed values and slightly higher respective Romberg ratios. They propose using the 95th percentile of 3D sway speed (closed eyes) in healthy controls as a threshold for abnormal sway in persons with multiple sclerosis, which amounts to 0.50°/s and compares well to our data (95th percentile = 0.47°/s; not explicitly reported here). Consistent with our findings, they reported no associations with age, sex, height, or BMI [26]. In studies with a broader age range, increased center of pressure sway paths in balance tasks were, however, reported to be associated with older age and male sex [9,10].

4.6. Z-Score Transformation and Visualization

Comprehensible data visualization greatly increases interpretability and usability of outcomes for experienced and technology-naïve users alike. Z-score transformations are well established considering, e.g., neuropsychological testing [38], and relate to visualizations used in the usual lab report format, which is highly familiar to clinical users. In instrumental motion analysis, such transformations and visualizations are used less frequently. Notably, however, z-scores are visualized alongside metric value and percentile representations for the commercially available Mobility Lab v2 system (APDM Inc., Portland, OR, USA) [30].

4.7. Limitations

Our sample is biased demographically towards a German, Caucasian, and urban population, which potentially influences motor behavior [25,39,40]. However, such biases can be counteracted by comparing our results with databases from other study sites, social or cultural groups [25].

Expanding the data set may generally lead to more stable estimates of normative values and associations with confounders. For expansion, more demographically and anthropometrically extreme data should be used where appropriate, e.g., from older subjects when investigating neurodegenerative diseases such as Parkinson's disease.

In terms of analysis, we restricted our modelling of confounding effects to linear associations. Further, advantages of z-scores are limited for non-normally distributed parameters, e.g., direct conversion into percentiles is not possible. They still serve the general purpose of normalization, but, depending on use case, other transformations could be explored. Lastly, the participant used for exemplification of the z-score visualizations was chosen at random and does not necessarily show representative motor behavior.

5. Conclusions

The reported normative values fill existing gaps in the literature of motion capture for various tasks assessing motor capacity as well as generally RGB-Depth camera-based motion analysis. The results will inform clinicians and researchers on how to effectively use and interpret the outcomes of this technology.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijerph192416989/s1, Figure S1. Histograms depicting distributions of age, height, weight, and BMI subdivided by sex. A: Distributions from all studies for motor tasks postural control (POCO), stepping in place (SIP), short comfortable speed walk (SCSW), and short line walk (SLW). B: Distributions from studies Valkinect and VIMS for motor tasks short maximum speed walk (SMSW), standing up and sitting down (SAS); Table S1. Overview of discard rates and exclusion reasons for recordings from studies ASD, Valkinect, and VIMS. Abbreviations: POCO: postural control; rec/s: recording/s; RR: Romberg ratio; SAS: standing up and sitting down; SCSW: short comfortable speed walk; SIP: stepping in place; SLW: short line walk; SMSW: short maximum speed walk; Table S2. Associations of spatiotemporal parameters with factors age, height, weight, sex and study, assessed with Pearson's correlation coefficients (r), independent samples t-tests or one-way ANOVA respectively. Statistics with respective p-values smaller than 0.05 are highlighted in bold face. Abbreviations: AP: anterior-posterior; n.u.: unitless; RR: Romberg ratio.

Author Contributions: H.M.R.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing—original draft, Writing—Review & Editing, Visualization; K.O.: Conceptualization, Methodology, Writing—Review & Editing, Software; S.R.: Conceptualization, Investigation, Writing—Review & Editing; C.F.: Conceptualization, Writing—Review & Editing, Supervision; R.R.: Data Curation, Writing—Review & Editing, Project administration; E.-M.D.: Data Curation, Writing—Review & Editing; B.B.: Resources, Data Curation, Writing—Review & Editing, Project administration; F.P.: Conceptualization, Resources, Writing—Review & Editing, Supervision, Project administration; T.S.-H.: Conceptualization, Methodology, Resources, Writing—original draft, Writing—Review & Editing, Supervision, Project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: All studies were approved by the local ethics committee (ASD: EA1/392/16; Valkinect: EA1/339/16, amendment 1; VIMS: EA1/163/12) and conducted in accordance with the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated and analyzed as part of this study are available upon reasonable request from the corresponding author for the purpose of replication of results. Participant consent at the time of enrolment did not comprise sharing nor anonymization of individual data.

Acknowledgments: We thank NeuroCure Clinical Research Center (NCRC) for administrative support. The NCRC is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC-2049—390688087 and Charité-BIH Clinical Study Center. R.R. has received a speaker honorarium from Roche independent from this study.

Conflicts of Interest: K.O. is a stakeholder and H.M.R. a paid employee at Motognosis GmbH. All other authors state no conflict of interest.

References

1. Christopher, A.; Kraft, E.; Olenick, H.; Kiesling, R.; Doty, A. The reliability and validity of the Timed Up and Go as a clinical tool in individuals with and without disabilities across a lifespan: A systematic review. *Disabil. Rehabil.* **2021**, *43*, 1799–1813. [CrossRef] [PubMed]
2. Motl, R.W.; Cohen, J.A.; Benedict, R.; Phillips, G.; LaRocca, N.; Hudson, L.D.; Rudick, R. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. *Mult. Scler.* **2017**, *23*, 704–710. [CrossRef] [PubMed]
3. Bohannon, R.W.; Wang, Y.C. Four-Meter Gait Speed: Normative Values and Reliability Determined for Adults Participating in the NIH Toolbox Study. *Arch. Phys. Med. Rehabil.* **2019**, *100*, 509–513. [CrossRef] [PubMed]
4. Bohannon, R.W.; Andrews, A.W. Normal walking speed: A descriptive meta-analysis. *Physiotherapy* **2011**, *97*, 182–189. [CrossRef]
5. Latorre, J.; Colomer, C.; Alcañiz, M.; Llorens, R. Gait analysis with the Kinect v2: Normative study with healthy individuals and comprehensive study of its sensitivity, validity, and reliability in individuals with stroke. *J. Neuroeng. Rehabil.* **2019**, *16*, 97. [CrossRef]
6. Mirelman, A.; Bernad-Elazari, H.; Nobel, T.; Thaler, A.; Peruzzi, A.; Plotnik, M.; Giladi, N.; Hausdorff, J.M. Effects of Aging on Arm Swing during Gait: The Role of Gait Speed and Dual Tasking. *PLoS ONE* **2015**, *10*, e0136043. [CrossRef]
7. Plate, A.; Sedunko, D.; Pelykh, O.; Schlick, C.; Ilmberger, J.R.; Bötzel, K. Normative data for arm swing asymmetry: How (a)symmetrical are we? *Gait Posture* **2015**, *41*, 13–18. [CrossRef]
8. Ganz, N.; Gazit, E.; Giladi, N.; Dawe, R.J.; Mirelman, A.; Buchman, A.S.; Hausdorff, J.M. Automatic Quantification of Tandem Walking Using a Wearable Device: New Insights Into Dynamic Balance and Mobility in Older Adults. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **2021**, *76*, 101–107. [CrossRef]
9. Goble, D.J.; Baweja, H.S. Normative Data for the BTrackS Balance Test of Postural Sway: Results from 16,357 Community-Dwelling Individuals Who Were 5 to 100 Years Old. *Phys. Ther.* **2018**, *98*, 779–785. [CrossRef]
10. Patti, A.; Bianco, A.; Şahin, N.; Sekulic, D.; Paoli, A.; Iovane, A.; Messina, G.; Gagey, P.M.; Palma, A. Postural control and balance in a cohort of healthy people living in Europe: An observational study. *Medicine* **2018**, *97*, e13835. [CrossRef]
11. Clark, R.A.; Pua, Y.H.; Oliviera, C.C.; Bower, K.J.; Thilarajah, S.; McGaw, R.; Hasanki, K.; Mentiplay, B. Reliability and concurrent validity of the Microsoft Xbox One Kinect for assessment of standing balance and postural control. *Gait Posture* **2015**, *42*, 210–213. [CrossRef]
12. Mancini, M.; Salarian, A.; Carlson-Kuhta, P.; Zampieri, C.; King, L.; Chiari, L.; Horak, F.B. ISway: A sensitive, valid and reliable measure of postural control. *J. Neuroeng. Rehabil.* **2012**, *9*, 59. [CrossRef] [PubMed]
13. Dalton, C.; Sciadas, R.; Nantel, J. Executive function is necessary for the regulation of the stepping activity when stepping in place in older adults. *Aging Clin. Exp. Res.* **2016**, *28*, 909–915. [CrossRef]

14. Garcia, R.K.; Nelson, A.J.; Ling, W.; Van Olden, C. Comparing stepping-in-place and gait ability in adults with and without hemiplegia. *Arch. Phys. Med. Rehabil.* **2001**, *82*, 36–42. [CrossRef] [PubMed]
15. Nantel, J.; de Solages, C.; Bronte-Stewart, H. Repetitive stepping in place identifies and measures freezing episodes in subjects. *Gait Posture* **2011**, *34*, 329–333. [CrossRef] [PubMed]
16. Van Lummel, R.C.; Ainsworth, E.; Lindemann, U.; Zijlstra, W.; Chiari, L.; Van Campen, P.; Hausdorff, J.M. Automated approach for quantifying the repeated sit-to-stand using one body fixed sensor in young and older adults. *Gait Posture* **2013**, *38*, 153–156. [CrossRef] [PubMed]
17. Weiss, A.; Herman, T.; Plotnik, M.; Brozgol, M.; Giladi, N.; Hausdorff, J.M. An instrumented timed up and go: The added value of an accelerometer for identifying fall risk in idiopathic fallers. *Physiol. Meas.* **2011**, *32*, 2003–2018. [CrossRef]
18. Weiss, A.; Herman, T.; Plotnik, M.; Brozgol, M.; Maidan, I.; Giladi, N.; Gurevich, T.; Hausdorff, J.M. Can an accelerometer enhance the utility of the Timed Up & Go Test when evaluating patients with Parkinson's disease? *Med. Eng. Phys.* **2010**, *32*, 119–125.
19. Kralj, A.; Jaeger, R.J.; Munih, M. Analysis of standing up and sitting down in humans: Definitions and normative data presentation. *J. Biomech.* **1990**, *23*, 1123–1138. [CrossRef]
20. Mourey, F.; Pozzo, T.; Rouhier-Marcer, I.; Didier, J.P. A kinematic comparison between elderly and young subjects standing up from and sitting down in a chair. *Age Ageing* **1998**, *27*, 137–146. [CrossRef]
21. Grinberg, Y.; Berkowitz, S.; Hershkovitz, L.; Malcay, O.; Kalron, A. The ability of the instrumented tandem walking tests to discriminate fully ambulatory people with MS from healthy adults. *Gait Posture* **2019**, *70*, 90–94. [CrossRef] [PubMed]
22. Otte, K.; Kayser, B.; Mansow-Model, S.; Verrel, J.; Paul, F.; Brandt, A.U.; Schmitz-Hübsch, T. Accuracy and Reliability of the Kinect Version 2 for Clinical Measurement of Motor Function. *PLoS ONE* **2016**, *11*, e0166532. [CrossRef]
23. Otte, K.; Ellermeyer, T.; Vater, T.S.; Voigt, M.; Kroneberg, D.; Rasche, L.; Krüger, T.; Röhling, H.M.; Kayser, B.; Mansow-Model, S.; et al. Instrumental Assessment of Stepping in Place Captures Clinically Relevant Motor Symptoms of Parkinson's Disease. *Sensors* **2020**, *20*, 5465. [CrossRef] [PubMed]
24. Morrison, C.; D'Souza, M.; Huckvale, K.; Dorn, J.F.; Burggraaff, J.; Kamm, C.P.; Steinheimer, S.M.; Kontschieder, P.; Criminisi, A.; Uitdehaag, B.; et al. Usability and Acceptability of ASSESS MS: Assessment of Motor Dysfunction in Multiple Sclerosis Using Depth-Sensing Computer Vision. *JMIR Hum. Factors* **2015**, *2*, e11. [CrossRef]
25. Otte, K.; Ellermeyer, T.; Suzuki, M.; Röhling, H.M.; Kuroiwa, R.; Cooper, G.; Mansow-Model, S.; Mori, M.; Zimmermann, H.; Brandt, A.U.; et al. Cultural bias in motor function patterns: Potential relevance for predictive, preventive, and personalized medicine. *EPMA J.* **2021**, *12*, 91–101. [CrossRef] [PubMed]
26. Behrens, J.R.; Mertens, S.; Krüger, T.; Grobelny, A.; Otte, K.; Mansow-Model, S.; Gusho, E.; Paul, F.; Brandt, A.U.; Schmitz-Hübsch, T. Validity of visual perceptive computing for static posturography in patients with multiple sclerosis. *Mult. Scler.* **2016**, *22*, 1596–1606. [CrossRef] [PubMed]
27. Behrens, J.; Pfüller, C.; Mansow-Model, S.; Otte, K.; Paul, F.; Brandt, A.U. Using perceptive computing in multiple sclerosis—The Short Maximum Speed Walk test. *J. Neuroeng. Rehabil.* **2014**, *11*, 89. [CrossRef] [PubMed]
28. Grobelny, A.; Behrens, J.R.; Mertens, S.; Otte, K.; Mansow-Model, S.; Krüger, T.; Gusho, E.; Bellmann-Strobl, J.; Paul, F.; Brandt, A.U.; et al. Maximum walking speed in multiple sclerosis assessed with visual perceptive computing. *PLoS ONE* **2017**, *12*, e0189281. [CrossRef] [PubMed]
29. Cho, A.B.; Otte, K.; Baskow, I.; Ehlen, F.; Maslahati, T.; Mansow-Model, S.; Schmitz-Hübsch, T.; Behnia, B.; Roepke, S. Motor signature of autism spectrum disorder in adults without intellectual impairment. *Sci. Rep.* **2022**, *12*, 7670. [CrossRef]
30. APDM Mobility Lab User Guide, V2R Version 3. Available online: https://share.apdm.com/documentation/MobilityLabUserGuide.pdf (accessed on 10 September 2022).
31. Holmstrom, L. Normative Data Used in Mobility Lab v2. Available online: https://support.apdm.com/hc/en-us/articles/214504686-Normative-Data-Used-in-Mobility-Lab-v2 (accessed on 10 September 2022).
32. Alves, S.A.; Ehrig, R.M.; Raffalt, P.C.; Bender, A.; Duda, G.N.; Agres, A.N. Quantifying Asymmetry in Gait: The Weighted Universal Symmetry Index to Evaluate 3D Ground Reaction Forces. *Front. Bioeng. Biotechnol.* **2020**, *8*, 579511. [CrossRef]
33. Röhling, H.M.; Althoff, P.; Arsenova, R.; Drebinger, D.; Gigengack, N.; Chorschew, A.; Kroneberg, D.; Rönnefarth, M.; Ellermeyer, T.; Rosenkranz, S.C.; et al. Proposal for Post Hoc Quality Control in Instrumented Motion Analysis Using Markerless Motion Capture: Development and Usability Study. *JMIR Hum. Factors* **2022**, *9*, e26825. [CrossRef] [PubMed]
34. Zijlstra, W.; Prokop, T.; Berger, W. Adaptability of leg movements during normal treadmill walking and split-belt walking in children. *Gait Posture* **1996**, *4*, 212–221. [CrossRef]
35. Schwesig, R.; Leuchte, S.; Fischer, D.; Ullmann, R.; Kluttig, A. Inertial sensor based reference gait data for healthy subjects. *Gait Posture* **2011**, *33*, 673–678. [CrossRef] [PubMed]
36. Schneiders, A.G.; Sullivan, S.J.; Gray, A.R.; Hammond-Tooke, G.D.; McCrory, P.R. Normative values for three clinical measures of motor performance used in the neurological assessment of sports concussion. *J. Sci. Med. Sport* **2010**, *13*, 196–201. [CrossRef]
37. Velázquez-Pérez, L.; Rodriguez-Labrada, R.; González-Garcés, Y.; Arrufat-Pie, E.; Torres-Vega, R.; Medrano-Montero, J.; Ramirez-Bautista, B.; Vazquez-Mojena, Y.; Auburger, G.; Horak, F.; et al. Prodromal Spinocerebellar Ataxia Type 2 Subjects Have Quantifiable Gait and Postural Sway Deficits. *Mov. Disord.* **2021**, *36*, 471–480. [CrossRef]
38. St-Hilaire, A.; Parent, C.; Potvin, O.; Bherer, L.; Gagnon, J.F.; Joubert, S.; Belleville, S.; Wilson, M.A.; Koski, L.; Rouleau, I.; et al. Trail Making Tests A and B: Regression-based normative data for Quebec French-speaking mid and older aged adults. *Clin. Neuropsychol.* **2018**, *32* (Suppl. 1), 77–90. [CrossRef]

39. Plouvier, S.; Carton, M.; Cyr, D.; Sabia, S.; Leclerc, A.; Zins, M.; Descatha, A. Socioeconomic disparities in gait speed and associated characteristics in early old age. *BMC Musculoskelet. Disord.* **2016**, *17*, 178. [CrossRef]
40. Al-Obaidi, S.; Wall, J.C.; Al-Yaqoub, A.; Al-Ghanim, M. Basic gait parameters: A comparison of reference data for normal subjects 20 to 29 years of age from Kuwait and Scandinavia. *J. Rehabil. Res. Dev.* **2003**, *40*, 361–366. [CrossRef]

Curriculum vitae

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Publication list

Steinert A, Sattler I, Otte K, Röhling H, Mansow-Model S, Müller-Werdan U. Using New Camera-Based Technologies for Gait Analysis in Older Adults in Comparison to the Established GAITRite System. *Sensors (Basel)*, 20(1):125, January 2020.
Impact Factor (2020): 3.576

Otte K, Ellermeyer T, Vater TS, Voigt M, Kroneberg D, Rasche L, Krüger T, Röhling HM, Kayser B, Mansow-Model S, Klostermann F, Brandt AU, Paul F, Lipp A, and Schmitz-Hübsch T. Instrumental assessment of stepping in place captures clinically relevant motor symptoms of Parkinson's disease. *Sensors (Basel)*, 20(19):5465, September 2020.
Impact Factor (2020): 3.576

Otte K, Ellermeyer T, Suzuki M, Röhling HM, Kuroiwa R, Cooper G, Mansow-Model S, Mori M, Zimmermann H, Brandt AU, Paul F, Hirano S, Kuwabara S, and Schmitz-Hübsch T. Cultural bias in motor function patterns: Potential relevance for predictive, preventive, and personalized medicine. *EPMA Journal*, 12(1):91–101, March 2021.
Impact Factor (2021): 8.836

van Kersbergen J, Otte K, de Vries NM, Bloem BR, Röhling HM, Mansow-Model S, van der Kolk NM, Overeem S, Zinger S, van Gilst MM. Camera-based objective measures of Parkinson's disease gait features. *BMC Research Notes*, 14(1):329, August 2021.
Impact factor (2021): /

Röhling HM, Althoff P, Arsenova R, Drebinger D, Gigengack N, Chorschew A, Kroneberg D, Rönnefarth M, Ellermeyer T, Rosenkranz SC, Heesen C, Behnia B, Hirano S, Kuwabara S, Paul F, Brandt AU, and Schmitz-Hübsch T. Proposal for post hoc quality control in instrumented motion analysis using markerless motion capture: Development and usability study. *JMIR Human Factors*, 9(2):e26825, April 2022.
Impact factor (2022): /

Röhling HM, Otte K, Rekers S, Finke C, Rust R, Dorsch EM, Behnia B, Paul F, and Schmitz-Hübsch T. RGB-depth camera-based assessment of motor capacity: Normative data for six standardized motor tasks. *International Journal of Environmental Research and Public Health*, 19(24):16989, December 2022.
Impact factor (2022): 4.614

Bertram J, Krüger T, Röhling HM, Jelusic A, Mansow-Model S, Schniepp R, Wuehr M, and Otte K. Accuracy and repeatability of the Microsoft Azure Kinect for clinical

measurement of motor function. *PLoS ONE*, 18(1):e0279697, January 2023.
Impact factor (2022): 3.7

Suzuki M, Hirano S, Otte K, Schmitz-Hübsch T, Izumi M, Tamura M, Kuroiwa R, Sugiyama A, Mori M, Röhling HM, Brandt AU, Murata A, Paul F, Kuwabara S. Digital motor biomarkers of cerebellar ataxia using an RGB-depth camera-based motion analysis system. *Cerebellum*, epub ahead of print, September 2023.
Impact factor (2022): 3.5

Dorsch EM, Röhling HM, Zocholl D, Hafermann L, Paul F, Schmitz-Hübsch T. Progression events defined by home-based assessment of motor function in multiple sclerosis: protocol of a prospective study. *Frontiers in Neurology*, 14:1258635, October 2023.
Impact factor (2022): 3.4

Acknowledgments

I have been very fortunate to be surrounded by inspiring and supportive people throughout my academic journey, and I would like to take this opportunity to thank them.

First, I would like to thank PD Dr. Tanja Schmitz-Hübsch for her continuous encouragement and scientific guidance. I genuinely admire her dedication, diligence, and thoughtfulness. I thank Prof. Dr. Friedemann Paul for his insightful feedback and for welcoming me into the kind research environment at the NCRC. On this note, I sincerely thank the incredible team at the NCRC for their support regarding the Valkinect study and for a wonderful time that will be remembered. I would also like to express my gratitude to my co-authors for their valuable contributions to our publications.

I am incredibly grateful to Dr. Karen Otte and Sebastian Mansow-Model for introducing me to the field of instrumental motion analysis and encouraging me to pursue a doctorate. They and my other Motognosis colleagues have wholeheartedly fostered my professional and personal development during our time together.

Many thanks go to BCCN year 2014 for a formative time that fueled my research interest. I especially want to thank Laura, a wonderful friend and source of academic inspiration since Lübeck times. I thank Jacki and Lara for their open ears, motivating words, and much-needed diversions.

I thank my parents and my sister for their invaluable consideration and reinforcement regarding my endeavors inside and outside of academia throughout the years. Finally, I thank Martin, who has kept me emotionally afloat with his endless patience and loving support.