

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 Drebinge^{1,2,3}; Norman Gigengack^{1,2,3}; Anna Chorschew^{1,2,3}; Daniel Kroneberg⁶, MD; Maria Rönnefarth^{6,7}, MD; Tobias Ellermeyer^{1,2,3,8}; Sina Cathérine Rosenkranz^{9,10}, MD; Christoph Heesen^{9,10}, MD; Behnoush Behnia¹¹, MD; Shigeki Hirano¹², MD, PhD; Satoshi Kuwabara¹², MD, PhD; Friedemann Paul^{1,2,3,6,13}, MD; Alexander Ulrich Brandt^{13,14}, MD; Tanja Schmitz-Hübsch^{1,2,3,13}, MD

¹Experimental 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

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

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

⁴Motognosis GmbH, Berlin, Germany

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

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

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

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

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

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

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

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

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

¹⁴Department 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](https://doi.org/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 Oprimis (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

^aHC: healthy controls.^bPwMS: people with multiple sclerosis.^c—: not available.^dEDSS: Expanded Disability Status Scale.^eN/A: not applicable.^fPASS-MS: Perceptive Assessment in Multiple Sclerosis.^gPOCO: Postural Control.^hPOCO-DUAL: Postural Control with Dual Task.ⁱSCSW: Short Comfortable Speed Walk.^jSMSW: Short Maximum Speed Walk.^kSLW: Short Line Walk.^lSIP: Stepping in Place.^mSAS: 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).

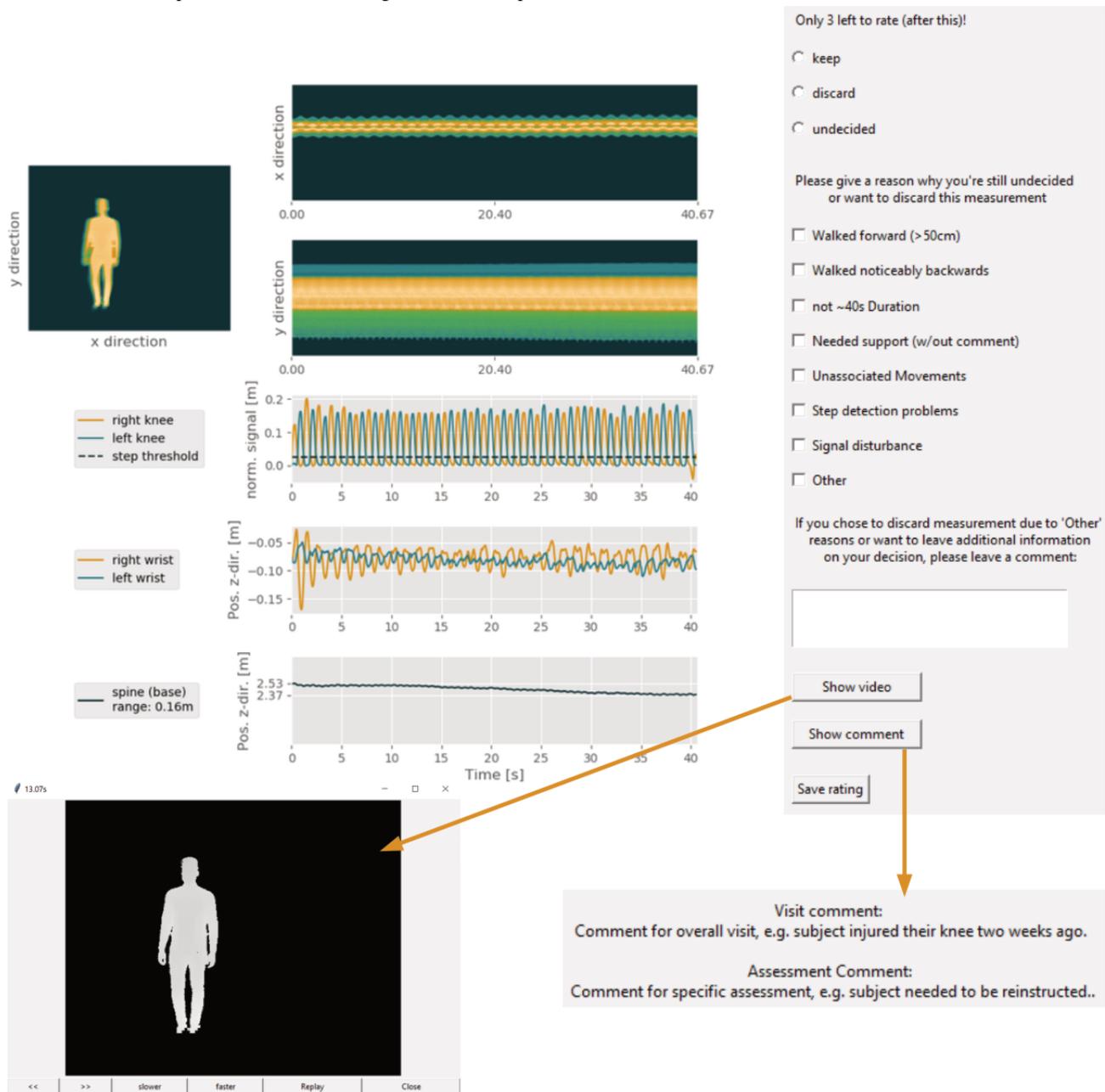
Results

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.

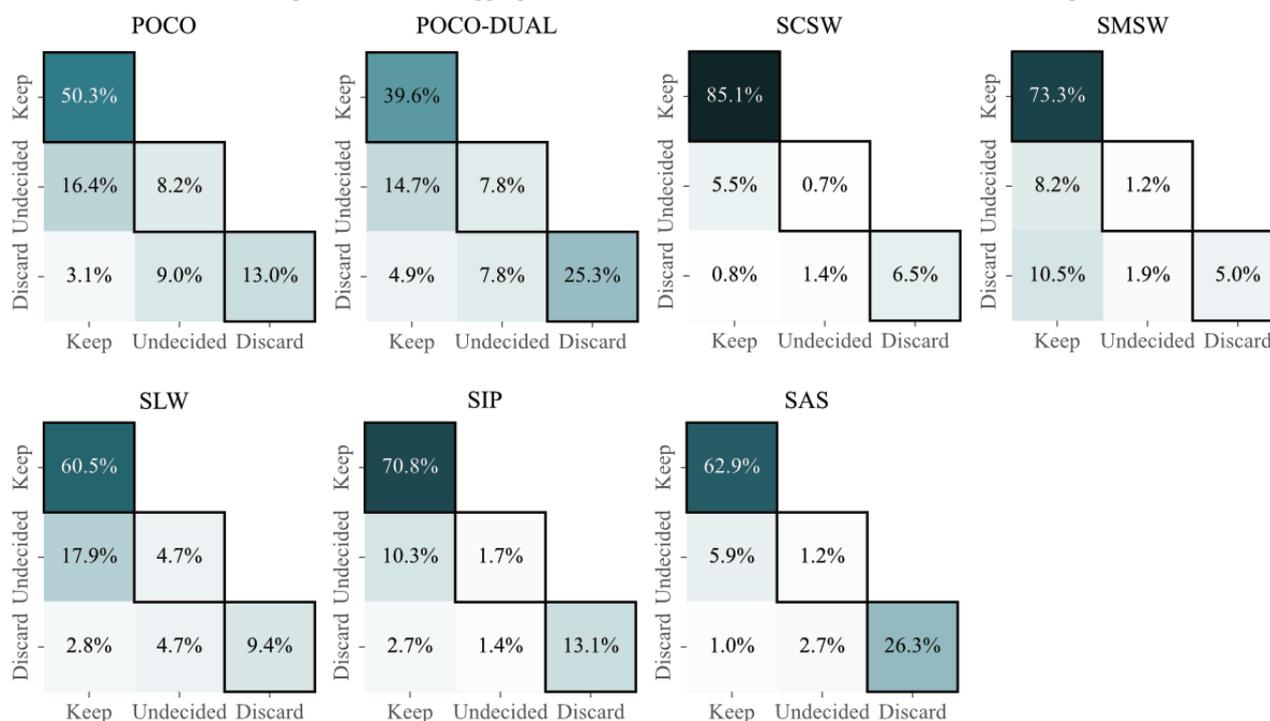


Rater Concordance and Usability of Recordings

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

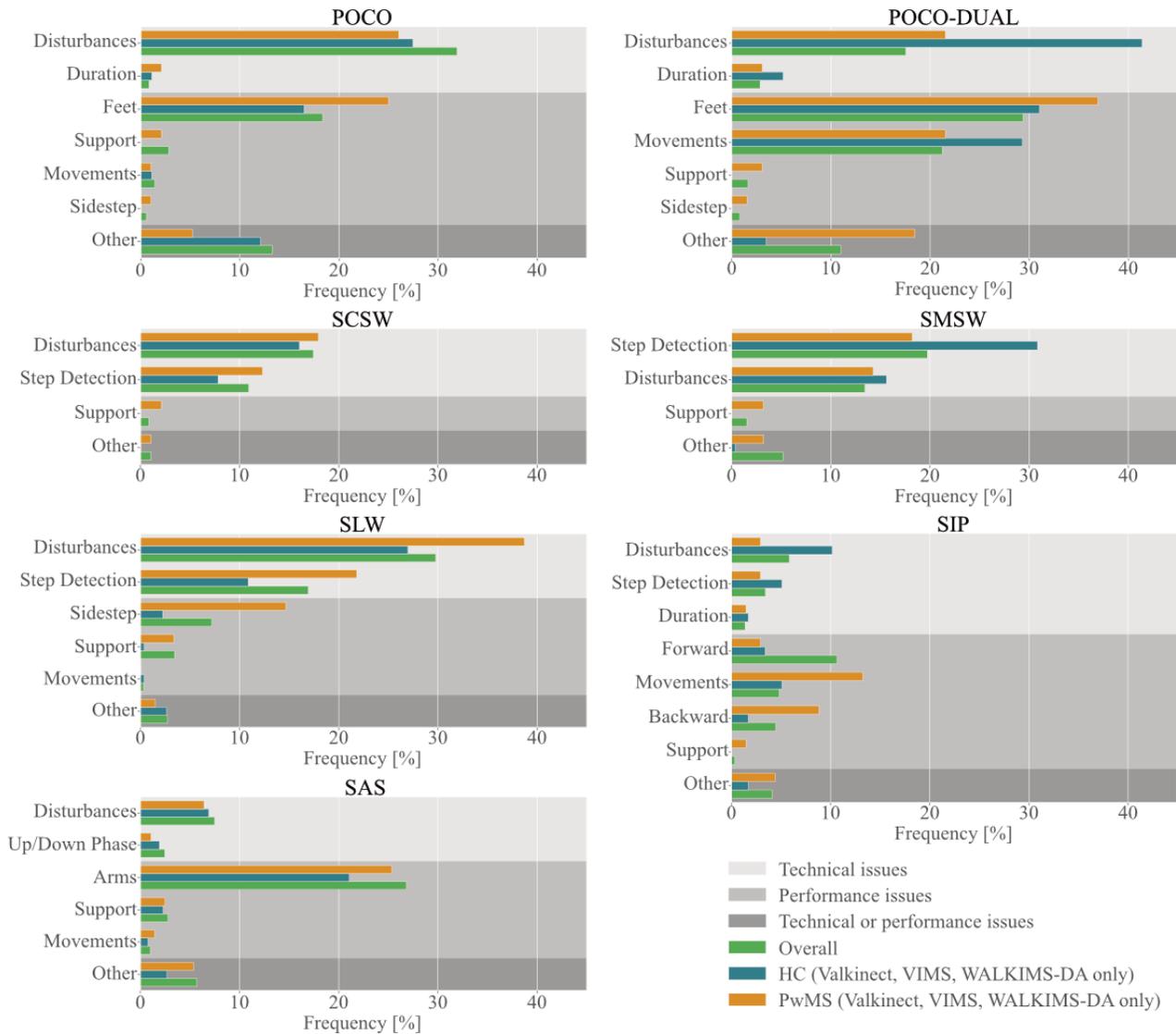
The main rating criteria compiled during QC pipeline development are listed below, with the respective tasks indicated in parentheses.

- 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.

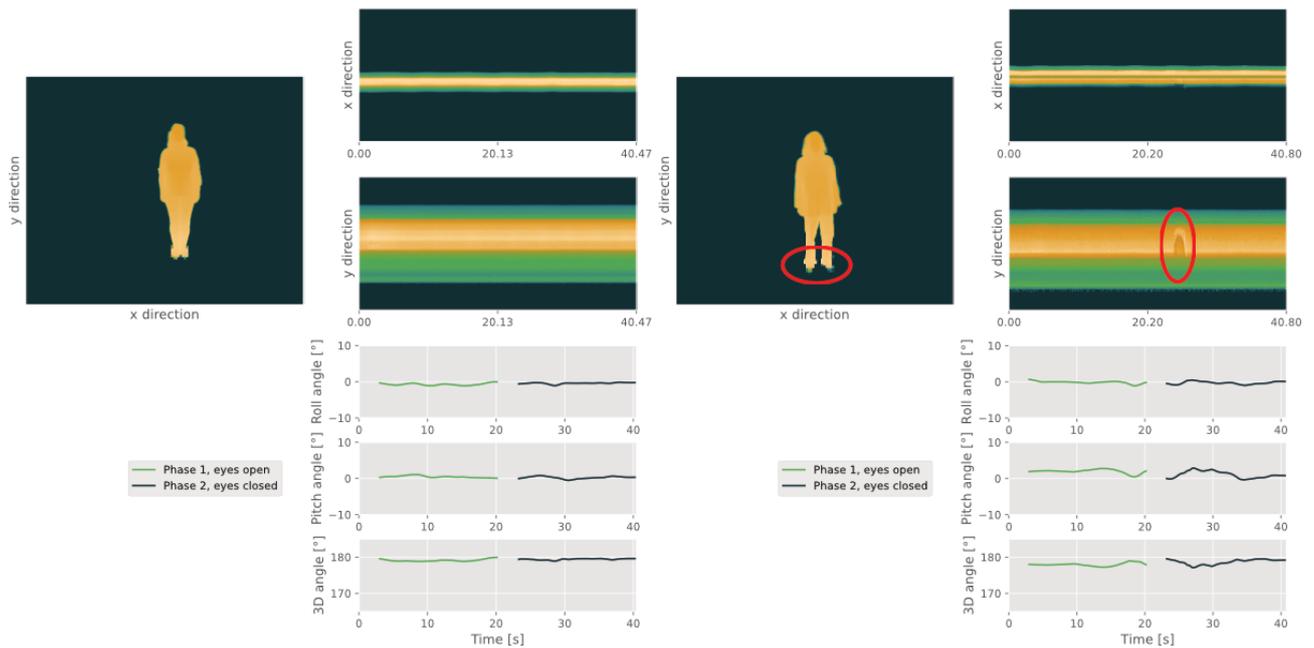
Figure 3. Selection frequencies of technical and performance-related rating criteria for all subjects as well as split by group for the 3 studies featuring healthy controls and people with multiple sclerosis. HC: healthy controls; POCO: Postural Control; POCO-DUAL: Postural Control with Dual Task; PwMS: 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.

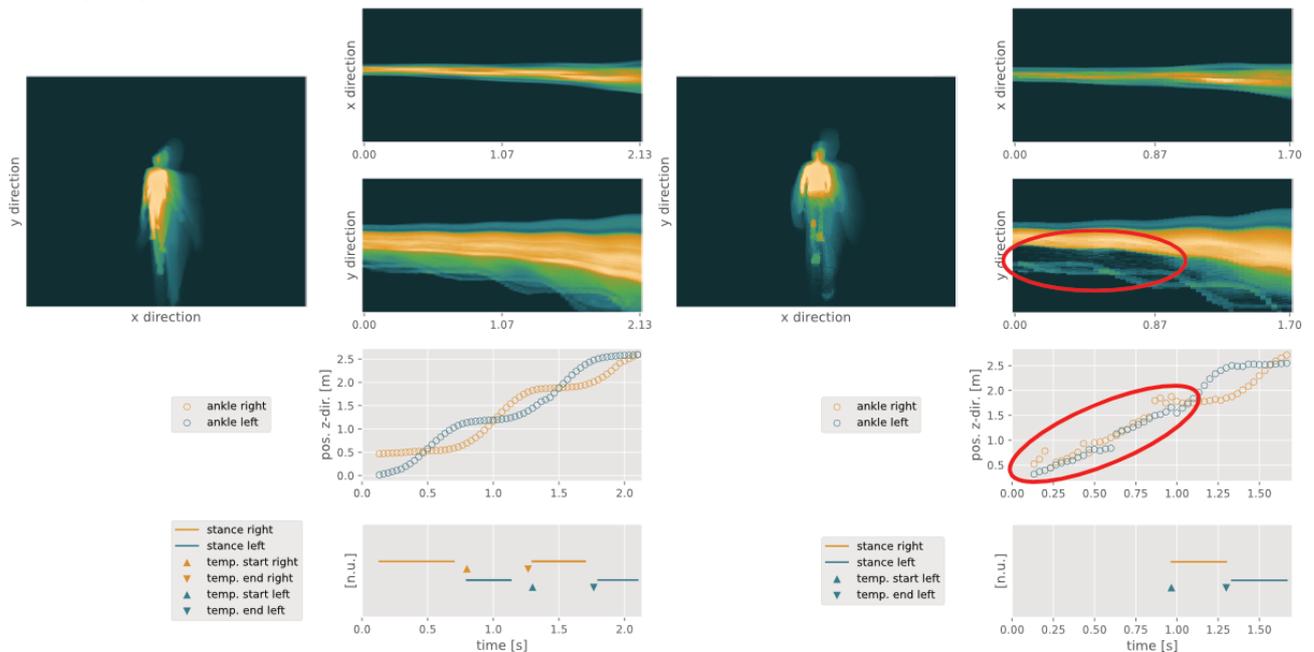
Figure 4. Left: quality control pipeline visualization screenshot of a high-quality Postural Control recording. Right: quality control pipeline visualization screenshot of a Postural Control recording featuring 2 frequently observed performance-related quality concerns, incorrect Feet positioning (according 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. rep. 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](https://doi.org/10.23736/S1973-9087.18.05306-6)] [Medline: [30311493](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.3390/s140203362)] [Medline: [24556672](https://pubmed.ncbi.nlm.nih.gov/24556672/)]
3. Pradhan C, Wuehr M, Akrami F, Neuhaeuser 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](https://doi.org/10.1016/j.jelekin.2015.01.004)] [Medline: [25725811](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1007/s11910-019-0997-9)] [Medline: [31485896](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1177/1352458520946005)] [Medline: [32749928](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1002/mds.28377)] [Medline: [33191498](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1371/journal.pone.0166532)] [Medline: [27861541](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1016/j.gaitpost.2004.10.005)] [Medline: [16274913](https://pubmed.ncbi.nlm.nih.gov/16274913/)]
9. Wren TAL, Tucker CA, Rethlefsen SA, Gorton GEIII, Öunpuu 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](https://doi.org/10.1016/j.gaitpost.2020.05.031)] [Medline: [32563727](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.2196/humanfactors.4129)] [Medline: [27025782](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1002/mds.27671)] [Medline: [30901495](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1177/1352458519857075)] [Medline: [31502913](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1002/mds.26673)] [Medline: [27273651](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/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](https://doi.org/10.3390/s20205920)] [Medline: [33092143](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.3390/s20010125)] [Medline: [31878177](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1186/1743-0003-9-59)] [Medline: [22913719](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1007/BF03327437)] [Medline: [16702791](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1016/j.gaitpost.2017.10.011)] [Medline: [29032000](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1016/j.gaitpost.2015.09.017)] [Medline: [26439183](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1186/s12874-019-0737-5)] [Medline: [31092212](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1109/JBHI.2019.2937574)] [Medline: [31449035](https://pubmed.ncbi.nlm.nih.gov/31449035/)]

23. Rampp A, Barth J, Schüle 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](https://doi.org/10.1109/TBME.2014.2368211)] [Medline: [25389237](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1371/journal.pone.0189281)] [Medline: [29244874](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.3389/fnagi.2018.00435)] [Medline: [30719002](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1186/1743-0003-11-89)] [Medline: [24886525](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1177/1352458515625807)] [Medline: [26814201](https://pubmed.ncbi.nlm.nih.gov/26814201/)]
28. Drebing 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](https://doi.org/10.3389/fneur.2020.00208)] [Medline: [32351439](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1038/s41598-019-53050-3)] [Medline: [31727918](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1007/s13167-021-00236-3)] [Medline: [33782636](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1212/WNL.33.11.1444)] [Medline: [6685237](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/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](https://doi.org/10.1016/j.parkreldis.2016.02.007)] [Medline: [26952699](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1016/j.gaitpost.2015.02.019)] [Medline: [25800649](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1016/j.gaitpost.2006.06.003)] [Medline: [16875823](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1016/j.gaitpost.2016.05.014)] [Medline: [27289221](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1016/j.gaitpost.2014.02.007)] [Medline: [24630463](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1371/journal.pone.0034823)] [Medline: [22536333](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1212/WNL.0000000000002774)] [Medline: [27225223](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.3174/ajnr.A4539)] [Medline: [26564433](https://pubmed.ncbi.nlm.nih.gov/26564433/)]
41. Motamedi S, Gawlik K, Ayadi N, Zimmermann HG, Asseger 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](https://doi.org/10.3389/fneur.2019.01117)] [Medline: [31824393](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.3174/ajnr.A6394)] [Medline: [31974079](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.2196/26825)

PMID:

©Hanna Marie Röhling, Patrik Althoff, Radina Arsenova, Daniel Drebinger, Norman Gigengack, Anna Chorschew, Daniel Kroneberg, Maria Rönnefarth, Tobias Ellermeyer, Sina Cathérine Rosenkranz, Christoph Heesen, Behnoush Behnia, Shigeki Hirano, Satoshi Kuwabara, Friedemann Paul, Alexander Ulrich Brandt, Tanja Schmitz-Hübsch. Originally published in JMIR Human Factors (<https://humanfactors.jmir.org>), 01.04.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on <https://humanfactors.jmir.org>, as well as this copyright and license information must be included.