bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A data-driven framework for the selection and validation of digital health metrics: use-case in neurological sensorimotor impairments

Christoph M. Kanzler, MSc[1], Mike D. Rinderknecht, PhD[1], Anne Schwarz, MSc[2,3], Ilse Lamers, PhD[4,5], Cynthia Gagnon, PhD[6], Jeremia Held, MSc[2,3], Peter Feys, PhD[4], Andreas R. Luft, MD[2,3], Roger Gassert, PhD[1], and Olivier Lambercy, PhD[1]

1 Rehabilitation Engineering Laboratory, Institute of Robotics and Intelligent Systems, Department of Health Sciences and Technology, ETH Zürich, Switzerland.
2 Division of Vascular Neurology and Rehabilitation, Department of Neurology, University Hospital and University of Zürich, Switzerland.
3 cereneo Center for Neurology and Rehabilitation, Vitznau, Switzerland.
4 REVAL, Rehabilitation Research Center, BIOMED, Biomedical Research Institute, Faculty of Medicine and Life Sciences, Hasselt University, Belgium.
5 Rehabilitation and MS center, Pelt, Belgium.
6 School of Rehabilitation, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Québec, Canada.

Corresponding author: Christoph M. Kanzler, Rehabilitation Engineering Laboratory, ETH Zürich, BAA C 307.1, Lengghalde 5, 8008 Zürich, Switzerland. christoph.kanzler@hest.ethz.ch. +41 44 510 72 34.
Keywords: feature selection, clinical endpoints, outcome measures, digital biomarkers, assessment, upper limb

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Background.
Digital health metrics have the potential to advance the monitoring and understanding of impaired body functions, for example in persons with neurological disorders. However, their integration into clinical research and practice is challenged by insufficient validation of the vast amount of existing and often abstract metrics. Here, we propose a data-driven framework to select and validate a clinically-relevant core set of digital health metrics extracted from a technology-aided assessment. As a use-case, this framework is applied to metrics extracted from the Virtual Peg Insertion Test (VPIT), a sensor-based assessment of upper limb sensorimotor impairments.

Methods.
The framework builds on a use-case specific pathophysiological motivation of digital health metrics to represent clinically-relevant impairments, models the influence of confounds from participant demographics, and evaluates the most important clinimetric properties (discriminant validity, structural validity, test-retest reliability, measurement error, learning effects). This approach was applied to 77 kinematic and kinetic metrics extracted from the VPIT, using data from 120 neurologically intact controls and 89 subjects with neurological disorders (post-stroke, multiple sclerosis, or hereditary ataxia). An exploratory factor analysis to discuss the initially proposed pathophysiological hypotheses was performed and the sensitivity of the metrics to clinically-defined disability levels was investigated.

Results.
Applied to the VPIT, the framework selected 10 (13.0%) clinically-relevant core metrics. These assess the severity of multiple sensorimotor impairments in a valid, reliable, and informative manner for all three disorders while being least susceptible to measurement error and learning effects. The digital health metrics of the VPIT provided additional clinical value by detecting impairments in neurological subjects that did not show any deficits according to conventional scales, and by covering several sensorimotor impairments of the arm and hand with a single assessment.

Conclusions.
The proposed framework could help to address the insufficient evaluation, standardization, and interpretability of digital health metrics. In the presented use-case, it allowed to establish validated core metrics for the VPIT, paving the way for its integration into clinical neurorehabilitation trials.

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Introduction

Assessments of impaired body functions, as observed in many diseases and disorders, are a fundamental part of the modern healthcare system [1]. Specifically, these assessments are essential to provide documentation for insurances, to individualize therapeutic interventions, and to shed light on the often unknown mechanisms underlying the impairments and their temporal evolution. An exemplary application scenario of assessments are neurological disorders, including stroke, multiple sclerosis (MS), and hereditary ataxic conditions, where impairments in the sensorimotor system are commonly present, for example when coordinating arm and hand during goal-directed activities [2–5]. In research studies, such deficits are often assessed by healthcare practitioners, who subjectively evaluate persons with impairments during multiple standardized tasks (referred to as *conventional scales*) [6–8]. While most of these scales are validated and their interpretation fairly well understood and documented, they often have a limited ability to detect fine impairments because of limited knowledge about behavioral variability, low resolution, and ceiling effects, leading to bias when attempting to model and better understand longitudinal changes in impairment severity [9–12].

Digital health metrics, herein defined as discrete one-dimensional metrics that are extracted from health-related sensor data, promise to overcome these shortcomings by proposing objective and traceable descriptions of human behaviour without ceiling effects and with high resolution [13]. This offers the potential to more sensitively characterize impairments and significantly reduce sample sizes required in resource-demanding clinical trials [14]. In the context of assessing sensorimotor impairments, a variety of digital health metics relying on kinematic or kinetic data have been successfully applied to characterize abnormal movement patterns [13, 15, 16]. However, the integration of digital health metrics into clinical routine and research is still inhibited by an insufficient evaluation of the vast amount of existing measures and the need for core sets of validated and clinically-relevant measures for the targeted impairments [13, 17–20]. Indeed, recent reviews reported the use of over 150 sensor-based metrics for quantifying upper limb sensorimotor impairments and highlighted a clear lack of evidence regarding their pathophysiological motivation and clinimetric properties [13, 21]. Especially the ability of a metric to detect impairments (discriminant validity) as well as the dependency to other metrics and the underlying information content (structural validity) are often not evaluated. Similarly, test-retest reliability, measurement error arising from intra-subject variability, and learning effects are only rarely considered, but their evaluation is fundamental to reliably and sensitively quantify impairments in an insightful manner [22]. Further, the influence of participant demographics, such as age, sex, and handedness, on the metrics is often not accurately modeled, but needs to be taken into account to remove possible confounds and provide an unbiased assessment. Most importantly, the high variability of clinimetric properties across behavioral tasks and sensor-based metrics motivates the need for a methodology to select metrics for a specific assessment task, starting from

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a large set of potential metrics that should be narrowed down to a clinically-relevant core set [13, 17, 23]. Unfortunately, existing approaches to select core sets often do not consider the pathophysiological interpretation of metrics or are rarely tailored to the specific requirements of digital health metrics (e.g,. sufficient clinimetric properties) [17, 24–28].

Hence, the objective of this work was to propose and apply a data-driven framework to select and validate digital health metrics, aimed at providing evidence that facilitates their clinical integration. The approach relies on i) a use-case specific pathophysiological motivation for sensor-based metrics to represent clinically-relevant impairments, considers ii) the modeling of confounds arising through participant demographics, and implements iii) data processing steps to quantitatively evaluate metrics based on the most important clinimetric properties (discriminant validity, structural validity, test-retest reliability, measurement error, and learning effects). Herein, we present this framework in the context of a use-case with the Virtual Peg Insertion Test (VPIT), an instrumented assessment of upper limb sensorimotor impairments consisting of a goal-directed manipulation task in a virtual environment [29–34]. For this purpose, 77 kinematic and kinetic metrics were extracted from VPIT data from a cohort of neurologically intact and affected subjects (stroke, MS, and hereditary ataxia). We hypothesized that the presented methodology would be able to reduce a large set of metrics to a core set with optimal clinimetric properties that allows assessing the severity of the targeted impairments in a robust and insightful manner.

Targeting this objective is important, as the proposed data-driven framework can easily be applied to metrics gathered with other digital health technologies. This will help addressing the lacking evaluation, standardization, and interpretability of digital health metrics, a necessary step to address their still limited clinical relevance [19, 20, 35]. Further, the presented use-case establishes a validated core set of metrics for the VPIT, paving the way for its integration into clinical trials in neurorehabilitation.

2 Methods

To objectively reduce a large set of digital health metrics to a clinically-relevant subset, we implemented a three-step process (Figure 1) considering the most important statistical requirements to sensitively and robustly monitor impairments in a longitudinal manner. These requirements were inspired from the COSMIN guidelines for judging the quality of metrics based on systematic reviews and related work on digital health metrics [13, 22, 36–38]. Further, two additional validation steps were implemented to improve the understanding of the selected core metrics (Figure 1). While this selection and validation framework is independent of a specific assessment platform (i.e., the initial set of metrics to be evaluated), the manuscript defines the framework in the context of the VPIT with the goal to provide specific instructions including a hands-on example, starting from the initial motivation of metrics to the selection of a

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

validated core set.

2.1 Virtual Peg Insertion Test

The VPIT is a digital health assessment combining a commercial haptic end-effector (PHANTOM Omni/Touch, 3D Systems, CA, USA), a custom-made handle with piezoresistive force sensors (CentoNewton40, EPFL, Switzerland), and a virtual reality (VR) environment, implemented in C++ and OpenGL on a Microsoft (Redmond, WA, USA) Windows laptop (Figure 1). The assessment features a goal-directed pick-and-place task that requires arm and hand movements while actively lifting the arm against gravity, thereby combining elements of the Nine Hole Peg Test (NHPT) and the Box and Block Test [39,40]. The VR environment displays a rectangular board with nine cylindrical pegs and nine corresponding holes arranged as a 3×3 matrix with the same dimensions as the NHPT (31.1×26.0×4.3 cm) [39]. The objective is to transport the virtual pegs into the holes by controlling a cursor through the 6D-movements (3D-position and 3D-angular orientation) of the haptic device, which provides up to 3.3 N of haptic feedback to render the virtual pegboard. A peg can be picked up by aligning the position of a cursor with the peg (alignment tolerance: 3.0 mm) and applying a grasping force above a 2 N threshold. The peg needs to be transported towards a hole while maintaining a grasping force of at least 2 N, and can be inserted in the hole by releasing the force below the threshold, once properly aligned with a hole. The holes in the board of the VR environment are rendered through reduced haptic impedance compared to other parts of the board. The pegs cannot be picked up anymore upon insertion in a hole and are perceived as transparent throughout the test (i.e., no collisions between pegs are possible). The default color of the cursor is yellow and changes after spatially aligning cursor and peg (orange), during the lifting of a peg (green), or after applying a grasping force above the threshold while not being spatially aligned with the peg (red). During the execution of the task, 6D-endpoint position, grasping forces, and interaction forces with the VR environment are recorded at 1 kHz.

2.2 Participants & procedures

The analysis presented in this work builds on data from different studies that included assessments with the VPIT [30, 41–43]. In addition, age-matched reference data was based on 120 neurologically intact subjects. Their handedness was evaluated using the Edinburgh Handedness Inventory and potential stereo vision deficits that might influence the perception of a virtual environment were screened using the Lang stereo test [44]. Sixty of these subjects were further tested a second time one to three days apart to evaluate test-retest reliability. Additionally, 53 post-stroke subjects, 28 MS subjects, and 8 subjects with autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS) were tested. Each subject was tested with the VPIT on both body sides if possible. The administered conventional assessments were dependent on the disease and the specific study. Commonly applied assessments were the Fugl-Meyer upper ex-

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tremity (FMA-UE) [9], the Nine Hole Peg Test (NHPT) [39], and the Action Research Arm Test (ARAT) [45]. Detailed exclusion criteria and ethical approval references are listed in the supplementary material (SM). All subjects gave informed written consent.

To perform the VPIT, participants were seated in a chair with backrest and without armrests in front of a personal computer with the haptic device being placed on the side of the tested limb. The initial position of the subjects (i.e., hand resting on the handle) was defined by a shoulder abduction angle of ≈45°, a shoulder flexion angle of ≈10°, and an elbow flexion angle of ≈90°. Subjects were familiarized with the task and subsequently performed five repetitions (i.e., inserting all nine pegs five times) per body side. Participants were instructed to perform the task as fast and accurately as possible.

2.3 Data preprocessing

Data preprocessing steps are required to optimize the quality of the sensor data and dissect the complex recorded movement patterns into distinct movement phases that can be related to specific sensorimotor impairments. First, temporal gaps larger than 50 samples in the recorded position, force, and haptic time-series were linearly interpolated. Such gaps can stem from a delayed communication between the soft- and hardware components during the data recordings. Subsequently, a 1D distance trajectory $d(t)$ was estimated from the 3D cartesian position trajectories p_x, p_y, and p_z by summing up their absolute first time-derivatives until timepoint t:

$$d(t) = \sum_{1}^{t} \|\dot{p}_x\| + \|\dot{p}_y\| + \|\dot{p}_z\| \tag{1}$$

Afterwards, velocity (first time-derivative) and jerk (third time-derivative) signals were derived from $d(t)$. Also, single grasping force and grip force rate (first time-derivative) trajectories were generated by averaging across the signals of the three piezoresistive sensors. All time-series were low-pass filtered initially and after each derivation using a zero-phase Butterworth filter (4nd order, cut-off frequency 8 Hz). Data from an entire peg were removed if it was dropped and not inserted into a hole before another peg was picked up.

To isolate rapid ballistic movements, the trajectories of each peg were segmented into the *transport* (i.e., ballistic movement while transporting the peg to a hole) and *return* (i.e., ballistic movement while returning the cursor to the next peg) phases (Figure SM1). The *transport* phase started at the last occasion the velocity exceeded a threshold $\theta_{vel,tp}$ after the peg was picked up and before maximum velocity $v_{max,tp}$ was reached. The threshold $\theta_{vel,tp}$ was set to 10% of $v_{max,tp}$ that occurred before the insertion of the peg into the next hole. The end of the *transport* was defined as the first time the velocity dropped below $\theta_{vel,tp}$ after $v_{max,tp}$. To ensure a robust segmentation, the *transport* phase of a peg was discarded in case the peg was taken at $v_{max,tp}$, the velocity never dropped below $\theta_{vel,tp}$ after $v_{max,tp}$ before releasing the peg, or the length of the

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

phase was below 0.1 s. The same criteria were applied to segment the *return* phase, which was defined as the main ballistic movement component between releasing a peg and picking up the next peg, given the maximal velocity $v_{max,rt}$ during return and $\theta_{vel,rt}$. For segmenting the *transport* and *return* phases, only the horizontal component of $d(t)$ was used [46].

To isolate the overshoot when reaching for a target as well as the precise position adjustments related to virtual object manipulations, the trajectories were additionally segmented into the *peg approach* and *hole approach* phases. The former was defined from the end of the *return* until the next peg was picked up. The latter was defined from the end of the *transport* until the current peg was inserted into a hole.

Further, grasping forces were additionally segmented into the *force buildup* (i.e., behaviour during the most rapid production of force) and *force release* phases (i.e., behaviour during the most rapid release of force), by first identifying the position of the maximum and minimum value in grip force rate between approaching and inserting each peg (Figure SM1). Subsequently, the start and end of the *force buildup* phase was defined as the last and first time the grip force rate was below 10% of its maximum before and after the maximum, respectively. Similarly, the start and end of the *force release* phase was determined based on the last and first time the grip force rate was above 10% of its minimum value before and after the minimum, respectively.

2.4 Pathophysiological motivation of digital health metrics

To facilitate the pathophysiological interpretation of sensor-based metrics for each use-case, it is of importance to describe the mechanisms underlying a specific disease, their effect on the assessed behavioral construct, and how metrics are expected to capture these abnormalities. Within the use-case of the VPIT, this pathophysiological motivation is implemented using the computation, anatomy, and physiology model as well as the clinical syndromes ataxia and paresis that are commonly present in neurological disorders [47, 48]. Leveraging these concepts allows to especially connect how inappropriately scaled motor commands and an inability to voluntarily activate spinal motor neurons affect upper limb movement behaviour. As the VPIT strives to capture multiple heterogeneous and clinically-relevant sensorimotor deficits, a variety of different movement characteristics were defined to describe commonly observed upper limb sensorimotor impairments in neurological disorders. Subsequently, an initial set of 77 metrics (Table 1 and 2) for the VPIT were proposed with the aim to describe these movement characteristics and the associated sensorimotor impairments. These metrics were preselected based on the available sensor data (i.e., end-effector kinematic, kinetics, and haptic interactions), recent systematic literature reviews as well as evidence-based recommendations [13, 21, 49], and the technical and clinical experience of the authors.

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.4.1 Movement smoothness

Goal-directed movements are executed by translating parameters such as target distance into neural commands of certain amplitude, which are transferred to peripheral muscles performing a movement [50]. The signals' amplitudes are chosen to minimize movement endpoint variance, which leads to smooth behaviour (i.e., bell-shaped velocity trajectories) [51]. These velocity trajectories can be modeled using a superposition of submovements and minimize the magnitude of the jerk trajectory [52]. In neurological subjects, more submovements with increased temporal shift and higher jerk magnitudes have been observed [53, 54], potentially due to disrupted feedforward control mechanisms. The temporal shift between subcomponents and the jerk magnitude was shown to reduce after receiving rehabilitation therapy [53], thereby highlighting their relevance to track recovery. We used the integrated jerk (referred to as *jerk*) normalized with respect to movement duration and length leading to a dimensionless metric to represent the intrinsic minimization of jerk [53]. The same metric was used with an additionally applied log transformation (*log jerk*) [55]. Additionally, the *spectral arc length* (i.e., metric describing spectral energy content) of the velocity trajectory should reflect the energy induced by jerky movements [55, 56]. Further, the number of peaks in the velocity profile (*number of velocity peaks*; MATLAB function *findpeaks*) was established as an inidicator for the number of submovements. Lastly, we calculated the time (*time to max. velocity*) and distance (*distance to max. velocity*) covered at peak velocity normalized with respect to the totally covered distance and time, respectively, to captures deviation from the typically observed bell-shaped velocity profile [57]. We calculated these metrics separately for *transport* and *return* as the *transport* requires precise grip force control, which could further affect feedforward control mechanisms.

2.4.2 Movement efficiency

Ballistic movements in healthy subjects tend to follow a trajectory similar to the shortest path between start and target [58]. Previous studies suggested that neurologically affected subjects instead perform movements less close to the optimal trajectory compared to healthy controls [59] and that this behaviour correlates with impairment severity, as measured by the FMA-UE [60]. This suboptimal movement efficiency results in general from abnormal sensorimotor control, for example due to from erroneous state estimates for feedforward control, abnormal muscle synergy patterns (e.g., during shoulder flexion and abduction), weakness, and missing proprioceptive cues [47, 59, 61]. We used the path length ratio (i.e., shortest possible distance divided by the actually covered distance) to represent inefficient movements [59]. Additionally, the *throughput* (ratio of target distance and target width divided by movement time) was used as an information theory-driven descriptor of movement efficiency [62, 63]. The metrics were extracted from the start of the *transport* phase until the current peg was released and from the start of the *return* phase until the next peg was

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

taken, as not only ballistic movements but also the endpoint error is of interest when describing the efficiency of movements.

2.4.3 Movement curvature

While movement efficiency describes the overall deviation from the shortest path, it does not account for the direction of the spatial deviation. This might, however, be relevant to better discriminate abnormal feedfoward control from flexor synergy pattern or weakness, as in the latter two cases the movements might be especially performed closer to the body. We therefore selected five additional metrics to analyze the spatial deviation from the optimal trajectory in the horizontal plane [31, 32]. The *initial movement angle* was defined as the angular deviation between the actual and optimal trajectory [61]. As this metric requires the definition of a specific timepoint in the trajectory to measure the deviation, and as multiple approaches were used in literature [57, 61, 62, 64], we explored three different ways to define the timepoint. This included the time at which 20% of the shortest distance between peg and hole was covered (*initial movement angle* θ_1), the time at which 20% of the actually covered distance between peg and hole was reached *initial movement angle* θ_2, and the time at which peak velocity was achieved (*initial movement angle* θ_3). Additionally, the *mean* and *maximal trajectory error* with respect to the ideal, straight trajectory were calculated. All metrics were estimated separately for *transport* and *return*.

2.4.4 Movement speed

The speed of ballistic movements in healthy subjects is mostly controlled by the tradeoff between speed and accuracy as described by Fitt's law, which is indirectly imposed through the concept of velocity-dependent neural noise [51, 63]. In neurologically affected subjects, increased speed can, for example, result from inappropriately scaled motor commands and disrupted feedforward control [47]. On the other hand, reduced speed can also stem from weakness (i.e., reduced ability to active spinal motor neurons leading to decreased strength) or spasticity (i.e., velocity-dependent increase in muscle tone), the latter resulting from upper motor neuron lesions, abnormally modulated activity in the supraspinal pathways, and thereby increased hyperexcitability of stretch reflexes [47, 65]. We calculated the mean (*velocity mean*) and maximum (*velocity max.*) values of the velocity trajectory to represent movement speed during the *transport* and *return* phases.

2.4.5 Endpoint error

To fully characterize the speed-accuracy tradeoff, we additionally analyzed the position error at the end of a movement. In neurological disorders, increased endpoint error (i.e., dysmetria) was commonly observed and can, for example, result from inappropriately scaled motor commands and thereby disrupted feedforward control [66, 67], but also from cognitive and proprioceptive deficits [68].

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dysmetria was found especially in post-stroke subjects with lateral-posterior thalamic lesions [68], is a common manifestation of intention tremor in MS [69], and is typically observed in subjects with cerebellar ataxia [70]. In the VPIT, the cumulative horizontal Euclidean distance between the cursor position and targeted peg or hole (*position error*) were calculated during the *peg approach* and *hole approach* phases, respectively. Further, the *jerk*, *log jerk*, and *spectral arc length* metrics were calculated during both phases, as a jerk index was shown previously to correlate with the severity of intention tremor in MS [71].

2.4.6 Haptic collisions

Haptic collisions describe the interaction forces between a subject and the virtual pegboard rendered through the haptic device. Haptic guidance can be used to ease inserting the virtual pegs into the holes, which have reduced haptic impedance. Previous studies indicated increased haptic collision forces in multiple neurological disorders and especially stroke subjects with sensory deficits [29, 72]. We additionally expected that collision forces during *transport* and *return* (i.e., phases during which haptic guidance is not required) could be increased due to arm weakness. In particular, neurological subjects can have a limited capability to lift their arm against gravity, leading to increased vertical haptic collisions [73]. The mean and max. vertical collision force (*haptic collisions mean* and *haptic collisions max.*) was calculated during *transport* and *return* to quantify haptic collision behaviour.

2.4.7 Number of successful movements

Subjects without neurological deficits can start and end goal-directed movements with ease. On the contrary, persons with neurological disorders can have a reduced ability to initiate and terminate ballistic movements with potentially heterogeneous underlying impairments including abnormal feedforward control, sensory feedback, spasticity, weakness, and fatigue [13, 47, 57]. Therefore, the metric *number of movement onsets* was defined based on the number of valid pegs, using the defined segmentation algorithm, when identifying the start of the *transport* and *return* phases. Analogously, *number of movement ends* was based on the sum of correctly segmented ends for the *transport* and *return* phases.

2.4.8 Object drops

Neurologically intact subjects can precisely coordinate arm movements and finger forces to transport objects. This ability can be reduced in neurological disorders and can potentially lead to the drop of an object during its transport [74]. Underlying mechanisms include for example distorted force control due to incorrectly scaled motor commands or distorted sensory feedback as well as reduced spatio-temporal coordination between arm and hand movements [47, 74]. In the VPIT, the number of virtual pegs that were dropped (*dropped pegs*) should represent object drops and thereby grip force control as well as the spatio-temporal

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

coordination of arm and hand movements. The metric was defined based on how often the grasping force dropped below a 2 N threshold (i.e., subjects still holding the handle) while lifting a virtual peg [32].

2.4.9 Grip force scaling and coordination

The precise scaling and spatio-temporal coordination of grasping forces is a key requirement for successful object manipulation and leads, in neurologically intact subjects, to single-peaked bell-shaped grip force rate profiles when starting to grasp objects [75]. Abnormal grip force scaling and decreased grip force coordination have been reported in neurological subjects, resulting in multi-peaked grip force rate profiles, and were attributed to, for example, distorted feedforward control, abnormal somatosensory feedback and processing, as well as the presence of the pathological flexor synergy [75–82]. Also, a reduction in applied grip force levels due to weakness can be expected depending on the neurological profile of a subject [47]. Further, a slowness of *force buildup* [77] and *force release* [78] has been reported, even though other studies showed that the ability to produce and maintain submaximal grip forces was preserved [74, 78]. Additionally, there is evidence suggesting that *force buildup* and *force release* have different neural mechanisms and that force control can further be decomposed into force scaling and motor coordination [78, 79].

To describe grip force scaling, we applied four metrics separately to the *transport*, *return*, *peg approach*, and *hole approach* phases. We calculated the mean (*grip force mean*) and maximum (*grip force max.*) value of the grasping force signal during each phase. Additionally, we estimated the mean absolute value (*grip force rate mean*) and absolute maximum (*grip force rate max.*) of the grip force rate time-series. Similarly, we characterized grip force coordination during the *transport*, *return*, *peg approach*, *hole approach*, *force buildup* and *force release* phases, for which we calculated the number of positive and negative extrema (*grip force rate number of peaks*) and the spectral arc length (*grip force rate spectral arc length*). For the *force buildup* and *force release* phases, which contain only the segments of most rapid force generation and release, respectively, we additionally calculated their duration (*force buildup/release duration*).

2.4.10 Overall disability

A single indicator expected to describe the subject-specific *overall disability* level was defined based on the *task completion time* (i.e., duration from first *transport* phase until insertion of last peg).

2.5 Data postprocessing

To reduce the influence of intra-subject variability, the grand median across pegs and repetitions was computed for each metric. Subsequently, the influence of possible confounds, which emerge from subject demographics not related to

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

neurological disorders, was modeled based on data from all neurologically intact subjects. This should allow to compensate for these factors when analyzing data from neurologically affected subjects. In more detail, the impact of age (in yrs), sex (male or female), tested body side (left or right), and handedness (performing the test with the dominant side: true or false) were used as fixed effects (i.e., one model slope parameter per independent variable) in a linear mixed effect model generated for each sensor-based metric [83]. Additionally, the presence of stereo vision deficits (true or false) was used as a fixed effect, as the perception of depth in the VR environments might influence task performance [84, 85]. A subject-specific random effect (i.e., one model intercept parameter per subject) was added to account for intra-subject correlations arising from including both tested body sides for each subject. A Box-Cox transformation was applied on each metric to correct for heteroscedasticity, as subjectively perceived through non-normally distributed model residuals in quantile-quantile plots [86]. Additionally, this transformation allows to capture non-linear effects with the linear models. The models were fitted using maximum likelihood estimation (MATLAB function *fitlme*) and defined as:

$$\begin{aligned} y_{i,j}^{intact} = \beta_{i,0} + \beta_{i,1}\,\text{age}_j + \beta_{i,2}\,\text{sex}_j + \beta_{i,3}\,\text{tested body side}_j + \\ \beta_{i,4}\,\text{handedness}_j + \beta_{i,5}\,\text{stereo vision deficits}_j + W_{i,j} + \epsilon_i, \end{aligned} \tag{2}$$

where:

$y_{i,j}^{intact}$	value of a metric i of neurologically intact subject j
β_i	model parameters
$W_{i,j}$	subject-specific intercept
ϵ_i	residual error.

For any subject being analyzed, the effect of all confounds on the sensor-based metric was removed based on the fitted models. This generated the value $\bar{y}_{i,j}$ of a metric without confounds arising from subject demographics:

$$\begin{aligned} \bar{y}_{i,j} = y_{i,j} - \beta_{i,1}\,\text{age}_j - \beta_{i,2}\,\text{sex}_j - \beta_{i,3}\,\text{tested body side}_j - \\ \beta_{i,4}\,\text{handedness}_j - \beta_{i,5}\,\text{stereo vision deficits}_j. \end{aligned} \tag{3}$$

Furthermore, the corrected values $\bar{y}_{i,j}$ were then expressed relative to all neurologically intact subjects (${\bar{y}_i}^{intact}$) with the goal to standardize the range of all metrics, which simplifies their physiological interpretation and enables the direct comparison of different metrics. Therefore, the normalized value $\hat{y}_{i,j}$ was defined relative to the median and variability d_i of all neurologically intact subjects:

$$\hat{y}_{i,j} = \frac{\bar{y}_{i,j} - \text{median}\left({\bar{y}_i}^{intact}\right)}{d_i}, \tag{4}$$

with the median absolute deviation (MAD) of all neurologically intact subjects being used as a variability measure [87]:

$$d_i = \text{median}\left(\left\|\bar{y}_{i,j}^{intact} - \text{median}\left({\bar{y}_i}^{intact}\right)\right\|\right), \tag{5}$$

The MAD was preferred over the standard deviation, as the former allows a more robust analysis that is independent of the underlying distribution of a

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

metric [87]. Lastly, the values $\hat{y}_{i,j}$ were divided by the maximal observed value in the included neurological population, such that the subject currently showing worst task-performance receives a score of 100%. In order to discriminate normal from abnormal behaviour based on the normalized values, a cut-off was defined based on the 95th percentile (i.e., imposed false positive detection rate of 5%) of each metric $\hat{y}_i^{intact}$ across all neurologically intact subjects.

2.6 Data-driven selection and validation of sensor-based metrics

The sensor-based metrics were reduced to a subset with optimal clinimetric properties based on three selection steps, followed by two additional validation steps. To evaluate the ability of this selection process to discriminate between physiologically-relevant information and random noise, the selection steps were additionally applied to a simulated random metric (*simulated Gaussian noise*) containing no physiologically relevant information. This metric was constructed by randomly drawing data from a log-normal distribution (mean 46.0, standard deviation 32.2, mimicking the distribution of the *total time* for the *reference population*) for each subject and tested body side.

Metric selection & validation: step 1
With the goal to better understand the influence of subject demographics on the sensor-based metric, simulated likelihood ratio tests (1000 iterations) between the full model and a reduced model without the fixed effect of interest were used to generate p-values that were interpreted based on a 5% significance level [88]. This allowed to judge whether a fixed effect influenced the sensor-based metric in a statistically significant manner. We removed metrics that were significantly influenced by stereo vision deficits, as we expected that the influence of stereo vision deficits can not always be compensated for, for example if their presence is not screened in a clinical setting.

As the performance of the presented confound correction process depends on the fit of the model to the data, we additionally removed metrics with low model quality according to the criteria $C1$ and $C2$, which describe the mean absolute estimation error (MAE) of the models and its variability [89]:

$$C1_i : \frac{MAE_i}{\text{range}\left(y_i^{intact}\right)} \leq 15\% \tag{6}$$

and

$$C2_i : \frac{MAE_i + 3\,\sigma_i}{\text{range}\left(y_i^{intact}\right)} \leq 25\% \tag{7}$$

where: $MAE = \frac{1}{n}\sum \left\|\epsilon_i^{intact}\right\|$
n = number of data points from neurologically intact subjects
σ_i = $\text{std}\left(\left\|\epsilon_i^{intact}\right\|\right)$
std = standard deviation.

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fulfilling both criteria leads to the selection of models with *moderate* and *good* quality according to the definition of Roy et al. [89]. Before the calculation of C1 and C2, data points with the 5% highest residuals were removed [89]. The criteria C1 and C2 were preferred over the more commonly used coefficient of determination R^2, because the magnitude of this metric is highly dependent on the distribution of the dependent variable, which prohibits the definition of a model quality threshold that is valid across metrics [89, 90].

Metric selection & validation: step 2

Receiver operating characteristic (ROC) analysis was used to judge the potential of a metric to discriminate between neurologically intact and affected subjects, which is a fundamental requirement to validate that the proposed metrics are sensitive to sensorimotor impairments [22, 91]. In more detail, a threshold was applied for each metric to classify subjects as being either neurologically intact or impaired. The threshold was varied across the range of all observed values for each metric and the true positive rate (number of subjects correctly classified as neurologically affected divided by the total number of neurologically affected subjects) and false positive rate (number of subjects incorrectly classified as neurologically affected divided by the total number of neurologically intact subjects) were calculated. The area under the curve (AUC) when plotting true positive rates against false positive rates was used as a quality criterion for each metric (Figure 2).

For metrics to be responsive to intervention-induced physiological changes and allow a meaningful tracking of longitudinal changes, it is fundamental to have low intra-subject variability, high inter-subject variability, and yield repeatable values across a test-retest sessions. Therefore, the data set with 60 neurologically intact subjects performing the VPIT protocol on two separate testing days was used to quantify test-retest reliability. Specifically, the intraclass correlation coefficient (ICC) was calculated to describe the ability of a metric to discriminate between subjects across multiple testing days (i.e., inter-subject variability) [92, 93]. The agreement ICC based on a two-way analysis of variance (ICC A,k) was applied while pooling data across both tested body sides. Further, the smallest real difference (SRD) was used to define a range of values for that the assessment cannot distinguish between measurement error and an actual change in the underlying physiological construct (i.e., intra-subject variability) [94]. For each metric i, the SRD was defined as

$$SRD_i = 1.96 \cdot \sqrt{2} \cdot \Sigma_i^{intact} \cdot \sqrt{1 - ICC_i} \tag{8}$$

where: Σ_i = std across repetitions, subjects, and testing days.

To directly relate the SRD to the distribution of a metric, it was further expressed relative to a metrics' range:

$$SRD\%_i = 100 \cdot \frac{SRD_i}{\text{range}\left(\hat{y}_i^{intact}\right)}. \tag{9}$$

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lastly, to distinguish task-related learning from physiological changes when testing subjects before and after receiving an intervention, the presence and strength of learning effects was calculated for each metric. For this purpose, a paired t-test was performed between data collected at test- and retest to check for a statistically significant difference between the days. Then, the strength (i.e., slope) of the learning effect was estimated by calculating the mean difference between test and retest and normalizing it with respect to the range of observed values:

$$\eta_i = 100 \cdot \frac{\text{mean}\left(\hat{y}^{intact}_{i,j,retest} - \hat{y}^{intact}_{i,j,test}\right)}{\text{range}\left(\hat{y}^{intact}_{i}\right)}. \tag{10}$$

Metrics passed this second selection step if the AUC did indicate acceptable, excellent, or outstanding discriminant ability (AUC$\geq$0.7) and they had at least acceptable reliability (i.e., ICC values above 0.7) [22, 91]. As no cutoff has been defined for the interpretation of the SRD% [95], we removed the metrics that had the 20% worst SRD% values. Hence, metric passed the evaluation (i.e., small measurement error relative to other metrics) if the SRD% was below 30.3 (80th-percentile). Similarly, no cutoff for the interpretation of learning effects was available. Hence, metrics passed the evaluation (i.e., no strong learning effects) if η was above -6.35 (20th-percentile) of observed values.

Metric selection & validation: step 3

The correlations between the metrics were analyzed with the goal to identify a set of metrics that contains little redundant information to simplify clinical interpretability. Therefore, a correlation matrix was constructed using partial Spearman correlations. This technique allows to describe the relation between two metrics and to simultaneously model all other metrics that could potentially influence the relationship between the two metrics of interest [96, 97]. Hence, this approach can help to exclude certain non-causal correlations. A pair of metrics with an absolute partial correlation ρ_p of at least 0.5 was considered for removal [98]. From this pair of metric, the one that had inferior psychometric properties (AUC, ICC, and SRD%) or was less accepted in literature was removed. To simplify the interpretation of the correlation results, we applied the analysis only to metrics that passed all previous selection steps. Additionally, this analysis was applied in an iterative manner, as the removal of certain metrics, which were previously modeled, can change the remaining intercorrelations. The correlation coefficients were interpreted according to Hinkle et al.: very high: $\rho_p \geq 0.9$; high: $0.7 \leq \rho_p < 0.9$; moderate: $0.5 \leq \rho_p < 0.7$; low: $0.3 \leq \rho_p < 0.5$; very low: $\rho_p < 0.3$ [98].

Further validation of metrics: step 1

To better identify the pathophysiological correlates of the metrics that passed all previous evaluation steps, exploratory factor analysis was applied [99–101]. This method tries to associate the variability observed in all metrics with k unobserved latent variables via factor loadings, which can be interpreted in light of the initial physiological motivation of the metrics. Exploratory factor analysis

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

was implemented using maximum likelihood common factor analysis followed by a *promax* rotation (MATLAB function *factoran*). For the interpretation of the emerged latent space, we only considered strong (absolute value≥0.5) factor loadings [99]. The number of factors k was estimated in a data-driven manner using parallel analysis (R function *fa.parallel*) [102]. This approach simulates a lower bound that needs to be fulfilled by the eigenvalue associated to each factor and has been shown to be advantageous compared to other more commonly used criteria, such as the Kaiser condition (i.e., eigenvalues>1 are retained) [100,101]. Also, the Kaiser-Meyer-Olkin value (KMO) was calculated to evaluate whether the data was mathematically suitable for the factor analysis.

Further validation of metrics: step 2

An additional clinically-relevant validation step evaluated the ability of the metrics to capture the severity of upper limb disability. For this purpose, each population was grouped according to their disability level as defined by commonly used clinical scores. Subsequently, the behaviour of the metrics across the subpopulations and the *reference population* were statistically analyzed. Stroke subjects were grouped according to the FMA-UE score (ceiling: FMA-UE=66; mild impairment: 54≤FMA-UE<66; moderate impairment: 35≤FMA-UE<54) [103]. MS subjects were split into three groups based on their ARAT score (full capacity: 55≤ARAT≤57; notable capacity: 43≤ARAT<55; limited capacity: 22≤ARAT<43) [104]. ARSACS subjects were divided into three different age-groups (young: 26≤age≤36; mid-age: 37≤age≤47; older-age: 48≤age≤58) due to the neurodegenerative nature of the disease [4]. A Kruskal-Wallis omnibus test followed by post-hoc tests (MATLAB functions *kruskalwallis* and *multcompare*) were applied to check for statistically significant differences between groups. Bonferroni corrections were applied in both cases.

3 Results

Data from 120 neurologically intact subjects of age 51.1 [34.6, 65.6] yrs (median [25th-percentile, 75th-percentile]; 60 male; 107 right hand dominant; 12 with stereo vision deficits) was acquired with 60 of them performing a test-retest session (age 48.8 [40.2, 60.2]; 34 male; 48 right hand dominant; time between sessions 5.0 [4.0, 6.5] days). Eighty-nine neurologically affected subjects (53 post-stroke with affected side FMA-UE 57 [49, 65], 28 MS with ARAT 52.0 [46.5, 56.0], 8 ARSACS with NHPT 43.5 [33.1, 58.7] s) were used for the selection and validation of the metrics. Their age was 56.2 [42.1, 65.3] yrs, 52 were male, 75 were right hand dominant, and for 35 stroke subjects, the right body side was most affected. In total, data from 43350 individual movements were recorded. Detailed demographic and the available clinical information for each neurologically affected subject can be found in Table SM1.

Selection of metrics: step 1

The influence of all potential confounds and the model quality for each sensor-

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: **Results for the data-driven selection of kinematic metrics.** The area under the curve (AUC, optimum at 1), intraclass correlation coefficient (ICC, optimum at 1), the smallest real difference (SRD%, optimum at 0), and η value (optimum at 0, worst at $-\inf$) were used to describe discriminative validity, test-retest reliability, measurement error, and learning effects, respectively. Metrics in bold fulfilled all evaluation criteria (AUC>0.7, ICC>0.7, SRD%<30.3, and η >-6.35). Metrics with insufficient model quality according to selection step 1 are annotated with a † and reported for completeness. mov: movement; TP: transport; RT: return; SPARC: spectral arc length; num: number.

Movement characteristic	Sensor-based metric	Validity: AUC	Reliability: ICC	Error: SRD%	Learning: η
Mov. smoothness TP	Jerk TP	0.80	0.69	23.10	-4.41
	Log jerk TP	**0.78**	**0.74**	**26.11**	**-4.82**
	SPARC TP†	0.84	0.83	23.78	-7.16
	Num. velocity peaks TP†	0.82	0.79	21.30	-6.36
	Distance to max. velocity TP†	0.44	0.74	33.64	2.42
	Time to max. velocity TP†	0.45	0.78	28.70	3.93
Mov. smoothness RT	Jerk RT	0.84	0.68	20.83	-4.70
	Log jerk RT	**0.73**	**0.75**	**25.33**	**-6.08**
	SPARC RT	**0.71**	**0.76**	**28.93**	**-1.57**
	Num. velocity peaks RT†	**0.76**	**0.70**	**23.27**	**-3.28**
	Distance to max. velocity RT	0.43	0.65	41.39	3.67
	Time to max. velocity RT	0.48	0.73	33.99	2.43
Mov. efficiency TP	**Path length ratio TP**	**0.89**	**0.76**	**24.24**	**-2.17**
	Throughput TP†	0.92	0.81	24.07	-12.18
Mov. efficiency RT	**Path length ratio RT**	**0.83**	**0.79**	**17.30**	**-3.61**
	Throughput RT	0.90	0.78	27.43	-13.21
Mov. curvature TP	Trajectory error mean TP	0.55	0.86	17.14	-0.60
	Trajectory error max. TP	0.57	0.86	15.84	-0.37
	Initial mov. angle TP θ_1†	0.67	0.90	13.56	-1.50
	Initial mov. angle TP θ_2†	0.67	0.90	13.29	-1.52
	Initial mov. angle TP θ_3	0.61	0.88	14.37	-2.06
Mov. curvature RT	Trajectory error mean RT	0.56	0.84	20.00	1.24
	Trajectory error max. RT	0.55	0.84	18.58	1.22
	Initial mov. angle RT θ_1	0.51	0.75	33.90	3.18
	Initial mov. angle RT θ_2	0.51	0.71	28.65	2.92
	Initial mov. angle RT θ_3	0.60	0.79	23.99	1.53
Mov. speed TP	Velocity mean TP	0.83	0.88	20.61	-9.99
	Velocity max. TP	0.83	0.87	18.57	-9.14
Mov. speed RT	Velocity mean RT	0.75	0.87	19.01	-7.60
	Velocity max. RT	**0.76**	**0.86**	**19.41**	**-6.27**
Endpoint error peg approach	Position error peg approach	0.86	0.64	29.54	-4.66
	Jerk peg approach	**0.74**	**0.72**	**27.65**	**-2.94**
	Log jerk peg approach	0.69	0.75	30.20	-8.36
	SPARC peg approach	0.78	0.64	46.55	-10.29
Endpoint error hole approach	Position error hole approach	0.94	0.76	31.29	-5.36
	Jerk hole approach	0.57	0.68	30.63	-4.84
	Log jerk hole approach	0.66	0.83	23.25	-6.53
	SPARC hole approach	**0.86**	**0.81**	**24.81**	**-5.72**
Haptic collisions TP	Haptic collisions mean TP	0.61	0.85	24.55	-3.99
	Haptic collisions max. TP	0.63	0.84	20.54	-1.08
Haptic collisions RT	Haptic collisions mean RT	0.61	0.72	25.32	-0.07
	Haptic collisions max. RT†	0.46	0.79	27.02	4.37
Number of movements	Number of mov. onsets	0.22	0.22	61.34	-0.82
	Number of mov. ends	0.09	0.29	57.01	0.00
Object drops	Number of dropped pegs	0.65	0.50	41.11	-3.20

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2: **Results for the data-driven selection of kinetic metrics.** The area under the curve (AUC, optimum at 1), intraclass correlation coefficient (ICC, optimum at 1), the smallest real difference (SRD%, optimum at 0), and η value (optimum at 0, worst at $-\inf$) were used to describe discriminative validity, test-retest reliability, measurement error, and learning effects, respectively. The *task completion time* and the *simulated Gaussian noise* metrics were evaluated in addition to the kinetic metrics. Rows in bold fulfilled all evaluation criteria (AUC>0.7, ICC>0.7, SRD%<30.3, and η >-6.35). Metrics with insufficient model quality according to selection step 1 are annotated with a † and reported for completeness. GF: grip force; TP: transport; RT: return; SPARC: spectral arc length; num: number.

Movement characteristic	Sensor-based metric	Validity: AUC	Reliability: ICC	Error: SRD%	Learning: η
GF scaling TP	GF mean TP	0.40	0.84	14.46	0.39
	GF max. TP	0.40	0.86	15.19	0.07
	GF rate mean TP	0.25	0.87	12.14	2.07
	GF rate max. TP	0.25	0.79	20.53	3.93
GF scaling RT	GF mean RT	0.49	0.76	27.62	0.17
	GF max. RT	0.45	0.66	37.61	2.80
	GF rate mean RT	0.07	**0.82**	27.79	5.87
	GF rate max. RT	0.29	0.48	34.05	7.19
GF scaling peg approach	GF mean peg approach	0.45	0.83	18.09	1.10
	GF max. peg approach	0.39	0.84	19.40	-0.72
	GF rate mean peg approach	0.18	0.88	14.76	3.54
	GF rate max. peg approach	0.32	0.84	19.52	0.74
GF scaling hole approach	GF mean hole approach	0.36	0.81	15.34	0.76
	GF max. hole approach	0.37	0.82	16.43	0.50
	GF rate mean hole approach	0.15	0.82	14.18	2.73
	GF rate max. hole approach	0.28	0.77	21.41	1.82
GF coord. TP	**GF rate num. peaks TP**	**0.74**	**0.81**	**20.59**	**-6.11**
	GF rate SPARC TP	**0.74**	**0.82**	**22.48**	**-5.71**
GF coord. RT	GF rate num. peaks RT	0.60	0.83	20.17	-4.16
	GF rate SPARC RT	0.64	0.78	23.81	-6.35
GF coord. peg approach	GF rate num. peaks peg approach	0.90	0.78	25.60	-12.25
	GF rate SPARC peg approach	0.90	0.83	22.99	-8.19
GF coord. hole approach	**GF rate num. peaks hole approach**	**0.91**	**0.81**	**24.29**	**-6.14**
	GF rate SPARC hole approach	**0.84**	**0.82**	**26.38**	**-5.94**
GF coord. buildup	GF rate num. peaks buildup†	0.15	0.44	57.70	0.77
	GF rate SPARC buildup†	0.56	0.79	28.62	-3.22
	GF buildup duration	0.70	0.82	21.36	-6.97
GF coord. release	GF rate num. peaks release†	0.44	0.48	56.80	1.78
	GF rate SPARC release	0.91	0.86	18.63	-6.78
	GF release duration	0.67	0.81	21.63	-2.78
Overall disability	Task completion time	0.91	0.78	26.16	-11.34
	Simulated Gaussian noise†	0.37	-0.07	117.04	0.25

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

based metric including p-values can be found in Table SM2 (example in Figure 2). For all metrics, 69.7%, 44.7%, 27.6%, 6.6%, and 7.9% were significantly influenced by age, sex, tested side, hand dominance, and stereo vision deficits, respectively. In more detail, *initial movement angle transport* θ_1, θ_2, θ_3, *number of movement ends*, *number of dropped pegs*, *grip force rate number of peaks buildup* were the metrics being altered by stereo vision deficits. The required quality of the models, according to the C1 and C2 criteria, were not fulfilled by thirteen (16.9%) of all metrics (including the *simulated Gaussian noise*, see SM for a detailed list).

Selection of metrics: step 2

Thirteen (16.9%) out of 77 metrics fulfilled the criteria of the validity, reliability, measurement error, and learning analysis (Figure 2, Table 1, and Table 2). The median AUC, ICC, SRD%, and η values of the 12 metrics that passed step 1 and step 2 were 0.77 [0.74, 0.85], and 0.80 [0.75, 0.82], 24.6 [21.5, 26.2], and -5.72 [-6.09, -3.27] respectively. The *simulated Gaussian noise* metric did not pass this evaluation step (AUC 0.37, ICC -0.07, SRD% 117.04, η 0.25).

Selection of metrics: step 3

The constructed partial correlation matrices can be found in Figure 3. Among the remaining metrics, *grip force rate number of peaks hole approach* was removed as it correlated ($\rho_p \geq 0.5$) with *grip force rate spectral arc length approach hole* and the latter metric is less influenced through confounds as it is independent of movement distance. Additionally, *spectral arc length hole approach* was discarded as it correlated with *grip force rate spectral arc length hole approach* and the latter metric is more directly related to hand function, which was not yet well covered by the other metrics. The remaining 10 metrics yielded absolute partial inter-correlations of 0.14 [0.06 0.24] (zero very high, zero high, zero moderate, six low, and 39 very low inter-correlations).

Further validation of metrics: step 1

The Kaiser-Meyer-Olkin value was 0.82, which indicated that the application of the factor analysis was suitable [105, 106]. According to the parallel analysis, the most likely number of underlying latent factors k was five (Figure SM2). The factor loadings can be found in Table 3. The metrics *path length ratio transport/return* and *jerk peg approach* had strong loadings on factor 1. The metrics *log jerk transport*, *log jerk return*, and *spectral arc length return* loaded strongly on factor 2. The metrics *grip force rate number of peaks transport* and *grip force rate spectral arc length transport* had strong loadings on factor 3, whereas *velocity max. return* and *grip force rate spectral arc length hole approach* loaded strongly on factor 4 and 5, respectively.

Further validation of metrics: step 2

The behaviour of all metrics across subject subpopulations with increasing disability level can be found in Figure 4, 5, and 6. All metrics indicated statistically

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3: **Structural validity: exploratory factor analysis.** Loadings of metrics on underlying latent factors extracted with exploratory factor analysis. The interpretation of each metric was physiologically motivated initially. Larger absolute loadings indicate a stronger contribution to a factor. Bold font indicate strong loadings (i.e., absolute loading of at least 0.5). Abbreviations: F1-5: data-driven latent factors. GF: grip force; coord: coordination; num: number; SPARC: spectral arc length.

Expected interpretation	**Sensor-based metric**	**F1**	**F2**	**F3**	**F4**	**F5**
Movement smoothness transport	Log jerk transport	0.09	**0.73**	0.21	-0.19	-0.05
Movement smoothness return	Log jerk return	-0.08	**0.86**	-0.11	0.02	0.02
	SPARC return	0.10	**0.59**	-0.10	0.23	-0.03
Movement efficiency transport	Path length ratio transport	**0.83**	0.08	-0.17	0.06	0.11
Movement efficiency return	Path length ratio return	**0.79**	-0.06	0.08	-0.14	0.04
Movement speed transport	Velocity max. return	-0.02	0.01	0.16	**0.90**	0.01
Endpoint error peg approach	Jerk peg approach	**0.72**	-0.04	0.12	0.07	-0.14
GF coord. transport	GF num. peaks transport	0.00	-0.06	**0.93**	0.11	-0.03
	GF rate SPARC transport	-0.08	0.19	**0.62**	0.00	0.11
GF coord. hole approach	GF rate SPARC hole approach	0.11	-0.02	0.02	0.01	**0.94**

significant differences between the neurologically intact and at least one of the neurologically affected subpopulations for each disorder, with the exception of *jerk peg approach* in MS subjects. Additionally, significant differences between subpopulations were found for *log jerk transport* in stroke subjects. Consistent trends (i.e., monotonically increasing medians across subpopulations) were found for all metrics except for *spectral arc length return*, *force rate spectral arc length approach hole*, and *force rate num. peaks approach hole*.

4 Discussion

In this work, we aimed to propose and apply a data-driven framework to select and validate digital health metrics, with the objective to facilitate their still lacking clinical integration. The approach considers i) the targeted impairments, ii) the influence of participant demographics, and iii) important clinimetric properties. As an example use-case, we implemented this framework with 77 kinematic and kinetic metrics extracted from the VPIT, a previously proposed sensor-based assessment of arm and hand sensorimotor impairments. For this purpose, the VPIT was administered to 120 neurologically intact and 89 neurologically affected subjects, yielding data from 43350 individual movements.

This objective methodology to identify a core set of validated metrics based on pathophysiological hypotheses and quantitative selection criteria can complement currently applied paradigms for selecting digital health metrics [17, 24–28, 38]. While consensus-based recommendations from groups of experts are indispensable for constructing high-level hypothesis (e.g., which body functions to assess in a given context), the selection of specific sensor-based metrics should

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

solely be implemented based on objective and data-driven evaluation criteria to avoid selection bias. Also, guidelines to pool data within systematic reviews, often intended for the selection of conventional assessments, need to be considered carefully in the context of digital health metrics. Compared to conventional assessments that often provide a single, intuitively understandable, task-specific metric (e.g., FMA-UE score), a plethora of abstract digital health metrics exists and the same metric (e.g., *log jerk*) can be extracted from all technologies sharing similar sensor data. However, for a meaningful interpretation of sensor-based metrics, it is essential to consider them in light of the assessment context, as data processing steps (e.g., filter design), assessment platform type (e.g., end-effector or camera-based system), task type (e.g,. goal-directed or explorative movements), and target population (e.g, neurological or muscoskeletal impairments) strongly influences the anticipated hypotheses and clinimetric properties [13]. This emphasizes the importance of a validation of each metric in its specific context (i.e., assessment platform, task, and target population). While objective metric selection algorithms leveraging on the nowadays existing big data sets are already well established in the machine learning domain (therein referred to as *feature selection algorithms*) [24], these usually rely on accurate ground truth (i.e., supervised learning) about the targeted impairment, which is unfortunately often not available in the healthcare domain. Hence, the proposed approach should be seen as an unsupervised metric selection framework aimed to provide a solid foundation of evidence that is required to better transfer research findings into clinical healthcare environments [20, 35].

4.1 Specific methodological contributions

In line with literature [39, 40, 57], the mixed effect model analysis (Table SM2) revealed that a high amount of all metrics were significantly modulated by age (69.7% of all metrics) and sex (44.7%), whereas a more selective influence was found for tested side (27.6%) and hand dominance (6.6%). For an accurate assessment of sensorimotor disability without confounds, it is therefore essential to account for these factors when comparing between neurologically intact and affected subjects with different demographics. The presented analysis adds an important methodological contribution to previous work that used linear models to compensate for confounds by additionally evaluating the quality of these models [37, 107–109]. This allowed to discard metrics for which the confounds could not be accurately modeled (16.8% of all metrics). Especially metrics that have mathematical support with two finite boundaries (e.g., 0% and 100%) received low model quality, which can result from skewness and heteroscedasticity that can not be corrected using variance-stabilizing transformations, such as the Box-Cox method. Such metrics should therefore be considered carefully and other modeling approaches, for example based on beta distributions, might be required to accurately compensate for the effect of measurement confounds [110].

Eighty-three percent of all metrics (Table 1) were discarded through the second selection step. It is fundamental to understand that these evaluation criteria (AUC, ICC, SRD%, η) are complementary to each other, focusing on different

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

components of intra-subject and inter-subject variability, which are all essential to sensitively monitor impairments. It is therefore not sufficient to solely consider a subset of these criteria, as often done in literature. Evaluating the validity of sensor-based metrics using a *reference population* and ROC analysis is superior to the more commonly applied correlations with conventional scales (concurrent validity) [21, 22]. A reason for this is that sensor-based approaches are being expected to provide complementary information to conventional scales that improves upon their limitations, thereby challenging the definition of accurate hypothesis about the correlation between conventional and sensor-based scales. Nevertheless, comparisons between metrics and conventional scales can help to better interpret sensor-based metrics or to test their sensitivity to impairment severity, as attempted in the last validation step. This analysis was not used as a criteria for metric selection as, to expect trends across subgroups, each sensor-based metric would require a carefully selected clinical counterpart that captures a similar physiological construct. Also, stepwise regression approaches that model conventional scales in order to select metrics have been extensively applied even though they have been considered bad practice due to statistical shortcomings [111–114].

Lastly, a simulated metric without relevant information content (*simulated Gaussian noise*) was rejected in the first and second selection step, thereby providing evidence that the framework allows to discard certain physiologically-irrelevant metrics.

4.2 A core set of validated metrics for the VPIT

Applying the proposed framework, ten almost independent metrics (Table 3) were identified as a validated core set for the VPIT and were able to reliably assess the severity of multiple sensorimotor impairments in arm and hand for subjects with mild to moderate disability levels (i.e., the target population of the VPIT). These metrics were related to the movement characteristics smoothness, efficiency, speed, endpoint error, and grip force coordination during specific phases of the task (*transport*, *return*, *peg approach*, *hole approach*). While these characteristics are generally expected to inform on abnormal feedforward control, impaired somatosensory feedback, increased muscle tone, abnormal flexor synergies, dysmetria, and weakness, the clustering of the metrics into five factors allows to further speculate about their interpretation (Table 3). The first factor was dominated by movement efficiency metrics (*path length ratio transport* and *return*), and the *jerk peg approach* as a descriptor for the endpoint error of a movement, thereby fully characterizing the speed-accuracy tradeoff that is a typical characteristic of goal-directed movements [51, 63]. The second factor contained metrics focusing on movement quality (smoothness) during *transport* and *return*, which is expected to describe impaired feedforward control of arm movements. Hence, it is unlikely that the first factor also informs on feedforward control. We therefore expect the movement efficiency metrics (first factor) to be rather related to flexor synergy patterns, weakness, proprioceptive deficits, and dysmetria. Among these impairments, weakness and proprioceptive deficits

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are most commonly observed in neurological disorders [2, 115]. The third factor focused on grip force coordination during *transport* (*grip force rate num. peaks transport* and *grip force rate spectral arc length transport*), which is expected to be related to abnormal feedforward control and impaired somatosensory feedback. The dissociation between factor one and three is interesting, as it suggests different control schemes underlying the regulation of arm movements and grip forces. A tight predictive coupling between the modulation of grip forces and rapid arm movements has been reported in neurologically intact subjects [116]. The factor analysis suggests that this predictive coupling might possibly be disrupted in neurologically affected subjects, potentially due to altered sensory feedback (e.g., proprioception) leading to inaccurate predictive internal models or abnormal neural transmission (e.g., corticospinal tract integrity) [47, 50]. Reduced corticospinal tract integrity can also lead to weakness and could affect movement speed, as described by factor four (*velocity max. return*) [47]. This factor might further be influenced by an altered inhibition of the supraspinal pathways, often resulting from upper motor neuron lesions, leading to increased muscle tone and thereby altered movement speed [65]. Lastly, the fifth factor covered grip force coordination during *hole approach*, thereby diverging from the coordination of grip forces during gross movements (*transport*) as described by factor 3 and focusing more on grip force coordination during precise position adjustments. This suggests that the two phases are differently controlled, potentially because the *hole approach* is more dominated by sensory and cognitive feedback loops guiding the precise insertion of the peg, whereas gross movements (*transport*) are more dominated through feedforward mechanisms [50]. Also, the physiological control origin of the two movement phases might differ, as gross movements are expected to be orchestrated by the reticulospinal tract, whereas precise control are more linked to the corticospinal tract [117].

Even though the *task completion time* did not pass the selection procedure due to strong learning effects, one might still consider to report this metric when using the VPIT in a cross-sectional manner as its intuitive interpretation allows to give an insightful first indication about the overall level of impairment that might potentially be interesting for both clinical personnel and the tested patient.

The added clinical value of the VPIT core metrics compared to existing conventional assessments is visible in Figure 4 and 5, as the former allowed to detect sensorimotor impairments in certain subjects that did not show any deficits according to the typically used conventional scales. Such a sensitive identification of sensorimotor impairments might allow to provide evidence for the potential of additional neurorehabilitation. Further, the identified core set of metrics can efficiently inform on multiple impairments, both sensory and motor, in arm and hand with a single task that can typically be performed within 15 minutes. This advances the state-of-the-art that mainly focused on the evaluation of arm movements [14, 57, 118], or required more complex or time-consuming measurement setups (e.g., optical motion capture) to quantify arm and hand movements while also neglecting grasping function [119]. Such a fine-grained evaluation covering multiple sensorimotor impairments can help

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to stratify subjects into homogeneous groups with low inter-subject variability. This is important to reduce the required number of subjects to demonstrate significant effects of novel therapies in clinical trials [14].

4.3 Limitations and future directions

The proposed methodology should be considered in light of certain limitations. Most importantly, the framework was especially designed for metrics aimed at longitudinally monitoring impairments and might need additional refinement when transferring it to other healthcare applications (e.g., screening of electronic health record data) with different clinical requirements or data types. Additionally, the definition of multiple cut-off values for the metric selection process influences the final core set of metrics. Even though most of the cut-offs were based on accepted definitions from the research community (e.g., COSMIN guidelines), we acknowledge that the optimality of these values needs to be further validated from a clinical point of view. To evaluate measurement error and learning effects, novel cut-offs were introduced based on the distribution of observed values for the VPIT with the goal to exclude metrics that showed highest measurement error and strongest learning effects. It is important to note that this only considers the relative and not the absolute level of measurement error. However, this can only be adequately judged using data recorded pre- and post-intervention, allowing to compare the measurement error (SRD%) to intervention-induced physiological changes (minimal important clinical difference) [22]. Hence, the rather high absolute level of observed measurement errors for the VPIT (up to 57.7% of the range of observed values) warrants further critical evaluation with longitudinal data. Also, it is important to note that, even though certain VPIT-based metrics did not pass the selection procedure, they might still prove to be valid and reliable for other assessment tasks and platforms, or more specific subject populations. In this context, it should be stressed that test-retest reliability, measurement error, and learning effects for the VPIT were evaluated with neurologically intact subjects and might require additional investigation in neurological populations.

5 Conclusions

We proposed a data-driven framework for selecting and validating digital health metrics based on the targeted impairments, the influence of participant demographics, and their clinimetric properties. In a use-case with the VPIT, the methodology enabled the selection and validation of a core set of ten kinematic and kinetic metrics out of 77 initially proposed metrics. The chosen metrics were able to accurately describe the severity of multiple sensorimotor impairments in a cross-sectional manner and have high potential to sensitively monitor neurorehabilitation and to individualize interventions. Additionally, an in-depth physiological motivation of these metrics and the interpretation based on an exploratory factor analysis allowed to better understand their relation to the

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

targeted impairments. Hence, this work makes an important contribution to implement digital health metrics as complementary endpoints for clinical trials and routine, next to the still more established conventional scales and patient reported metrics [120]. We urge researchers and clinicians to capitalize on the promising properties of digital health metrics and further contribute to their validation and acceptance, which in the long-term will lead to a more thorough understanding of disease mechanisms and enable novel applications (e.g., personalized predictions of rehabilitation outcomes) with the potential to improve healthcare quality.

Ethics approval and consent to participate

All experimental procedure were approved by the responsible local Ethic Committees (approval numbers in SM) and all subjects gave written informed consent prior to participation in the experiments.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Funding

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 688857 (SoftPro), from the Swiss State Secretariat for Education, Research and Innovation (15.0283-1), from the P&K Puhringer Foundation, by the James S. McDonnell Foundation (90043345, 220020220), and by the Canadian Institutes of Health Research in partnership with the Fondation de l'Ataxie Charlevoix-Saguenay (Emerging Team Grant TR2-119189). CG holds a career-grant-funding from Fonds de recherche en santé du Québec (22193, 31011).

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Contributions

Study design: CK, AS, IL, CG, JH, PF, ARL, RG, OL. Data collection: CMK, AS, JH, CG, IL, OL. Data analysis: CMK. Data interpretation: CMK, MDR, RG, OL. Manuscript writing: CMK, RG, OL. Manuscript review: MDR, AS, IL, JH, PF, RG, OL. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Marie-Christine Fluet, Sascha Motazedi Tabrizi, Werner Popp, Joachim Cerny, Isabelle Lessard, Caroline Lavoie, and Meret Branscheidt for help during data collection and the insightful discussions.

References

[1] World Health Organization: International Classification of Functioning, Disability and Health: ICF. World Health Organization, Geneva (2001)

[2] Lawrence, E.S., Coshall, C., Dundas, R., Stewart, J., Rudd, a.G., Howard, R., Wolfe, C.D.: Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. Stroke - A Journal of Cerebral Circulation **32**(6), 1279–1284 (2001)

[3] Kister, I., Bacon, T.E., Chamot, E., Salter, A.R., Cutter, G.R., Kalina, J.T., Herbert, J.: Natural history of multiple sclerosis symptoms. International Journal of MS Care **15**(3), 146–158 (2013)

[4] Gagnon, C., Desrosiers, J., Mathieu, J.: Autosomal recessive spastic ataxia of charlevoix-saguenay: upper extremity aptitudes, functional independence and social participation. International Journal of Rehabilitation Research **27**(3), 253–256 (2004)

[5] Yozbatiran, N., Baskurt, F., Baskurt, Z., Ozakbas, S., Idiman, E.: Motor assessment of upper extremity function and its relation with fatigue, cognitive function and quality of life in multiple sclerosis patients. Journal of the Neurological Sciences **246**(1-2), 117–122 (2006)

[6] Lamers, I., Kelchtermans, S., Baert, I., Feys, P.: Upper limb assessment in multiple sclerosis: A systematic review of outcome measures and their psychometric properties. Archives of Physical Medicine and Rehabilitation **95**(6), 1184–1200 (2014)

[7] Santisteban, L., Térémetz, M., Bleton, J.-P., Baron, J.-C., Maier, M.A., Lindberg, P.G.: Upper Limb Outcome Measures Used in Stroke Rehabilitation Studies: A Systematic Literature Review. Plos One **11**(5), 1932–6203 (2016)

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[8] Burridge, J., Alt Murphy, M., Buurke, J., Feys, P., Keller, T., Klamroth-Marganska, V., Lamers, I., McNicholas, L., Prange, G., Tarkka, I., Timmermans, A., Hughes, A.-M.: A Systematic Review of International Clinical Guidelines for Rehabilitation of People With Neurological Conditions: What Recommendations Are Made for Upper Limb Assessment? Frontiers in Neurology **10**(June), 1–14 (2019). doi:10.3389/fneur.2019.00567

[9] Gladstone, D.J., Danells, C.J., Black, S.E.: The Fugl-Meyer Assessment of Motor Recovery after Stroke: A Critical Review of Its Measurement Properties. Neurorehabilitation and Neural Repair **16**(3), 232–240 (2002)

[10] Chen, H.-M., Chen, C.C., Hsueh, I.-P., Huang, S.-L., Hsieh, C.-L.: Test-Retest Reproducibility and Smallest Real Difference of 5 Hand Function Tests in Patients With Stroke. Neurorehabilitation and Neural Repair **23**(5), 435–440 (2009)

[11] Hawe, R.L., Scott, S.H., Dukelow, S.P.: Taking Proportional Out of Stroke Recovery. Stroke **50**(1), 204–211 (2018)

[12] Hope, T.M.H., Friston, K., Price, C.J., Leff, A.P., Rotshtein, P., Bowman, H.: Recovery after stroke: not so proportional after all? Brain **142**(1), 15–22 (2019). doi:10.1093/brain/awy302

[13] Schwarz, A., Kanzler, C.M., Lambercy, O., Luft, A.R., Veerbeek, J.M.: Systematic review on kinematic assessments of upper limb movements after stroke. Stroke **50**(3), 718–727 (2019)

[14] Krebs, H.I., Krams, M., Agrafiotis, D.K., Di Bernardo, A., Chavez, J.C., Littman, G.S., Yang, E., Byttebier, G., Dipietro, L., Rykman, A., McArthur, K., Hajjar, K., Lees, K.R., Volpe, B.T.: Robotic measurement of arm movements after stroke establishes biomarkers of motor recovery. Stroke **45**(1), 200–204 (2014)

[15] Shull, P.B., Jirattigalachote, W., Hunt, M.A., Cutkosky, M.R., Delp, S.L.: Quantified self and human movement: A review on the clinical impact of wearable sensing and feedback for gait analysis and intervention. Gait and Posture **40**(1), 11–19 (2014)

[16] Eskofier, B., Lee, S., Baron, M., Simon, A., Martindale, C., Gaßner, H., Klucken, J.: An Overview of Smart Shoes in the Internet of Health Things: Gait and Mobility Assessment in Health Promotion and Disease Monitoring. Applied Sciences **7**(10), 986 (2017)

[17] Kwakkel, G., Lannin, N.A., Borschmann, K., English, C., Ali, M., Churilov, L., Saposnik, G., Winstein, C., van Wegen, E.E.H., Wolf, S.L., Krakauer, J.W., Bernhardt, J.: Standardized Measurement of Sensorimotor Recovery in Stroke Trials: Consensus-Based Core Recommendations from the Stroke Recovery and Rehabilitation Roundtable. Neurorehabilitation and Neural Repair **31**(9), 784–792 (2017)

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[18] Mathews, S.C., McShea, M.J., Hanley, C.L., Ravitz, A., Labrique, A.B., Cohen, A.B.: Digital health: a path to validation. NPJ digital medicine **2**(1), 1–9 (2019)

[19] Shirota, C., Balasubramanian, S., Melendez-Calderon, A.: Technology-aided assessments of sensorimotor function: current use, barriers and future directions in the view of different stakeholders. Journal of NeuroEngineering and Rehabilitation **16**(1), 53 (2019)

[20] Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G., King, D.: Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine **17**(1), 195 (2019)

[21] Tran, V.D., Dario, P., Mazzoleni, S.: Kinematic measures for upper limb robot-assisted therapy following stroke and correlations with clinical outcome measures: A review. Medical Engineering & Physics **53**, 13–31 (2018)

[22] Prinsen, C.A.C., Mokkink, L.B., Bouter, L.M., Alonso, J., Patrick, D.L., de Vet, H.C.W., Terwee, C.B.: COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research **27**(5), 1147–1157 (2018)

[23] Shishov, N., Melzer, I., Bar-Haim, S.: Parameters and Measures in Assessment of Motor Learning in Neurorehabilitation; A Systematic Review of the Literature. Frontiers in Human Neuroscience **11**(1) (2017)

[24] Saeys, Y., Inza, I., Larrañaga, P.: A review of feature selection techniques in bioinformatics. Bioinformatics **23**(19), 2507–2517 (2007)

[25] Williamson, P.R., Altman, D.G., Blazeby, J.M., Clarke, M., Devane, D., Gargon, E., Tugwell, P.: Developing core outcome sets for clinical trials: Issues to consider. Trials **13**, 1–8 (2012)

[26] Boers, M., Kirwan, J.R., Wells, G., Beaton, D., Gossec, L., D'Agostino, M.A., Conaghan, P.G., Bingham, C.O., Brooks, P., Landewé, R., March, L., Simon, L.S., Singh, J.A., Strand, V., Tugwell, P.: Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. Journal of Clinical Epidemiology **67**(7), 745–753 (2014)

[27] Kirkham, J.J., Davis, K., Altman, D.G., Blazeby, J.M., Clarke, M., Tunis, S., Williamson, P.R.: Core Outcome Set-STAndards for Development: The COS-STAD recommendations. PLoS Medicine **14**(11), 1–10 (2017)

[28] Kwakkel, G., Wegen, E.V., Burridge, J.H., Winstein, C., van Dokkum, L., Murphy, M.A., Levin, M.F., Krakauer, J.: Standardized measurement of quality of upper limb movement after stroke: Consensus-based core recommendations from the Second Stroke Recovery and Rehabilitation Roundtable. in press (2019)

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[29] Fluet, M., Lambercy, O., Gassert, R.: Upper limb assessment using a Virtual Peg Insertion Test. In: Proceedings of the IEEE International Conference on Rehabilitation Robotics (ICORR), pp. 1–6 (2011)

[30] Lambercy, O., Fluet, M.-C., Lamers, I., Kerkhofs, L., Feys, P., Gassert, R.: Assessment of upper limb motor function in patients with multiple sclerosis using the virtual peg insertion test: a pilot study. In: Proceedings of the IEEE International Conference on Rehabilitation Robotics (ICORR), pp. 1–6 (2013)

[31] Hofmann, P., Held, J.P., Gassert, R., Lambercy, O.: Assessment of movement patterns in stroke patients: A case study with the virtual peg insertion test. In: Proceedings of the International Convention on Rehabilitation Engineering & Assistive Technology (i-CREATe). 2016, pp. 14–14 (2016)

[32] Tobler-Ammann, B.C., de Bruin, E.D., Fluet, M.-C., Lambercy, O., de Bie, R.A., Knols, R.H.: Concurrent validity and test-retest reliability of the Virtual Peg Insertion Test to quantify upper limb function in patients with chronic stroke. Journal of NeuroEngineering and Rehabilitation **13**(1), 8 (2016)

[33] Kanzler, C.M., Gomez, S.M., Rinderknecht, M.D., Gassert, R., Lambercy, O.: Influence of arm weight support on a robotic assessment of upper limb function. In: Proceedings of the 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob), pp. 1–6 (2018)

[34] Kanzler, C.M., Catalano, M.G., Piazza, C., Bicchi, A., Gassert, R., Lambercy, O.: An Objective Functional Evaluation of Myoelectrically-Controlled Hand Prostheses: A Pilot Study Using the Virtual Peg Insertion Test. In: IEEE 16th International Conference on Rehabilitation Robotics (ICORR), pp. 392–397 (2019)

[35] Car, J., Sheikh, A., Wicks, P., Williams, M.S.: Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom. BMC Medicine **17**(1), 143 (2019)

[36] Subramanian, S.K., Yamanaka, J., Chilingaryan, G., Levin, M.F.: Validity of movement pattern kinematics as measures of arm motor impairment poststroke. Stroke **41**(10), 2303–2308 (2010)

[37] Rinderknecht, M.D., Lambercy, O., Raible, V., Liepert, J., Gassert, R.: Age-based model for metacarpophalangeal joint proprioception in elderly. Clinical Interventions in Aging **12**, 635–643 (2017)

[38] Prinsen, C.A.C., Vohra, S., Rose, M.R., Boers, M., Tugwell, P., Clarke, M., Williamson, P.R., Terwee, C.B.: How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. Trials **17**(1), 1–10 (2016)

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[39] Mathiowetz, V., Weber, K., Kashman, N., Volland, G.: Adult norms for the nine hole peg test of finger dexterity. The Occupational Therapy Journal of Research **5**(1), 24–38 (1985)

[40] Mathiowetz, V., Volland, G., Kashman, N., Weber, K.: Adult norms for the box and block test of manual dexterity. American Journal of Occupational Therapy **39**(6), 386–391 (1985)

[41] Gagnon, C., Lavoie, C., Lessard, I., Mathieu, J., Brais, B., Bouchard, J.P., Fluet, M.C., Gassert, R., Lambercy, O.: The Virtual Peg Insertion Test as an assessment of upper limb coordination in ARSACS patients: A pilot study. Journal of the Neurological Sciences **347**(1-2), 341–344 (2014)

[42] Feys, P., Coninx, K., Kerkhofs, L., Weyer, T.D., Truyens, V., Maris, A., Lamers, I.: Robot-supported upper limb training in a virtual learning environment: a pilot randomized controlled trial in persons with MS. Journal of NeuroEngineering and Rehabilitation **12**(1), 1–12 (2015)

[43] Lamers, I., Raats, J., Spaas, J., Meuleman, M., Kerkhofs, L., Schouteden, S., Feys, P.: Intensity-dependent clinical effects of an individualized technology-supported task-oriented upper limb training program in Multiple Sclerosis: A pilot randomized controlled trial. Multiple Sclerosis and Related Disorders **34**(November 2018), 119–127 (2019)

[44] Lang, J.I., Lang, T.J.: Eye screening with the lang stereotest. American Orthoptic Journal **38**(1), 48–50 (1988)

[45] Lang, C.E., Bland, M.D., Bailey, R.R., Schaefer, S.Y., Birkenmeier, R.L.: Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making. Journal of Hand Therapy **26**(2), 104–115 (2013)

[46] Johansson, G.M., Häger, C.K.: A modified standardized nine hole peg test for valid and reliable kinematic assessment of dexterity post-stroke. Journal of NeuroEngineering and Rehabilitation, 1–11 (2019)

[47] Sathian, K., Buxbaum, L.J., Cohen, L.G., Krakauer, J.W., Lang, C.E., Corbetta, M., Fitzpatrick, S.M.: Neurological Principles and Rehabilitation of Action Disorders: Common Clinical Deficits. Neurorehabilitation and Neural Repair **25**(5), 21–32 (2011)

[48] Frey, S.H., Fogassi, L., Grafton, S., Picard, N., Rothwell, J.C., Schweighofer, N., Corbetta, M., Fitzpatrick, S.M.: Neurological Principles and Rehabilitation of Action Disorders: Computation, Anatomy, and Physiology (CAP) Model. Neurorehabilitation and Neural Repair **25**(5), 6–20 (2011)

[49] Nordin, N., Xie, S.Q., Wünsche, B., Wunsche, B.: Assessment of movement quality in robot- assisted upper limb rehabilitation after stroke: a review. Journal of NeuroEngineering and Rehabilitation **11**(1), 137 (2014)

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[50] Scott, S.H.: Optimal feedback control and the neural basis of volitional motor control. Nature Reviews Neuroscience **5**(7), 532–546 (2004)

[51] Harris, C.M., Wolpert, D.M.: Signal-dependent noise determines motor planning. Nature **394**(6695), 780–4 (1998)

[52] Flash, T., Hogan, N.: The coordination of arm movements: an experimentally confirmed mathematical model. The Journal of Neuroscience **5**(7), 1688–1703 (1985)

[53] Rohrer, B., Fasoli, S., Krebs, H.I., Hughes, R., Volpe, B., Frontera, W.R., Stein, J., Hogan, N.: Movement smoothness changes during stroke recovery. The Journal of Neuroscience **22**(18), 8297–8304 (2002)

[54] Pellegrino, L., Coscia, M., Muller, M., Solaro, C., Casadio, M.: Evaluating upper limb impairments in multiple sclerosis by exposure to different mechanical environments. Scientific Reports **8**(1), 2110 (2018)

[55] Balasubramanian, S., Melendez-Calderon, A., Burdet, E.: A robust and sensitive metric for quantifying movement smoothness. IEEE Transactions on Biomedical Engineering **59**(8), 2126–2136 (2012)

[56] Balasubramanian, S., Melendez-Calderon, A., Roby-Brami, A., Burdet, E.: On the analysis of movement smoothness. Journal of NeuroEngineering and Rehabilitation **12**(1), 112 (2015)

[57] Coderre, A.M., Amr Abou Zeid, Dukelow, S.P., Demmer, M.J., Moore, K.D., Demers, M.J., Bretzke, H., Herter, T.M., Glasgow, J.I., Norman, K.E., Bagg, S.D., Scott, S.H.: Assessment of Upper-Limb Sensorimotor Function of Subacute Stroke Patients Using Visually Guided Reaching. Neurorehabilitation and Neural Repair **24**(6), 528–541 (2010)

[58] de Graaf, J.B., Sittig, A.C., Denier van der Gon, J.J.: Misdirections in slow goal-directed arm movements and pointer-setting tasks. Experimental Brain Research **84**(2), 434–8 (1991)

[59] Cirstea, M.C., Levin, M.F.: Compensatory strategies for reaching in stroke. Brain **123**(5), 940–953 (2000)

[60] Otaka, E., Otaka, Y., Kasuga, S., Nishimoto, A., Yamazaki, K., Kawakami, M., Ushiba, J., Liu, M.: Clinical usefulness and validity of robotic measures of reaching movement in hemiparetic stroke patients. Journal of NeuroEngineering and Rehabilitation **12**(1), 66 (2015)

[61] Reinkensmeyer, D.J., Iobbi, M.G., Kahn, L.E., Kamper, D.G., Takahashi, C.D.: Modeling reaching impairment after stroke using a population vector model of movement control that incorporates neural firing-rate variability. Neural Computation **15**(11), 2619–2642 (2003)

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[62] Mottet, D., Van Dokkum, L.E.H., Froger, J., Gouaïch, A., Laffont, I.: Trajectory formation principles are the same after mild or moderate stroke. PLoS ONE **12**(3), 1–17 (2017)

[63] Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology **47**(6), 381 (1954)

[64] Galea, J.M., Miall, R.C.: Concurrent adaptation to opposing visual displacements during an alternating movement. Experimental Brain Research **175**(4), 676–688 (2006)

[65] Mukherjee, A., Chakravarty, A.: Spasticity Mechanisms – for the Clinician. Frontiers in Neurology **1**(December), 1–10 (2010)

[66] Kurtzke, J.F.: Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology **33**(11), 1444–1452 (1983)

[67] Fahn, S., Tolosa, E., Marín, C.: Clinical rating scale for tremor. Parkinson's disease and movement disorders **2**, 271–280 (1993)

[68] Kim, J.S.: Delayed onset mixed involuntary movements after thalamic stroke Clinical, radiological and pathophysiological findings. Brain **124**(2), 299–309 (2001)

[69] Alusi, S.H., Worthington, J., Glickman, S., Bain, P.G.: A study of tremor in multiple sclerosis. Brain **124**(Pt 4), 720–730 (2001)

[70] Manto, M.: Mechanisms of human cerebellar dysmetria: Experimental evidence and current conceptual bases. Journal of NeuroEngineering and Rehabilitation **6**(1), 1–18 (2009)

[71] Carpinella, I., Cattaneo, D., Ferrarin, M.: Quantitative assessment of upper limb motor function in Multiple Sclerosis using an instrumented Action Research Arm Test. Journal of NeuroEngineering and Rehabilitation **11**(1), 1–16 (2014)

[72] Bardorfer, A., Munih, M., Zupan, A., Primožič, A.: Upper limb motion analysis using haptic interface. IEEE/ASME Transactions on Mechatronics **6**(3), 253–260 (2001). doi:10.1109/3516.951363

[73] Beer, R.F., Given, J.D., Dewald, J.P.A.: Task-dependent weakness at the elbow in patients with hemiparesis. Archives of Physical Medicine and Rehabilitation **80**(7), 766–772 (1999)

[74] Quinn, L., Reilmann, R., Marder, K., Gordon, A.M.: Altered movement trajectories and force control during object transport in Huntington's disease. Movement Disorders **16**(3), 469–480 (2001)

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[75] Forssberg, H., Kinoshita, H., Eliasson, A.C., Johansson, R.S., Westling, G., Gordon, A.M.: Development of human precision grip i: Basic coordination of force. Experimental Brain Research **90**(2), 393–398 (1992)

[76] Hermsdörfer, J., Hagl, E., Nowak, D.A., Marquardt, C.: Grip force control during object manipulation in cerebral stroke. Clinical Neurophysiology **114**(5), 915–929 (2003)

[77] Wenzelburger, R., Kopper, F., Frenzel, A., Stolze, H., Klebe, S., Brossmann, A., Kuhtz-Buschbeck, J., Gölge, M., Illert, M., Deuschl, G.: Hand coordination following capsular stroke. Brain **128**(1), 64–74 (2005)

[78] Lindberg, P.G., Roche, N., Robertson, J., Roby-Brami, A., Bussel, B., Maier, M.A.: Affected and unaffected quantitative aspects of grip force control in hemiparetic patients after stroke. Brain Research **1452**, 96–107 (2012)

[79] Allgöwer, K., Hermsdörfer, J.: Fine motor skills predict performance in the Jebsen Taylor Hand Function Test after stroke. Clinical Neurophysiology **128**(10), 1858–1871 (2017)

[80] Iyengar, V., Santos, M.J., Ko, M., Aruin, A.S.: Grip Force Control in Individuals With Multiple Sclerosis. Neurorehabilitation and Neural Repair **23**(8), 855–861 (2009)

[81] Gordon, A., Duff, S.: Fingertip forces during object manipulation in children with hemiplegic cerebral palsy, I: anticipatory scaling. Developmental Medicine and Child Neurology **33**(3), 225–231 (1991)

[82] Lan, Y., Yao, J., Dewald, J.P.A.: The Impact of Shoulder Abduction Loading on Volitional Hand Opening and Grasping in Chronic Hemiparetic Stroke. Neurorehabilitation and Neural Repair **31**(6), 521–529 (2017)

[83] Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.-S.S.: Generalized linear mixed models: a practical guide for ecology and evolution. Trends in ecology & evolution **24**(3), 127–135 (2009)

[84] Fluet, M.C., Lambercy, O., Gassert, R.: Effects of 2D/3D visual feedback and visuomotor collocation on motor performance in a virtual peg insertion test. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), pp. 4776–4779 (2012)

[85] Gerig, N., Mayo, J., Baur, K., Wittmann, F., Riener, R., Wolf, P.: Missing depth cues in virtual reality limit performance and quality of three dimensional reaching movements. PLoS ONE **13**(1), 1–18 (2018)

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[86] Box, G.E., Cox, D.R.: An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological), 211–252 (1964)

[87] Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L.: Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology **49**(4), 764–766 (2013)

[88] Andersen, L.M.: Obtaining reliable likelihood ratio tests from simulated likelihood functions. PLoS ONE **9**(10) (2014)

[89] Roy, K., Das, R.N., Ambure, P., Aher, R.B.: Be aware of error measures. Further studies on validation of predictive QSAR models. Chemometrics and Intelligent Laboratory Systems **152**, 18–33 (2016)

[90] Hamilton, D.F., Ghert, M., Simpson, A.H.R.W.: Interpreting regression models in clinical outcome studies. Bone & Joint Research **4**(9), 152–153 (2015)

[91] Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X.: Applied Logistic Regression. John Wiley & Sons, New Jersey (2013)

[92] Lexell, J.E., Downham, D.Y.: How to Assess the Reliability of Measurements in Rehabilitation. American Journal of Physical Medicine & Rehabilitation **84**(September), 719–723 (2005)

[93] de Vet, H.C.W., Terwee, C.B., Knol, D.L., Bouter, L.M.: When to use agreement versus reliability measures. Journal of Clinical Epidemiology **59**(10), 1033–1039 (2006)

[94] Beckerman, H., Roebroeck, M.E., Lankhorst, G.J., Becher, J.G., Bezemer, P.D., Verbeek, A.L.M.: Smallest real difference, a link between reproducibility and responsiveness. Quality of Life Research **10**(7), 571–578 (2001)

[95] Smidt, N., Van der Windt, D.A., Assendelft, W.J., Mourits, A.J., Devill, W.L., De Winter, A.F., Bouter, L.M.: Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. Archives of Physical Medicine and Rehabilitation **83**(8), 1145–1150 (2002)

[96] Baba, K., Shibata, R., Sibuya, M.: Partial correlation and conditional correlation as measures of conditional independence. Australian & New Zealand Journal of Statistics **46**(4), 657–664 (2004)

[97] Kenett, D.Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R.N., Ben-Jacob, E.: Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. PLoS ONE **5**(12), 1–14 (2010)

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[98] Hinkle, D.E., Wiersma, W., Jurs, S.G.: Applied Statistics for the Behavioral Sciences. Houghton Mifflin, Boston (1988)

[99] Costello, A.B., Osborne, J.W.: Best Practices in Exploratory Factor Analysis : Four Recommendations for Getting the Most From Your Analysis. Practical Assessment, Research & Education **10**, 1–9 (2005)

[100] Hayton, J.C., Allen, D.G., Scarpello, V.: Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. Organizational Research Methods **7**(2), 191–205 (2004)

[101] Franklin, S.B., Gibson, D.J., Robertson, P.A., Pohlmann, J.T., Fralish, J.S.: Parallel Analysis: a method for determining significant principal components. Journal of Vegetation Science **6**(1), 99–106 (2006)

[102] Cattell, R.: Factors in Factor Analysis. Psychometrika **30**(2), 179–185 (1965)

[103] Woytowicz, E.J., Rietschel, J.C., Goodman, R.N., Conroy, S.S., Sorkin, J.D., Whitall, J., McCombe Waller, S.: Determining Levels of Upper Extremity Movement Impairment by Applying a Cluster Analysis to the Fugl-Meyer Assessment of the Upper Extremity in Chronic Stroke. Archives of Physical Medicine and Rehabilitation **98**(3), 456–462 (2017)

[104] Hoonhorst, M.H., Nijland, R.H., Van Den Berg, J.S., Emmelot, C.H., Kollen, B.J., Kwakkel, G.: How Do Fugl-Meyer Arm Motor Scores Relate to Dexterity According to the Action Research Arm Test at 6 Months Poststroke? Archives of Physical Medicine and Rehabilitation **96**(10), 1845–1849 (2015)

[105] Kaiser, H.F.: A second generation little jiffy. Psychometrika **35**(4), 401–415 (1970)

[106] Kaiser, H.F.: An index of factorial simplicity. Psychometrika **39**(1), 31–36 (1974)

[107] Kalisch, T., Kattenstroth, J.-C., Kowalewski, R., Tegenthoff, M., Dinse, H.: Age-related changes in the joint position sense of the human hand. Clinical Interventions in Aging **7**, 499 (2012)

[108] Herter, T.M., Scott, S.H., Dukelow, S.P.: Systematic changes in position sense accompany normal aging across adulthood. Journal of NeuroEngineering and Rehabilitation **11**(1), 1–12 (2014)

[109] Tyryshkin, K., Coderre, A.M., Glasgow, J.I., Herter, T.M., Bagg, S.D., Dukelow, S.P., Scott, S.H.: A robotic object hitting task to quantify sensorimotor impairments in participants with stroke. Journal of NeuroEngineering and Rehabilitation **11**(1), 47 (2014)

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[110] Verkuilen, J., Smithson, M.: Mixed and mixture regression models for continuous bounded responses using the beta distribution. Journal of Educational and Behavioral Statistics **37**(1), 82–113 (2011)

[111] Derksen, S., Keselman, H.J.: Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. British Journal of Mathematical and Statistical Psychology **45**(2), 265–282 (1992)

[112] Steyerberg, E.W., Eijkemans, M.J.C., Habbema, J.D.F.: Stepwise selection in small data sets. Journal of Clinical Epidemiology **52**(10), 935–942 (1999)

[113] Harrell, F.E.: Regression Modeling Strategies. Springer Series in Statistics, vol. 27, pp. 83–85. Springer, New York, NY (2001)

[114] Whittingham, M.J., Stephens, P.A., Bradbury, R.B., Freckleton, R.P.: Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology **75**(5), 1182–1189 (2006)

[115] Dukelow, S.P., Herter, T.M., Moore, K.D., Demers, M.J., Glasgow, J.I., Bagg, S.D., Norman, K.E., Scott, S.H.: Quantitative assessment of limb position sense following stroke. Neurorehabilitation and Neural Repair **24**(2), 178–187 (2010)

[116] Flanagan, R.J., Wing, A.M.: Modulation of grip force with load force during point-to-point arm movements. Experimental Brain Research **95**(1), 301–324 (1993)

[117] Baker, S.N.: The primate reticulospinal tract, hand function and functional recovery. Journal of Physiology **589**(23), 5603–5612 (2011)

[118] Colombo, R., Pisano, F., Micera, S., Mazzone, A., Delconte, C., Carrozza, M.C., Dario, P., Minuco, G.: Assessing Mechanisms of Recovery During Robot-Aided Neurorehabilitation of the Upper Limb. Neurorehabilitation and Neural Repair **22**(1), 50–63 (2008)

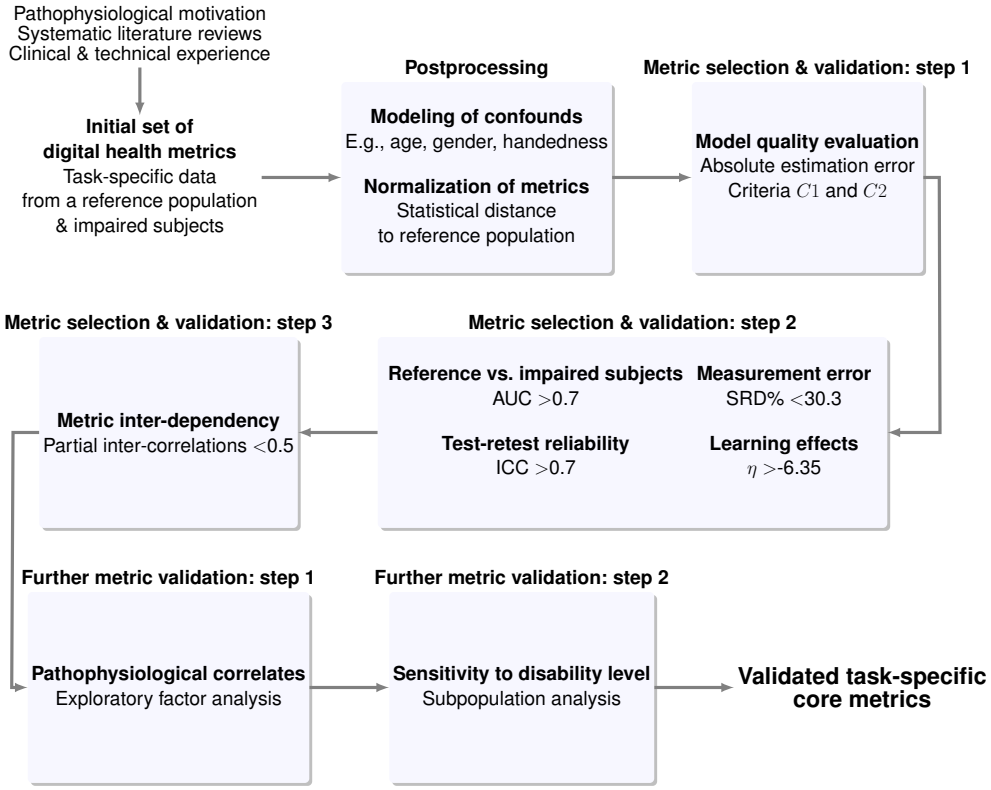
[119] Alt Murphy, M., Willén, C., Sunnerhagen, K.S.: Movement kinematics during a drinking task are associated with the activity capacity level after stroke. Neurorehabilitation and Neural Repair **26**(9), 1106–1115 (2012)

[120] Lamers, I., Feys, P.: Patient reported outcome measures of upper limb function in multiple sclerosis: A critical overview. SAGE Publications Sage UK: London, England (2018)

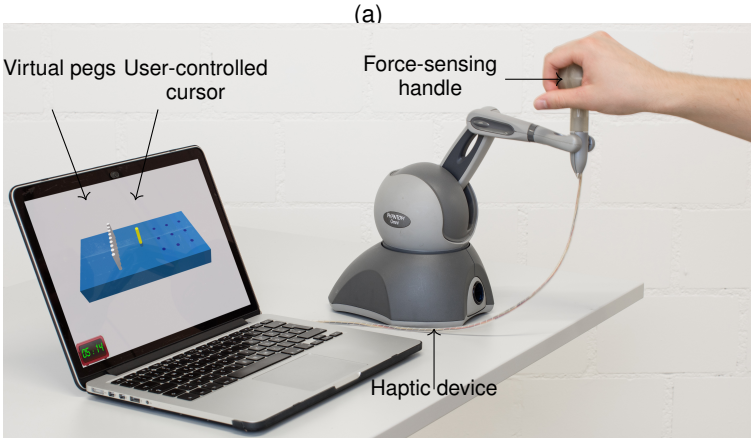
bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



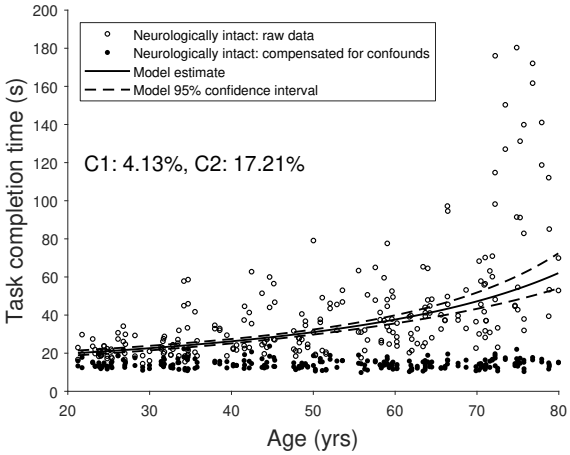
(a)



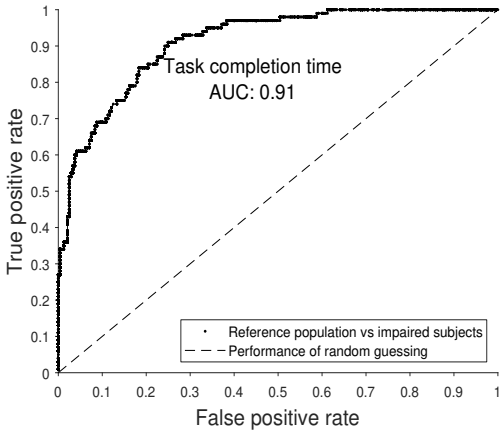
(b)

Figure 1: **Overview of the data-driven framework and the Virtual Peg Insertion Test (VPIT).** (a) The frameworks allows to select a core set of validated digital health metrics. Criteria $C1$ and $C2$ defining model quality; ROC: receiver operating characteristics; AUC: area under curve; ICC: intra-class correlation; SRD%: smallest real difference; η strength of learning effects; (b) as a use-case, the framework was applied to data recorded with the VPIT, a sensor-based upper limb sensorimotor assessment requiring the coordination of arm and hand movements as well as grip forces. The test combines a commercial haptic device, a handle instrumented with force sensors, and a virtual pegboard.

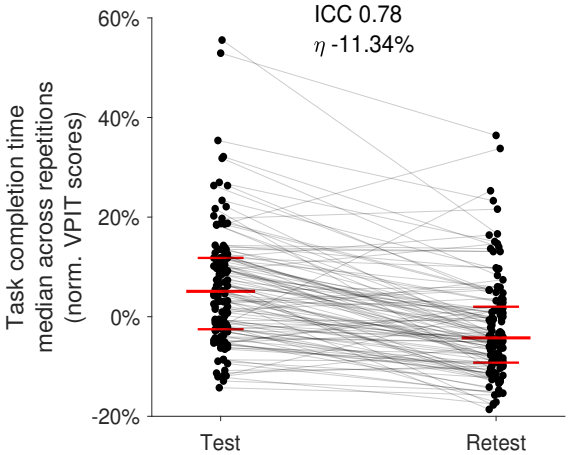
bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



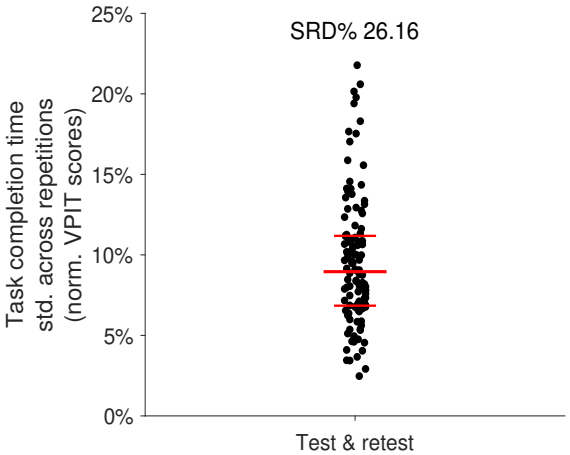
(a) Modeling of confounds



(b) Discriminant validity



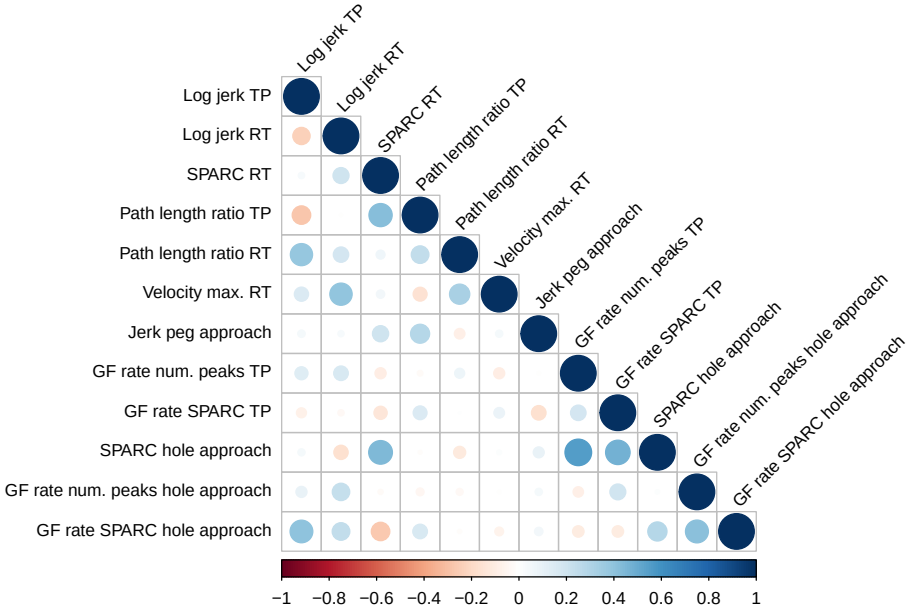
(c) Test-retest reliability & learning effect



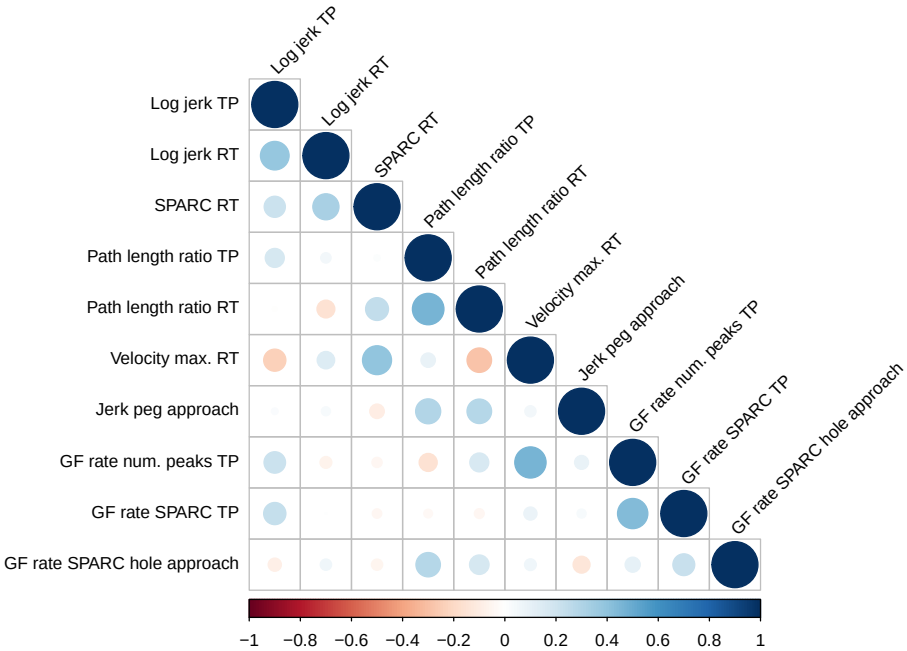
(d) Intra-subject variability (measurement error)

Figure 2: **Data-driven selection and validation of metrics: example of task completion time**. **(a)** the influence of age, sex, tested body side, handedness, and stereo vision deficits on each digital health metrics was removed using data from neurologically intact subjects and mixed effect models (model quality criteria $C1$ and $C2$). Models were fitted in a Box-Cox-transformed space and back-transformed for visualization. Metrics with low model quality ($C1$ >15% or $C2$ >25%) were removed. **(b)** The ability of a metric to discriminate between neurologically intact and affected subjects (discriminant validity) was evaluated using the area under the curve value (AUC). Metrics with AUC<0.7 were removed. **(c)** Test-retest reliability was evaluated using the intra-class correlation coefficient (ICC) indicating the ability of a metric to discriminate between subjects across testing days. Metrics with ICC<0.7 were removed. Additionally, metrics with strong learning effects (η >-6.35) were removed. The long horizontal red line indicates the median, whereas the short ones represent the 25th- and 75th- percentile. **(d)** Measurement error was defined using the smallest real difference (SRD%), indicating a range of values for that the assessment cannot discriminate between measurement error and physiological changes. The distribution of the intra-subject variability was visualized, as it strongly influences the SRD. Metrics with SRD%>30.3 were removed.

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(a) First iteration



(b) Last iteration

Figure 3: **Partial correlation analysis.** The objective was to remove redundant information. Therefore, partial Spearman correlations were calculated between all combination of metrics while controlling for the potential influence of all other metrics. Pairs of metrics were considered for removal if the correlation was equal or above 0.5 The process was done in an iterative manner and the first and the last iterations are presented.

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

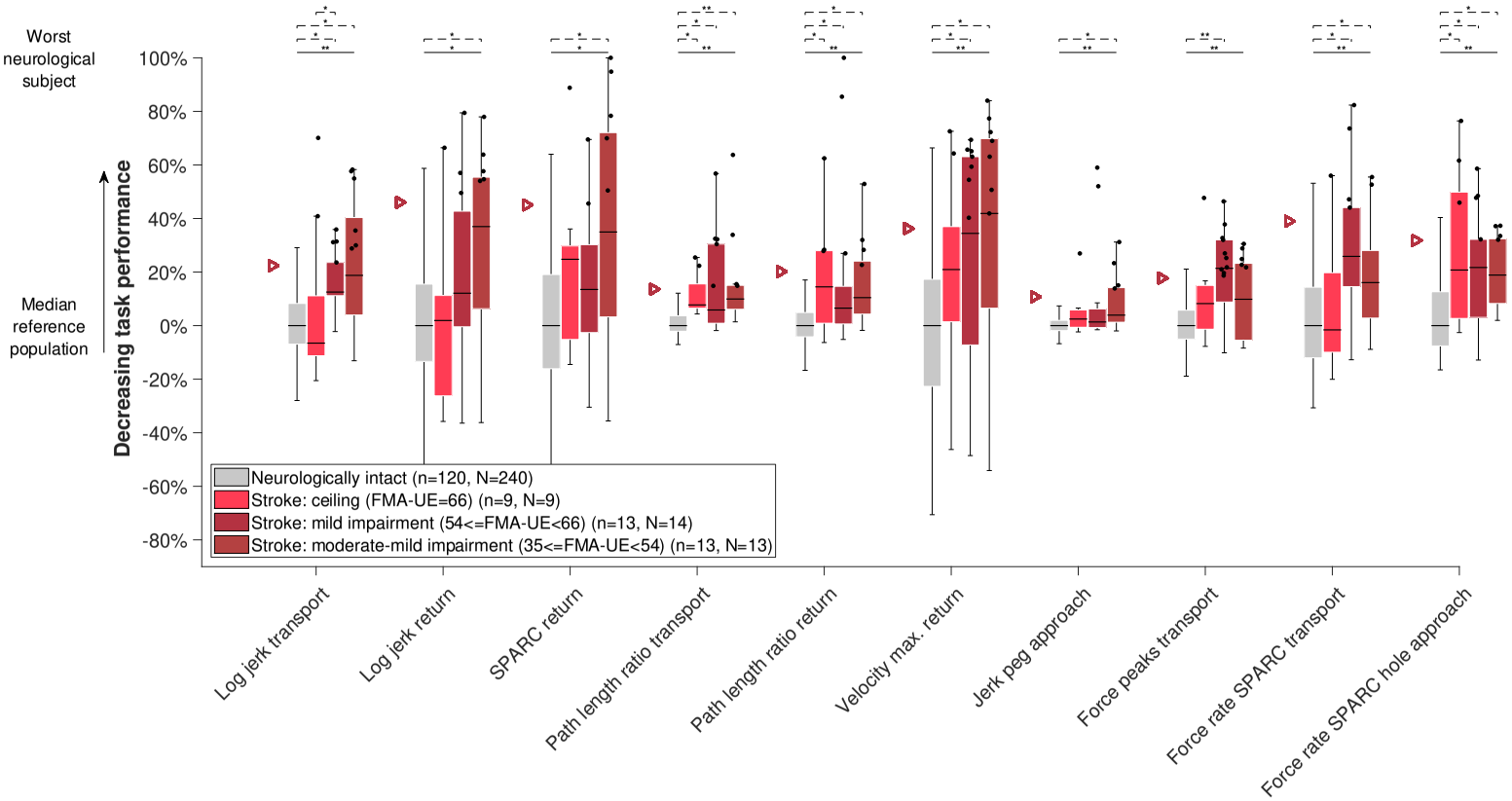


Figure 4: **Sensitivity of metrics to disability severity in stroke subjects.** The subject population was grouped according to their clinical disability level and compared between the subpopulations and to neurologically intact subjects. The vertical axis in the sensorimotor profile indicates task performance based on the distance to the reference population. In the box plots, the median is visualized through the black horizontal line, the interquartile range (IQR) through the boxes, and the minimum and maximum value within 1.5 IQR of the lower and upper quartile, respectively, through the whiskers. Single data points above the 95th-percentile (indicated with triangles) of neurologically intact subjects are defined as showing abnormal behaviour and are represented with black dots. Solid and dashed horizontal black lines above the box plots indicate results of the omnibus and post-hoc statistical tests, respectively. Only significant p-values after Bonferroni correction were visualized (*indicates $p < 0.05$ and **$p < 0.001$). The value n refers to the number of subjects in that group and N to the number of data points. Only subjects with available clinical scores were used for the analysis. For the *jerk peg approach*, one outlier data point was not visualized to maintain a meaningful representation. SPARC: spectral arc length.

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

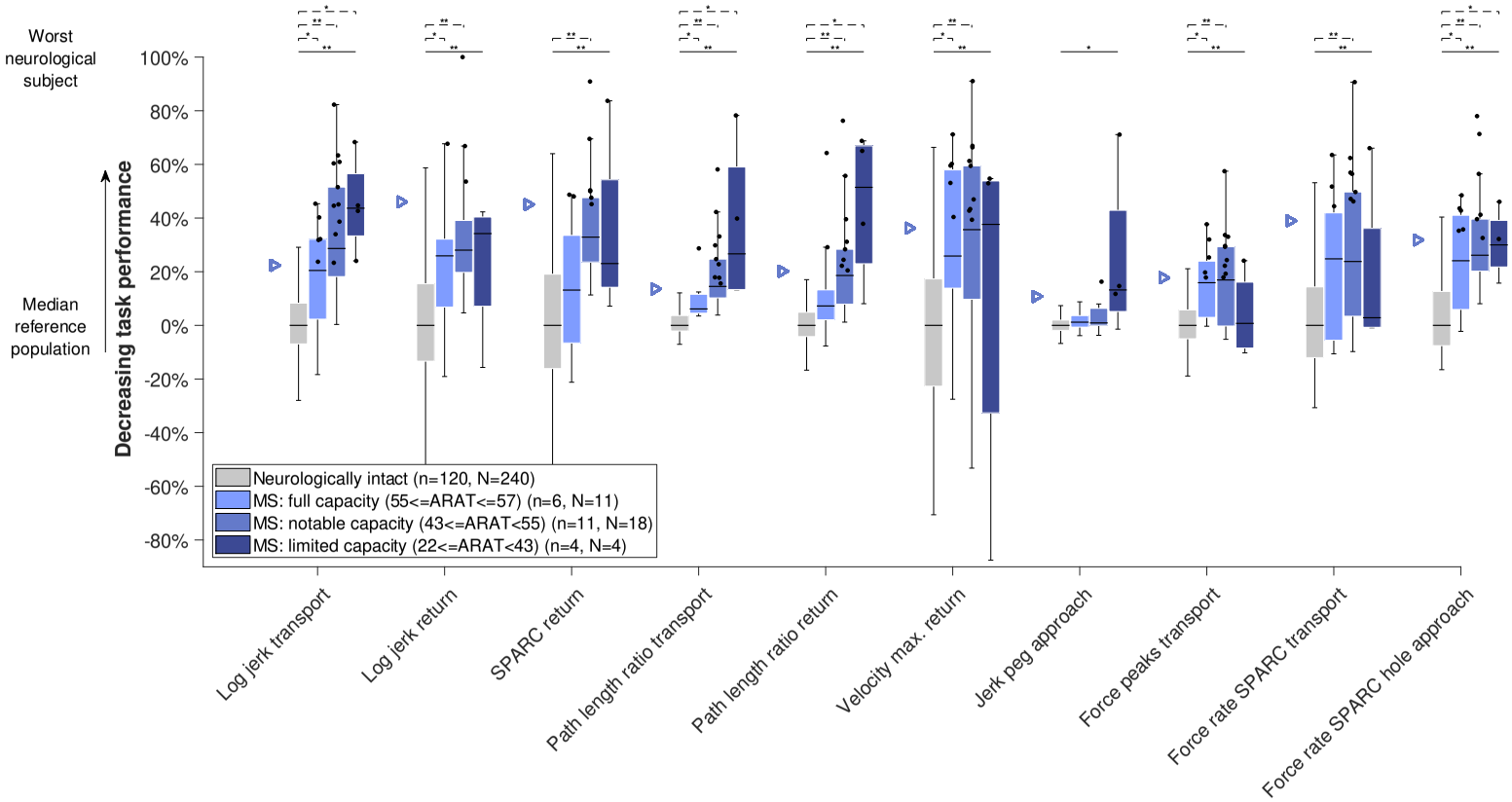


Figure 5: **Sensitivity of metrics to disability severity in MS subjects.** See Figure 4 for a detailed description.

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

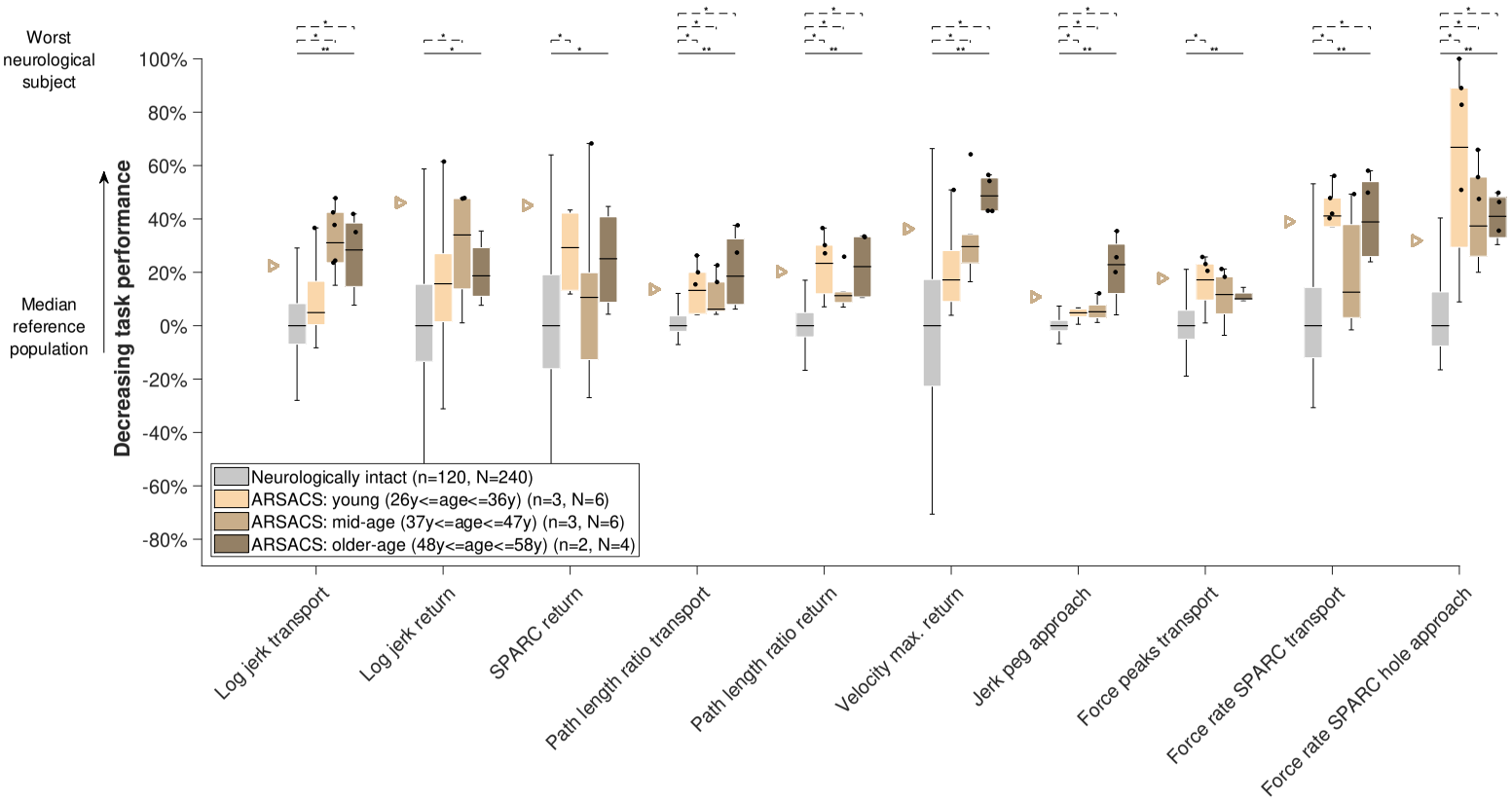


Figure 6: **Sensitivity of metrics to disability severity in ARSACS subjects.** See Figure 4 for a detailed description.

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Supplementary Material: Methods

1.1 Participants

Neurologically intact subjects were recruited at ETH Zurich (Zurich, Switzerland). Stroke patients were tested at the University Hospital of Zurich (Zurich, Switzerland), the cereneo Center for Neurology and Rehabilitation (Vitznau, Switzerland), and the Zentrum für ambulante Rehabilitation (ZAR, Zurich, Switzerland) as part of the Study of Motor Learning and Acute Recovery Time Course in Stroke (SMARTS) or the synergy-based open-source foundations and technologies for prosthetics and rehabilitation (SoftPro). Multiple sclerosis (MS) patients were recruited at Hasselt University (Hasselt, Belgium) and at the Rehabilitation and MS Center Overpelt (Overpelt, Belgium), some of them as part of the individualised technology-supported and robot-assisted virtual learning environment(I-TRAVLE) study. Exclusion criteria involved the inability to lift the arm against gravity, to flex/extend the fingers, and the presence of any concomitant disease affecting the upper limb. The studies involving stroke patients additionally used increased muscle tone, severe sensory deficits, hemorrhagic infarct, traumatic brain injury as exclusion criteria. MS patients had to be diagnosed according to the McDonald criteria. All clinical assessments were performed within the same or few days of the Virtual Peg Insertion Test (VPIT) assessment. Experimental procedures were approved by the local Ethics Committees: neurologically intact subjects subjects: EK2010-N-40; stroke patients: EKNZ-2016-02075, KEK-ZH 2011-0268; MS patients: CME2013/314, ML9521 (S55614), B322201318078; ARSACS patients: 2012-012.

2 Supplementary Material: Results

The metrics that did not fullfil the required quality of the models, according to the C1 and C2 criteria, were *spectral arc length transport*, *number of velocity peaks transport*, *distance to max. velocity transport*, *time to max. velocity transport*, *number of velocity peaks return*, *throughput transport*, *initial movement angle transport* θ_1, *initial movement angle* θ_2, *collision force max. return*, *grip force rate number of peaks buildup*, *grip force rate spectral arc length buildup*, *grip force rate number of peaks release*, and *simulated Gaussian noise*.

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table SM1: **Detailed demographics and clinical information for each body side of each included neurologically impaired subject.**

Disease	Age (yrs)	Sex	Tested side	Affected side	Dominant side	Chronicity (yrs)	FMA-UE (0-66)	ARAT (0-57)	NHPT (s)	EDSS (0-10)
Stroke	67	Male	Right	Left	Right	2.09	66	57	23.25	-
Stroke	55	Male	Left	Left	Right	1.69	54	56	33.25	-
Stroke	55	Male	Right	Left	Right	1.69	66	57	21.85	-
Stroke	55	Male	Left	Right	Right	2.01	65	57	22.82	-
Stroke	55	Male	Right	Right	Right	2.01	49	55	29.28	-
Stroke	52	Male	Left	Left	Right	2.74	55	52	35.36	-
Stroke	52	Male	Right	Left	Right	2.74	65	57	20.99	-
Stroke	73	Male	Left	Right	Right	0.89	62	-	-	-
Stroke	69	Female	Right	Left	Right	0.86	61	57	20.32	-
Stroke	67	Male	Left	Left	Right	2.42	50	-	-	-
Stroke	67	Male	Right	Left	Right	2.42	66	-	-	-
Stroke	40	Female	Left	Right	Right	0.77	56	45	-	-
Stroke	40	Female	Right	Right	Right	0.77	49	49	-	-
Stroke	71	Male	Left	Left	Left	4.49	40	35	196.69	-
Stroke	71	Male	Right	Left	Left	4.49	65	57	15.03	-
Stroke	59	Female	Left	Left	Right	4.35	50	47	17.70	-
Stroke	59	Female	Right	Left	Right	4.35	66	57	12.57	-
Stroke	88	Female	Left	Left	Right	1.65	37	39	42.17	-
Stroke	88	Female	Right	Left	Right	1.65	63	-	14.33	-
Stroke	69	Female	Left	Right	Right	0.58	63	57	19.81	-
Stroke	69	Female	Right	Right	Right	0.58	44	39	49.16	-
Stroke	59	Female	Left	Right	Right	1.94	66	57	21.50	-
Stroke	59	Female	Right	Right	Right	1.94	57	56	21.63	-
Stroke	50	Female	Right	Left	Right	4.83	64	-	-	-
Stroke	61	Male	Left	Right	Right	8.70	66	56	24.51	-
Stroke	61	Male	Right	Right	Right	8.70	38	42	34.95	-
Stroke	59	Male	Left	Left	Right	1.64	46	40	40.84	-
Stroke	59	Male	Right	Left	Right	1.64	63	57	14.85	-
Stroke	69	Male	Left	Left	Right	0.51	53	51	23.08	-
Stroke	69	Male	Right	Left	Right	0.51	63	56	13.67	-
Stroke	55	Male	Left	Left	Right	1.45	59	57	28.08	-
Stroke	55	Male	Right	Left	Right	1.45	66	57	18.50	-
Stroke	42	Male	Left	Left	Right	0.48	39	30	-	-
Stroke	42	Male	Right	Left	Right	0.48	65	57	20.47	-
Stroke	51	Female	Left	Right	Right	0.97	66	57	21.01	-
Stroke	51	Female	Right	Right	Right	0.97	61	57	25.70	-
Stroke	58	Male	Left	Right	Right	0.48	62	57	23.33	-
Stroke	58	Male	Right	Right	Right	0.48	42	53	26.00	-
Stroke	46	Male	Left	Left	Right	1.05	57	42	24.03	-
Stroke	46	Male	Right	Left	Right	1.05	66	57	23.09	-
Stroke	76	Male	Left	Right	Right	2.74	66	55	39.73	-
Stroke	76	Male	Right	Right	Right	2.74	60	54	29.19	-
Stroke	53	Female	Left	Right	Right	2.98	66	57	22.99	-
Stroke	53	Female	Right	Right	Right	2.98	58	55	20.67	-
Stroke	62	Male	Left	Right	Right	14.65	66	57	19.58	-
Stroke	62	Male	Right	Right	Right	14.65	34	33	154.00	-
Stroke	62	Male	Left	Right	-	-	-	57	24.60	-
Stroke	62	Male	Right	Right	-	-	-	43	86.00	-

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Disease	Age (yrs)	Sex	Tested side	Affected side	Dominant side	Chronicity (yrs)	FMA-UE (0-66)	ARAT (0-57)	NHPT (s)	EDSS (0-10)
Stroke	54	Female	Left	Left	Right	1.00	66	57	-	-
Stroke	54	Female	Right	Left	Right	1.00	66	57	-	-
Stroke	67	Male	Left	Left	Right	0.46	66	57	-	-
Stroke	67	Male	Right	Left	Right	0.46	66	57	-	-
Stroke	52	Male	Left	Left	Right	0.23	66	57	-	-
Stroke	52	Male	Right	Left	Right	0.23	66	57	-	-
Stroke	46	Male	Left	Right	Right	0.23	66	57	-	-
Stroke	71	Male	Left	Left	Right	0.23	64	57	-	-
Stroke	71	Male	Right	Left	Right	0.23	66	57	-	-
Stroke	48	Male	Left	Right	Right	0.02	57	57	-	-
Stroke	48	Male	Right	Right	Right	0.02	66	47	-	-
Stroke	45	Female	Right	Left	Right	0.02	66	57	-	-
Stroke	55	Female	Right	Left	Right	0.08	66	57	-	-
Stroke	65	Male	Left	Left	Right	0.23	60	-	-	-
Stroke	65	Male	Right	Left	Right	0.02	62	53	-	-
Stroke	43	Male	Left	Right	Right	0.46	66	57	-	-
Stroke	43	Male	Right	Right	Right	0.46	66	56	-	-
Stroke	41	Female	Right	Left	Right	0.02	64	-	-	-
Stroke	35	Male	Left	Left	Right	0.46	61	57	-	-
Stroke	35	Male	Right	Left	Right	0.02	64	57	-	-
Stroke	76	Male	Right	Left	Left	0.23	66	57	-	-
Stroke	86	Male	Right	Left	Right	0.02	62	56	-	-
Stroke	50	Male	Left	Left	Left	1.00	65	57	-	-
Stroke	49	Male	Right	Left	Left	0.23	66	57	-	-
Stroke	74	Male	Left	Right	Right	0.02	66	57	-	-
Stroke	81	Female	Left	Right	Right	0.23	66	57	-	-
Stroke	65	Female	Left	Left	Right	0.23	66	56	-	-
Stroke	65	Female	Right	Left	Right	0.23	66	57	-	-
Stroke	21	Male	Left	Right	Right	0.02	63	57	-	-
Stroke	21	Male	Right	Right	Right	0.02	66	56	-	-
Stroke	87	Female	Left	Right	Left	0.02	66	57	-	-
Stroke	87	Female	Right	Right	Left	0.02	50	29	-	-
Stroke	54	Male	Left	Right	Left	0.46	66	57	-	-
Stroke	54	Male	Right	Right	Left	0.46	54	57	-	-
Stroke	57	Male	Left	Left	Right	0.02	66	57	-	-
Stroke	57	Male	Right	Left	Right	0.02	61	57	-	-
Stroke	70	Female	Right	Left	Right	0.53	66	-	16.46	-
Stroke	57	Male	Right	Right	-	0.48	66	-	22.61	-
Stroke	73	Male	Left	Left	Right	0.53	63	-	28.55	-
Stroke	56	Male	Left	Left	Right	0.03	25	-	60.81	-
Stroke	63	Male	Left	Right	Right	0.48	66	-	14.33	-
ARSACS	41	Female	Left	Both	Right	-	-	-	28.59	-
ARSACS	41	Female	Right	Both	Right	-	-	-	37.14	-

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Disease	Age (yrs)	Sex	Tested side	Affected side	Dominant side	Chronicity (yrs)	FMA-UE (0-66)	ARAT (0-57)	NHPT (s)	EDSS (0-10)
ARSACS	29	Male	Left	Both	Right	-	-	-	56.98	-
ARSACS	29	Male	Right	Both	Right	-	-	-	40.34	-
ARSACS	56	Female	Left	Both	Left	-	-	-	83.59	-
ARSACS	56	Female	Right	Both	Left	-	-	-	95.20	-
ARSACS	37	Male	Left	Both	Left	-	-	-	36.36	-
ARSACS	37	Male	Right	Both	Left	-	-	-	46.72	-
ARSACS	26	Female	Left	Both	Right	-	-	-	-	-
ARSACS	26	Female	Right	Both	Right	-	-	-	-	-
ARSACS	37	Female	Left	Both	Right	-	-	-	-	-
ARSACS	37	Female	Right	Both	Right	-	-	-	-	-
ARSACS	31	Male	Left	Both	Right	-	-	-	29.88	-
ARSACS	31	Male	Right	Both	Right	-	-	-	23.52	-
ARSACS	58	Male	Left	Both	Right	-	-	-	60.43	-
ARSACS	58	Male	Right	Both	Right	-	-	-	47.33	-
MS	52	Female	Left	Both	Right	29.00	-	37	45.25	7.0
MS	52	Female	Right	Both	Right	29.00	-	47	24.75	7.0
MS	69	Male	Right	Both	Right	19.00	-	44	140.27	7.5
MS	25	Female	Left	Both	Right	6.00	-	52	29.35	6.0
MS	25	Female	Right	Both	Right	6.00	-	53	29.62	6.0
MS	42	Female	Left	Both	Right	1.00	-	56	27.81	4.0
MS	42	Female	Right	Both	Right	1.00	-	54	20.48	4.0
MS	59	Female	Left	Both	Left	5.00	-	49	27.76	7.0
MS	56	Female	Left	Both	Right	10.00	-	49	33.72	7.0
MS	56	Female	Right	Both	Right	10.00	-	29	89.79	7.0
MS	65	Male	Left	Both	Left	19.00	-	52	39.90	8.0
MS	63	Female	Left	Both	Right	8.00	-	57	20.84	4.5
MS	63	Female	Right	Both	Right	8.00	-	54	35.04	4.5
MS	76	Female	Left	Both	Right	38.00	-	43	27.01	5.0
MS	76	Female	Right	Both	Right	38.00	-	34	34.46	5.0
MS	60	Male	Left	Both	Right	21.00	-	52	31.48	7.0
MS	60	Male	Right	Both	Right	21.00	-	53	25.29	7.0
MS	42	Female	Right	Both	Right	21.00	-	39	74.39	7.5
MS	46	Male	Left	Both	Right	11.00	-	55	30.58	5.5
MS	46	Male	Right	Both	Right	11.00	-	56	23.23	5.5
MS	70	Female	Left	Both	Right	37.00	-	53	29.86	6.0
MS	70	Female	Right	Both	Right	37.00	-	45	53.21	6.0
MS	36	Female	Right	Both	Right	6.76	61	56	22.87	7.5
MS	40	Male	Right	Both	Left	12.55	53	44	56.17	7.5
MS	35	Male	Left	Both	Right	0.97	65	57	22.90	4.5
MS	35	Male	Right	Both	Right	0.97	65	52	24.90	4.5
MS	52	Female	Left	Both	Right	9.66	62	56	35.13	5.5
MS	52	Female	Right	Both	Right	9.66	65	56	29.31	5.5
MS	65	Female	Right	Both	Both	14.48	61	52	55.37	7.5
MS	53	Male	Right	Both	Right	9.66	62	56	23.93	2.5
MS	59	Male	Left	Both	Right	1.93	63	56	28.70	4.0
MS	59	Male	Right	Both	Right	1.93	63	55	47.49	4.0
MS	35	Female	Left	Both	Right	14.48	62	51	50.17	7.5
MS	38	Male	Left	Both	Right	2.90	-	-	-	3.5
MS	38	Male	Right	Both	Right	2.90	-	-	-	3.5
MS	66	Female	Left	Both	Left	16.42	-	-	-	7.5
MS	66	Female	Right	Both	Left	16.42	-	-	-	7.5
MS	22	Male	Left	Both	Right	3.86	-	-	-	6.5
MS	22	Male	Right	Both	Right	3.86	-	-	-	6.5
MS	38	Female	Left	Both	Right	8.69	-	-	-	7.0
MS	38	Female	Right	Both	Right	8.69	-	-	-	7.0
MS	61	Male	Left	Both	Right	5.79	-	-	20.00	5.0
MS	61	Male	Right	Both	Right	5.79	-	-	26.27	5.0
MS	63	Female	Right	Both	Right	6.76	-	-	28.00	6.0
MS	63	Male	Right	Both	Right	29.93	-	-	16.00	3.0

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table SM2: **Influence of potential confounds on each sensor-based metric**. For each metric, a mixed effect model was fitted to the Box-Cox-transformed outcome measure. Bold entries indicate that the fixed effect contributed in a statistically significant manner to model quality according to a simulated likelihood ratio test or that model quality was at least moderate according to the criteria $C1$ and $C2$. Abbreviations: SE: standard error; SD: standard deviation. R^2: adjusted coefficient of determination. SPARC: spectral arc length.

Sensor-based metric	**Fixed effects**												**Subject-specific effects**	$\mathbf{R^2}$	**Model quality**
	(Intercept)		Age		Sex		Tested side		Hand dominance		Stereo vision		(Intercept)		
	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	SD		C1 (%), C2 (%)
Jerk transport	**-30.59 (2.50)**	**0.001**	**0.08 (0.03)**	**0.007**	-0.47 (0.91)	0.604	0.41 (1.09)	0.691	0.78 (1.09)	0.5	-0.59 (1.57)	0.718	3.21	0.20	**5.87, 19.07**
Log jerk transport	**0.27 (0.00)**	**0.001**	**0.00 (0.00)**	**0.005**	0.00 (0.00)	0.734	0.00 (0.00)	0.793	0.00 (0.00)	0.519	0.00 (0.00)	0.879	0.00	0.43	**6.68, 20.55**
Spectral arc length transport	**0.16 (0.00)**	**0.001**	**0.00 (0.00)**	**0.001**	0.00 (0.00)	0.929	**0.00 (0.00)**	**0.027**	0.00 (0.00)	0.995	0.00 (0.00)	0.467	0.00	0.40	8.32, 27.27
Number of velocity peaks transport	-0.05 (0.07)	0.444	**0.00 (0.00)**	**0.001**	0.00 (0.02)	0.872	-0.05 (0.03)	0.142	0.01 (0.03)	0.783	-0.02 (0.04)	0.655	0.07	0.30	8.48, 34.79
Distance to max. velocity transport	**13.77 (0.41)**	**0.001**	0.00 (0.00)	0.613	0.02 (0.14)	0.884	-0.41 (0.21)	0.055	-0.08 (0.21)	0.724	-0.04 (0.25)	0.873	0.23	0.01	9.88, 29.99
Time to max. velocity transport	**14.83 (0.58)**	**0.001**	0.00 (0.01)	0.877	-0.09 (0.21)	0.677	-0.24 (0.28)	0.410	0.08 (0.28)	0.775	-0.18 (0.35)	0.607	0.54	0.03	10.54, 31.32
Jerk return	**-9.88 (0.69)**	**0.001**	**0.03 (0.01)**	**0.001**	-0.32 (0.26)	0.220	-0.26 (0.28)	0.346	0.10 (0.28)	0.724	-0.29 (0.44)	0.505	1.00	0.36	**7.11, 24.36**
Log jerk return	**0.24 (0.00)**	**0.001**	**0.00 (0.00)**	**0.002**	0.00 (0.00)	0.4	0.00 (0.00)	0.293	0.00 (0.00)	0.751	0.00 (0.00)	0.690	0.00	0.49	**7.23, 22.71**
Spectral arc length return	**0.25 (0.01)**	**0.001**	**0.00 (0.00)**	**0.028**	0.00 (0.00)	0.077	0.00 (0.00)	0.096	0.00 (0.00)	0.887	0.00 (0.00)	0.418	0.01	0.42	**6.29, 18.82**
Number of velocity peaks return	0.12 (0.09)	0.189	**0.00 (0.00)**	**0.003**	-0.06 (0.03)	0.065	0.03 (0.04)	0.455	0.04 (0.04)	0.374	-0.07 (0.05)	0.255	0.10	0.17	11.52, 36.43
Distance to max. velocity return	**632.23 (63.79)**	**0.001**	-0.61 (0.68)	0.380	-6.84 (22.91)	0.780	41.14 (29.75)	0.178	25.56 (29.75)	0.395	56.23 (39.31)	0.167	69.24	0.08	**7.22, 22.56**
Time to max. velocity return	**86.45 (7.70)**	**0.001**	-0.04 (0.08)	0.622	1.09 (2.79)	0.727	4.35 (3.47)	0.213	-0.03 (3.47)	0.992	6.21 (4.78)	0.198	9.15	0.12	**7.76, 24.86**
Path length ratio transport	**-3.11 (0.32)**	**0.001**	**0.02 (0.00)**	**0.001**	**0.32 (0.12)**	**0.009**	**0.35 (0.13)**	**0.009**	-0.09 (0.13)	0.5	-0.11 (0.20)	0.585	0.46	0.47	**3.01, 11.43**
Throughput transport	**2.36 (0.18)**	**0.001**	**-0.02 (0.00)**	**0.001**	**-0.36 (0.07)**	**0.001**	-0.03 (0.08)	0.690	0.02 (0.08)	0.805	0.04 (0.12)	0.729	0.24	0.65	8.27, 25.53
Path length ratio return	**-2.80 (0.29)**	**0.001**	**0.02 (0.00)**	**0.001**	**0.40 (0.11)**	**0.004**	**0.59 (0.11)**	**0.001**	-0.08 (0.11)	0.479	0.19 (0.19)	0.317	0.46	0.60	**2.95, 11.10**
Throughput return	**2.59 (0.18)**	**0.001**	**-0.03 (0.00)**	**0.001**	**-0.38 (0.07)**	**0.001**	**-0.36 (0.08)**	**0.001**	0.03 (0.08)	0.701	-0.17 (0.12)	0.133	0.23	0.69	**7.77, 22.93**
Trajectory error mean transport	**0.25 (0.00)**	**0.001**	0.00 (0.00)	0.930	0.00 (0.00)	0.332	**0.00 (0.00)**	**0.001**	0.00 (0.00)	0.976	0.00 (0.00)	0.896	0.00	0.54	**7.61, 22.75**
Trajectory error max transport	**0.25 (0.00)**	**0.001**	0.00 (0.00)	0.269	0.00 (0.00)	0.208	**0.00 (0.00)**	**0.001**	0.00 (0.00)	0.933	0.00 (0.00)	0.280	0.00	0.55	**7.35, 21.98**
Initial movement angle transport θ_1	-0.01 (1.99)	0.998	0.04 (0.02)	0.059	-1.28 (0.72)	0.105	**5.41 (0.90)**	**0.001**	-0.15 (0.90)	0.876	**3.65 (1.24)**	**0.006**	2.34	0.39	8.25, 25.74
Initial movement angle transport θ_2	0.08 (2.03)	0.963	0.04 (0.02)	0.093	-1.33 (0.73)	0.082	**5.42 (0.92)**	**0.001**	-0.23 (0.92)	0.812	**3.74 (1.26)**	**0.004**	2.35	0.39	8.29, 25.89
Initial movement angle transport θ_3	-0.63 (1.45)	0.675	0.03 (0.02)	0.097	-0.77 (0.53)	0.165	**4.59 (0.62)**	**0.001**	0.34 (0.62)	0.584	**1.97 (0.91)**	**0.044**	1.89	0.46	**7.48, 23.81**
Trajectory error mean return	**0.25 (0.00)**	**0.001**	0.00 (0.00)	0.3	0.00 (0.00)	0.191	**0.00 (0.00)**	**0.003**	0.00 (0.00)	0.431	0.00 (0.00)	0.168	0.00	0.40	**4.55, 14.40**
Trajectory error max return	**0.25 (0.00)**	**0.001**	0.00 (0.00)	0.432	0.00 (0.00)	0.385	**0.00 (0.00)**	**0.006**	0.00 (0.00)	0.360	0.00 (0.00)	0.333	0.00	0.39	**5.68, 17.36**
Initial movement angle return θ_1	**14.38 (3.10)**	**0.001**	-0.05 (0.03)	0.170	1.00 (1.12)	0.382	-2.25 (1.43)	0.126	-0.30 (1.43)	0.843	1.17 (1.92)	0.563	3.51	0.11	**5.59, 18.87**
Initial movement angle return θ_2	**14.00 (3.38)**	**0.001**	-0.05 (0.04)	0.188	1.24 (1.22)	0.284	-2.13 (1.56)	0.187	-0.23 (1.56)	0.891	1.27 (2.09)	0.555	3.75	0.10	**5.58, 18.76**
Initial movement angle return θ_3	**10.12 (1.99)**	**0.001**	-0.04 (0.02)	0.071	0.74 (0.72)	0.324	**-2.49 (0.89)**	**0.012**	0.00 (0.89)	1	0.88 (1.24)	0.473	2.41	0.20	**3.89, 12.19**
Velocity mean transport	**19.94 (1.15)**	**0.001**	**-0.09 (0.01)**	**0.001**	**-1.06 (0.44)**	**0.017**	**0.94 (0.34)**	**0.006**	-0.34 (0.34)	0.304	1.05 (0.75)	0.184	2.09	0.79	**6.07, 20.78**
Velocity max transport	**22.05 (1.10)**	**0.001**	**-0.09 (0.01)**	**0.001**	-0.76 (0.42)	0.081	0.48 (0.33)	0.141	-0.37 (0.33)	0.280	1.24 (0.72)	0.114	1.99	0.78	**5.67, 17.27**
Velocity mean return	**5.34 (0.20)**	**0.001**	**-0.01 (0.00)**	**0.001**	0.02 (0.08)	0.8	-0.03 (0.06)	0.581	-0.08 (0.06)	0.150	0.17 (0.13)	0.198	0.36	0.70	**5.94, 18.44**
Velocity max return	**6.40 (0.19)**	**0.001**	**-0.01 (0.00)**	**0.001**	0.05 (0.07)	0.482	-0.11 (0.05)	0.054	-0.07 (0.05)	0.176	0.17 (0.12)	0.169	0.35	0.76	**5.84, 18.05**
Position error peg approach	**3.54 (0.14)**	**0.001**	**0.01 (0.00)**	**0.001**	**0.19 (0.05)**	**0.001**	0.06 (0.06)	0.339	-0.08 (0.06)	0.205	0.01 (0.09)	0.932	0.19	0.42	**3.09, 12.73**
Jerk peg approach	**-13.55 (0.81)**	**0.001**	**0.07 (0.01)**	**0.001**	**1.48 (0.29)**	**0.001**	**1.91 (0.38)**	**0.001**	-0.15 (0.38)	0.694	0.15 (0.50)	0.772	0.88	0.52	**2.57, 16.38**
Log jerk peg approach	**0.23 (0.00)**	**0.001**	**0.00 (0.00)**	**0.001**	**0.00 (0.00)**	**0.001**	**0.00 (0.00)**	**0.001**	0.00 (0.00)	0.927	0.00 (0.00)	0.969	0.00	0.68	**7.77, 23.63**
SPARC peg approach	**0.33 (0.01)**	**0.001**	**0.00 (0.00)**	**0.001**	**0.02 (0.01)**	**0.001**	**0.01 (0.01)**	**0.035**	0.00 (0.01)	0.715	0.00 (0.01)	0.698	0.02	0.50	**3.59, 12.74**
Position error hole approach	**3.01 (0.21)**	**0.001**	**0.02 (0.00)**	**0.001**	**0.40 (0.08)**	**0.001**	**0.56 (0.11)**	**0.001**	0.12 (0.11)	0.315	0.09 (0.13)	0.477	0.16	0.40	**1.25, 8.41**
Jerk hole approach	**-17.57 (1.26)**	**0.001**	**0.07 (0.01)**	**0.001**	**1.83 (0.46)**	**0.001**	0.09 (0.52)	0.859	-0.37 (0.52)	0.476	-0.35 (0.80)	0.658	1.78	0.43	**1.97, 9.31**
Log jerk hole approach	**0.22 (0.00)**	**0.001**	0.00 (0.00)	0.161	0.00 (0.00)	0.629	0.00 (0.00)	0.304	0.00 (0.00)	0.341	0.00 (0.00)	0.565	0.00	0.51	**6.78, 22.32**
SPARC hole approach	**0.39 (0.02)**	**0.001**	**0.00 (0.00)**	**0.001**	**0.02 (0.01)**	**0.006**	0.01 (0.01)	0.515	0.00 (0.01)	0.991	-0.01 (0.01)	0.490	0.03	0.51	**6.57, 22.17**
Number of movement onsets	**16.86 (0.12)**	**0.001**	0.00 (0.00)	0.118	0.02 (0.04)	0.703	-0.01 (0.06)	0.869	0.01 (0.06)	0.855	0.05 (0.07)	0.532	0.12	0.06	**1.21, 8.48**
Number of movement ends	**16.43 (0.17)**	**0.001**	**0.00 (0.00)**	**0.025**	0.05 (0.06)	0.389	-0.03 (0.09)	0.721	-0.04 (0.09)	0.655	**0.41 (0.10)**	**0.001**	0.00	0.05	**3.84, 21.98**
Dropped pegs	**0.26 (0.00)**	**0.001**	**0.00 (0.00)**	**0.001**	0.00 (0.00)	1	0.00 (0.00)	0.374	**0.00 (0.00)**	**0.002**	**0.00 (0.00)**	**0.001**	0.00	0.24	**3.93, 17.99**

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sensor-based metric	Fixed effects												Subject-specific effects	R^2	Model quality
	(Intercept)		Age		Sex		Tested side		Hand dominance		Stereo vision		(Intercept)		
	Estimate (SE)	*p*-value	Estimate (SE)	*p*-value	Estimate (SE)	*p*-value	Estimate (SE)	*p*-value	Estimate (SE)	*p*-value	Estimate (SE)	*p*-value	SD		C1 (%), C2 (%)
Haptic collisions mean transport	**-1.50 (0.32)**	**0.001**	**0.01 (0.00)**	**0.001**	-0.17 (0.12)	0.177	-0.14 (0.11)	0.215	0.12 (0.11)	0.307	0.23 (0.21)	0.289	0.54	0.55	**7.05, 24.00**
Haptic collisions max transport	-0.47 (0.26)	0.079	**0.01 (0.00)**	**0.001**	**-0.25 (0.10)**	**0.020**	-0.09 (0.09)	0.317	-0.01 (0.09)	0.888	0.20 (0.17)	0.198	0.42	0.52	**6.91, 21.97**
Haptic collisions mean return	**0.25 (0.00)**	**0.001**	**0.00 (0.00)**	**0.019**	0.00 (0.00)	0.070	0.00 (0.00)	0.702	**0.00 (0.00)**	**0.027**	0.00 (0.00)	0.915	0.00	0.50	**6.07, 20.78**
Haptic collisions max return	0.57 (0.41)	0.197	0.00 (0.00)	0.786	**-0.51 (0.15)**	**0.002**	**-0.34 (0.16)**	**0.039**	0.10 (0.16)	0.573	-0.11 (0.26)	0.679	0.60	0.35	7.64, 26.05
Force mean transport	**2.34 (0.20)**	**0.001**	**-0.01 (0.00)**	**0.012**	**-0.27 (0.08)**	**0.003**	0.03 (0.05)	0.558	0.09 (0.05)	0.078	-0.05 (0.14)	0.716	0.39	0.80	**4.21, 15.86**
Force max. transport	**2.59 (0.21)**	**0.001**	**-0.01 (0.00)**	**0.012**	**-0.29 (0.08)**	**0.001**	0.04 (0.05)	0.444	0.06 (0.05)	0.237	-0.06 (0.14)	0.688	0.42	0.81	**3.73, 14.13**
Force rate mean transport	**3.30 (0.36)**	**0.001**	**-0.01 (0.00)**	**0.002**	**-0.50 (0.14)**	**0.001**	0.18 (0.09)	0.053	0.06 (0.09)	0.528	-0.16 (0.24)	0.485	0.68	0.80	**2.70, 10.44**
Force rate max. transport	5.72 (0.38)	**0.001**	-0.01 (0.00)	**0.001**	-0.65 (0.15)	**0.001**	0.14 (0.09)	0.125	0.16 (0.09)	0.082	-0.24 (0.25)	0.374	0.74	**0.84**	**2.98, 10.49**
Force mean return	0.34 (0.22)	0.139	**0.01 (0.00)**	**0.001**	0.03 (0.08)	0.755	**0.25 (0.11)**	**0.035**	0.09 (0.11)	0.458	-0.03 (0.13)	0.823	0.16	0.17	**5.58, 18.73**
Force max. return	**0.26 (0.00)**	**0.001**	**0.00 (0.00)**	**0.002**	0.00 (0.00)	0.790	0.00 (0.00)	0.122	0.00 (0.00)	0.448	0.00 (0.00)	0.747	0.00	0.47	**4.14, 14.57**
Force rate mean return	**0.97 (0.20)**	**0.001**	0.00 (0.00)	0.563	-0.10 (0.08)	0.203	**0.18 (0.07)**	**0.017**	0.08 (0.07)	0.225	0.01 (0.13)	0.955	0.35	0.53	**6.48, 22.78**
Force rate max. return	1.55 (0.11)	**0.001**	0.00 (0.00)	**0.001**	-0.11 (0.04)	**0.008**	0.14 (0.04)	**0.002**	0.02 (0.04)	0.546	-0.04 (0.07)	0.523	0.18	**0.61**	**4.98, 18.57**
Force mean peg approach	**1.96 (0.21)**	**0.001**	**-0.01 (0.00)**	**0.026**	**-0.27 (0.08)**	**0.002**	-0.10 (0.07)	0.110	0.08 (0.07)	0.245	-0.16 (0.14)	0.228	0.37	0.65	**3.54, 12.43**
Force max. peg approach	**2.64 (0.22)**	**0.001**	0.00 (0.00)	0.130	**-0.24 (0.08)**	**0.006**	0.01 (0.05)	0.791	0.04 (0.05)	0.469	-0.12 (0.15)	0.444	0.42	0.80	**3.84, 14.20**
Force rate mean peg approach	**3.68 (0.28)**	**0.001**	**-0.02 (0.00)**	**0.001**	**-0.52 (0.11)**	**0.001**	-0.08 (0.07)	0.267	0.07 (0.07)	0.312	-0.27 (0.19)	0.159	0.54	0.83	**3.13, 11.09**
Force rate max. peg approach	**5.17 (0.31)**	**0.001**	**-0.01 (0.00)**	**0.009**	**-0.44 (0.12)**	**0.001**	0.00 (0.08)	0.955	0.07 (0.08)	0.403	-0.13 (0.21)	0.540	0.61	0.82	**3.13, 10.87**
Force mean hole approach	**1.82 (0.15)**	**0.001**	0.00 (0.00)	0.064	**-0.17 (0.06)**	**0.006**	-0.01 (0.04)	0.843	0.06 (0.04)	0.160	-0.07 (0.10)	0.467	0.28	0.76	**4.84, 18.59**
Force max. hole approach	**1.99 (0.16)**	**0.001**	0.00 (0.00)	0.204	**-0.17 (0.06)**	**0.012**	0.01 (0.04)	0.870	0.08 (0.04)	0.067	-0.10 (0.11)	0.361	0.30	0.75	**5.33, 20.21**
Force rate mean hole approach	**30.63 (4.59)**	**0.001**	**-0.21 (0.05)**	**0.001**	**-6.80 (1.78)**	**0.001**	0.92 (1.15)	0.427	**3.91 (1.15)**	**0.002**	-3.56 (3.06)	0.253	8.89	0.83	**4.57, 15.92**
Force rate max. hole approach	**5.25 (0.36)**	**0.001**	**-0.01 (0.00)**	**0.011**	**-0.48 (0.14)**	**0.001**	0.01 (0.10)	0.951	**0.26 (0.10)**	**0.010**	-0.16 (0.24)	0.524	0.67	0.73	**5.81, 21.28**
Force rate number of peaks transport	**1.17 (0.19)**	**0.001**	**0.02 (0.00)**	**0.001**	0.12 (0.07)	0.069	-0.07 (0.08)	0.421	0.09 (0.08)	0.259	-0.14 (0.12)	0.256	0.25	0.47	**7.11, 23.64**
Force rate spectral arc length transport	**0.42 (0.02)**	**0.001**	**0.00 (0.00)**	**0.001**	0.00 (0.01)	0.627	-0.02 (0.01)	0.075	-0.01 (0.01)	0.297	-0.01 (0.02)	0.606	0.04	0.42	**6.97, 23.19**
Force rate number of peaks return	**1.86 (0.12)**	**0.001**	**0.00 (0.00)**	**0.002**	**-0.15 (0.04)**	**0.002**	-0.06 (0.04)	0.173	0.02 (0.04)	0.722	-0.12 (0.07)	0.116	0.18	0.48	**7.53, 23.94**
Force rate spectral arc length return	**0.63 (0.04)**	**0.001**	**0.00 (0.00)**	**0.001**	-0.03 (0.02)	0.075	0.02 (0.02)	0.176	0.02 (0.02)	0.155	-0.02 (0.03)	0.513	0.06	0.43	**8.11, 24.87**
Force rate number of peaks peg approach	**1.24 (0.01)**	**0.001**	**0.00 (0.00)**	**0.001**	**0.02 (0.00)**	**0.001**	**0.01 (0.00)**	**0.007**	0.00 (0.00)	0.941	0.01 (0.01)	0.275	0.01	0.66	**3.30, 14.23**
Force rate spectral arc length peg approach	**0.80 (0.01)**	**0.001**	**0.00 (0.00)**	**0.001**	**0.02 (0.00)**	**0.001**	0.01 (0.00)	0.237	0.00 (0.00)	0.563	0.01 (0.01)	0.182	0.01	0.48	**3.86, 15.46**
Force rate number of peaks hole approach	**0.82 (0.10)**	**0.001**	**0.01 (0.00)**	**0.001**	**0.13 (0.04)**	**0.006**	0.02 (0.04)	0.702	-0.01 (0.04)	0.851	0.02 (0.06)	0.772	0.16	0.57	**2.17, 9.08**
Force rate spectral arc length hole approach	**0.26 (0.03)**	**0.001**	**0.00 (0.00)**	**0.001**	**0.04 (0.01)**	**0.001**	0.00 (0.01)	0.781	-0.01 (0.01)	0.376	-0.01 (0.02)	0.594	0.05	0.62	**2.28, 11.56**
Force rate number of peaks buildup	0.00 (0.02)	0.987	**0.00 (0.00)**	**0.008**	0.00 (0.01)	0.494	0.02 (0.01)	0.062	**0.02 (0.01)**	**0.032**	**-0.03 (0.01)**	**0.012**	0.02	0.12	4.58, 35.23
Force rate spectral arc length buildup	**0.04 (0.00)**	**0.001**	**0.00 (0.00)**	**0.001**	**0.00 (0.00)**	**0.006**	0.00 (0.00)	0.733	0.00 (0.00)	0.262	0.00 (0.00)	0.177	0.00	0.40	6.64, 25.87
Force buildup duration	**-2.66 (0.19)**	**0.001**	**0.01 (0.00)**	**0.002**	**0.26 (0.07)**	**0.001**	-0.01 (0.06)	0.878	0.01 (0.06)	0.904	-0.16 (0.12)	0.223	0.33	0.61	**6.16, 19.68**
Force rate number of peaks release	0.00 (0.01)	0.795	0.00 (0.00)	0.226	0.00 (0.00)	0.819	0.01 (0.00)	0.207	0.00 (0.00)	0.748	0.00 (0.01)	0.687	0.00	0.02	6.87, 71.30
Force rate spectral arc length release	**1.04 (0.04)**	**0.001**	**0.00 (0.00)**	**0.001**	**0.05 (0.01)**	**0.001**	-0.02 (0.01)	0.214	0.00 (0.01)	0.882	0.02 (0.02)	0.325	0.05	0.56	**3.53, 12.34**
Force release duration	**-2.32 (0.18)**	**0.001**	0.00 (0.00)	0.170	**0.17 (0.07)**	**0.012**	0.12 (0.07)	0.124	-0.04 (0.07)	0.617	-0.13 (0.11)	0.279	0.25	0.31	**8.05, 24.38**
Task completion time	**1.13 (0.01)**	**0.001**	**0.00 (0.00)**	**0.001**	**0.02 (0.00)**	**0.001**	**0.01 (0.00)**	**0.033**	0.00 (0.00)	0.665	0.00 (0.01)	0.633	0.01	0.73	**4.13, 17.21**
Simulated Gaussian noise	**4.42 (0.12)**	**0.001**	0.00 (0.00)	0.851	0.04 (0.04)	0.302	-0.10 (0.06)	0.118	-0.02 (0.06)	0.784	0.03 (0.07)	0.651	0.10	0.03	12.00, 35.29

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

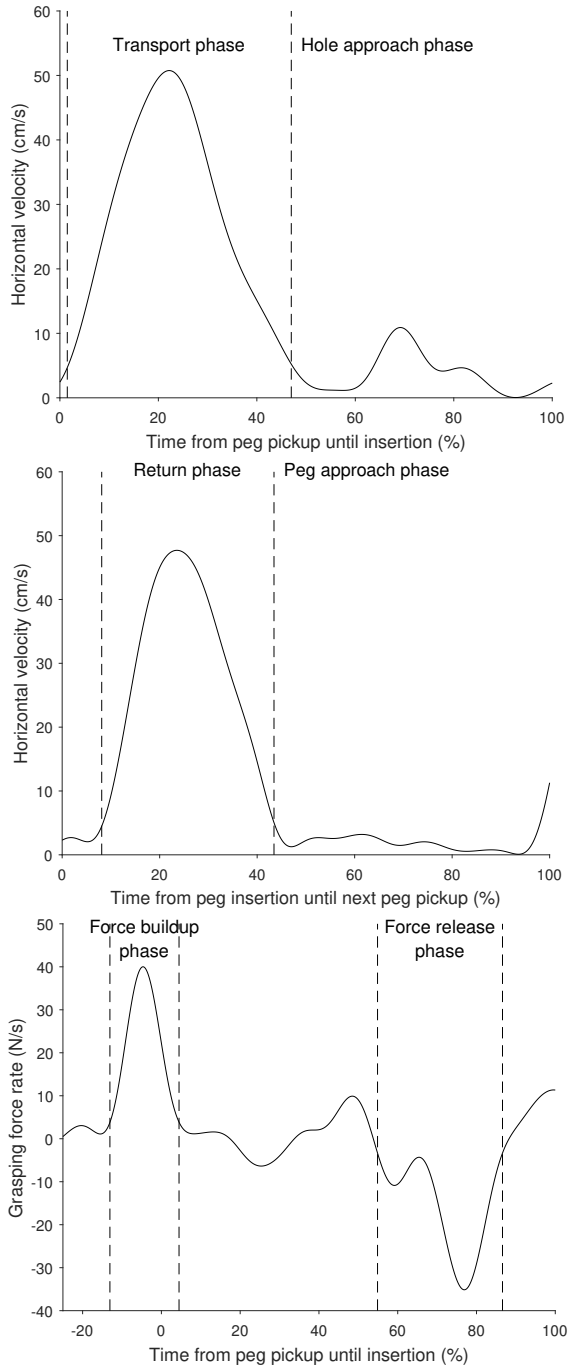


Figure SM1: Temporal segmentation of kinematic and kinetic trajectories. Representative example from one neurologically intact subject (49yrs, female, tested hand left, dominant hand right).

bioRxiv preprint doi: https://doi.org/10.1101/544601; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

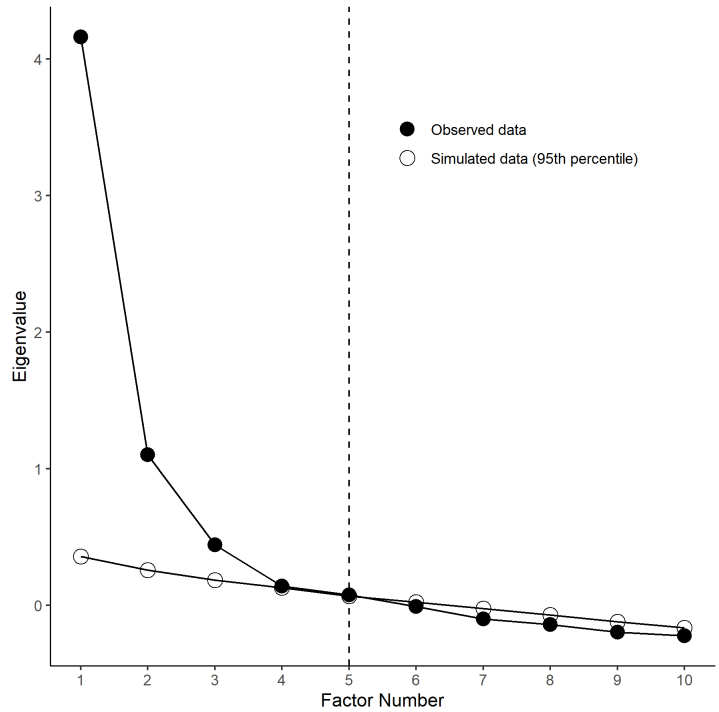


Figure SM2: Scree plot for estimating the number of latent variables k in the factor analysis. Parallel analysis was used to simulate a lower bound of an eigenvalues magnitude that each eigenvalue in the observed data needs to fulfill. The chosen number of factors was set to five accordingly.