Received: 25 September 2021 | Accepted: 24 October 2021

DOI: 10.1111/ene.15162

european journal of neurology
the official journal of the european academy of neurology

ORIGINAL ARTICLE

Smartphone-derived keystroke dynamics are sensitive to relevant changes in multiple sclerosis

Ka-Hoo Lam[1] | James Twose[2] | Hannah McConchie[2] | Giovanni Licitra[2] | Kim Meijer[2] | Lodewijk de Ruiter[1] | Zoë van Lierop[1] | Bastiaan Moraal[3] | Frederik Barkhof[3,4] | Bernard Uitdehaag[1] | Vincent de Groot[5] | Joep Killestein[1]

[1]Department of Neurology, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam University Medical Centers, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[2]Neurocast B.V., Amsterdam, The Netherlands

[3]Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam University Medical Centers, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[4]Queen Square Institute of Neurology and Centre for Medical Image Computing, University College London, London, UK

[5]Department of Rehabilitation Medicine, MS Center Amsterdam, Amsterdam Neuroscience, Amsterdam University Medical Centers, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Correspondence
Ka-Hoo Lam, Department of Neurology, MS Center Amsterdam, Amsterdam Neuroscience, De Boelelaan 1117, Amsterdam University Medical Centers, Vrije Universiteit Amsterdam, 1081HV Amsterdam, The Netherlands.
Email: k.lam1@amsterdamumc.nl

Funding information
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: the collaboration project is co-funded by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health [grant number LSHM16060-SGF], and Stichting MS Research [grant number 16-946 MS] to stimulate public–private partnerships as well as by a contribution from Biogen (unrestricted funding)

Abstract

Background: To investigate smartphone keystroke dynamics (KD), derived from regular typing, on sensitivity to relevant change in disease activity, fatigue, and clinical disability in multiple sclerosis (MS).

Methods: Preplanned interim analysis of a cohort study with 102 MS patients assessed at baseline and 3-month follow-up for gadolinium-enhancing lesions on magnetic resonance imaging, relapses, fatigue and clinical disability outcomes. Keyboard interactions were unobtrusively collected during typing using the Neurokeys App. From these interactions 15 keystroke features were derived and aggregated using 16 summary and time series statistics. Responsiveness of KD to clinical anchor-based change was assessed by calculating the area under the receiver operating characteristic curve (AUC). The optimal cut-point was used to determine the minimal clinically important difference (MCID) and compared to the smallest real change (SRC). Commonly used clinical measures were analyzed for comparison.

Results: A total of 94 patients completed the follow-up. The five best performing keystroke features had AUC-values in the range 0.72–0.78 for change in gadolinium-enhancing lesions, 0.67–0.70 for the Checklist Individual Strength Fatigue subscale, 0.66–0.79 for the Expanded Disability Status Scale, 0.69–0.73 for the Ambulation Functional System, and 0.72–0.75 for Arm function in MS Questionnaire. The MCID of these features exceeded the SRC on group level. KD had higher AUC-values than comparative clinical measures for the study outcomes, aside from ambulatory function.

Conclusions: Keystroke dynamics demonstrated good responsiveness to changes in disease activity, fatigue, and clinical disability in MS, and detected important change beyond measurement error on group level. Responsiveness of KD was better than commonly used clinical measures.

KEYWORDS
biometry, multiple sclerosis, pattern recognition, physiological, ROC curve, smartphone

INTRODUCTION

In multiple sclerosis (MS) inflammatory disease mechanisms lead to neurological deficits, accumulation of disability, and associated disabling symptoms such as fatigue and cognitive dysfunction [1–3]. Therefore, the treatment of MS is focused on suppressing inflammatory disease activity and disability progression, and managing the associated symptoms. Given that inflammatory activity is often subclinical [4], magnetic resonance imaging (MRI) is essential to evaluate the initiation, escalation, or discontinuation of disease-modifying therapies [5]. However, frequent neuroimaging is infeasible due to the relatively long scan times resulting in high costs of repeated assessments within a short time frame. Likewise, limitations in frequency of clinical assessment and the lack of standardization of current measures impedes the optimal monitoring of disability progression and fatigue [6–8].

Therefore, new biomarkers in MS are needed. Serum-based and electrophysiological markers have recently emerged [9], as well as technology-based biomarkers such as the collection and analysis of keystroke dynamics (KD). Analysis of KD has the advantage of passive and unobtrusive monitoring of typing behavior collected remotely through smartphone usage. Timing-related keystroke features were recently shown to be reliable and valid for measuring clinical disability outcomes in MS in a cross-sectional setting [10]. In addition to timing-related keystroke features, the use of emoji (i.e., graphical emoticons) may also be analyzed to capture the sentiment of messages [11]. In this preplanned interim analysis, we intended to expand on the use of KD and analyze the ability of KD to detect clinically relevant changes over time (i.e., responsiveness) [12].

Objective

To investigate the responsiveness of KD to detect short-term change in disease activity, self-reported fatigue, and clinical disability in MS.

METHODS

Prospective cohort study at the MS Center of the Amsterdam University Medical Centers, location VU University Medical Center. The study design and baseline analysis have been reported previously [10]. In short, from August 2018 to December 2019 participants were recruited to use the Neurokeys keyboard app [13] on their own smartphones in the everyday environment for 1 year. Clinical outcomes were assessed at baseline and every 3 months for a total of five clinical visits. Reported here are the results of the preplanned interim analysis of the baseline and 3-month follow-up visit in order to address responsiveness of KD to short-term change in clinical outcomes. Patients with MS were consecutively recruited until a sample size of 100 patients was reached. Eligibility criteria included: use of a smartphone, age between 18 and 65 years, a definite diagnosis of MS, a baseline Expanded Disability Status Scale (EDSS) score below 7.5, no visual or upper extremity deficits that seriously interfere with smartphone use, and no significant mood or sleep disorder at baseline based on medical history-taking by a screening physician. The study received ethical approval (METc VUmc, reference 2017.576) and conformed to legislation regarding data privacy and medical devices (Dutch Health and Youth Care Inspectorate, reference VGR2006948). All patients gave written informed consent. The study was registered at trialregister.nl (NL7070).

Study outcomes and anchors for clinically relevant change

The study outcomes assessed at baseline and follow-up at 3 months were: (1) disease activity, (2) fatigue, and (3) clinical disability. Clinical measures were chosen for the study outcomes based on whether change in the measure can be directly related to clinical relevancy, to which the responsiveness of KD is investigated. Using this anchor-based method, for each study outcome patients were anchored (i.e., stratified) as having ‘clinically relevant change’ (improvement or worsening) or ‘no clinically relevant change’ according to the clinical anchors.

Disease activity was only assessed in patients with relapsing-remitting MS (RRMS) subtype. Radiological disease activity was determined with MRI for new or enlarged T2 lesions and gadolinium-enhancing (Gd+) T1 lesions at baseline and 3-month follow-up visit. Since KD were analyzed in the prespecified windows of 7 days within the baseline and 3-month follow-up visits, only Gd+ lesions were considered for the responsiveness analyses due to the known temporal association with the acute inflammatory phase of lesions [5]. Any amount of change in Gd+ lesions between baseline and follow-up was anchored as relevant change [14]. Clinical disease activity was assessed at the 3-month follow-up visit for the occurrences of relapses. Relapses were defined as new or worsened neurological deficits persisting for at least 24 h and in the absence of fever, infection, or an explanation other than MS. Patients who had relapses with no or partial recovery during the follow-up visit were anchored as having relevant change. Patients who had a relapse with full recovery at the follow-up visit were anchored as having no clinically relevant change compared to baseline.

Fatigue was assessed using the Checklist Individual Strength Fatigue subscale (CIS-F) [15]. A CIS-F score below 35 was used to indicate non-severe fatigue, and a score of 35 or higher as severe fatigue [16]. Patients moving from one side of the cut-off to the other were defined as having clinically relevant change in fatigue.

For clinical disability, the overall severity of disability and commonly affected domains in MS were assessed: ambulatory function, arm function, and information processing speed. The EDSS was used for the overall severity of disability due to MS. Clinically relevant change in EDSS was defined as follows: ≥1.5-point increase if baseline EDSS is 0, ≥1-point increase if baseline EDSS is 1.0–5.5, and ≥0.5-point increase if baseline EDSS is ≥6.0 [17]. For ambulatory function and arm function, the EDSS Ambulation Functional System (FS) score and Arm function in MS Questionnaire (AMSQ), respectively, were used since these measures assess the function in daily living [18]. An EDSS Ambulation FS score change of ≥1 was anchored as clinically relevant. An AMSQ score change of ≥15 points was anchored as clinically relevant [19]. Information processing speed was assessed with the Symbol Digit Modalities Test (SDMT) [20]. A 4-point change or more on the SDMT was used as anchor for clinically relevant change [20].

Keystroke dynamics and features

Patients installed the Neurokeys app on their own smartphone which replaced the native keyboard with the intelligent Neurokeys keyboard [13]. During the study follow-up, Neurokeys continuously and passively collected timestamped keystroke data any time the keyboard was used during regular typing. A total of 15 keystroke features were derived from the keystroke data based on (also illustrated in Figure 1a):

- Keystrokes: absolute amount of any keys used.
- Typing sessions: duration and count of typing sessions (defined as one successive period of activation (i.e., 'keyboard up') and inactivation (i.e., 'keyboard down') of the keyboard).
- Alphanumeric keys: the time a key is pressed down (Hold Time, HT), the latency between a key release and the next key press (Flight Time, FT), the latency between successive key presses (Press-Press Latency, PPL), and releases (Release-Release Latency, RRL).
- Backspace key: the time before (Pre-Correction Slowing, Pre-CS), during (Correction Duration, CD), and after (Post-Correction Slowing, Post-CS) the use of backspace keys.
- Punctuation marks: the latency between the use of a punctuation mark and a next key (After Punctuation Pause, APP).
- Use of graphical emoticons ('emojis'): emojis were assigned an Emoji Sentiment Score (ESS) each for negativity, neutrality, positivity and polarity of emojis based on an emoji sentiment ranking [11].

To enable comparison between the KD gathered per typing session and the clinical measures, each keystroke feature was aggregated over 14-day periods (Figure 1b). First, the typing sessions were aggregated per day by calculation of 16 summary statistics (i.e., vectors) for each feature in order to aggregate the high sample rate data while retaining meaningful information. These included the mean and median (indicators of the central tendency), standard deviation (SD), skewness and kurtosis (indicators of the dispersion), minimum and maximum (indicators of the range), and time series aggregation methods (to capture changes within the day); see also the Appendix S1. Next, the vectors for each feature were aggregated based on the windows of 7 days on either side of the clinical visit date by taking the median value, which is more robust to outliers compared to the mean value [21].

Comparative responsiveness

Keystroke features were compared to existing clinical measures to contextualize their responsiveness to the study outcomes. For each study outcome, commonly used clinical measures were chosen as benchmarks. Responsiveness of KD to change in disease activity was benchmarked with the EDSS [22]. The Fatigue Severity Scale (FSS) and Modified Fatigue Impact Scale (MFIS) were used as clinical benchmarks for the responsiveness to fatigue [23,24]. Responsiveness to change in clinical disability was benchmarked with the Timed 25-foot Walk Test (TWT) and Nine-Hole Peg Test (NHPT) [25,26]. For information processing speed, the EDSS was chosen as the best available clinical benchmark.

Statistical analysis

Descriptives

Categorical variables were summarized as frequencies with percentages. Continuous variables were summarized by the mean and SD if normally distributed, otherwise the median and interquartile range (IQR) were used. For each study outcome, only patients with complete data at baseline and follow-up were analyzed.

Responsiveness

Responsiveness is the clinimetric property of a measure that represents its ability to detect clinically relevant change over time [12]. For each outcome measure, the correspondence between absolute changes in keystroke features and anchor-based change in study outcomes was assessed with receiver operating characteristic (ROC) curves by plotting the true positive rate (sensitivity) against the false

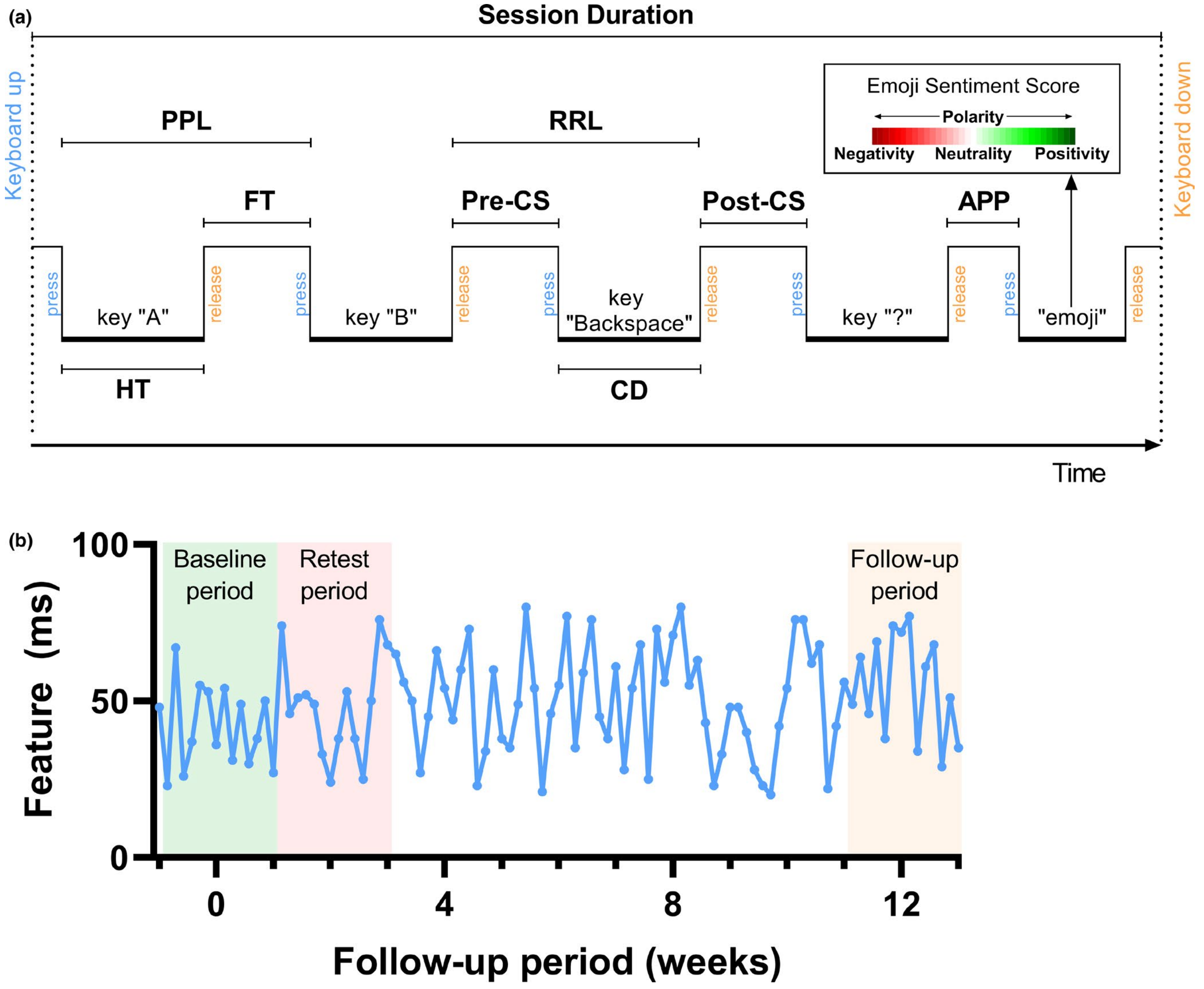

FIGURE 1 Schematic representation of the keystroke features and aggregation periods. (a) Schematic representation of the keystroke features: APP, After Punctuation Pause; CD, Correction Duration; FT, Flight Time; HT, Hold Time; Post-CS, Post-Correction Slowing; PPL, Press-Press Latency; Pre-CS, Pre-Correction Slowing; RRL, Release-Release Latency [11]. (b) The keystroke data for each keystroke feature were aggregated per day using statistical and time series methods (see Appendix S1). The daily keystroke data were aggregated into three 14-day periods shown here by taking the median value. For the responsiveness the baseline and follow-up period were analyzed; for the quantification of measurement error the baseline and retest period were analyzed [Colour figure can be viewed at wileyonlinelibrary.com]

positive rate (1 − specificity). The area under the ROC curve (AUC) was then calculated where values of ≥0.70 were indicative for adequate responsiveness [12]. AUC-values of keystroke features were also compared to AUC-values of the clinical measures selected as benchmarks.

Minimal clinically important difference and measurement error

The smallest change in keystroke features that most optimally distinguished clinically relevant change, that is, the minimal clinically important difference (MCID), was determined from the ROC analysis. This most optimal trade-off between sensitivity and specificity was found with the maximum value of the Youden's index [19]. Since score changes may be due to measurement error instead of real change, the MCID should be equal to or higher than the smallest real change (SRC) of the instrument [12]. The SRC is the smallest amount of change that can be reliably distinguished from measurement error. The SRC can be calculated at the individual (i.e., SRC_{ind}) and group (i.e., SRC_{group}) level. SRC_{ind} was calculated from the naturally occurring variability of KD during a stable period, that is, between a baseline and retest period (see Figure 1b): $SRC_{ind} = 1.96 \times SD$. The SRC at group level was calculated as: $SRC_{group} = SRC_{ind}/\sqrt{n}$. As measurement error can be reduced with repeated measurements, the minimum number of patients or repeated measurements within individuals needed to attain SRC = MCID was calculated using the ratio between the $SRC_{ind}$and the MCID: $n = \left(\frac{SRC_{ind}}{MCID}\right)^2$ [27].

RESULTS

Among the 102 patients included, six dropped out of the study before the follow-up visit and two patients did not complete the follow-up visit due to the COVID-19 pandemic. The clinical and demographical characteristics at baseline of the remaining 94 patients are summarized in Table 1. The study population had a mean (±SD) age of 46.7 (10.4) years, 72.3% were female, median (IQR) disease duration since diagnosis was 6.0 (3.0–2.4) years, and median (IQR) EDSS was 3.5 (2.5–4.0).

Change between baseline and follow-up

The mean (±SD) duration between baseline and 3-month follow-up was 96.5 (±11.2) days. The clinical study outcomes at baseline and 3-month follow-up are shown in Table 2. For the disease activity outcome, 54 patients had a RRMS subtype of which six patients had no MRI with gadolinium administration (gadolinium allergy, *n* = 2; breastfeeding/pregnancy wish, *n* = 2; omission of gadolinium administration, *n* = 2) and were excluded from the disease activity outcome analysis. Of the remaining 48 patients with RRMS, five developed a relapse: three of whom also had a change in radiological disease activity and two were radiologically stable. The three patients with clinically and radiologically active MS had no or partial recovery of the relapse at follow-up and were regarded as changed for the disease activity outcome. Of the two patients with a relapse but radiologically stable MS, one had full and the other had partial recovery of the relapse and were regarded as stable and changed, respectively, for the disease activity outcome. Twelve patients had MRI activity without a relapse.

Responsiveness of keystroke features

The AUC-values of the keystroke features in distinguishing clinically relevant change for each study outcome are summarized with histograms in Figure 2. For each study outcome the five keystroke features with the highest AUC-values are summarized in Table 3. For all study outcomes at least one keystroke feature had adequate responsiveness (i.e., AUC ≥0.70) to clinically relevant change, except for information processing speed (AUC 0.29–0.68). The range of AUC-values of keystroke features in distinguishing relevant change in disease activity was 0.20–0.78 when only Gd+ lesions were considered, and 0.25–0.76 when Gd+ lesions and relapses were combined. Keystroke features responsive to change in disease activity were mostly timing-related. For fatigue the range of AUC-values was 0.32–0.70, with negative emoji sentiments and duration of typing sessions being adequately responsive. The range of AUC-values of keystroke features for clinically relevant change in EDSS was 0.32–0.79, with the most responsive features being a blend of emoji sentiments, a timing-related feature, and the length of words. Finally, predominantly timing-related features were most responsive to ambulatory and arm function domains of clinical disability, with AUC-values in the ranges 0.27–0.73 and 0.20–0.75, respectively.

TABLE 1 Baseline patient characteristics split between multiple sclerosis subtype

Characteristic	RRMS	SPMS	PPMS
Participants, *n* (row %)	54 (57.4)	29 (30.9)	11 (11.7)
Age, years, mean (SD)	42.0 (9.8)	53.3 (6.5)	52.5 (9.4)
Sex, *n* (%)			
Female	44 (81.5)	16 (55.2)	8 (72.7)
Male	10 (18.5)	13 (44.8)	3 (27.3)
Level of education, *n* (%)			
Low	2 (3.7)	0 (0.0)	1 (9.1)
Middle	15 (27.8)	13 (44.8)	3 (27.3)
High	37 (68.5)	16 (55.2)	7 (63.6)
Disease duration, years, median (IQR)			
Since diagnosis	4.5 (2.6–10.4)	12.1 (7.3–23.3)	2.9 (0.8–4.7)
Since onset	8.3 (4.4–14.8)	13.9 (9.6–27.9)	6.1 (4.0–11.8)
EDSS, median (IQR)	3.0 (2.5–4.0)	4.0 (3.5–6.0)	4.0 (3.0–5.5)
DMT use, *n* (%)	43 (79.6)[a]	12 (41.4)[b]	2 (18.2)[c]

Abbreviations: DMT, disease-modifying therapy; EDSS, Expanded Disability Status Scale; IQR, interquartile range; PPMS, primary progressive multiple sclerosis; RRMS, relapsing-remitting multiple sclerosis; SD, standard deviation; SPMS, secondary progressive multiple sclerosis.

[a] Dimethyl fumarate, *n* = 10; interferon beta, *n* = 8; glatiramer acetate, *n* = 7; teriflunomide, *n* = 6; ocrelizumab, *n* = 5; fingolimod, *n* = 4; natalizumab, *n* = 2; alemtuzumab, *n* = 1.

[b] Dimethyl fumarate, *n* = 4; glatiramer acetate, *n* = 3; interferon beta, teriflunomide, natalizumab, fingolimod, and ocrelizumab, each *n* = 1.

[c] Ocrelizumab, *n* = 2.

TABLE 2 Clinical outcome measures at baseline and follow-up, and proportion of clinically relevant change

Clinical outcome	*n*	Baseline	3-month follow-up	Patients with clinically relevant change (%)
Disease activity	48[a]			
Gd+ lesions, *n* patients (%)		13 (27.1)[b]	5 (10.4)[c]	31.3
New/enlarged T2 lesions without Gd-enhancement, *n* patients (%)		n.a.	1 (2.1)[d]	n.a.
Relapses, *n* (%)		n.a.	4 (8.3)	8.3
CIS-F, mean (SD)	90	34.1 (11.8)	35.2 (11.6)	25.6
EDSS, median (IQR)	93	3.5 (2.5–4.0)	3.5 (2.5–4.5)	20.4
EDSS Ambulation FS, median (range)	93	1.0 (0–10)	1.0 (0–11)	32.3
AMSQ, median (IQR)	89	35.5 (31.0–45.5)	36.0 (32.0–45.0)	9.8
SDMT, mean (SD)	93	54.1 (10.4)	57.1 (10.5)	51.6

Abbreviations: Ambulation FS, Ambulation Functional System; AMSQ, Arm function in MS Questionnaire; CIS-F, Checklist Individual Strength Fatigue subscale; EDSS, Expanded Disability Status Scale; G+, gadolinium-enhancing; IQR, interquartile range; SD, standard deviation; SDMT, Symbol Digit Modalities Test.

[a] Only patients with relapsing-remitting multiple sclerosis (RRMS).

[b] 1 Gd+ lesion, *n* = 9; 2 Gd+ lesions, *n* = 2; 6 Gd+ lesions, *n* = 2.

[c] 1 Gd+ lesion, *n* = 2; 2 Gd+ lesions, *n* = 1; 4 Gd+ lesions, *n* = 1; 10 Gd+ lesions, *n* = 1.

[d] 2 new T2 lesions.

Comparative responsiveness

When comparing the AUC-values of the keystroke features and clinical benchmarks, only for ambulatory function (AUC = 0.80) did the clinical benchmark have higher responsiveness than the keystroke features (AUC = 0.73). For the remaining study outcomes, keystroke features were more responsive than the clinical benchmarks for 14%–35% of the investigated features (see also Figure 2).

MCID and measurement error

The ROC-curves for each study outcome are shown in Figure 3. The MCID based on the most optimal cut-point for change in radiological disease activity (Gd+ lesions) was higher than the $\mathrm{SRC}_{\mathrm{ind}}$ for After Punctuation Pause feature. This was also true for the features Post-Correction Slowing for arm function (AMSQ) and Flight Time for ambulatory function (EDSS Ambulation FS). In the remaining features with adequate responsiveness, the MCID was lower than the $\mathrm{SRC}_{\mathrm{ind}}$. When considering the measurement error at group level, the MCID exceeded the $\mathrm{SRC}_{\mathrm{group}}$ in 87.9% of keystroke features with an AUC ≥0.70. Table 3 also shows the minimum number (*n*) of patients or repeated measures within individuals needed to detect clinical relevant change beyond measurement error (i.e., MCID = SRC).

DISCUSSION

Assessment tools should be able to measure change in an outcome of interest in the order of magnitude that is considered clinically meaningful [28]. Hence, we analyzed KD for responsiveness to clinically anchored relevant change in clinical outcomes in patients with MS. We found responsiveness of KD to change in disease activity, fatigue, ambulatory, and arm function. Moreover, a significant amount of keystroke features were more responsive to clinical outcomes aside from ambulatory function, compared to commonly used clinical measures. The majority of the responsive keystroke features were able to distinguish this important change from measurement error at group level. At the individual level, three keystroke features were reliably responsive for radiological disease activity, arm function, and ambulation. For the other best-performing keystroke features, two to four repeated measures are sufficient to reduce the measurement error to such an extent that important change can be detected for the individual patient.

Timing-related keystroke features had the highest responsiveness to change in arm function and ambulation, whereas emoji-based features were found to be the most responsive to change in fatigue and cognition. This may be rationalized by the fact that arm function and ambulation are predominantly motor-based outcomes and associated with physical performance of typing. Fatigue and cognition, on the other hand, are non-motor outcomes and relate more strongly to non-physical aspects of typing. Emoji-based features were also responsive to the more general measure of disability, the EDSS, which includes both physical and non-physical functioning. When also looking at the aggregation methods of the most responsive keystroke features for the EDSS outcome, four of the five features were the mean absolute change and the unpredictability of fluctuations (i.e., ‘stability-based’ aggregations). This could be interpreted as the instability of typing or emoji use is indicative of change in overall disability. To a lesser extent, emoji-based features were also responsive in the other study outcomes. This may be similar to the interrelatedness seen in clinical MS outcomes, such

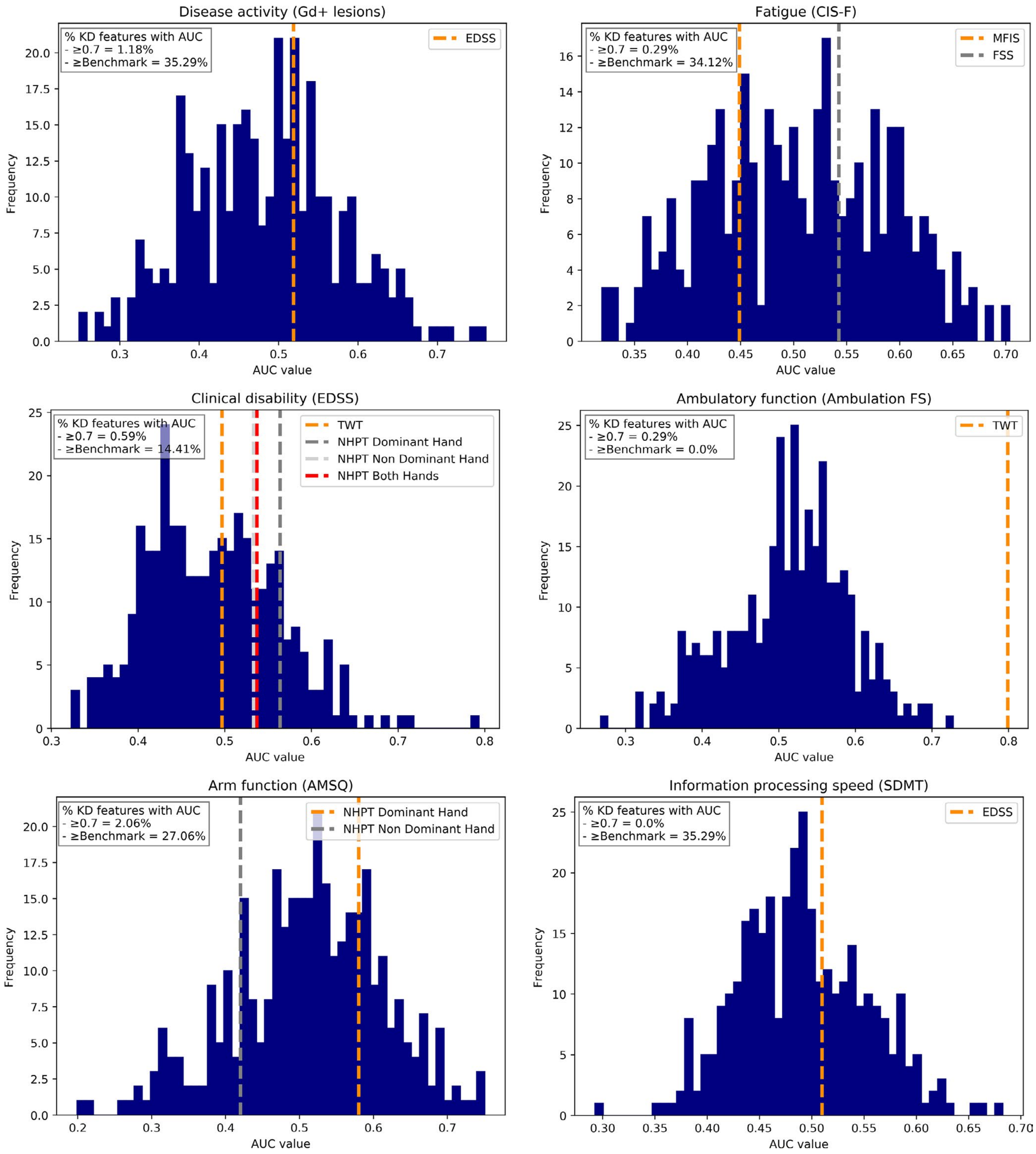

FIGURE 2 Histograms of area under the curve (AUC)-values for all keystroke features for each study outcome. AUC-values of comparative clinical benchmarks are shown as dotted lines. Ambulation FS, EDSS Ambulation Functional System; AMSQ, Arm function in MS Questionnaire; AUC, area under the curve; CIS-F, Checklist Individual Strength Fatigue subscale; EDSS, Expanded Disability Status Scale; FSS, Fatigue Severity Scale; Gd+ lesions, gadolinium-enhancing lesions; KD, keystroke dynamics; MFIS, Modified Fatigue Impact Scale; NHPT, Nine-Hole Peg Test; SDMT, Symbol Digit Modalities Test; TWT, Timed 25-foot Walk Test [Colour figure can be viewed at wileyonlinelibrary.com]

as the EDSS, arm function, and cognition measures. Disease activity is more complex as both physical and non-physical domains may be affected or, in the majority of cases, may only be apparent on MRI. Correspondingly, keystroke features most responsive to change in disease activity were mostly timing-related, but also consisted of more complex time-series aggregation methods and an emoji-based

TABLE 3 Area under the curve values of the clinical benchmarks and the top five keystroke features for each clinical anchor

	Clinical benchmark		Keystroke dynamics					
Clinical anchor	**Measure**	**AUC (95% CI)**	**Feature**[a]	**AUC (95% CI)**	**MCID**	**SRC_{ind}**	**$\text{SRC}_{\text{group}}$**	**n**[b]
Disease activity								
Radiological and/or clinical	EDSS	0.52 (0.34–0.70)	$\text{PPL}_{\text{last maximum}}$	0.76 (0.60–0.91)	0.16	0.22	0.02	2
			$\text{APP}_{\text{first maximum}}$	0.75 (0.58–0.92)	0.24	0.26	0.03	2
			$\text{RRL}_{\text{last maximum}}$	0.71 (0.53–0.89)	0.16	0.21	0.02	2
			ES-neutrality[c]	0.71 (0.53–0.89)	0.01	0.29	0.04	841
			$\text{Session duration}_{\text{absolute energy}}$	0.71 (0.53–0.89)	0.70	0.86	0.09	2
Radiological	EDSS	0.54 (0.35–0.72)	$\text{PPL}_{\text{last maximum}}$	0.78 (0.62–0.95)	0.20	0.23	0.02	2
			$\text{RRL}_{\text{last maximum}}$	0.74 (0.56–0.92)	0.18	0.22	0.02	2
			$\text{Session duration}_{\text{absolute energy}}$	0.74 (0.55–0.92)	0.70	0.87	0.10	2
			$\text{APP}_{\text{first maximum}}$	0.73 (0.52–0.93)	0.28	0.26	0.03	1
			$\text{ES-polarity}_{\text{first maximum}}$	0.72 (0.54–0.91)	0.22	0.25	0.03	2
Fatigue								
CIS-F	FSS	0.54 (0.40–0.68)	$\text{ES-negativity}_{\text{first maximum}}$	0.70 (0.56–0.84)	0.19	0.28	0.03	3
	MFIS	0.45 (0.31–0.58)	Session duration[d]	0.70 (0.56–0.84)	0.01	0.09	0.01	81
			$\text{ES-negativity}_{\text{last maximum}}$	0.69 (0.54–0.83)	0.25	0.43	0.05	3
			$\text{ES-negativity}_{\text{approximate entropy}}$	0.69 (0.51–0.86)	0.09	0.14	0.02	3
			$\text{ES-positivity}_{\text{strike below mean}}$	0.67 (0.52–0.83)	1.00	1.68	0.20	3
Clinical disability								
EDSS	TWT	0.50 (0.34–0.66)	$\text{ES-neutrality}_{\text{approximate entropy}}$	0.79 (0.63–0.96)	0.11	0.14	0.02	2
	NHPT-D	0.56 (0.42–0.71)	$\text{Word length}_{\text{first maximum}}$	0.72 (0.60–0.83)	0.11	0.19	0.02	3
			$\text{ES-positivity}_{\text{approximate entropy}}$	0.70 (0.52–0.88)	0.11	0.16	0.02	3
			$\text{FT}_{\text{approximate entropy}}$	0.68 (0.56–0.80)	0.04	0.08	0.01	4
			$\text{ES-neutrality}_{\text{mean absolute change}}$	0.66 (0.52–0.80)	0.16	0.54	0.06	12
Ambulation FS	TWT	0.80 (0.70–0.90)	$\text{RRL}_{\text{absolute energy}}$	0.73 (0.62–0.84)	3.44	6.86	0.73	4
			$\text{Pre-CS}_{\text{absolute sum of changes}}$	0.70 (0.58–0.81)	2.70	5.77	0.62	5
			$\text{FT}_{\text{strike above mean}}$	0.69 (0.58–0.80)	1.50	1.23	0.13	1
			$\text{FT}_{\text{absolute energy}}$	0.69 (0.57–0.80)	2.72	7.30	0.78	8
			$\text{ES-polarity}_{\text{strike above mean}}$	0.69 (0.55–0.83)	1.00	1.34	0.16	2

(Continues)

TABLE 3 (Continued)

	Clinical benchmark		Keystroke dynamics					
Clinical anchor	**Measure**	**AUC (95% CI)**	**Feature**[a]	**AUC (95% CI)**	**MCID**	**SRC_{ind}**	**SRC_{group}**	***n***[b]
AMSQ	NHPT	0.58 (0.36–0.80)	Pre-CS_{median}	0.75 (0.59–0.92)	113.10	259.54	27.67	6
			Pre-CS_{mean}	0.74 (0.54–0.95)	186.42	281.46	30.00	3
			ES-$negativity_{mean}$	0.74 (0.62–0.86)	0.03	0.05	0.01	3
			Post-CS_{median}	0.73 (0.53–0.93)	135.75	126.24	13.46	1
			ES-$neutrality_{strike\ above\ mean}$	0.72 (0.54–0.90)	1.00	1.50	0.18	3
SDMT	EDSS	0.51 (0.39–0.63)	ES-$positivity_{first\ maximum}$	0.68 (0.56–0.81)	0.09	0.28	0.03	10
			ES-$polarity_{first\ maximum}$	0.66 (0.53–0.79)	0.14	0.26	0.03	4
			$PPL_{strike\ below\ mean}$	0.66 (0.54–0.78)	3.00	5.67	0.60	4
			$CD_{strike\ above\ mean}$	0.63 (0.51–0.75)	1.50	12.90	1.37	74
			Keystroke count	0.63 (0.51–0.75)	3.50	13.29	1.42	15

Abbreviations: AMSQ, Arm function in MS Questionnaire; APP, After Punctuation Pause; AUC, area under the curve; CD, Correction Duration; CI, confidence interval; CIS-F, Checklist Individual Strength Fatigue subscale; EDSS, Expanded Disability Status Scale; ES, Emoji Sentiment; FSS, Fatigue Severity Scale; FT, Flight Time; MCID, minimal clinically important difference; MFIS, Modified Fatigue Impact Scale; NHPT-D, Nine-Hole Peg Test dominant hand; Post-CS, Post-Correction Slowing; PPL, Press-Press Latency; Pre-CS, Pre-Correction Slowing; RRL, Release-Release Latency; SDMT, Symbol Digit Modalities Test; SRC_{group}, smallest real change group; SRC_{ind}, smallest real change individual; TNR, true negative rate; TPR, true positive rate; TWT, Timed 25-foot Walk Test.

[a] Five keystroke features with the highest AUC-values.

[b] Number of patients or repeated measures needed to detect clinical relevant change beyond measurement error on the individual level (i.e., MCID = SRC).

[c] Time reversal asymmetry statistic.

[d] Mean value of a central approximation of the second derivative.

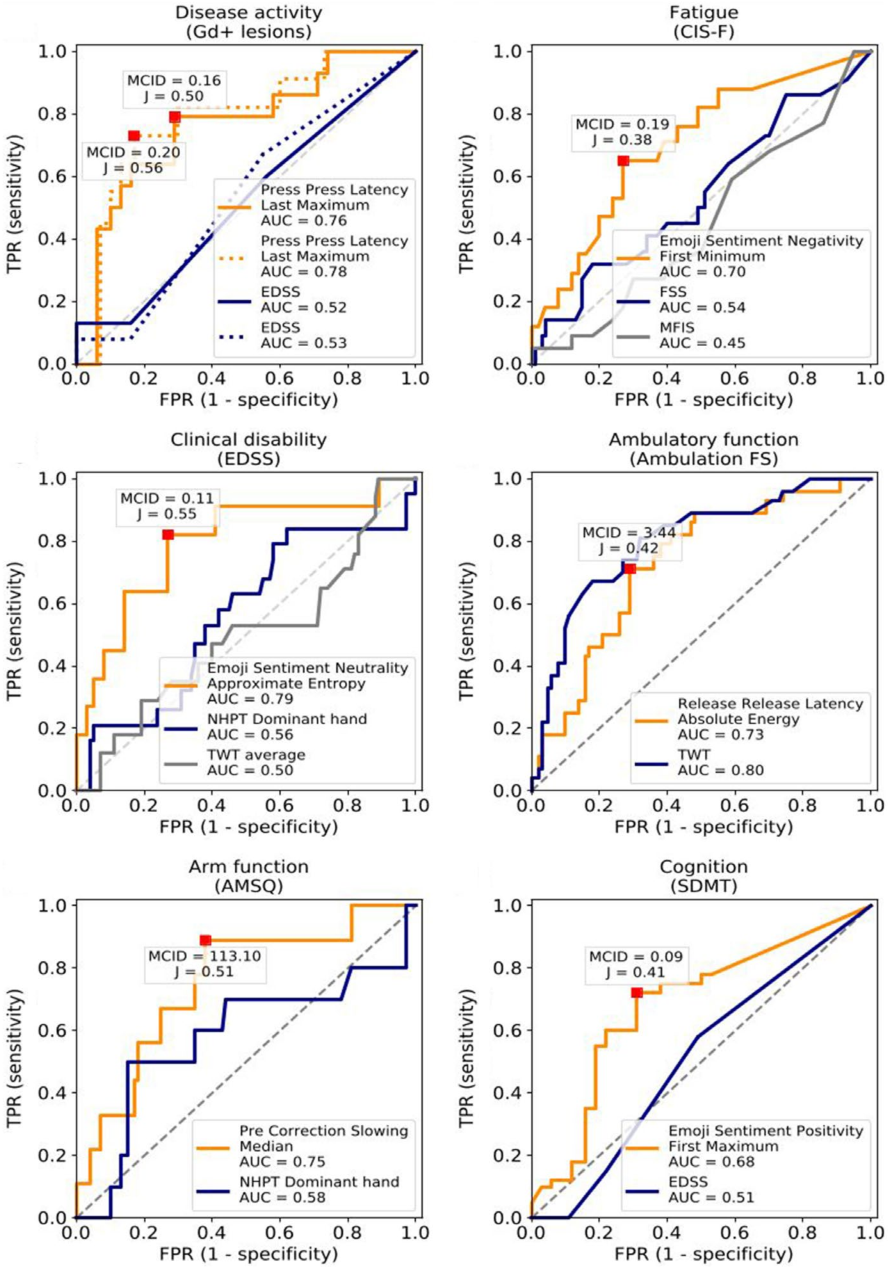

FIGURE 3 Receiver operator characteristic (ROC)-curves for keystroke features with highest area under the curve (AUC)-value for each study outcome. ROC curves of change in keystroke dynamics (KD) (orange curve), clinical benchmarks (blue and grey curves), and the classification of clinically relevant change for all study outcomes. The optimal value (Youden's J statistic) of the keystroke features is shown as a red square with the corresponding minimal clinically important difference (MCID). For disease activity (A), the solid line represents the ROC-curve for radiological and/or clinical activity, and the dotted line represents the ROC-curve for radiological activity only. Ambulation FS, EDSS Ambulation Functional System; AMSQ, Arm function in MS Questionnaire; AUC, area under the curve; CIS-F, Checklist Individual Strength Fatigue subscale; EDSS, Expanded Disability Status Scale; FPR, false positive rate; FSS, Fatigue Severity Scale; Gd+ lesions, gadolinium-enhancing lesions; MCID, minimal clinically important difference; MFIS, Modified Fatigue Impact Scale; NHPT, Nine-Hole Peg Test; SDMT, Symbol Digit Modalities Test; TPR, true positive rate; TWT, Timed 25-foot Walk Test [Colour figure can be viewed at wileyonlinelibrary.com]

feature. To summarize, timing-related keystroke features have high potential as a responsive biomarker for physical functioning and disease activity in MS. Emoji sentiment analysis is of interest for non-physical domains and overall disability. For overall disability, the stability of KD is also of importance when looking at responsiveness to change.

Technological solutions, such as digitalized adaptations, that are able to validly and reliably measure ‘gold standards’ in MS are emerging [29]. However, analysis of responsiveness or even longitudinal assessment of outcome measures in MS remain scarce. Nonetheless, it is important to examine other available clinical and technological biomarkers to contextualize our findings within MS. For classifying patients with RRMS who were stable or had change in disease activity, the Press-Press Latency was most responsive with an AUC-value of 0.76. When only looking at change in gadolinium-enhancing lesions the AUC-value reached 0.78. After Punctuation Pause had an AUC-value of 0.73, but in contrast to the other features could distinguish important change from measurement error on the individual level (i.e., MCID ≥ SRC_{ind}). Currently MRI is the most commonly used biomarker to monitor (subclinical) disease activity in MS. Another promising and more recently emerging method is serum neurofilament light chain (NfL). Serum NfL was associated with gadolinium-enhancing lesions, with each contrast-enhancing lesion corresponding to a 17.8% increase in sNfL [30]. Serum and cerebrospinal fluid NfL for assessment of disease activity (clinical relapse or gadolinium-enhancing lesions) had AUC-values of 0.66 and 0.77, respectively [31] the latter being comparable to our findings with KD.

To date, fatigue is conventionally assessed with patient-reported measures which are valid and reliable, albeit lacking responsiveness [7]. Our current study did indeed show poor responsiveness of the comparative patient-reported fatigue measures (FSS and MFIS). Regarding KD, it can be hypothesized that timing-related features relate to fatigability, as latency metrics in typing are performance-based. Supporting this, we found that the duration of typing session was adequately responsive (AUC = 0.70) to fatigue; other timing-related features, however, were not. In our previous cross-sectional results, poor correlation between timing-related features and fatigue was explained by differing constructs between perceived fatigue and performance-based fatigability [10]. In this manner, subjective features such as emoji use during typing were more prevalent in the most responsive features to change in fatigue. All in all, the use of KD as an outcome measure may strengthen trials investigating interventions on fatigue, in the current absence of sufficient responsive fatigue measures in MS. Considering the extent of disease burden of fatigue and fatigability in MS, further investigations on KD utilizing performance-based fatigability measures are of interest.

For clinical disability outcomes, Release-Release Latency was responsive to change in ambulatory function as assessed by the EDSS Ambulation Function System (AUC = 0.73). The responsiveness of the keystroke feature was found to be lower than the clinical TWT (AUC = 0.80) in our study, despite an earlier report of poor responsiveness of the TWT for patient-perceived walking function [32]. This was not surprising given that typing performance is most likely not reflected by ambulatory function and can be seen as differing constructs. Pre-Correction Slowing (i.e., the latency prior to backspaces) was responsive to change in arm function measured with the AMSQ, with sensitivity and specificity of 0.89 and 0.62, respectively. Where longitudinal studies on responsiveness of disease activity and fatigue measures are extremely limited, the responsiveness of clinical disability measures has been studied more extensively with the TWT, NHPT, and, to a lesser extent, the SDMT. These studies reported low correlations and sensitivity [33–35], and the sensitivity depended on the chosen cut-offs with lower thresholds potentially reflecting noise rather than reliable change [36], and inability to detect important change at the individual level [28]. For most of the keystroke features with highest responsiveness to ambulation, arm function, and cognition, the cut-off value of important change was also within the range of noise on the individual level. Flight Time and Post-Correction Slowing, however, were adequately responsive and could detect important change at the individual level for ambulation and arm function, respectively.

In MS, electronic- and device-based measures exist to facilitate screening, monitoring, self-management, treatment, and education [37]. Digitalized versions of gold standard tasks have been shown to have improved responsiveness on some measures. For instance, a manual finger and foot tapping task had higher sensitivity and specificity to detect 1-year clinical disease progression in comparison to the current gold standards of 9HPT and TWT within an untreated PPMS population [38]. The sensitivity of change detection increased with cases of increased disease severity, or increased assessment duration. Wearables have also been shown to be responsive to clinical change. Within a sample of patients with RRMS, pressure sensors detected a decline in gait performance over 12 months in the absence of EDSS change [39]. In our study, KD were also more responsive than nearly all clinical outcome measures, even despite the shorter duration of follow-up and our cohort having milder disability compared to the aforementioned studies.

Some limitations of our study should be considered when interpreting our findings. In our current study the sample sizes for patients with clinically relevant improvement and worsening in clinical disability measures were relatively small due to the relatively short period of follow-up for disability progression measures. Where changes within a 3-month period are highly relevant for disease activity and fatigue, change in clinical disability is less expected in a cohort with relatively mildly affected patients with MS, outside of any intervention. Hence, improvement and worsening were stratified similarly as changed. We also acknowledge the existence of multiple methods to calculate the MCID, and that different methods may yield different MCID-values [40]. Future studies should apply a method incorporating a point-estimate measure to enable an anchor-based method to calculate the MCID.

From our analysis of a large amount of keystroke features and aggregation methods, we demonstrated responsiveness of KD to important outcomes in MS. The results of this study can guide the future direction in the selection of keystroke features and aggregation methods. Future studies can then select features and aggregation methods a priori and develop composites of keystroke features, possibly coupled with clinical measures, for external validation in order to implement the use of KD in research and clinical practice.

CONCLUSIONS

Short-term change in smartphone KD were found to be sensitive to clinically relevant change in disease activity, patient-reported fatigue, ambulation, and arm function. A large number of keystroke features were responsive for each outcome measure, and most were more responsive than clinical measures. Based on this exploration of the responsiveness of KD, future directions should focus on the standardization and selection of keystroke features to address the external validity in order to further integrate this biomarker into both clinical and research practice in MS. Besides adequate responsiveness, the continuous and passive remote acquisition of objective data in the everyday environment makes this technology-based biomarker highly relevant for disease monitoring and thus helpful in the management of MS.

ACKNOWLEDGMENTS

The authors would like to thank all the patients for their participation in this study.

CONFLICT OF INTEREST

K.H. Lam has no conflicts of interest. J. Twose, H. McConchie, G. Licitra, and K.A. Meijer are employees of Neurocast B.V. (industry partner). L.R.J. de Ruiter, Z.Y.G.J. van Lierop, and B. Moraal have no conflicts of interest. F. Barkhof acts as a consultant to Biogen-Idec, Bayer-Schering, Merck-Serono, Roche, Novartis, IXICO, and Combinostics. He has received sponsorship from EU-H2020, NWO, SMSR, EU-FP7, Novartis, and GE Healthcare. He is on the editorial boards of *Radiology*, *Neuroradiology*, *Multiple Sclerosis Journal*, and *Neurology*. He is supported by the NIHR Biomedical Research Center at UCLH. B.M.J. Uitdehaag received consultancy fees from Biogen Idec, Genzyme, Merck Serono, Novartis, Roche, and Teva. V. de Groot has no conflicts of interest. J. Killestein has accepted speaker and consultancy fees from Merck, Biogen, Teva, Genzyme, Roche, and Novartis.

AUTHOR CONTRIBUTIONS

Ka-Hoo Lam: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal). **James Twose:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Visualization (equal); Writing-review & editing (equal). **Hannah McConchie:** Formal analysis (equal); Project administration (equal); Writing-review & editing (equal). **Giovanni Licitra:** Formal analysis (equal); Methodology (equal); Project administration (equal); Writing-review & editing (equal). **Kim Meijer:** Project administration (equal); Writing-review & editing (equal). **Lodewijk De Ruiter:** Investigation (equal); Writing-review & editing (equal). **Zoë Van Lierop:** Investigation (equal); Writing-review & editing (equal). **Bastiaan Moraal:** Investigation (equal); Writing-review & editing (equal). **Frederik Barkhof:** Methodology (equal); Writing-review & editing (equal). **Bernard M. J. Uitdehaag:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Supervision (equal); Writing-review & editing (equal). **Vincent De Groot:** Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Supervision (equal); Writing-review & editing (equal). **Joep Killestein:** Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Supervision (equal); Writing-review & editing (equal).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Ka-Hoo Lam https://orcid.org/0000-0003-0926-1445
Hannah McConchie https://orcid.org/0000-0002-6422-4447

REFERENCES

1. Sormani MP, Gasperini C, Romeo M, et al. Assessing response to interferon-β in a multicenter dataset of patients with MS. *Neurology*. 2016;87(2):134-140.
2. Chiaravalloti ND, DeLuca J. Cognitive impairment in multiple sclerosis. *Lancet Neurol*. 2008;7(12):1139-1151.
3. Mäurer M, Comi G, Freedman MS, et al. Multiple sclerosis relapses are associated with increased fatigue and reduced health-related quality of life – a post hoc analysis of the TEMSO and TOWER studies. *Mult Scler Relat Disord*. 2016;7:33-40.
4. Barkhof F, Scheltens P, Frequin ST, et al. Relapsing-remitting multiple sclerosis: sequential enhanced MR imaging vs clinical findings in determining disease activity. *AJR Am J Roentgenol*. 1992;159(5):1041-1047.
5. Wattjes MP, Rovira À, Miller D, et al. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis–establishing disease prognosis and monitoring patients. *Nat Rev Neurol*. 2015;11(10):597-606.
6. Uitdehaag BMJ. Disability outcome measures in phase III clinical trials in multiple sclerosis. *CNS Drugs*. 2018;32(6):543-558.
7. Rietberg MB, Van Wegen EE, Kwakkel G. Measuring fatigue in patients with multiple sclerosis: reproducibility, responsiveness and concurrent validity of three Dutch self-report questionnaires. *Disabil Rehabil*. 2010;32(22):1870-1876.
8. Faissner S, Plemel JR, Gold R, Yong VW. Progressive multiple sclerosis: from pathophysiology to therapeutic strategies. *Nat Rev Drug Discovery*. 2019;18(12):905-922.
9. Barro C, Leocani L, Leppert D, Comi G, Kappos L, Kuhle J. Fluid biomarker and electrophysiological outcome measures for progressive MS trials. *Mult Scler J*. 2017;23(12):1600-1613.
10. Lam KH, Meijer KA, Loonstra FC, et al. Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis. *Mult Scler J*. 2020;27(9):1421-1431.
11. Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of emojis. *PLoS One*. 2015;10(12):e0144296.
12. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge University Press; 2011.
13. Neurocast BV. [Internet]. 2018 [cited 2020 July 15]. https://neurokeys.app/. Accessed July 15, 2020.
14. Bermel RA, Naismith RT. Using MRI to make informed clinical decisions in multiple sclerosis care. *Curr Opin Neurol*. 2015;28(3):244-249.
15. Vercoulen JH, Hommes OR, Swanink CM, et al. The measurement of fatigue in patients with multiple sclerosis. A multidimensional comparison with patients with chronic fatigue syndrome and healthy subjects. *Arch Neurol*. 1996;53(7):642-649.
16. Worm-Smeitink M, Gielissen M, Bloot L, et al. The assessment of fatigue: psychometric qualities and norms for the checklist individual strength. *J Psychosom Res*. 2017;98:40-46.
17. Kappos L, Butzkueven H, Wiendl H, et al. Greater sensitivity to multiple sclerosis disability worsening and progression events using a roving versus a fixed reference value in a prospective cohort study. *Mult Scler J*. 2018;24(7):963-973.
18. Mokkink LB, Knol DL, van der Linden FH, Sonder JM, D'Hooghe M, Uitdehaag BMJ. The arm function in multiple sclerosis questionnaire (AMSQ): development and validation of a new tool using IRT methods. *Disabil Rehabil*. 2015;37(26):2445-2451.
19. van Munster CE, Kaya L, Obura M, Kalkers NF, Uitdehaag BM. Minimal clinically important difference of improvement on the arm function in Multiple Sclerosis Questionnaire (AMSQ). *Mult Scler*. 2020;26(4):505-508.
20. Benedict RH, DeLuca J, Phillips G, LaRocca N, Hudson LD, Rudick R. Validity of the symbol digit modalities test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler*. 2017;23(5):721-733.
21. Kafadar K. John Tukey and robustness. *Statist Sci*. 2003;18(3):319-331.
22. Sormani MP, Bonzano L, Roccatagliata L, Mancardi GL, Uccelli A, Bruzzi P. Surrogate endpoints for EDSS worsening in multiple sclerosis. A meta-analytic approach. *Neurology*. 2010;75(4):302-309.
23. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol*. 1989;46(10):1121-1123.

24. Kos D, Kerckhofs E, Nagels G, et al. Assessing fatigue in multiple sclerosis: Dutch modified fatigue impact scale. *Acta Neurol Belg*. 2003;103(4):185-191.
25. Hobart J, Blight AR, Goodman A, Lynn F, Putzki N. Timed 25-foot walk: direct evidence that improving 20% or greater is clinically meaningful in MS. *Neurology*. 2013;80(16):1509-1517.
26. Feys P, Lamers I, Francis G, et al. The Nine-Hole Peg test as a manual dexterity performance measure for multiple sclerosis. *Mult Scler*. 2017;23(5):711-720.
27. de Boer MR, de Vet HCW, Terwee CB, Moll AC, Völker-Dieben HJM, van Rens GHMB. Changes to the subscales of two vision-related quality of life questionnaires are proposed. *J Clin Epidemiol*. 2005;58(12):1260-1268.
28. de Groot V, Beckerman H, Uitdehaag BM, et al. The usefulness of evaluative outcome measures in patients with multiple sclerosis. *Brain*. 2006;129(Pt 10):2648-2659.
29. Inojosa H, Schriefer D, Ziemssen T. Clinical outcome measures in multiple sclerosis: a review. *Autoimmun Rev*. 2020;19(5):102512.
30. Barro C, Benkert P, Disanto G, et al. Serum neurofilament as a predictor of disease worsening and brain and spinal cord atrophy in multiple sclerosis. *Brain*. 2018;141(8):2382-2391.
31. Novakova L, Zetterberg H, Sundström P, et al. Monitoring disease activity in multiple sclerosis using serum neurofilament light protein. *Neurology*. 2017;89(22):2230-2237.
32. Baert I, Freeman J, Smedal T, et al. Responsiveness and clinically meaningful improvement, according to disability level, of five walking measures after rehabilitation in multiple sclerosis: a European multicenter study. *Neurorehabil Neural Repair*. 2014;28(7):621-631.
33. Goldman MD, LaRocca NG, Rudick RA, et al. Evaluation of multiple sclerosis disability outcome measures using pooled clinical trial data. *Neurology*. 2019;93(21):e1921-e1931.
34. van Winsen LM, Kragt JJ, Hoogervorst EL, Polman CH, Uitdehaag BM. Outcome measurement in multiple sclerosis: detection of clinically relevant improvement. *Mult Scler J*. 2010;16(5):604-610.
35. Kragt JJ, Thompson AJ, Montalban X, et al. Responsiveness and predictive value of EDSS and MSFC in primary progressive MS. *Neurology*. 2008;70(13 Pt 2):1084-1091.
36. Schwid SR, Goodman AD, Apatoff BR, et al. Are quantitative functional measures more sensitive to worsening MS than traditional measures? *Neurology*. 2000;55(12):1901-1903.
37. Yousef A, Jonzzon S, Suleiman L, Arjona J, Graves JS. Biosensing in multiple sclerosis. *Expert Rev Med Devices*. 2017;14(11):901-912.
38. Tanigawa M, Stein J, Park J, Kosa P, Cortese I, Bielekova B. Finger and foot tapping as alternative outcomes of upper and lower extremity function in multiple sclerosis. *Mult Scler J Exp Transl Clin*. 2017;3(1):2055217316688930.
39. Galea MP, Cofré Lizama LE, Butzkueven H, Kilpatrick TJ. Gait and balance deterioration over a 12-month period in multiple sclerosis patients with EDSS scores ≤3.0. *NeuroRehabilitation*. 2017;40(2):277-284.
40. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther*. 2012;20(3):160-166.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

App S1

How to cite this article: Lam K-H, Twose J, McConchie H, et al. Smartphone-derived keystroke dynamics are sensitive to relevant changes in multiple sclerosis. *Eur J Neurol*. 2022;29:522–534. https://doi.org/10.1111/ene.15162