



# Associations between smartphone keystroke dynamics and cognition in MS

Digital Health  
Volume 8: 1–12  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076221143234  
journals.sagepub.com/home/dhj



Michelle H Chen<sup>1,2</sup> , Alex Leow<sup>3</sup>, Mindy K Ross<sup>3</sup>, John DeLuca<sup>4,5</sup>,  
Nancy Chiaravalloti<sup>4,5</sup> , Silvana L Costa<sup>4,5</sup>, Helen M Genova<sup>4,5</sup>,  
Erica Weber<sup>4,5</sup>, Faraz Hussain<sup>3</sup> and Alexander P Demos<sup>3</sup>

## Abstract

**Objective:** Examine the associations between smartphone keystroke dynamics and cognitive functioning among persons with multiple sclerosis (MS).

**Methods:** Sixteen persons with MS with no self-reported upper extremity or typing difficulties and 10 healthy controls (HCs) completed six weeks of remote monitoring of their keystroke dynamics (i.e., how they typed on their smartphone keyboards). They also completed a comprehensive neuropsychological assessment and symptom ratings about fatigue, depression, and anxiety at baseline.

**Results:** A total of 1,335,787 keystrokes were collected, which were part of 30,968 typing sessions. The MS group typed slower ( $P < .001$ ) and more variably ( $P = .032$ ) than the HC group. Faster typing speed was associated with better performance on measures of processing speed ( $P = .016$ ), attention ( $P = .022$ ), and executive functioning (cognitive flexibility:  $P = .029$ ; behavioral inhibition:  $P = .002$ ; verbal fluency:  $P = .039$ ), as well as less severe impact from fatigue ( $P < .001$ ) and less severe anxiety symptoms ( $P = .007$ ). Those with better cognitive functioning and less severe symptoms showed a stronger correlation between the use of backspace and autocorrection events ( $P < .001$ ).

**Conclusion:** Typing speed may be sensitive to cognitive functions subserved by the frontal-subcortical brain circuits. Individuals with better cognitive functioning and less severe symptoms may be better at monitoring their typing errors. Keystroke dynamics have the potential to be used as an unobtrusive remote monitoring method for real-life cognitive functioning among persons with MS, which may improve the detection of relapses, evaluate treatment efficacy, and track disability progression.

## Keywords

Multiple sclerosis, typing, digital phenotyping, mHealth, passive monitoring

Submission date: 4 August 2022; Acceptance date: 17 November 2022

## Introduction

Multiple sclerosis (MS) is a progressive, demyelinating disease of the central nervous system. The most common disease course of MS is relapsing-remitting, which is characterized by episodes of relapses (i.e., sudden worsening of neurological functioning lasting for at least 24 h) followed by periods of remission (i.e., neurological functioning returning to baseline). However, even among persons with relapsing-remitting MS, levels of disability increase over time, and most individuals eventually convert to a progressive disease course.<sup>1</sup>

<sup>1</sup>Institute for Health, Health Care Policy and Aging Research, Rutgers University, New Brunswick, NJ, USA

<sup>2</sup>Department of Neurology, Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ, USA

<sup>3</sup>Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, USA

<sup>4</sup>Kessler Foundation, East Hanover, NJ, USA

<sup>5</sup>Department of Physical Medicine and Rehabilitation, New Jersey Medical School, Rutgers University, Newark, NJ, USA

### Corresponding authors:

Michelle H Chen, Institute for Health, Health Care Policy and Aging Research, Rutgers University, 112 Paterson St, New Brunswick, NJ 08901, USA.

Email: michelle.chen2@rutgers.edu

Alex Leow, Department of Psychiatry, University of Illinois at Chicago, 1601 W. Taylor St., SPHPI MC 912, Chicago, IL 60612, USA.

Email: alexfeuillet@gmail.com



Cognitive decline may indicate increased disease activity (e.g., clinical relapse)<sup>2</sup> and disability progression<sup>3</sup> among persons with MS. Therefore, routine monitoring of cognition is a vital component of comprehensive MS care.<sup>4</sup> Unfortunately, the current gold standard approach of evaluating cognitive functioning—neuropsychological assessment—is time-consuming (i.e., taking hours) and prone to practice effects (i.e., performing better on repeated cognitive assessments due to familiarity with the test materials),<sup>5</sup> which makes it impractical to be conducted more than once per year. Thus, there is a need to develop continuous monitoring methods that are sensitive to real-time changes in cognition (e.g., during a relapse), which may complement traditional neuropsychological exams and lead to more timely interventions.

Digital phenotyping is an emerging field that aims to characterize real-life internal state and behavior at the individual level through digital devices such as smartphones. Using digital phenotyping, researchers have been able to objectively quantify real-life behaviors and situations, such as social isolation (based on call and text logs, global positional system [GPS]), physical activity and sleep (through accelerometer, gyroscope, screen on/off status), and situational contexts (using cameras, GPS).<sup>6,7</sup> Keystroke dynamics (i.e., how individuals type on smartphone keyboards) is a promising digital phenotyping approach and may be used to infer real-life cognitive functioning continuously and unobtrusively. Keystroke dynamics (e.g., slower typing speed, typing inconsistency) have been associated with motor functioning among individuals with Parkinson's<sup>8,9</sup> and Huntington's<sup>10</sup> diseases (see recent meta-analysis<sup>11</sup> on the associations between keystroke dynamics and fine motor functions) and depressive and manic symptoms among individuals with bipolar disorder.<sup>12,13</sup> In MS, keystroke dynamics have been linked with performances on motor and cognitive tests, brain lesions, and neurological disability.<sup>14,15</sup>

Although these studies show keystroke dynamics as a potential clinical tool for remote monitoring, there has not been a comprehensive investigation into the associations between keystroke dynamics and various cognitive domains. Existing keystroke studies in MS have only used a brief cognitive screener measuring processing speed.<sup>14,15</sup> Studies in populations with cognitive impairment as a primary feature (e.g., mild cognitive impairment and dementia) focus on classification between clinical and control groups<sup>16,17</sup>; thus, associations between keystroke features and specific cognitive domains remain unknown. The goal of the current study was to provide proof-of-concept evidence that keystroke dynamics have the potential to be used as an unobtrusive remote monitoring method for real-life cognitive functioning. Specifically, the study explored differences in keystroke dynamics between persons with MS and healthy controls (HCs) and associations between keystroke dynamics and a *range* of cognitive and symptom measures.

## Methods

### Participants

Adult participants with MS and HCs were recruited from six parent studies with comprehensive neuropsychological assessment. Across these studies, exclusion criteria were any neurological or neurodevelopmental conditions other than MS, the presence of MS relapse symptoms within a month prior to study participation, and use of drugs and medications that may affect cognitive functioning. One of the studies was a cognitive rehabilitation study, so it only included individuals with MS who scored below cutoff on an episodic memory measure. An additional exclusion criterion for the current study was severe upper extremity motor difficulties that interfered with typing. Initially, only individuals who owned an iPhone were recruited because the study app (BiAffect) was only available for iOS devices. Subsequently, an Android version of the app was developed, and individuals with Android devices were also recruited.

### Procedures

The study was approved by the Kessler Foundation Institutional Review Board. Since the current investigation used baseline neuropsychological test data that were collected for the parent studies, participants only needed to complete six weeks of remote keystroke monitoring for the current study. Informed consent was obtained either verbally by phone or via an online survey using Research Electronic Data Capture (REDCap) tools.<sup>18,19</sup> Participants were instructed on downloading and using the BiAffect app on their personal smartphone. For iPhone users, the app was available for download from the Apple app store. For Android users, they were emailed an apk file, which they used to install the app on their phones. There were no active tasks that participants needed to complete; participants were told to simply type on their phones as they would normally type for six weeks. Using participants' personal smartphones, instead of study phones, ensured that we were capturing their typical typing patterns. To protect their privacy, participants were told that the app only tracked the metadata associated with their typing (such as typing speed and errors) and not the content of what they typed (e.g., words, sentences). They were also informed that the app would not access any of their personal information on their phones. Keystroke data were monitored weekly by study personnel. If there were missing data for several days, study personnel would call the participant to troubleshoot any technical issues, and data collection time would be lengthened to make up for the missing days with the permission of the participant.

### Smartphone app

Keystroke dynamics was monitored using the BiAffect app. For more details on the app, refer to previous publications

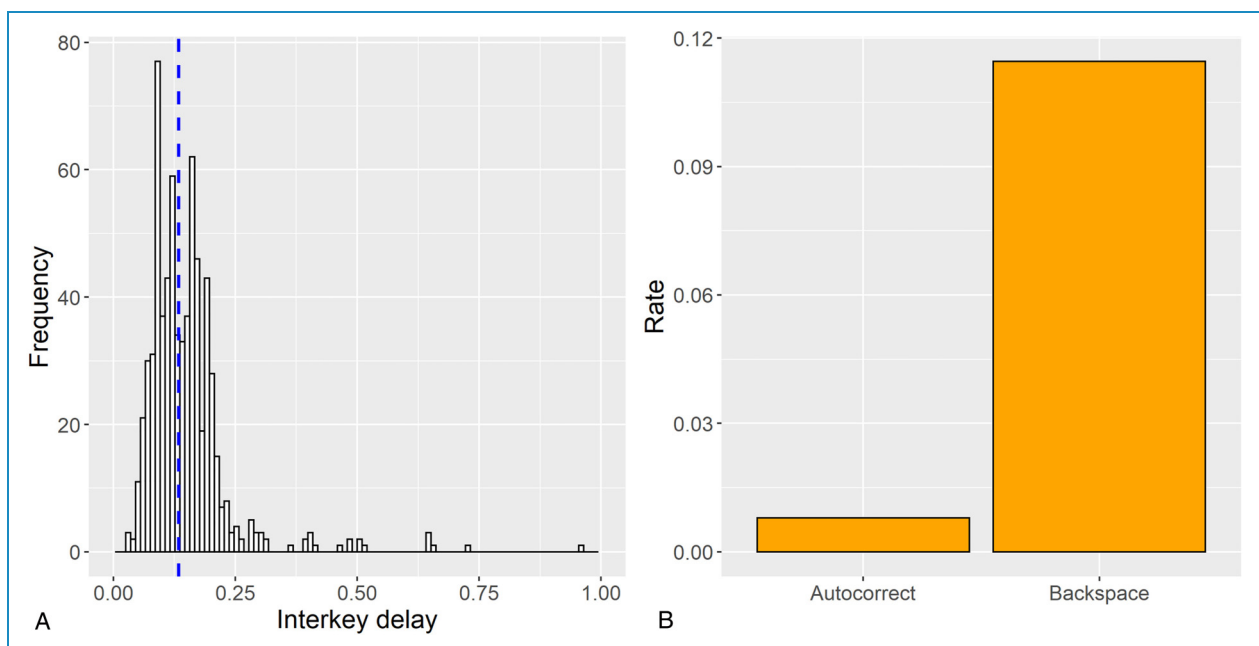
on the current<sup>20,21</sup> and previous pilot<sup>12,13,22</sup> versions of the app. In brief, BiAffect consists of a custom virtual keyboard that temporarily replaces the native smartphone keyboard. The app records each keystroke, with information on keypress type (e.g., alphanumeric, backspace, autocorrection), timestamp, relative distance between two consecutive keystrokes, distance between the keystroke and the center of the keyboard, and accelerometer. Keystrokes were organized into typing sessions. Operationally, one typing session was initiated when the keyboard was activated and terminated after 6 s of inactivity, or at the time of keyboard deactivation. The 6-s threshold was implemented to avoid collecting very long sessions with just the accelerometer but no keypresses, which could drain the participant's phone battery. Latency between consecutive keystrokes was calculated as interkey delay, and median interkey delay within each session was used to represent typing speed for that session (median instead of mean was used because interkey delay was positively skewed). We only included transitions between alphanumeric characters when computing session-level median interkey delay to accurately characterize typical typing speed (latencies may be larger when using backspace, punctuations, and special characters). For typing speed variability within the session, we calculated the median absolute deviation (MAD) of interkey delay. We also derived rates of autocorrection events and use of backspaces within each session to

signify the frequency of typing errors and self-monitoring of typing errors, respectively. See Figure 1 for an illustration of keystroke metrics.

In addition, we evaluated whether each typing session was conducted using two hands or one hand only based on our previously validated algorithm.<sup>20</sup> To determine this, interkey distance (distance from one keystroke to the next on the keyboard) was regressed on interkey delay. If the coefficient was positive and statistically significant ( $p < 0.05$ ), the typing session was labeled as one-handed. The intuition was that when typing with one hand, it would take longer to transition between keystrokes that were further away from each other on the keyboard. The rest of the sessions were labeled as two-handed because it would be quicker to press keys on opposite ends of the keyboard with two hands. To ensure stable estimates, only typing sessions with at least 20 keystrokes were used to derive the handedness measure.

### Neuropsychological and symptom rating measures

Since we used neuropsychological test and symptom rating data from six parent studies, the studies varied slightly in the measures used. Cognitive domains assessed by neuropsychological tests included processing speed (Symbol Digit Modalities Test [SDMT]<sup>23</sup>), attention and working memory (digit span<sup>24,25</sup>), cognitive flexibility (Trail-Making Test



**Figure 1.** Keystroke metrics. Panel A illustrates the frequency distribution of interkey delay (latency between consecutive keypresses) during transitions between alphanumeric characters in a sample typing session, with the blue dotted line denoting the median. Given the skewness of interkey delay, the median value is used to represent typing speed for each typing session. Median absolute deviation is used to represent variability in typing speed. Panel B illustrates rates of autocorrection and backspace in a sample typing session. In multilevel models, rates of autocorrection and backspace are represented with the total number of autocorrection events and backspaces used as the outcomes and offset by a log-transformed character count for each typing session.

[TMT]<sup>26</sup>), behavioral inhibition (Delis-Kaplan Executive Function System [D-KEFS]<sup>27</sup> Color-Word Interference Test [CWIT]), verbal fluency (Controlled Oral Word Association Test<sup>28</sup>/animal fluency<sup>29</sup> or D-KEFS<sup>27</sup> Verbal Fluency Test), and episodic memory (California Verbal Learning Test-II,<sup>30</sup> Rey Auditory Verbal Learning Test,<sup>31</sup> or Hopkins Verbal Learning Test-Revised<sup>32</sup>). Symptom rating scales included the Modified Fatigue Impact Scale (MFIS),<sup>33</sup> Chicago Multiscale Depression Inventory (CMDI),<sup>34</sup> and State-Trait Anxiety Inventory (STAI).<sup>35</sup> In order to compare different measures for each domain across studies, demographically adjusted z scores were calculated based on published norms. To reduce the total

number of comparisons, z-scores for different parts of the same test were averaged into a composite (e.g., immediate and delayed recall of an episodic memory measure); we verified that findings were similar when individual components were run in separate models. Supplemental Table 1 lists the measures used in each of the six parent studies. Supplemental Table 2 provides additional details on each measure, including test versions and data transformations.

### Statistical analysis

All data analyses were conducted in R<sup>36</sup> version 4.1.3. Group differences (MS vs. HC) in demographics were

**Table 1.** Summary of demographic, clinical, and keystroke dynamics.

	MS (n=16)	HC (n=10)	t/ $\chi^2$ /MLM estimate, p
Age: mean years (SD)	48.06 (9.55)	46.70 (17.13)	-0.23, 0.822
Sex			Fisher's, 0.644
Female: number (%)	13 (81)	7 (70)	
Male: number (%)	3 (19)	3 (30)	
Education: mean years (SD)	16.00 (2.03)	15.60 (2.63)	-0.41, 0.687
MS disease duration: mean years (SD)	14.42 (9.41)	N/A	N/A
MS disease course			N/A
Relapsing-remitting: number (%)	10 (62)	N/A	
Primary progressive: number (%)	3 (19)	N/A	
Secondary progressive: number (%)	3 (19)	N/A	
Operating system			Fisher's, 0.625
iOS: number (%)	14 (87)	8 (80)	
Android: number (%)	2 (13)	2 (20)	
Number of days of data collection: mean number (SD)	40.88 (16.56)	30.6 (11.63)	-1.86, 0.076
Proportion of time using one hand to type: %*	38	19	542.70, p < 0.001
Number of characters per typing session: mean number (SD)	38.01 (50.02)	63.30 (76.39)	0.01, 0.107
Median interkey delay (typing speed) per session: seconds (SD)*	0.46 (0.40)	0.26 (0.21)	-1.45, <0.001
Interkey delay median absolute deviation per session: seconds (SD)*	0.19 (0.22)	0.11 (0.14)	0.11, 0.032

Note. Welch's two-sample t test and Pearson's chi-squared test (or Fisher's exact tests for count data for comparisons with expected frequency <5) were used to examine group differences in demographic variables. MLM with keystroke sessions nested within persons were used to examine group differences in keystroke dynamics, adjusted for age, sex, years of formal education, operating system, and use of one or two hands (see Supplemental Table 3 for more details).

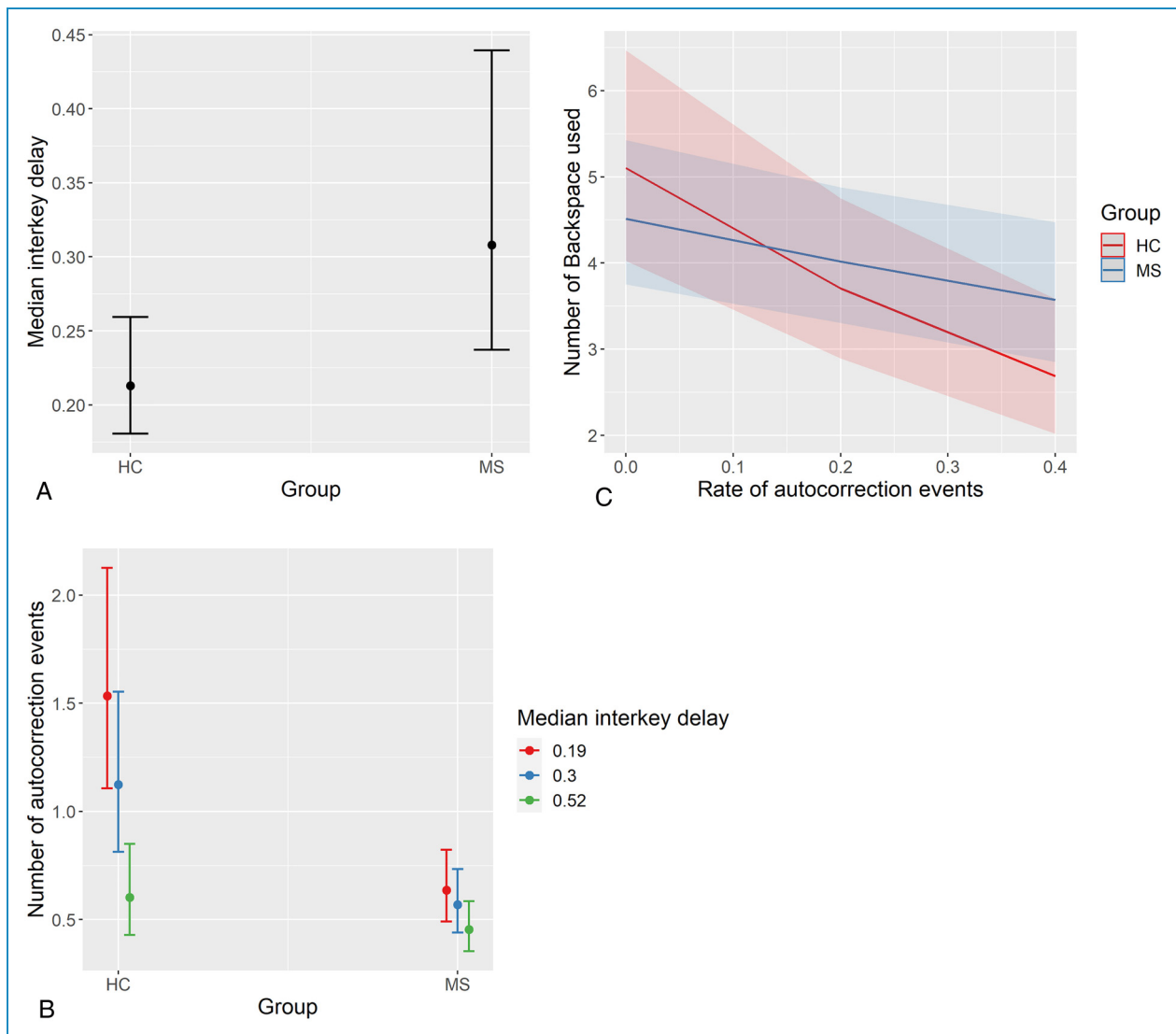
MS: multiple sclerosis; HC: healthy control; SD: standard deviation; MLM: generalized multilevel model.

\*Statistically significant difference between MS and HC groups.

evaluated using Welch's *t* test for continuous variables and Pearson's chi-squared test for categorical variables (or Fisher's exact tests for count data for comparisons with expected frequency of less than five). Multilevel models (MLMs) were used to model keystroke dynamics within typing sessions (level 1) which were nested within subjects (level 2). Supplemental Table 3 details all MLM model specifications. Because median interkey delay yielded non-normal residuals, we used a Gamma distribution as it is more suitable for nonzero, positively skewed reaction time-type data.<sup>37</sup> A Gaussian distribution was used for interkey delay MAD because it contained zero values, so we were

unable to use Gamma distribution. For rates of autocorrection events and use of backspaces, we performed Poisson generalized MLMs, with the count of autocorrection/backspace events per session as dependent variables and offset with the log-transformed character count per session (in order to signify proportions).

Group differences in keystroke dynamics were determined using a level-2 diagnostic group fixed effect. Associations between keystroke dynamics and neuropsychological test and symptom rating scores across both MS and HC groups were assessed with level 2 neuropsychological test/symptom scores as fixed effects (each



**Figure 2.** Group differences in keystroke dynamics. Panel A displays differences between MS and HC groups in median interkey delay (typing speed); the MS group typed significantly slower. Panel B displays moderating effect of median interkey delay on number of autocorrection events (offset with number of characters per session); interkey delay is split into lower, median, and upper quartiles for visualization. The HC group had more autocorrection events than the MS group, but only among those who typed faster. Panel C displays the moderating effect of diagnostic group on the relationship between backspace and autocorrection rates; the relationship was steeper among HC than MS. Plots represent predicted values from multilevel models, and error bars/bands represent 95% confidence intervals. MS: multiple sclerosis; HC: healthy control.

score in separate model). Since the number of autocorrection events was dependent on typing speed (faster speed, more autocorrection events), we examined interactions

**Table 2.** Associations between median interkey delay (typing speed) per session and baseline neuropsychological test performance and symptom ratings.

Measure	N	Median interkey delay main effect
SDMT	25	Estimate = 0.34, $p = 0.016^*$
Digit span	25	Estimate = 0.62, $p = 0.022^*$
TMT composite	16	Estimate = 0.35, $p = 0.029^*$
CWIT	19	Estimate = 1.27, $p = 0.002^*$
Verbal fluency composite	18	Estimate = 0.46, $p = 0.039^*$
Episodic memory composite	19	Estimate = 0.06, $p = 0.89$
MFIS	15	Estimate = $-1.87$ , $p < 0.001^*$
CMDI	22	Estimate = $-0.33$ , $p = 0.090$
STAI composite	19	Estimate = $-0.74$ , $p = 0.007^*$

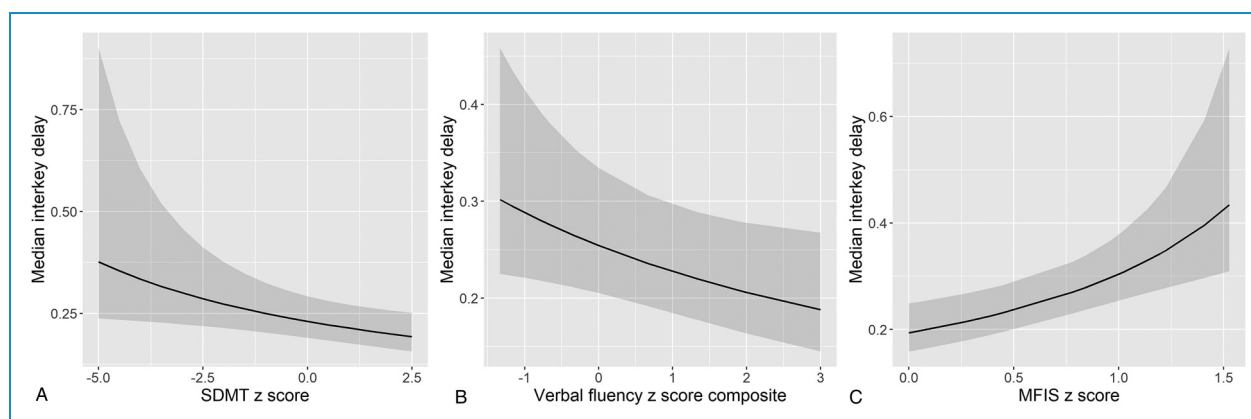
*Note.* Estimates were from generalized multilevel models with Gamma distributions. Analyses were adjusted for age, sex, years of formal education, operating system, and use of one or two hands (see Supplemental Table 3 for more details). All neuropsychological and symptom scores are standardized z scores based on demographics (see Supplemental Table 2 for more information).

SDMT: Symbol Digit Modalities Test; TMT: Trail-Making Test; CWIT: Color-Word Interference; MFIS: Modified Fatigue Impact Scale; CMDI: Chicago Multiscale Depression Inventory; STAI: State-Trait Anxiety Inventory. \*Statistically significant associations.

between the main predictor (group, neuropsychological test, or symptom rating) and median interkey delay in predicting autocorrection events. To elucidate how individuals monitored their typing errors, we also tested the interaction between rate of autocorrection events and each main predictor (group, neuropsychological test, or symptom rating) in predicting number of backspaces used. All MLMs included a random intercept for the subject and adjusted for age, sex, years of formal education, operating system, and use of one or two hands as fixed effects when feasible (some covariates were removed for some models due to failed model convergence). Since a portion of our subjects ( $n=6$ ) underwent remote instead of in-person neuropsychological assessment, we examined whether remote test administration moderated these associations (see Supplemental Table 4).

## Results

Thirty-three participants were recruited. Seven dropped out prematurely due to difficulties with using the BiAffect keyboard. Specifically, they noted that the autocorrection algorithm of the BiAffect keyboard was less accurate than the native smartphone keyboard. This resulted in a final sample of 16 persons with MS and 10 HCs. There were no demographic differences between those who dropped out and those who remained in the study ( $P > .05$ ). A total of 1,335,787 keystrokes were collected, which were part of 30,968 typing sessions. Mean number of days with recorded keystrokes was 37 days (no significant differences between groups,  $P > .05$ ). The ICCs were 65%, 33%, and 37% for median interkey delay, rate of autocorrection event, and rate of backspace used, respectively, which justified the use of MLMs. Lower median interkey delay (faster typing speed) was associated with more autocorrection events (estimate =  $-1.34$ ,  $P < .001$ ) and greater use of



**Figure 3.** Associations between median interkey delay and neuropsychological performance/symptom rating. Plots represent predicted values from multilevel models, and error bands represent 95% confidence intervals. Faster typing speed was associated with better performance on neuropsychological tests (e.g., SDMT and verbal fluency in Panels A and B and less severe symptom ratings (e.g., MFIS in Panel C). SDMT: Symbol Digit Modalities Test; MFIS: Modified Fatigue Impact Scale.

**Table 3.** Moderating effects of median interkey delay on associations between autocorrection events per session and baseline neuropsychological test performance and symptom ratings.

Measure	N	Neuropsychological performance/ symptom rating $\times$ Median interkey delay
SDMT	25	Estimate = $-0.34$ , $p < 0.001^*$
Digit span	25	Estimate = $-0.62$ , $p < 0.001^*$
TMT composite	16	Estimate = $-0.75$ , $p < 0.001^*$
CWIT	19	Estimate = $-1.60$ , $p < 0.001^*$
Verbal fluency composite	18	Estimate = $-1.07$ , $p < 0.001^*$
Episodic memory composite	19	Estimate = $-0.53$ , $p = 0.002^*$
MFIS	15	Estimate = $-0.16$ , $p = 0.502$
CMDI	22	Estimate = $-0.60$ , $p < 0.001^*$
STAI composite	19	Estimate = $0.40$ , $p < 0.001^*$

Note. Models were generalized multilevel models with keystroke sessions nested within persons and Poisson distributions. Estimates were from interactions between neuropsychological test performance/symptom rating and median interkey delay, predicting autocorrection events per session. Models were offset with the log-transformed number of characters per session and adjusted for age, sex, years of formal education, operating system, and use of one or two hands (see Supplemental Table 3 for more details). All neuropsychological and symptom scores are standardized z scores based on demographics (see Supplemental Table 2 for more information).

SDMT: Symbol Digit Modalities Test; TMT: Trail-Making Test; CWIT: Color-Word Interference; MFIS: Modified Fatigue Impact Scale; CMDI: Chicago Multiscale Depression Inventory; STAI: State-Trait Anxiety Inventory. \*Statistically significant interactions.

backspace ( $-0.12$ ,  $P < .01$ ). More autocorrection events were associated with less frequent backspace used (estimate =  $-0.96$ ,  $P < .001$ ).

### Differences between MS and HC groups

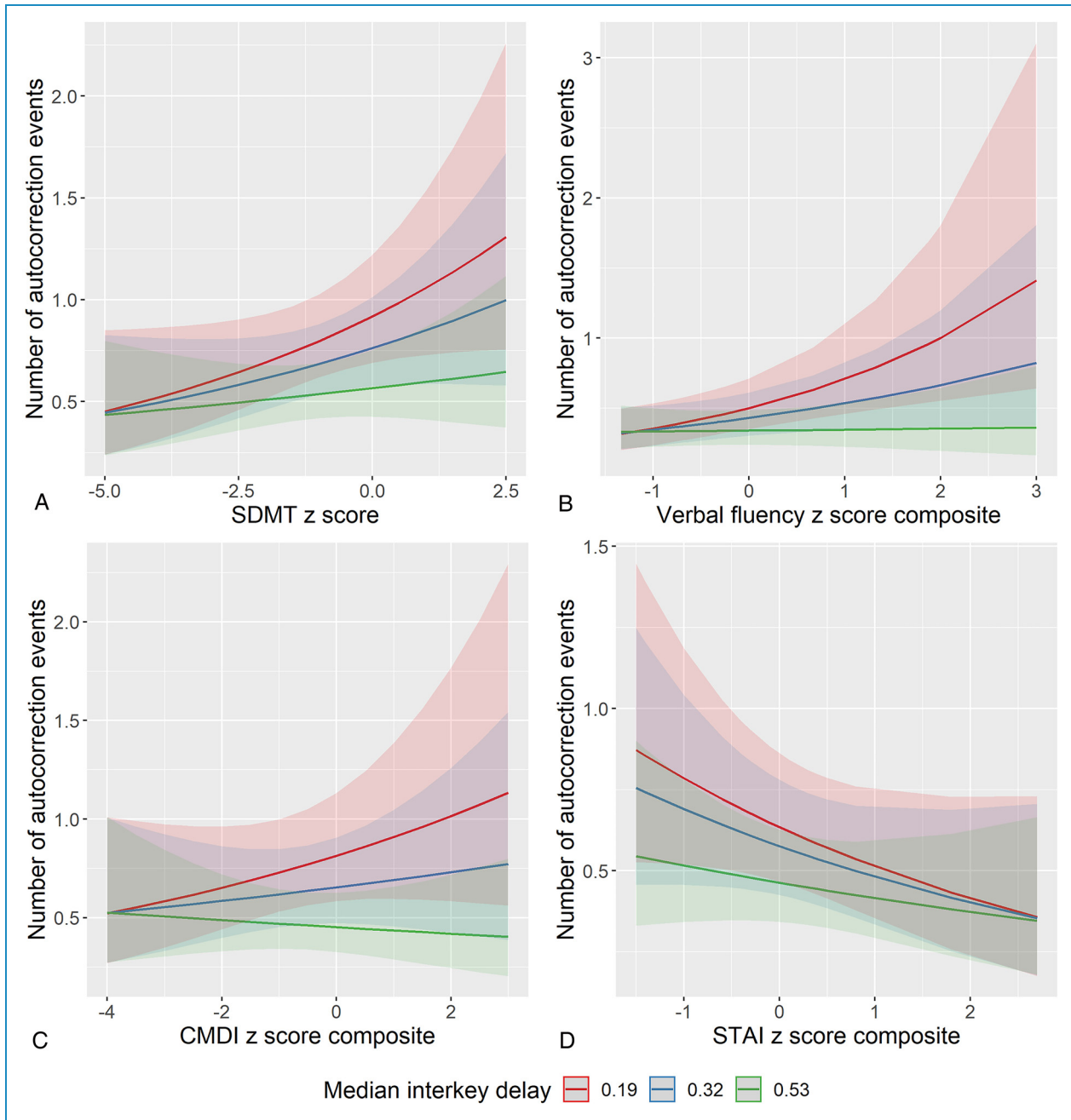
Table 1 summarizes demographic, clinical, and keystroke dynamics for MS and HC groups. There were no significant demographic differences between groups ( $P > .05$ ). A higher proportion of typing sessions were performed with one hand (as opposed to two hands) in the MS group, compared to the HC group (38% in MS vs. 19% in HC;  $\chi^2(1) = 542.70$ ,  $P < .001$ ). Adjusting for demographics, operating system, and typing with one or two hands, the MS group typed slower (longer median interkey delay; estimate =  $-1.45$ ,  $P < .001$ ) and more variably (higher interkey delay

MAD; estimate =  $0.11$ ,  $P = .032$ ) within sessions compared to the HC group (see Figure 2). The HC group had more autocorrection events than the MS group, but this was evident only among fast typers (group  $\times$  median interkey delay interaction estimate =  $1.82$ ,  $P < .001$ ; see Figure 2). The negative association between autocorrection and backspace rates was steeper for HC than MS group (group  $\times$  rate of autocorrection events interaction estimate =  $1.02$ ,  $P = .002$ ; see Figure 2).

### Associations between keystroke dynamics and baseline neuropsychological test performance and symptom ratings

Table 2 summarizes associations between median interkey delay per session and baseline neuropsychological test performance and symptom rating. Models were adjusted for demographics, operating system, and typing with one or two hands. Higher median interkey delay (slower typing speed) was associated with worse performance on the SDMT (estimate =  $.34$ ,  $P = .016$ ), digit span (estimate =  $.62$ ,  $P = .022$ ), TMT composite (estimate =  $.35$ ,  $P = .029$ ), CWIT (estimate =  $1.27$ ,  $P = .002$ ), and verbal fluency composite (estimate =  $.46$ ,  $P = .039$ ) scores (see Figure 3). Median interkey delay was not significantly associated with the episodic memory composite score ( $P > .05$ ). Regarding symptom ratings, higher median interkey delay was associated with more severe impact from fatigue (MFIS; estimate =  $-1.87$ ,  $P < .001$ ) and more severe anxiety symptoms (STAI composite; estimate =  $-.74$ ,  $P = .007$ ). There was also a trend of higher median interkey delay correlating with more severe depressive symptoms (CMDI; estimate =  $-.33$ ,  $P = .090$ ). Sensitivity analyses did not reveal moderating effects from remote neuropsychological test administration ( $P > .05$ ; see Supplemental Table 4).

There was a significant moderating effect of median interkey delay on the associations between rates of autocorrection and neuropsychological test performance/symptom rating (see Table 3 and Figure 4). Among fast typers (lower median interkey delay), more autocorrection events (more typing errors) were associated with better performance on the SDMT (interaction estimate =  $-.34$ ,  $P < .001$ ), digit span (interaction estimate =  $-.62$ ,  $P < .001$ ), TMT composite (interaction estimate =  $-.75$ ,  $P < .001$ ), CWIT (interaction estimate =  $-1.60$ ,  $P < .001$ ), and verbal fluency composite (interaction estimate =  $-1.07$ ,  $P < .001$ ) scores. Fewer autocorrection events were associated with better performance on the episodic memory composite score, and this relationship was stronger among those who typed slower (interaction estimate =  $-.53$ ,  $P = .002$ ). More autocorrection events were associated with more severe depressive symptoms (CMDI; interaction estimate =  $-.60$ ,  $P < .001$ ) and less severe anxiety symptoms (STAI composite; interaction estimate =  $.40$ ,  $P < .001$ ) among fast typers. There was no significant



**Figure 4.** Moderating effects of median interkey delay on associations between autocorrection events and neuropsychological performance/symptom rating. To visualize moderating effects, interkey delay is split into lower, median, and upper quartiles. Plots represent predicted values from multilevel models, and error bands represent 95% confidence intervals. Better neuropsychological performance was associated with more autocorrection events among fast typers only (Panels A and B). More severe depressive symptoms were associated with more autocorrection events (Panel C), and more severe anxiety symptoms were associated with fewer autocorrection events (Panel D); these relationships were only found among fast typers. SDMT: Symbol Digit Modalities Test; CMDI: Chicago Multiscale Depression Inventory; STAI: State-Trait Anxiety Inventory.

moderating effect of median interkey delay on the association between autocorrection and MFIS, but there was a significant main effect such that more autocorrection events were associated with less profound impact from fatigue (estimate =  $-.64$ ,  $P = .030$ ).

There was a significant moderating effect of neuropsychological performance/symptom rating on the association between rate of autocorrection events and use of backspace (see Table 4). Those with better cognitive functioning and less severe symptoms showed a stronger



**Table 4.** Moderating effects of baseline neuropsychological test performance/symptom ratings on associations between autocorrection events and use of backspace per session.

Measure	N	Neuropsychological performance/ symptom rating $\times$ rate of autocorrection events
SDMT	25	Estimate = $-0.32$ , $p < 0.001^*$
Digit span	25	Estimate = $0.04$ , $p = 0.750$
TMT composite	16	Estimate = $-0.70$ , $p < 0.001^*$
CWIT	19	Estimate = $-0.93$ , $p = 0.003^*$
Verbal fluency composite	18	Estimate = $-0.64$ , $p < 0.001^*$
Episodic memory composite	19	Estimate = $0.33$ , $p = 0.409$
MFIS	15	Estimate = $2.39$ , $p < 0.001^*$
CMDI	22	Estimate = $0.33$ , $p = 0.035^*$
STAI composite	19	Estimate = $0.80$ , $p = 0.001^*$

*Note.* Models were generalized multilevel models with keystroke sessions nested within persons and Poisson distributions. Estimates were from interactions between neuropsychological test performance/symptom rating and rate of autocorrection events per session, predicting backspace events per session. Models were offset with the log-transformed number of characters per session and adjusted for age, sex, years of formal education, operating system, and use of one or two hands (see Supplemental Table 3 for more details). All neuropsychological and symptom scores are standardized z scores based on demographics (see Supplemental Table 2 for more information).

SDMT: Symbol Digit Modalities Test; TMT: Trail-Making Test; CWIT: Color-Word Interference; MFIS: Modified Fatigue Impact Scale; CMDI: Chicago Multiscale Depression Inventory; STAI: State-Trait Anxiety Inventory. \*Statistically significant interactions.

negative association between autocorrection and backspace frequency (see Figure 5). This was evident for a range of neuropsychological tests and symptom inventories including the SDMT (interaction estimate =  $-0.32$ ,  $P < .001$ ), TMT composite (interaction estimate =  $-0.70$ ,  $P < .001$ ), CWIT (interaction estimate =  $-0.93$ ,  $P = .003$ ), verbal fluency composite (interaction estimate =  $-0.64$ ,  $P < .001$ ), MFIS (interaction estimate =  $2.39$ ,  $P < .001$ ), CMDI (interaction estimate =  $.33$ ,  $P < .001$ ), and STAI composite (interaction estimate =  $.80$ ,  $P < .001$ ).

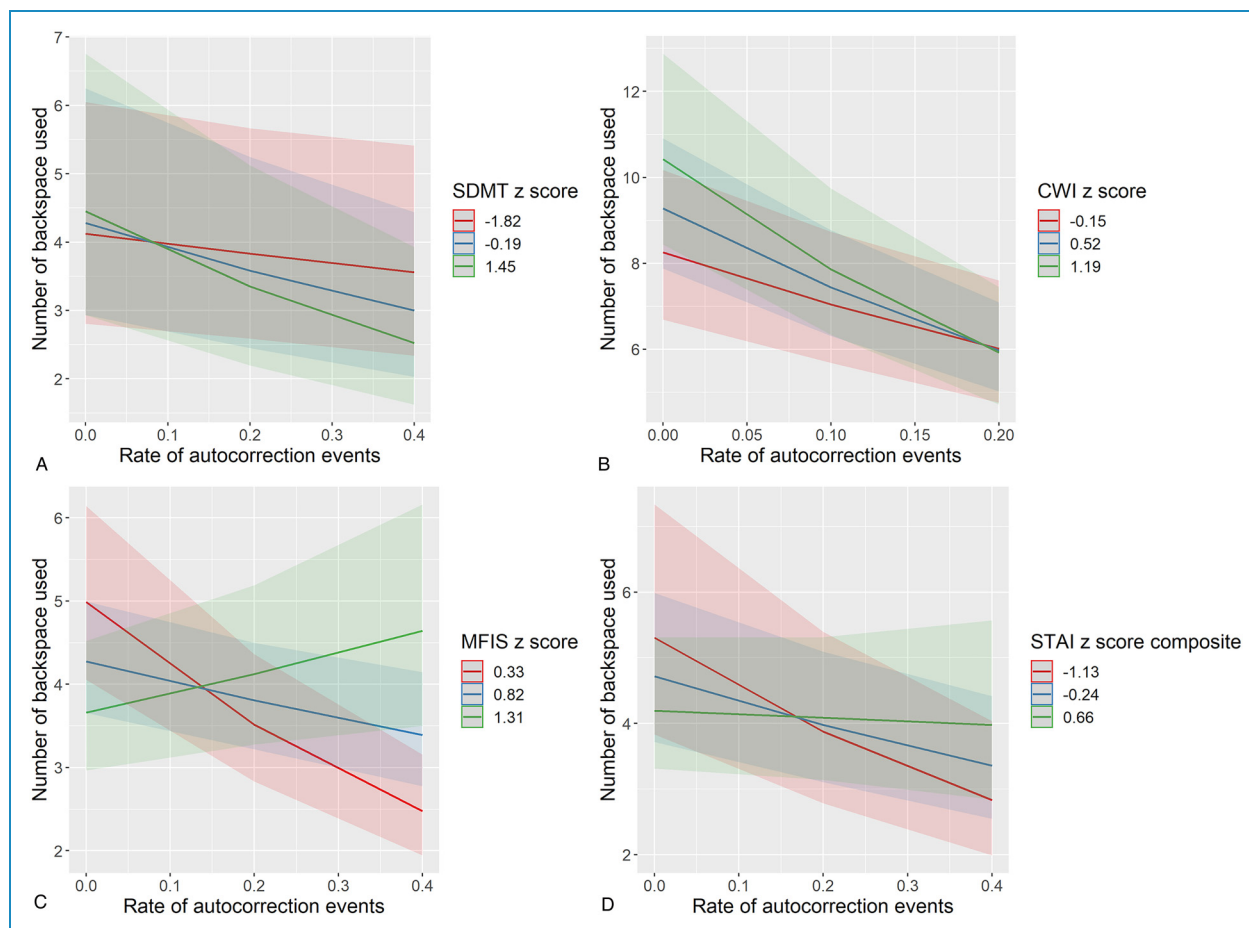
## Discussion

The current study examined the associations between smartphone keystroke dynamics and cognitive performance/symptom rating. The MS group typed slower than the HC group. Faster typing speed was associated with

better performance on measures of processing speed, attention, and executive functioning (i.e., higher-order cognitive abilities such as cognitive flexibility, behavioral inhibition, and verbal fluency as used in the current study), as well as less severe impact from fatigue and less severe anxiety symptoms (and a statistical trend with less severe depressive symptoms). Our results are consistent with and expanded findings from previous MS studies,<sup>14,15</sup> which showed significant associations between typing speed and performance on the SDMT. In terms of typing accuracy, there was a speed/accuracy trade-off, such that individuals with faster typing speed had more typing errors (autocorrection). Thus, because the HC group on average as well as those with better cognitive functioning and less severe symptoms typed faster, they also had more typing errors. Finally, more frequent use of backspaces was related to fewer typing errors, which confirmed our conceptualization that frequency of backspaces can be used as a measure of self-monitoring of typing errors. This relationship was stronger in the HC group on average as well as those with better cognitive functioning and less severe symptoms, suggesting that these individuals were better at monitoring their errors compared to their counterparts.

Episodic memory was the only cognitive construct that was not associated with typing speed. This was not surprising given different brain regions underlying memory compared to other cognitive domains assessed in this study. Memory consolidation is dependent on hippocampal functions,<sup>38</sup> whereas processing speed, attention, and executive functioning are reliant on frontal-subcortical brain circuits.<sup>39,40</sup> Therefore, it appears that typing speed may be more indicative of frontal-subcortical brain functions. Interestingly, we found that individuals who typed slower showed a negative relationship between episodic memory performance and autocorrection events. It is possible that for individuals who did not favor speed over accuracy, intact memory function was critical for making fewer typing errors. More research needs to be conducted to fully delineate the relationships among typing speed, typing errors, and cognitive functioning.

Based on our findings, keystroke dynamics have the potential to be used as an unobtrusive, remote monitoring method for real-life cognitive functioning. The high dimensional nature of keystroke data may be more sensitive to subtle changes in cognition and enable longitudinal comparisons to the individual's baseline.<sup>41</sup> Because it is a direct examination of a real-life behavior (i.e., typing), keystroke dynamics may be more ecologically valid than traditional lab-/clinic-based neuropsychological tests.<sup>42,43</sup> Remote cognitive monitoring via keystroke dynamics may lead to more timely identification of relapses and thus interventions. The continuous nature of keystroke monitoring may enable real-time assessment of treatment efficacy, allowing clinicians to modify medication regimen without requiring the patient to come in for an in-person visit. This will be especially helpful for physically disabled persons with MS, who often have difficulties arranging transportation and attending in-person



**Figure 5.** Moderating effects of neuropsychological performance/symptom rating on associations between autocorrection events and backspace use. To visualize moderating effects, each neuropsychological test, and symptom inventory is split into the mean and one standard deviation above and below the mean. Plots represent predicted values from multilevel models, and error bands represent 95% confidence intervals. Those with better cognitive functioning (Panels A and B) or less severe symptom ratings (Panels C and D) exhibited a stronger negative correlation between backspace and autocorrection frequencies. SDMT: Symbol Digit Modalities Test; CWI: Color-Word Interference; MFIS: Modified Fatigue Impact Scale; STAI: State-Trait Anxiety Inventory.

medical appointments. Finally, the unobtrusive nature of keystroke monitoring may be less prone to bias and is conducive to long-term monitoring with limited patient burden, which may be used to document the progression of cognitive disability. Cognitive decline over time is a major predictor of reduced/loss of employment and receipt of disability benefits among persons with MS.<sup>44</sup>

There are several limitations that require discussion. Firstly, given this was a proof-of-concept study, the sample size was small. Secondly, while we did exclude individuals with self-reported upper extremity or typing difficulties, we did not include objective measures of upper extremity motor skills due to the unavailability of such data in the parent studies. Thirdly, we only examined cross-sectional associations between keystroke dynamics and *baseline* neuropsychological test scores. We are currently conducting a larger trial with keystroke monitoring combined with real-time assessments of cognitive and motor performances (delivered through a smartphone)

as well as real-time symptom ratings using ecological momentary assessment, which will allow us to establish real-time correspondence between keystroke dynamics and cognitive/motor performance and symptom ratings. Fourthly, because we used neuropsychological data from six parent studies with different test batteries, we had to transform the data to make them comparable, which could have decreased their sensitivity. Finally, although the BiAffect app keyboard looked very similar to the native smartphone keyboards, some participants reported that it was more difficult to use (e.g., the autocorrection algorithm was not as robust).

## Conclusions

The current study provided proof-of-concept evidence that keystroke dynamics have the potential to be used as an unobtrusive remote monitoring method for real-life cognitive functioning, which may improve the detection of MS

relapses, evaluate treatment efficacy, and track disability progression among persons with MS.

**Acknowledgements:** The authors thank Dr Joshua Sandry (Montclair State University, NJ) for helping with recruitment and providing neuropsychological test data for this study.

**Contributorship:** MHC designed, obtained funding, analyzed data, and wrote the full draft of the manuscript. AL headed the team that created the study app and managed app data, obtained funding, and provided feedback on data analysis and manuscript draft. MKR and FH managed and processed the app data. DL and APD provided feedback for the analysis and manuscript draft. NC, SLC, HMG, and EW obtained funding, provided neuropsychological test data, and provided feedback for the manuscript draft.

**Declaration of Conflicting Interests:** The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Alex Leow is a cofounder of KeyWise AI, has served as a consultant for Otsuka US, and is currently on the medical board of Buoy Health. These entities had no involvement in the collection, analysis and interpretation of data; writing of the report; or in the decision to submit the article for publication.

**Ethical approval:** The study was approved by the Kessler Foundation Institutional Review Board.

**Funding:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by Robert Wood Johnson Foundation (a New Venture Fund/MoodChallenge for ResearchKit); National Multiple Sclerosis Society (MB-1606-08779; RG-1802-30111; RG-1901-33304; CA 1069-A-7); National Institutes of Health/Eunice Kennedy Shriver National Institute of Child Health and Human Development (1R01HD095915); National Academy of Neuropsychology; and New Jersey Commission on Spinal Cord Research (CSCR16ERG018). The funding sources had no involvement in the collection, analysis and interpretation of data; writing of the report; or in the decision to submit the article for publication.

**Guarantor:** MHC.

**ORCID iDs:** Michelle H. Chen  <https://orcid.org/0000-0002-5689-0044>

Nancy Chiaravalloti  <https://orcid.org/0000-0003-2943-7567>

**Supplemental Material:** Supplemental material for this article is available online.

## References

1. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology* 2014; 83: 278–286.
2. Benedict RH, Pol J, Yasin F, et al. Recovery of cognitive function after relapse in multiple sclerosis. *Mult Scler J* 2021; 27: 71–78.
3. Heled E, Aloni R and Achiron A. Cognitive functions and disability progression in relapsing-remitting multiple sclerosis: a longitudinal study. *Appl Neuropsychol Adult* 2021; 28: 210–219.
4. Kalb R, Beier M, Benedict RH, et al. Recommendations for cognitive screening and management in multiple sclerosis care. *Mult Scler J* 2018; 24: 1665–1680.
5. Calamia M, Markon K and Tranel D. Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *Clin Neuropsychol* 2012; 26: 543–570.
6. Torous J, Onnela J and Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. *Transl Psychiatry* 2017; 7: e1053.
7. Insel TR. Digital phenotyping: technology for a new science of behavior. *JAMA* 2017; 318: 1215–1216.
8. Iakovakis D, Hadjidimitriou S, Charisis V, et al. Motor impairment estimates via touchscreen typing dynamics toward Parkinson's disease detection from data harvested in-the-wild. *Front ICT* 2018; 5: 28.
9. Arroyo-Gallego T, Ledesma-Carbayo MJ, Butterworth I, et al. Detecting motor impairment in early Parkinson's disease via natural typing interaction with keyboards: validation of the neuroQWERTY approach in an uncontrolled at-home setting. *J Med Internet Res* 2018; 20: e9462.
10. Lang C, Gries C, Lindenberg KS, et al. Monitoring the motor phenotype in Huntington's disease by analysis of keyboard typing during real life computer use. *J Huntingtons Dis* 2021; 10: 259–268.
11. Alfalahi H, Khandoker AH, Chowdhury N, et al. Diagnostic accuracy of keystroke dynamics as digital biomarkers for fine motor decline in neuropsychiatric disorders: a systematic review and meta-analysis. *Sci Rep* 2022; 12: 1–24.
12. Zulueta J, Piscitello A, Rasic M, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a biaffected digital phenotyping study. *J Med Internet Res* 2018; 20: e241.
13. Stange JP, Zulueta J, Langenecker SA, et al. Let your fingers do the talking: passive typing instability predicts future mood outcomes. *Bipolar Disord* 2018; 20: 285–288.
14. Lam K, Meijer K, Loonstra F, et al. Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis. *Mult Scler J* 2021; 27: 1421–1431.
15. Lam KH, 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.
16. Chen R, Jankovic F, Marinsek N, et al. (eds) Developing measures of cognitive impairment in the real world from consumer-grade multimodal sensor streams. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2019.
17. Ntracha A, Iakovakis D, Hadjidimitriou S, et al. Detection of mild cognitive impairment through natural language and touchscreen typing processing. *Front Digit Health* 2020; 2: 567158.
18. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; 42: 377–381.

19. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform* 2019; 95: 103208.
  20. Vesel C, Rashidisabet H, Zulueta J, et al. Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: a BiAffect iOS study. *J Am Med Inform Assoc* 2020; 27: 1007–1018.
  21. Bennett CC, Ross MK, Baek E, et al. Predicting clinically relevant changes in bipolar disorder outside the clinic walls based on pervasive technology interactions via smartphone typing dynamics. *Pervasive Mob Comput* 2022; 83: 101598.
  22. Ross MK, Demos AP, Zulueta J, et al. Naturalistic smartphone keyboard typing reflects processing speed and executive function. *Brain Behav* 2021; 11: e2363.
  23. Smith A. *Symbol digit modalities test*. Los Angeles, CA: Western Psychological Services, 1982.
  24. Wechsler D. *Wechsler adult intelligence scale*. 4th ed. San Antonio, TX: Pearson, 2008.
  25. Germine L, Nakayama K, Duchaine BC, et al. Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychon Bull Rev* 2012; 19: 847–857.
  26. Reitan RM. Validity of the trail making test as an indicator of organic brain damage. *Percept Mot Skills* 1958; 8: 271–276.
  27. Delis DC, Kaplan E and Kramer JH. *Delis-Kaplan Executive Function System*. 2001.
  28. Ruff R, Light R, Parker S, et al. Benton Controlled Oral Word Association Test: reliability and updated norms. *Arch Clin Neuropsychol* 1996; 11: 329–338.
  29. Tombaugh TN, Kozak J and Rees L. Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Arch Clin Neuropsychol* 1999; 14: 167–177.
  30. Delis DC, Kramer JH, Kaplan E, et al. *California verbal learning test—second edition (CVLT-II)*. San Antonio, TX: The Psychological Corporation, 2000.
  31. Schmidt M. *Rey auditory verbal learning test: A handbook*. Los Angeles, CA: Western Psychological Services, 1996.
  32. Benedict RH, Schretlen D, Groninger L, et al. Hopkins verbal learning test—revised: normative data and analysis of interform and test-retest reliability. *Clin Neuropsychol* 1998; 12: 43–55.
  33. Learmonth Y, Dlugonski D, Pilutti L, et al. Psychometric properties of the fatigue severity scale and the Modified Fatigue Impact Scale. *J Neurol Sci* 2013; 331: 102–107.
  34. Nyenhuis DL and Luchetta T. The development, standardization, and initial validation of the Chicago Multiscale Depression Inventory. *J Pers Assess* 1998; 70: 386–401.
  35. Spielberg CD. *State-Trait Anxiety Inventory for adults*. 1983.
  36. R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2010.
  37. Lo S and Andrews S. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front Psychol* 2015; 6: 1171.
  38. Tulving E and Markowitsch HJ. Episodic and declarative memory: role of the hippocampus. *Hippocampus* 1998; 8: 198–204.
  39. Cummings JL. Frontal-subcortical circuits and human behavior. *Arch Neurol* 1993; 50: 873–880.
  40. Duering M, Gesierich B, Seiler S, et al. Strategic white matter tracts for processing speed deficits in age-related small vessel disease. *Neurology* 2014; 82: 1946–1950.
  41. Parsons T and Duffield T. Paradigm shift toward digital neuropsychology and high-dimensional neuropsychological assessments. *J Med Internet Res* 2020; 22: e23777.
  42. Thomas NWD, Beattie Z, Marcoe J, et al. An ecologically valid, longitudinal, and unbiased assessment of treatment efficacy in Alzheimer disease (the EVALUATE-AD trial): proof-of-concept study. *JMIR Res Protoc* 2020; 9: e17603.
  43. Chaytor N and Schmitter-Edgecombe M. The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. *Neuropsychol Rev* 2003; 13: 181–197.
  44. Morrow SA, Drake A, Zivadinov R, et al. Predicting loss of employment over three years in multiple sclerosis: clinically meaningful cognitive decline. *Clin Neuropsychol* 2010; 24: 1131–1145.
-