

## Remote Monitoring of Treatment Response in Parkinson's Disease: The Habit of Typing on a Computer

Michele Matarazzo, MD,<sup>1,2,3,4</sup> Teresa Arroyo-Gallego, MSc,<sup>5,6,7</sup> Paloma Montero, MD,<sup>2,8</sup> Verónica Puertas-Martín, PhD,<sup>2</sup> Ian Butterworth, MSc,<sup>9</sup> Carlos S. Mendoza, PhD,<sup>9</sup> María J. Ledesma-Carbayo, PhD,<sup>5</sup> María José Catalán, MD,<sup>8</sup> José Antonio Molina, MD,<sup>2</sup> Félix Bermejo-Pareja, MD,<sup>2</sup> Juan Carlos Martínez-Castrillo, MD,<sup>10</sup> Lydia López-Manzanares, MD,<sup>11</sup> Araceli Alonso-Cánovas, MD,<sup>10</sup> Jaime Herreros Rodríguez, MD,<sup>12</sup> Ignacio Obeso, PhD,<sup>1,3</sup> Pablo Martínez-Martín, MD,<sup>3,13</sup> José Carlos Martínez-Ávila, PhD,<sup>14</sup> Agustín Gómez de la Cámara, MD,<sup>14</sup> Martha Gray, PhD,<sup>6,9</sup> José A. Obeso, MD,<sup>1,3</sup> Luca Giancardo, PhD,<sup>6,15\*</sup> and Álvaro Sánchez-Ferro, MD<sup>1,2,3,6\*</sup>

<sup>1</sup>HM-CINAC, Hospital Universitario HM Puerta del Sur, Móstoles and Medical School, CEU-San Pablo University, Madrid, Spain

<sup>2</sup>Neurology Department, Instituto de Investigación del Hospital 12 de Octubre, Madrid, Spain

<sup>3</sup>Centro de Investigación Biomédica en Red, Enfermedades Neurodegenerativas, Madrid, Spain

<sup>4</sup>Pacific Parkinson's Research Centre, University of British Columbia, Vancouver, British Columbia, Canada

<sup>5</sup>Biomedical Image Technologies, Universidad Politécnica de Madrid and CIBERBBN, Madrid, Spain

<sup>6</sup>Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

<sup>7</sup>nQ Medical Inc., Cambridge, Massachusetts, USA

<sup>8</sup>Movement Disorders Unit, Hospital Clínico San Carlos, Madrid, Spain

<sup>9</sup>Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

<sup>10</sup>Movement Disorders Unit, Hospital Ramón y Cajal, Madrid, Spain

<sup>11</sup>Movement Disorders Unit, Hospital de la Princesa, Madrid, Spain

<sup>12</sup>Neurology Department, Hospital Infanta Leonor, Madrid, Spain

<sup>13</sup>Area of Applied Epidemiology, National Centre of Epidemiology, Carlos III Institute of Health, Madrid, Spain

<sup>14</sup>Clinical Research Unit, Instituto de Investigación Hospital 12 de Octubre, Consorcio de Investigación Biomédica en Red de Epidemiología y Salud Pública, Hospital Universitario 12 de Octubre, Madrid, Spain

<sup>15</sup>Center for Precision Health, School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, Texas, USA

**ABSTRACT: Objective:** The recent advances in technology are opening a new opportunity to remotely evaluate motor features in people with Parkinson's disease (PD). We hypothesized that typing on an electronic device, a habitual behavior facilitated by the nigrostriatal dopaminergic pathway, could allow for objectively and nonobtrusively monitoring parkinsonian features and response to medication in an at-home setting.

**Methods:** We enrolled 31 participants recently diagnosed with PD who were due to start dopaminergic treatment and 30 age-matched controls. We remotely monitored their typing pattern during a 6-month (24 weeks) follow-up period before and while dopaminergic medications were being titrated. The typing data were used to develop a novel algorithm based on recursive neural networks and detect participants' responses to medication. The latter were defined by the Unified Parkinson's Disease Rating Scale-III (UPDRS-III)

minimal clinically important difference. Furthermore, we tested the accuracy of the algorithm to predict the final response to medication as early as 21 weeks prior to the final 6-month clinical outcome.

**Results:** The score on the novel algorithm based on recursive neural networks had an overall moderate kappa agreement and fair area under the receiver operating characteristic (ROC) curve with the time-coincident UPDRS-III minimal clinically important difference. The participants classified as responders at the final visit (based on the UPDRS-III minimal clinically important difference) had higher scores on the novel algorithm based on recursive neural networks when compared with the participants with stable UPDRS-III, from the third week of the study onward.

**Conclusions:** This preliminary study suggests that remotely gathered unsupervised typing data allows for the accurate detection and prediction of drug response

\*Correspondence to: Dr. Álvaro Sánchez-Ferro, Centro Integral de Neurociencias A.C., Avenida del Emperador Carlos V, 70, E-28938 Madrid, Spain; E-mail: aferro.hmcinac@hnhospitales.com

Luca Giancardo and Álvaro Sánchez-Ferro contributed equally.

**Funding agency:** This study was conducted with research funds from Michael J. Fox Foundation for Parkinson's Research Grant 10860.

**Relevant conflicts of interests/financial disclosures:** Nothing to report.

**Received:** 11 February 2019; **Revised:** 3 May 2019; **Accepted:** 13 May 2019

**Published online** 00 Month 2019 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.27772

Current evaluation standards in Parkinson's disease (PD), such as the Unified Parkinson's Disease Rating Scale (UPDRS),<sup>1</sup> are very useful, but have important limitations. As recently pointed out, these scales report a semiquantitative and subjective score, nonsensitive to subtle motor changes.<sup>2-4</sup> In addition, these assessments typically require the patient to travel to the clinic, and they need to be performed by a trained specialist, representing an additional burden for the patient and hence being time and cost consuming.

For these reasons, several attempts are underway to complement traditional standards with more objective, quantitative, and continuous outcome measures.<sup>5,6</sup>

Notably, in the recent decades we have been witnessing an exponential adoption of smart technologies, such as computers, smartphones, and tablets. This natural interaction with keyboards and touch screens is probably driven by habitual-directed movements, whose control is regulated by nigro-striatal activity.<sup>7-9</sup> Hence, we set to ascertain whether a natural interaction with keyboards would enable a new method to remotely detect and monitor parkinsonian motor signs nonobtrusively by analyzing the characteristics of free-text typing. Such an approach could have advantages over existing solutions,<sup>10-12</sup> because (1) it can extract motor information from the natural interaction of the patients or study participants with their devices without requiring active collaboration; (2) it could virtually reach any person who is typing with an internet-connected device, opening a window on the motor skills and parkinsonian signs of an enormous number of individuals; (3) it can acquire data remotely without requiring to attend a clinic; and (4) participants could be monitored longitudinally in a quasi-continuous manner.

We have previously shown that data collected from an in-clinic typing task accurately differentiated early PD patients from sex- and age-matched healthy controls,<sup>13</sup> and replicated this result using at-home, unsupervised data.<sup>14</sup> Recently, we demonstrated similar performance with data acquired during typing on a touch-screen smartphone.<sup>15</sup> In the present study, our aim is to detect the response to medication in PD by using remotely gathered, unsupervised typing data in an at-home, everyday-life setting as an additional step to this new digital care model. Thus, we designed a prospective naturalistic study enrolling early PD patients who were going to start dopaminergic medication and followed them for 6 months. The specific goal of this study was to validate a novel approach applied to remotely gathered typing data for (1) detecting response to medication in an early PD population and (2) predicting which PD patients will respond

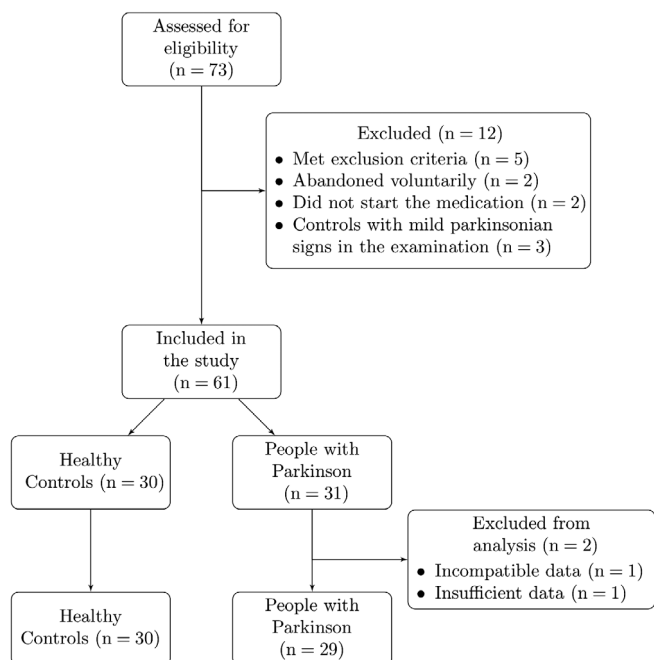
to the drugs at the final visit based on the typing data collected at home up to 21 weeks in advance.

## Materials and Methods

### Study Participants

Between March 2015 and June 2016, 31 consecutive early PD patients were recruited from 7 hospitals in Madrid, Spain (Fig. 1), according to prespecified inclusion and exclusion criteria that are detailed in the Supplementary Materials. A total of 30 age- and sex-matched healthy controls (HC) were enrolled after ruling out the existence of parkinsonism, hand deformities, cognitive impairment, sleep problems, or any other potential confounders (eg, use of psychoactive medication, drug abuse, or a serious medical condition).

The sample size was prespecified to detect at least 15 participants with response to medication, according to a previous definition of response (ie, decrease of at least 5 points in the total UPDRS-III score).<sup>16</sup> We a priori estimated a responder rate of 50% based on previous information from various randomized clinical trials.<sup>17-19</sup> For this, we targeted participants who were prescribed dopamine agonist or levodopa. We did not exclude participants



**FIG. 1.** Flowchart of the study sample.

already on rasagiline, but evaluated them *off* medication. We expected a negligible confounding motor effect of this drug based on the main pivotal trials that reported the UPDRS change at 12 to 36 weeks was  $\leq 0.11$  points.<sup>20</sup>

Besides the participants who met the exclusion criteria, two additional PD participants were excluded from the final analyses, in one case because the typing data were insufficient to generate a score, and in the other case because the participant's laptop had an operating system that was not compatible with our software (Fig. 1). Finally, five participants who got worse based on the definition of response as described later were also not included in the model training and typing analyses.

The post hoc evaluation of typing consistency was done to define the minimum amount of data needed to obtain a reliable score. For this reason, a typing day was defined as at least 10 valid windows per day, where a valid window was represented by a data sequence of at least 30 keystrokes within 90-second time interval. Then we defined the “consistent typers” as participants with at least one typing day in 80% of every possible 15 days rolling windows during the entire follow-up period (an overview of typing activity and consistency is available in Supporting Information Fig. 1). The analyses were conducted in both consistent and nonconsistent typers to evaluate the impact of typing frequency on the method's diagnostic performance.

### Study Design

A summary of the study design can be found in the Supplementary Materials, Supporting Information Fig. 2, and in [clinicaltrials.gov](https://clinicaltrials.gov) (NCT02522065). Succinctly, all of the participants included in the study received a complete evaluation by a movement disorder specialist (M.M., P.M.E.) at baseline that included the UPDRS-III, the Purdue Pegboard test, and other standards.

At the baseline visit, the neuroQWERTY software was installed in the participants' laptops. The participants were invited to freely type for at least 20 minutes per day during the whole duration of the study. The software ran in the background of the laptop, capturing the typing data—press/release timestamps of keystrokes—that was automatically sent to a remote server located at Massachusetts Institute of Technology (Boston, Massachusetts). The privacy of the typing data was assured by encryption of keystroke information, anonymization of the participants, and authentication for accessing the data.<sup>14</sup>

To obtain at-home (*off*) baseline typing data before the participants started the newly prescribed dopaminergic medications, the PD participants were instructed to delay the start of the new drug 7 days after the initial baseline visit. Further follow-up visits were scheduled flexibly at weeks 4, 8, 16, and 24 after starting the medication with the same assessments that were conducted at

baseline (for further information, see the Supplementary Materials).

### Standard Protocol Approvals and Patient Consents

All of the experimental protocols were approved by the Massachusetts Institute of Technology (no. 1412006804), HM Puerta del Sur University Hospital, Spain (no. 15.05.796-GHM), 12 de Octubre University Hospital, Spain (no. CEIC:14/090), and Clínico San Carlos University Hospital, Spain (no.14/136-E). All of the participants provided written informed consent prior to study enrollment.

### Definition of Drug Response: UPDRS-III Minimal Clinically Important Difference

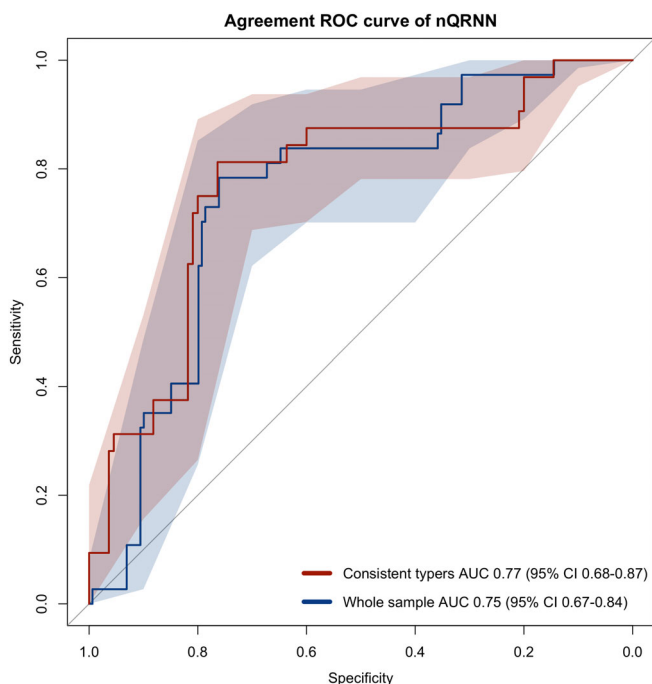
To classify participants as improved, not changed, or worsened, we calculated the minimal clinically important difference (MCID) of UPDRS-III for this study.<sup>21</sup> The relevant cut-off for our cohort was established in  $\pm 5$  points. There were only five participants who worsened, and they have not been included in the results (see limitations in the discussion). In terms of response, we compared those participants who did not change (UPDRS-III change ranging from  $-5$  to  $+5$  points) to those participants who improved (UPDRS-III scores that were lower by more than 5 points at follow-up).

### Classification Modeling: neuro QWERTY Recurrent Neural Network (nQRNN)

We used a machine-learning model (nQRNN) that receives as input typing features derived from the hold time, that is, the time required to press and release each key on a participant's laptop, regardless of the text typed. The typing features are encoded as “Key Hold Time Distribution” matrices<sup>22</sup> joined with the encoding previously described.<sup>13,14</sup> nQRNN outputs the probability of each patient of being a responder or nonresponder and were employed to generate the plots in Figures 2 and 3. nQRNN architecture is based on hierarchical layers of long short-term memory units (a type of recurrent neural network) trained using a nested cross-validation approach to avoid overfitting and a previously described optimization algorithm known as RMSprop (i.e., Root Mean Square Prop). This type of software architecture is known to be an effective predictive model for complex time-series data.<sup>23</sup> More details are available in the Supplementary Materials.

### Data Analyses

The following two different types of analysis were performed: (1) the agreement of the nQRNN-based with the time-coincident UPDRS-III MCID-based classification of response and (2) the prediction of whether the patients would be classified as responders or nonresponders at the final visit (according to UPDRS-III



**FIG. 2.** The ROC curves of nQRNN for UPDRS-III MCID-based classification of responders. The plot shows the ROC curves of the nQRNN for the binary classification of the patients as “improved” (ie, responders) and “not changed” according to the MCID of the UPDRS-III. The blue line is obtained plotting the whole sample of the study. The red line is obtained plotting only the participants classified as “consistent typers.” The shaded areas represent the 95% CI. AUC, area under the ROC curve; CI, confidence interval; MCID, minimal clinically important difference; nQRNN, novel algorithm based on recursive neural networks; ROC, receiver operating characteristic; UPDRS-III, Unified Parkinson’s Disease Rating Scale part III. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

MCID) based on the at-home nQRNN score obtained from the previous weekly timepoints throughout the study. The score was calculated from the third week after the medication was prescribed onward, when there was sufficient data to conduct the described analyses. The analyses were replicated in both consistent typers and in the whole study cohort. More details on the analyses conducted are provided in the Supplementary Materials.

### Evaluation of Cognition as Possible Confounder

To rule out the possible confounding effect of cognition, we computed the Spearman correlation between the Montreal Cognitive Assessment test (MoCA) score and the nQRNN score.<sup>24,25</sup> Moreover, we analyzed the response classification of participants who fulfilled the Movement Disorder Society criteria of PD mild cognitive impairment according to the results of a complete neuropsychological battery.<sup>26</sup>

## Data Collection, Database Processing, and Statistical Analysis

Software characteristics have been previously described.<sup>13</sup> Clinical data were collected using the Research Electronic Data Capture software (REDCap).<sup>27</sup> The nQRNN was developed in the Python<sup>28</sup> and Keras framework. Database processing and statistical analysis were performed using R 3.3.2.<sup>29</sup> Baseline characteristics were compared using a nonparametric approach (Mann-Whitney *U* test for continuous variables and Fisher’s exact test for categorical variables; multiple group comparisons at baseline were done using the Kruskal-Wallis test with post hoc pairwise Mann-Whitney *U* test in the continuous variables). The comparisons between nQRNN scores of the improved and not-changed groups of the prediction analysis were performed using the Mann-Whitney *U* test. Receiver operating characteristics (ROC) curve analysis was used for the agreement of at-home nQRNN score with UPDRS-III-based response classification as well as for the prediction analysis of final response. Cohen’s  $\kappa$  was used as measure of agreement, and Cohen’s effect size was calculated for agreement and prediction analyses. Significance was defined as a 2-sided type I error below the 5% probability.

## Results

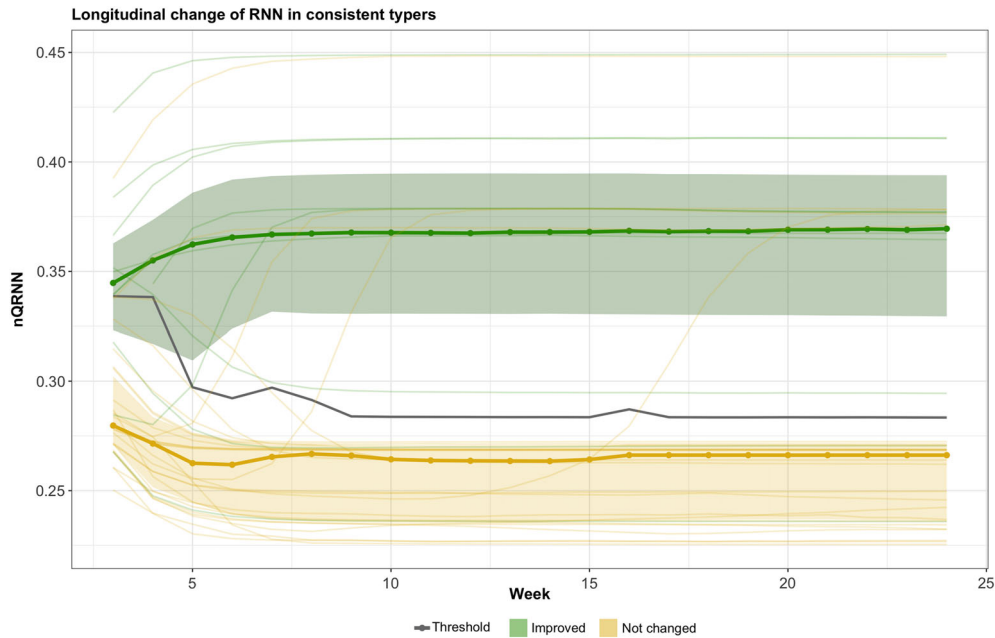
### Comparability and Characterization of the Studied Cohort

Baseline demographic characteristics were similar between the PD ( $N = 29$ ) and HC ( $N = 30$ ) groups, as shown in the Table 1. As expected, statistically significant differences were observed in the motor performance of the PD and controls (eg, UPDRS-III). The levodopa equivalent daily dose was also statistically different between the 2 groups.

The median MoCA score was 28 (interquartile range [IQR]: 27-29) in the HC group and 27 (IQR = 26-28) in the PD group, with one-point difference statistically significant ( $P = .049$ ).

Based on the participants’ final response to medication (HC, PD who improved, and PD who did not change), further between-group baseline comparisons were done (Supporting Information Table 1). Overall, statistically significant differences were observed only in motor performance-dependent variables. Pairwise comparisons found that those differences were the result of differences between PD participants and HC, as expected. The only statistically significant difference between the two PD groups was in the Purdue Assembly task ( $P = .01$ ).

The median disease duration at recruitment of the PD group was 13.9 months (IQR = 10.4-32.4). Most of the PD participants were in stage 2 of Hoehn and Yahr scale ( $n = 19/29$ , 65.5%). The rest were in stage 1 ( $n = 7/29$ , 24.1%) or 2.5 ( $n = 3/29$ , 10.3%); none of



**FIG. 3.** Longitudinal nQRNN change in consistent typers. The thick lines and shaded areas represent the longitudinal median nQRNN scores and their interquartile ranges for the Parkinson’s disease patients who were finally classified as “improved” (green line) and for the Parkinson’s disease patients and controls who were classified as “not changed” (yellow line). The thin lines represent the single participants and their score over time. The gray line represents the best possible threshold according to the Youden method, computed for each week interval, which maintained a stable value of 0.28 since week seven after starting the medication. nQRNN, novel algorithm based on recursive neural networks. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

**TABLE 1.** Demographic and baseline characteristics

Variable	Group		P value
	Healthy controls, n = 30	PD patients, n = 29	
Age	63.00 (56.48-69.44)	59.78 (54.19-68.60)	.476
Sex, woman	16 (53.3)	14 (48.3)	.797
Handedness, right	28 (93.3)	29 (100.0)	.492
Alcohol	8 (26.7)	4 (13.8)	.333
Tobacco	3 (10.0)	3 (10.3)	1.000
Hypertension	10 (33.3)	9 (31.0)	1.000
Diabetes mellitus	5 (16.7)	2 (6.9)	.424
Dyslipemia	8 (26.7)	7 (24.1)	1.000
Computer use, y	20.00 (10.00-25.00)	20.00 (12.00-20.00)	.982
Weekly computer use, d	7.00 (5.00-7.00)	7.00 (4.00-7.00)	.402
Education, y	15.00 (12.00-18.00)	18.00 (12.00-20.00)	.322
MoCA	28.00 (27.00-29.00)	27.00 (26.00-28.00)	.049 <sup>a</sup>
LED	0.00 (0.00-0.00)	0.00 (0.00-0.00)	.009 <sup>a</sup>
UPDRS-III	1.00 (0.00-2.00)	19.00 (17.00-26.00)	<.001 <sup>a</sup>
Purdue right	14.83 (14.00-16.00)	12.33 (11.33-15.33)	.003 <sup>a</sup>
Purdue left	13.50 (12.67-15.25)	11.33 (10.67-12.67)	<.001 <sup>a</sup>
Purdue both	11.17 (10.67-12.67)	9.00 (8.00-11.00)	<.001 <sup>a</sup>
Purdue assembly	27.67 (23.25-30.00)	24.67 (17.67-29.67)	.075

Quantitative variables are represented as “median (interquartile range)” and qualitative variables as “n (%).” LED, levodopa equivalent dose; MoCA, Montreal Cognitive Assessment test; UPDRS-III, Unified Parkinson’s Disease Rating Scale part III.

<sup>a</sup>Statistically significant difference.

them were in stages 3, 4 or 5. The tremor-dominant phenotype of PD was the most frequent (n = 15/29, 51.7%).<sup>30</sup> Six PD participants (20.7%) were already receiving rasagiline at recruitment.

Of the sample, 95% had a 6-month follow-up. Two of the PD participants and one HC dropped out before the 24-week visit for personal reasons not related to the study. During the 6-month follow-up, there was a progressive increase in the median levodopa equivalent daily dose of the PD group from 0 at baseline to 340 (IQR = 220-400) at week 24 (Supporting Information Fig. 3).

The number of PD participants who responded to medication, according to the 5-point UPDRS-III MCID, increased during the study as the medication was titrated (Supporting Information Fig. 4). At the final visit (ie, week 24), the PD participants who responded to medication were 51.9%. Of the whole sample, 37 (66.1%) participants were consistent typers, including 20 (69.0%) controls and 17 (63.0%) PD patients.

### Evaluating the Concurrent and Discriminant Validity of nQRNN

There was a moderate significant correlation (Spearman  $\rho = 0.33$ ;  $P = .02$ ) between nQRNN and the time-coincident UPDRS-III  $\Delta$  at the final endpoint visit. The correlation with different nonmotor measures such as MoCA was nonsignificant (Spearman  $\rho = 0.10$ ;  $P = .49$ ).

## Analyzing the Agreement and Accuracy of nQRNN to Detect Drug Response

The area under the ROC curve (AUC) for all the aggregated data (ie, all timepoints) to classify the participants as “improved” or “not changed” using the nQRNN was 0.77 (95% confidence interval, 0.68-0.87) in consistent typers and 0.75 (95% CI, 0.67-0.84) for the whole sample (Fig. 2). The  $\kappa$  agreement was moderate for both consistent typers and the whole cohort (0.55 and 0.47, respectively). A large Cohen’s *d* effect size was observed (1.26 for the consistent typers and 0.92 for the whole sample). The overall balanced accuracy was 76.5% in the whole sample and 77.4% in the consistent typers. Supporting Information Table 2 shows the results for each timepoint and the results of all available data aggregated in the whole sample and in consistent typers.

## Predicting Response to Medication With nQRNN Score at Home

Considering only consistent typers, the nQRNN achieved a good prediction of the final classification since week 3 after the treatment was started, showing stable median scores in both groups from week 7 onward (Fig. 3). The nQRNN score was higher in responders when compared with nonresponders for every week analyzed (ie, from week 3 to week 24), with *P* values <.005 for consistent typers (Supporting Information Table 3). When adjusted for multiple comparisons with a Bonferroni correction, weeks 20 to 24 did not reach statistical significance.

The longitudinal ROC curve analysis for predicting response to medication showed AUCs >0.80 for the entire period analyzed (from week 3 to week 24) in consistent typers, whereas the AUCs considering the whole cohort oscillated between 0.69 and 0.75 (0.73-0.75 after the sixth week of treatment). The nQRNN threshold, calculated on a weekly basis using Youden’s method, was stable from week 7 onward (Fig. 3). Supplementary data, including AUCs and Cohen’s *d* effect sizes, are available in Supporting Information Table 4.

The AUC of other baseline characteristics (age, computer use, UPDRS-III, and PD Questionnaire-39) were not statistically significant for the prediction of the final classification, confirming that the results of the nQRNN are not the result of baseline group differences (Supporting Information Table 5).

## Evaluation of Cognition as Confounding

A statistically significant difference was observed between PD participants and HC in the MoCA score, but such a difference was not noted between the 3 groups analyzed (PD participants who improved, PD participants who did not change, and HC). Spearman  $\rho$  between MoCA and nQRNN scores showed a nonsignificant correlation (see Results). Moreover, the PD participants classified as PD mild cognitive impairment

were evenly distributed between the 3 groups (1 in the improved group, 2 in the not changed group, and 2 in the worsened group).

## Discussion

Medicine and neurology are moving toward a new model of care based on objective data collected remotely (ie, ecologically valid) and nonintrusively (ie, not requiring the active cooperation of the patients).<sup>5,6,10,31,32</sup> This approach will allow doctors or drug makers to make informed decisions on PD diagnosis or therapy remotely.

In this new scenario, we investigated the preliminary validity of free unconstrained typing at home as a proxy of drug response in PD. In a longitudinal prospective naturalistic study with a lengthy follow-up, we have shown that a recurrent neural network algorithm accurately detected (ie, AUC = 0.75) the response to dopaminergic therapy in an early PD population, with a moderate  $\kappa$  agreement and large Cohen’s *d* effect size, compared to time-coincident in-clinic UPDRS-III classification. Furthermore, we showed that the remote monitoring of motor signs of PD using nonintrusive, free typing information is feasible and has good compliance considering that only two participants (3.3%) were excluded because of insufficient data.

The possibility of remotely monitoring the response to medication and the motor status of PD patients can be a major step for improving the management of the disease in clinical practice and making decisions on further changes of treatment or on the planning of future follow-up visits. Moreover, our score predicted from the third week after starting the drug, which PD participants responded to medication at the final visit. The classification became stable at week 7 with an nQRNN threshold of 0.28 that remained the same until study completion. Therefore, we were able to anticipate the clinical response to medication as early as 21 weeks in advance using uniquely remotely gathered typing data. These findings suggest that our tool may be sensitive to subtle motor changes, being able to detect people who are responding to medication at an earlier stage using remote, objective data. This could be crucial, for example, in supporting go/no-go decisions in early intervention trials, reducing the cost of developing new compounds and also potentially being helpful when adjusting treatments in clinical practice.

Our study has some limitations that should be considered. First, although we used a nested cross-validation approach that allowed us to test the generalizability of our model in a limited dataset, our cohort does not provide a complete representation of all nuances of PD progression, cognitive states, coexisting conditions, and typing habits. However, a machine-learning model such as nQRNN is able to learn from

new examples by design.<sup>33</sup> Therefore, there is the potential for fine tuning the nQRNN performance by increasing the dataset or even adding other data modalities (eg, touch screen or mouse clicks, among other possibilities). Second, because of the size of our dataset, we could not train the model to predict 3 distinct types of progressions, and we focused on “improved” and “not changed.” We are confident that future studies including a larger number of participants with heterogeneous types of PD progressions could overcome this issue that is critical for the translation of our approach to everyday practice and clinical trials. In our model, the “not changed” group included the HC. HC are indeed a good sample of stable motor status (as it has been confirmed by a UPDRS-III change that was always below UPDRS-III MCID); however, in a drug trial or in clinical practice HC do not necessarily need to be included.

As expected, the accuracy of agreement and prediction were higher in the consistent typers subgroup. Consistent typers are more likely to produce typing data evenly during the disease progression, leading to a more accurate prediction. However, the results obtained including nonconsistent typers are still significant, which was an unexpected result in light of the limited amount of typing data available for the nonconsistent typers subgroup. This is particularly important as currently the age group of people affected by PD may be less active users of technology. Finally, participants with other medical conditions, such as hand deformities or other neurological issues, were excluded from our study. The impact of these possible confounders on our outcome score still needs to be assessed.

In conclusion, we report on a pilot study on a novel technological approach to monitor motor features of PD and drug response remotely and ecologically in an accurate way, reflecting the underlying effects of basal ganglia neurodegeneration on a habitual task, such as typing. We show that this approach is feasible and suggest that it could be useful in everyday clinical practice and could complement the current standard outcomes for improving the efficacy of clinical trials in PD, helping to reduce the burden for participants and investigators and to assess in a more time- and cost-efficient way the response to medication. ■

## References

1. Fahn S, Elton R. Unified Parkinson's Disease Rating Scale (UPDRS). In: Fahn S, Marsden CD, Calne DB, et al., eds. *Recent Developments in Parkinson's Disease, Vol 2*. Florham Park, NJ: Macmillan Health Care Information, 1987:153-163.
2. Kiebertz K, Olanow CW. Advances in clinical trials for movement disorders. *Mov Disord* 2015;30:1580-1587. <https://doi.org/10.1002/mds.26371>
3. Schneider SA, Drude L, Kasten M, et al. A study of subtle motor signs in early Parkinson's disease. *Mov Disord* 2012;27:1563-1566. <https://doi.org/10.1002/mds.25161>

4. Maetzler W, Domingos J, Srulijes K, et al. Quantitative wearable sensors for objective assessment of Parkinson's disease. *Mov Disord* 2013;28:1628-1637. <https://doi.org/10.1002/mds.25628>
5. Sánchez-Ferro Á, Elshehabi M, Godinho C, et al. New methods for the assessment of Parkinson's disease (2005 to 2015): a systematic review. *Mov Disord* 2016;31:1283-1292. <https://doi.org/10.1002/mds.26723>
6. Ferreira JJ, Godinho C, Santos AT, et al. Quantitative home-based assessment of Parkinson's symptoms: the SENSE-PARK feasibility and usability study. *BMC Neurol* 2015;15:89. <https://doi.org/10.1186/s12883-015-0343-z>
7. Redgrave P, Rodriguez M, Smith Y, et al. Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. *Nat Rev Neurosci* 2010;11:760-772. <https://doi.org/10.1038/nrn2915>
8. Catalan MJ, Ishii K, Honda M, et al. A PET study of sequential finger movements of varying length in patients with Parkinson's disease. *Brain* 1999;122 ( Pt 3):483-495. <https://doi.org/10.1093/brain/122.3.483>
9. Lehéricy S, Bardinet E, Tremblay L, et al. Motor control in basal ganglia circuits using fMRI and brain atlas approaches. *Cereb Cortex* 2006;16:149-161. <https://doi.org/10.1093/cercor/bhi089>
10. Giovannoni G, van Schalkwyk J, Fritz VU, et al. Bradykinesia akinesia inco-ordination test (BRAIN TEST): an objective computerised assessment of upper limb motor function. *J Neurol Neurosurg Psychiatry* 1999;67:624-629. <https://doi.org/10.1136/jnnp.67.5.624>
11. Hasan H, Athauda DS, Foltynie T, et al. Technologies assessing limb bradykinesia in Parkinson's disease. *J Parkinsons Dis* 2017;7:65-77. <https://doi.org/10.3233/JPD-160878>
12. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord* 2008;23:2129-2170. <https://doi.org/10.1002/mds.22340>
13. Giancardo L, Sánchez-Ferro Á, Arroyo-Gallego T, et al. Computer keyboard interaction as an indicator of early Parkinson's disease. *Sci Rep* 2016;6:34468. <https://doi.org/10.1038/srep34468>
14. 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:e89. <https://doi.org/10.2196/jmir.9462>
15. Arroyo-Gallego T, Ledesma-Carbayo MJ, Sanchez-Ferro A, et al. Detection of motor impairment in Parkinson's disease via mobile touchscreen typing. *IEEE Trans Biomed Eng* 2017;64:1994-2002. <https://doi.org/10.1109/TBME.2017.2664802>
16. Schrag A, Sampaio C, Counsell N, et al. Minimal clinically important change on the Unified Parkinson's Disease Rating Scale. *Mov Disord* 2006;21:1200-1207. <https://doi.org/10.1002/mds.20914>
17. Jankovic J, Watts RL, Martin W, et al. Transdermal rotigotine: double-blind, placebo-controlled trial in Parkinson disease. *Arch Neurol* 2007;64:676-682. <https://doi.org/10.1001/archneur.64.5.676>
18. Poewe W, Rascol O, Barone P, et al. Extended-release pramipexole in early Parkinson disease: a 33-week randomized controlled trial. *Neurology* 2011;77:759-766. <https://doi.org/10.1212/WNL.0b013e31822affb0>
19. Kiebertz K, Parkinson Study Group PramiBID Investigators. Twice-daily, low-dose pramipexole in early Parkinson's disease: a randomized, placebo-controlled trial. *Mov Disord* 2011;26:37-44. <https://doi.org/10.1002/mds.23396>
20. Olanow CW, Rascol O, Hauser RA, et al. A double-blind, delayed-start trial of rasagiline in Parkinson's disease. *N Engl J Med* 2009;361:1268-1278. <https://doi.org/10.1056/NEJMoa0809335>
21. Sánchez-Ferro Á, Matarazzo M, Martínez-Martín P, et al. Minimal clinically important difference for UPDRS-III in daily practice. *Mov Disord Clin Pract* 2018;5:448-450. <https://doi.org/10.1002/mdc3.12632>
22. Giancardo L, Sánchez-Ferro A, Butterworth I, et al. Psychomotor impairment detection via finger interactions with a computer

- keyboard during natural typing. *Sci Rep* 2015;5:1-8. <https://doi.org/10.1038/srep09678>
23. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-444. <https://doi.org/10.1038/nature14539>
  24. Wissler C. The Spearman correlation formula. *Science* 1905;22:309-311. <https://doi.org/10.1126/science.22.558.309>
  25. Spearman C. The proof and measurement of association between two things. By C. Spearman, 1904. *Am J Psychol* 2012;100:441-471.
  26. Litvan I, Goldman JG, Tröster AI, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Mov Disord* 2012;27:349-356. <https://doi.org/10.1002/mds.24893>
  27. 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. <https://doi.org/10.1016/j.jbi.2008.08.010>
  28. van Rossum G, de Boer J. Interactively testing remote servers using the Python programming language. *CWI Q* 1991;4:283-303.
  29. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2016.
  30. Stebbins GT, Goetz CG, Burn DJ, et al. How to identify tremor dominant and postural instability/gait difficulty groups with the movement disorder society unified Parkinson's disease rating scale: comparison with the Unified Parkinson's Disease Rating Scale. *Mov Disord* 2013;28:668-670. <https://doi.org/10.1002/mds.25383>
  31. Taylor Tavares AL, Jefferis GSXE, Koop M, et al. Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation. *Mov Disord* 2005;20:1286-1298. <https://doi.org/10.1002/mds.20556>
  32. Noyce AJ, Nagy A, Acharya S, et al. Bradykinesia-akinesia incoordination test: validating an online keyboard test of upper limb function. *PLoS ONE* 2014;9:e96260. <https://doi.org/10.1371/journal.pone.0096260>
  33. Bishop CM. *Pattern Recognition and Machine Learning*. 1st ed. New York, NY: Springer, 2006.

## Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.