

1 **A generalizable and open-source algorithm for real-life monitoring**
2 **of tremor in Parkinson's disease**

3

4 Nienke A. Timmermans^{1*}, Roberta Terranova^{1,2*}, Diogo C. Soriano^{3,4}, Hayriye Cagnan^{3,5},
5 Yordan P. Raykov⁶, Ioan Gabriel Bucur⁷, Bastiaan R. Bloem¹, Rick C. Helmich¹, Luc J.W. Evers¹

6

7 ¹ *Department of Neurology, Radboud University Medical Center, Donders Institute for Brain,*
8 *Cognition and Behavior, Center of Expertise for Parkinson and Movement Disorders,*
9 *Nijmegen, The Netherlands*

10 ² *Department of Medical, Surgical Sciences and Advanced Technologies G.F. Ingrassia,*
11 *University of Catania, Catania, Italy*

12 ³ *MRC Brain Network Dynamics Unit, Nuffield Department of Clinical Neurosciences,*
13 *University of Oxford, Oxford, United Kingdom*

14 ⁴ *Federal University of ABC, São Bernado do Campo, Brazil*

15 ⁵ *Department of Bioengineering, Imperial College London, London, United Kingdom*

16 ⁶ *School of Mathematical Sciences, University of Nottingham, Nottingham, United Kingdom*

17 ⁷ *Institute for Computing and Information Sciences, Radboud University, Nijmegen, The*
18 *Netherlands*

19 ** These two authors contributed equally to this work.*

20

21 Corresponding author:

22 Nienke A. Timmermans

23 Nienke.timmermans@radboudumc.nl

Abstract

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Wearable sensors can objectively and continuously monitor daily-life tremor in Parkinson's Disease (PD). We developed an open-source algorithm for real-life monitoring of PD tremor which achieves generalizable performance across different wrist-worn devices. We achieved this using a unique combination of two independent, complementary datasets. The first was a small, but extensively video-labeled gyroscope dataset collected during unscripted activities at home (n=24 PD; n=24 controls). We used this to train and validate a logistic regression tremor detector based on cepstral coefficients. The second was a large, unsupervised dataset (n=517 PD; n=50 controls, data collected for 2 weeks with a different device), used to externally validate the algorithm. Results show that our algorithm can reliably quantify real-life PD tremor. Weekly aggregated tremor time and power showed excellent test-retest reliability and moderate correlation to MDS-UPDRS rest tremor scores. This opens possibilities to support clinical trials and individual tremor management with wearable technology.

Introduction

Parkinson's Disease (PD) is the fastest-growing neurodegenerative disease with a prevalence of 11.8 million people worldwide in 2021¹. There is currently no cure for PD, but treatments are available to alleviate symptoms. Tremor is one of the cardinal motor symptoms of PD, occurring in approximately 75% of persons with PD^{2,3}. It is experienced as one of the most bothersome symptoms, especially in early disease stages^{4,5}. The typical Parkinsonian tremor is a 3-7 Hz rest tremor in the upper limb, although it may also occur in the lower limb or jaw^{6,7}. Re-emergent tremor, pure postural tremor and kinetic tremor are also frequently observed⁶.

Adequate assessment of tremor severity and its context-dependency could allow for more personalized treatment, and more efficient clinical trials evaluating new treatments^{3,8}. However, the highly variable expression of tremor complicates its evaluation during the typically brief, episodic clinical visits⁴. In many patients, the expression of tremor is influenced by emotional and cognitive stress, voluntary movements, and timing of treatments⁹. Furthermore, the currently used Movement Disorder Society - Unified Parkinson's Disease Rating Scale (MDS-UPDRS) for clinical tremor assessment is limited by inter- and intra-rater variability¹⁰. A home diary is another tool for tremor assessment, but suffers from poor compliance and recall bias¹¹.

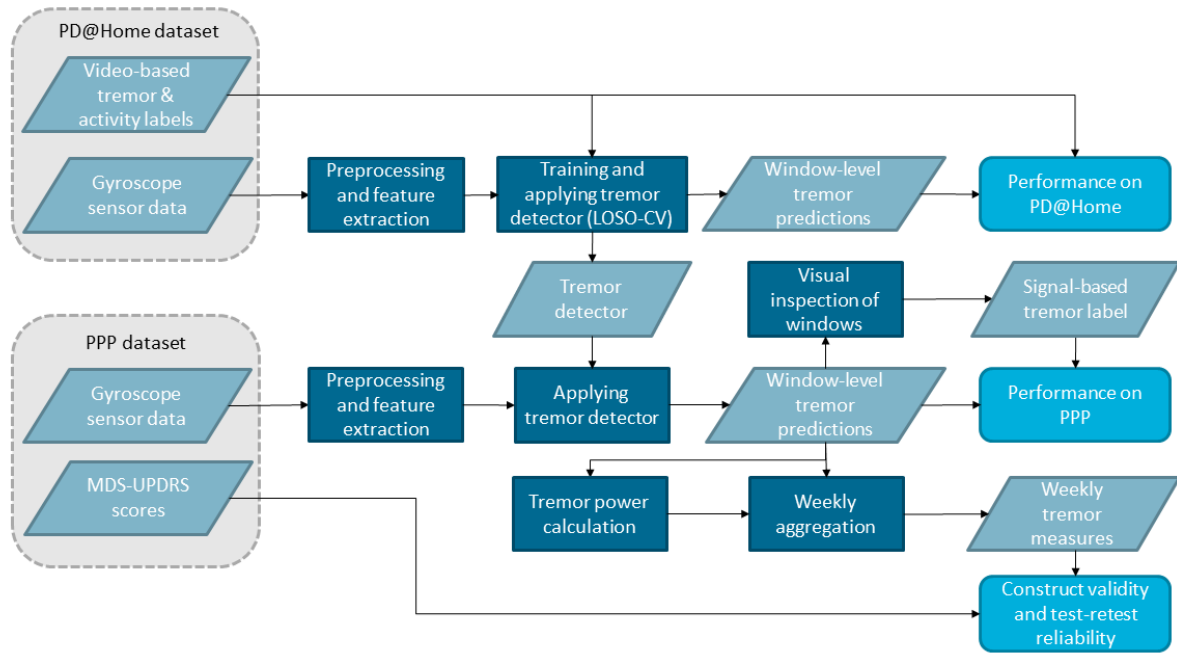
These limitations could be overcome by using wearable sensors that continuously and objectively measure tremor presence and severity in daily life^{12,13}. Several algorithms have been developed for remote monitoring of tremor in PD, mostly using a wrist-worn accelerometer, gyroscope, or a combination thereof¹⁴⁻²⁶. However, these wearable sensors and algorithms have not yet been introduced widely into clinical practice and research^{12,13,27,28}.

Two critical aspects must be addressed to enable robust monitoring of tremor in daily life. First, training of supervised machine learning algorithms relies on accurately labeled datasets based on concurrent video recordings, which are often collected in highly controlled environments²⁶. However, algorithms trained on these datasets are usually not generalizable to data collected in real life²⁹. Second, there is a need for open-source algorithms with generalizable performance across different sensor devices and across study populations^{30,31}. This would facilitate its adoption in clinical trials and care, and facilitate bring-your-own-device solutions. However, external validation of tremor detection algorithms on datasets using different devices is often lacking, despite potential performance degradation due to varying sensor positions or sensitivity to noise^{21,32}.

Here, we address these issues by developing and validating an open-source algorithm for real-life monitoring of rest tremor in PD. The algorithm is trained on video-labeled data collected with a wrist-worn gyroscope sensor, during unscripted daily life activities

1 (Parkinson@Home Validation Study) ³³. We subsequently applied the algorithm to free-living
 2 data collected during the Personalized Parkinson Project (PPP) using another wrist-worn
 3 gyroscope sensor ³⁴. Using this dataset, we assessed the generalizability of the developed
 4 algorithm. Finally, we determined the construct validity and test-retest reliability of several
 5 sensor-derived weekly aggregated tremor measures.

6 Results



7

8 **Figure 1: Overview of the development and validation of the tremor detection algorithm using the PD@Home and**
 9 **PPP datasets.** The tremor detector was trained and applied to obtain window-level tremor predictions, which were
 10 subsequently aggregated to obtain weekly tremor measures. LOSO-CV = leave-one-subject-out cross-validation.

11 Figure 1 shows a high-level overview of the steps taken to develop and validate the tremor
 12 detection algorithm. The performance of the algorithm on the PD@Home and PPP datasets
 13 will be described below, as well as the construct validity and test-retest reliability of weekly
 14 aggregated tremor measures.

15

16 Tremor detection

17 Performance on PD@Home

18 Details of the tremor detection algorithm can be found in the methods section. Briefly, we
 19 trained a logistic regression classifier to detect tremor based on scale-insensitive mel-
 20 frequency cepstral coefficients (MFCCs) that capture the periodicity in wrist gyroscope
 21 signals. After applying the logistic regression classifier, we filtered out detected tremor
 22 windows with a frequency peak outside the rest tremor range of 3-7 Hz ⁷, since we aimed for

1 a high specificity for rest and re-emergent tremor. We also filtered out all windows with
 2 detected non-tremor arm movements to further increase the specificity for rest tremor.

3 Table 1 shows the sensitivity and specificity of the tremor detection algorithm on the
 4 PD@Home dataset, for different subgroups and types of activities. The specificity in the
 5 tremor PD group was lower than in the non-tremor PD and non-PD control groups, and this
 6 difference remained after correcting for differences in the prevalence of the four activities
 7 that all subjects performed (sitting, standing, gait and postural transitions). Filtering out
 8 windows with detected non-tremor arm movements (e.g. arm swing during gait) increased
 9 the specificity for tremor from 96% to 97%, corresponding to a 25% reduction in false
 10 positives (see Supplementary Table 1 for the performance without filtering out windows with
 11 detected non-tremor arm movements). Although this comes at the cost of a slightly
 12 decreased overall sensitivity, it reduces the number of false positive tremor windows when
 13 measuring over a longer time period where often the non-tremor class is more prevalent.

14 Activities in the control group that were sometimes still misclassified as tremor included
 15 brushing teeth, washing the hands or dishes, stirring in a cup of tea, nodding and typing. In
 16 one control participant, a 3 Hz tremor was observed while the participant was holding a
 17 spoon. In addition, some 4-second predicted tremor windows showed a 3-5 Hz oscillatory
 18 gyroscope signal, but were not accompanied by rhythmic, oscillatory movement visible on the
 19 video recording.

20

21 *Table 1: Tremor detection performance on PD@Home (average across leave-one-subject out cross-validation folds, and*
 22 *standard deviation between brackets), for different subgroups and different types of activities. The number of participants*
 23 *with data from the different sub-classes is indicated between brackets. Annotations for "Significant upper limb activity" and*
 24 *"Periodic" were only available for the 8 PD patients with tremor.*

	Sensitivity	Specificity	Weighted specificity across sitting, standing, gait and postural transitions
Overall (n=48)	0.61 (0.20)	0.97 (0.05)	
Subgroup			
Tremor PD (n=8)	0.61 (0.20)	0.91 (0.04)	0.92 (0.05)
Non-tremor PD (n=16)	-	0.96 (0.06)	0.97 (0.05)
Non-PD controls (n=24)	-	0.99 (0.01)	0.99 (0.01)
Type of activity			
Sitting (n=8/n=48)	0.62 (0.28)	0.96 (0.07)	
Standing (n=7/n=48)	0.48 (0.34)	0.99 (0.03)	
Gait (n=7/n=48)	0.01 (0.03)	1 (0)	-
Postural transitions (n=48)	-	1 (0.01)	
Running/Exercising (n=5)	-	1 (0)	
Cycling (n=14)	-	1 (0)	-
Driving motorized vehicle (n=2)	-	1 (0.01)	
Significant upper limb activity (n=8)	-	0.98 (0.03)	-

Periodic activities (n=5)	-	0.98 (0.04)	-
---------------------------	---	-------------	---

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Generalizability to PPP

To assess the generalizability of the developed tremor detection algorithm to PPP, we visually inspected and annotated a sample of 7160 windows among 179 PD participants from the PPP (Table 2). Since all windows annotated as ‘doubt’ could be either tremor or non-tremor, we considered the two extreme scenarios (all doubt labels are considered ‘annotated tremor’ or ‘annotated non-tremor’) to obtain a performance range. The positive predictive value was 0.80 – 0.94, and the negative predictive value was 0.88 – 0.96. Taking into account the prevalence of predicted tremor and non-tremor in each subject, the average sensitivity of the tremor detector on the PPP dataset was 0.37 – 0.58, and the average specificity was 0.98 – 1. When only considering participants with clinically observed rest tremor (MDS-UPDRS 3.17 \geq 1), the average sensitivity was 0.44 – 0.64 and the average specificity was 0.98 – 0.99 (Supplementary Table 2), which is comparable to the performance in PD@Home.

Table 2: Annotations based on visual inspection of a sample of predicted tremor and predicted non-tremor windows among 179 PD participants from the PPP.

	Predicted tremor	Predicted non-tremor
Annotated tremor	1,622	625
Annotated non-tremor	67	3,954
Annotated doubt	232	660
Total	1,921	5,239

17
18
19
20
21
22
23
24
25
26

Weekly tremor measures

From the amount of tremor windows detected during daytime (08:00 am - 10:00 pm) and the tremor power per window, we derived four weekly aggregated tremor measures: tremor time, median tremor power, modal tremor power, and 90th percentile of tremor power. Tremor time reflects the number of detected tremor windows while the arm was at rest or in a stable posture. The median and modal tremor power represent the typical tremor severity, whereas the 90th percentile of tremor power captures the maximal tremor severity.

Construct validity

The construct validity of weekly tremor measures was assessed in three ways. First, we compared the weekly tremor measures between three groups of PD subjects with different clinical tremor severity ratings (MDS-UPDRS 3.17 ON) of the device-sided arm. Tremor time increased across these three groups of PD subjects, and was larger in PD group 0 compared to the control group (see Figure 2a, using Dunn’s test with Bonferroni correction). Figure 2b shows the distribution of median tremor power across the different PD groups. The median tremor power in PD group 2 was larger than in PD group 0 and 1, but PD group 0 and 1 did

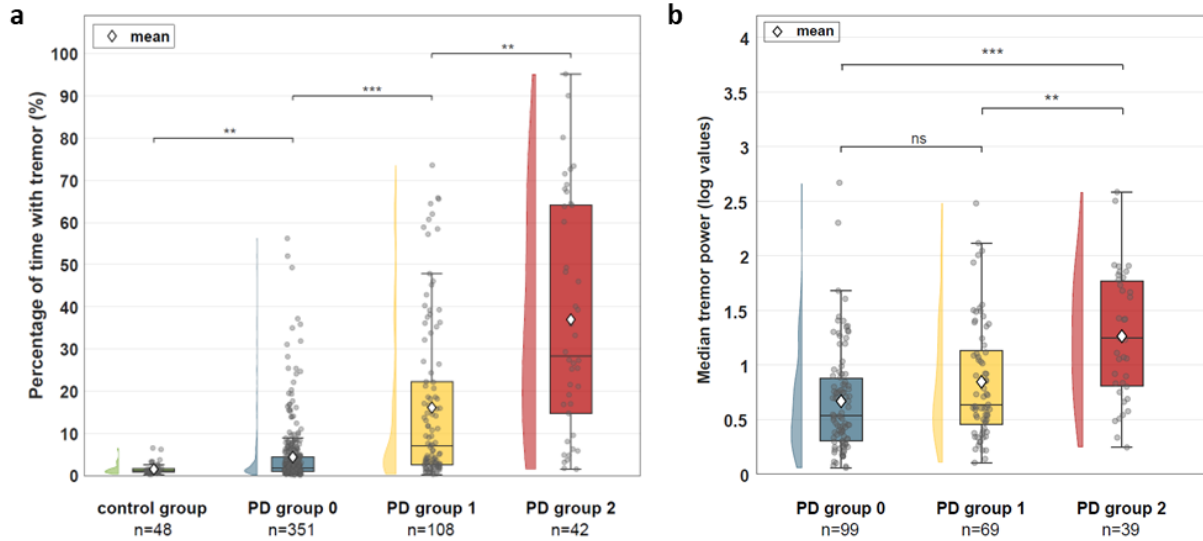
1 not differ significantly. Group differences for the mode and 90th percentile of tremor power
2 are visualized in Supplementary Figure 1. When using the MDS-UPDRS 3.17 score assessed in
3 OFF instead of ON to stratify PD subjects, all group comparisons for the four weekly tremor
4 measures were significantly different (Supplementary Figure 2).

5 We then assessed the association between the weekly tremor measures and clinical
6 severity ratings of different tremor types in different body parts using univariable correlations
7 (Spearman's rank correlation) and multivariable linear regression (Figure 3). Weekly tremor
8 time correlated positively with all MDS-UPDRS part III tremor scores and with the patient-
9 reported tremor score of MDS-UPDRS part II. Correlation was highest with the rest tremor
10 constancy score (MDS-UPDRS 3.18) and rest tremor severity score (MDS-UPDRS 3.17) in the
11 device-sided arm, followed by the patient-reported tremor score (MDS-UPDRS 2.10) and
12 postural tremor severity score in the device-sided arm (MDS-UPDRS 3.15). The weekly tremor
13 power measures also positively correlated with these clinical scores. To assess the added
14 value of filtering out non-tremor arm movements, we also assessed correlations without
15 filtering (Supplementary Figure 3). The correlation of the 90th percentile of tremor power and
16 clinical rest tremor constancy score decreased ($p < 0.01$), but the other correlations did not
17 significantly change.

18 As expected, multivariable linear regression showed that rest tremor severity (MDS-UPDRS
19 3.17) in the device-sided arm was the strongest predictor for all weekly tremor measures
20 (Figure 3b). Rest tremor severity in the non-device-sided arm and device-sided leg, and
21 postural tremor severity (MDS-UPDRS 3.15) in the device-sided arm were also significant
22 predictors of weekly tremor time. To investigate why tremor time in PD group 0 (with clinically
23 no rest tremor in the device-sided arm) was larger than in the control group, we systematically
24 assessed the weekly tremor time in subgroups by incrementally excluding subjects from PD
25 group 0 (see Figure 4). After excluding subjects with a rest tremor severity score (MDS-UPDRS
26 3.17) ≥ 1 in the device-sided leg and subjects with a postural tremor severity score (MDS-
27 UPDRS 3.15) ≥ 1 in the device-sided arm, the subgroups did not differ significantly from the
28 control group.

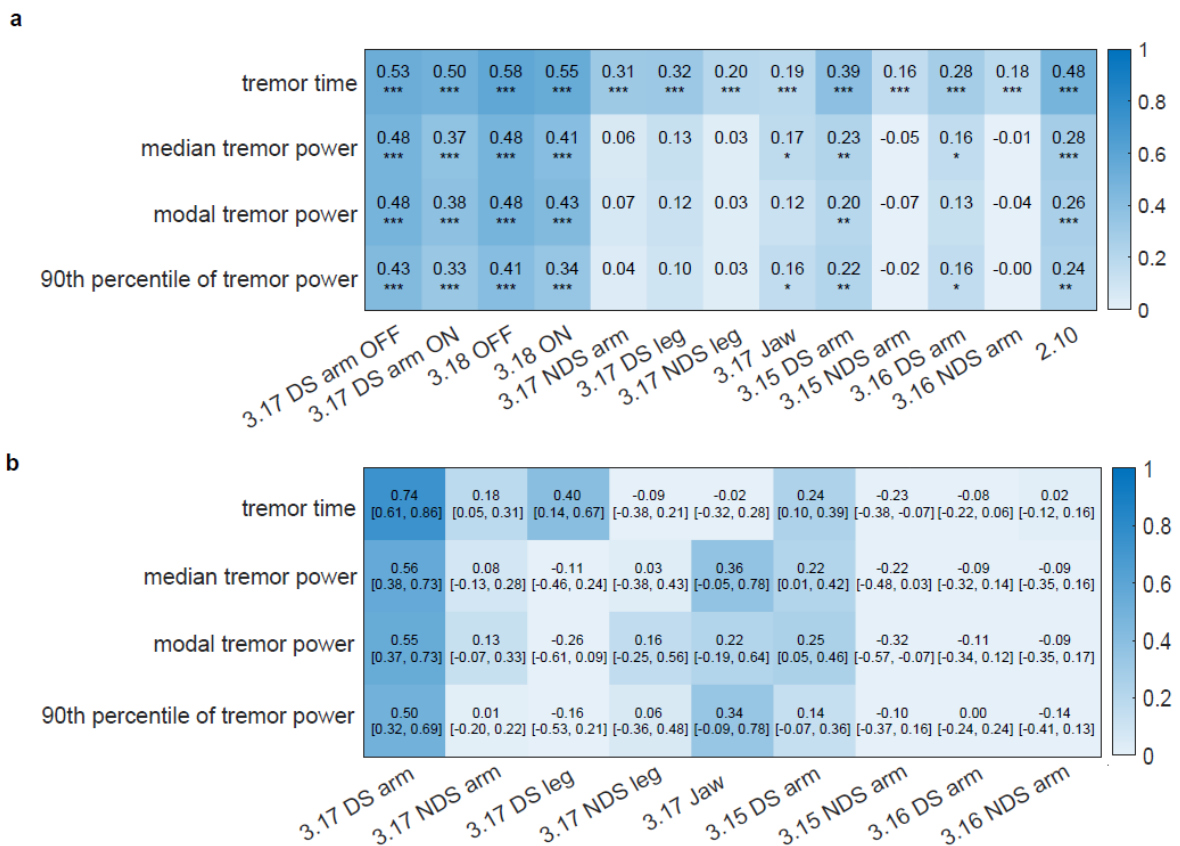
30 **Test-retest reliability**

31 In 495 PD subjects, the first and second week of data collection were valid and therefore
32 used to assess the test-retest reliability of weekly tremor time. Tremor power measures were
33 computed for the first two weeks in 187 PD subjects with at least 3.5% tremor time in both
34 weeks. Excellent test-retest reliability was found for all weekly tremor measures, with ICC for
35 tremor time of 0.98 (95%-CI 0.97 – 0.98), ICC for median tremor power of 0.96 (95%-CI 0.95
36 – 0.97), ICC for modal tremor power of 0.94 (95%-CI 0.92 – 0.95) and ICC for 90th percentile
37 of tremor power of 0.95 (95%-CI 0.93 – 0.96).



1
2
3
4
5
6
7
8
9

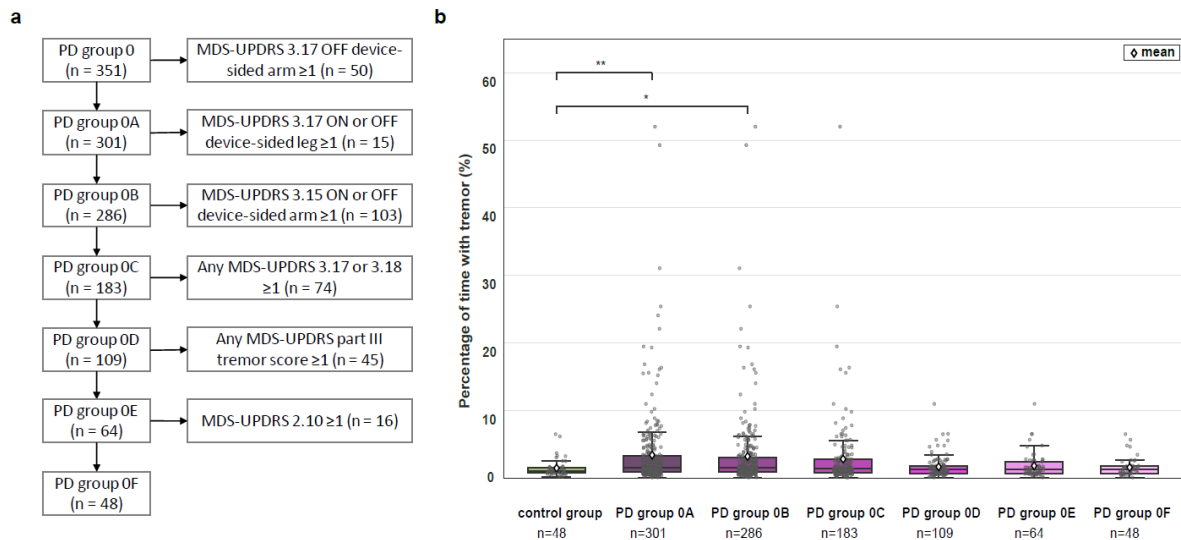
Figure 2: Group differences in weekly tremor measures. Three groups of PD participants with different clinical tremor severity were used (groups 0, 1 and 2 with MDS-UPDRS 3.17 of 0, 1 and ≥ 2 assessed in ON motor state in the device-sided arm). Non-PD controls were added for tremor time. **a:** Tremor time was calculated as the number of detected tremor windows divided by all windows without non-tremor arm movements during daytime (08:00 am – 10:00 pm), and expressed as percentage. **b:** Median tremor power was calculated across all detected tremor windows during daytime, but only assessed if the tremor time was $\geq 3.5\%$. For both measures, the first week of collected data of PPP was used. The number of subjects in each subgroup is indicated. Significant differences (using Dunn’s test with Bonferroni correction) between subsequent groups are shown (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).



10
11
12
13

Figure 3: Univariable correlation and multivariable linear regression of weekly tremor measures with clinical tremor scores. In **a** Spearman’s correlation coefficients are shown with their significance level (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, corrected using the false discovery rate method). Correlations with the rest tremor severity (3.17) and constancy (3.18)

1 scores in the device-sided arm were assessed for ON and OFF scores separately. The other scores were averaged over ON
 2 and OFF motor states. In **b** the beta coefficients of the multivariable linear regression are shown with their 95% confidence
 3 intervals. Here, each weekly tremor measure was used as outcome and MDS-UPDRS tremor scores (averaged over ON and
 4 OFF conditions) as independent predictors. Beta coefficients were standardized by dividing them by the standard deviation
 5 of the weekly tremor measure, so that they represent the expected change in the measure (in standardized units) due to an
 6 increase in each MDS-UPDRS score of 1, with all other scores unchanged. The rest tremor constancy score (3.18) and
 7 patient-reported tremor score (2.10) were not included in **b**, since these were not assessed for each body part separately.
 8 3.15 = postural tremor severity, 3.16 = kinetic tremor severity, 3.17 = rest tremor severity, 3.18 = rest tremor constancy, DS
 9 = device-sided, NDS = non-device-sided.



10
 11 **Figure 4: Tremor time in PD participants without clinically observed rest tremor in the device-sided arm.** **a:** Flowchart
 12 showing the incremental exclusion of subjects with clinical tremor scores ≥ 1 from PD group 0 (with MDS-UPDRS 3.17 ON of
 13 0 in the device-sided arm). The order was based on the items' predictive value for weekly tremor time in the multivariable
 14 regression. **b:** Tremor time (as percentage of daytime without non-tremor arm movements) measured in the first week of
 15 data collection of PPP in controls and different subgroups of PD group 0. The number of subjects in each subgroup is
 16 indicated. Significant differences (using Dunnett's test) with the control group are shown (* $p < 0.05$, ** $p < 0.01$, ***
 17 $p < 0.001$).

18 Discussion

19
 20 We developed and validated a generalizable algorithm for real-life monitoring of rest
 21 tremor in persons with PD, using a unique combination of two independent and
 22 complementary datasets: a small, but extensively video-labeled dataset at home, and a large,
 23 unsupervised dataset. By training a relatively simple logistic regression algorithm for tremor
 24 detection based on scale-insensitive MFCCs, we achieved generalizable performance when
 25 applied to a new dataset using a different wrist-worn device. This approach facilitates the use
 26 of our algorithm on different datasets and devices. In addition, we showed that weekly
 27 aggregated measures for tremor time and tremor power are reliable and clinically valid.

28 Our tremor detection algorithm had a sensitivity of 0.61 and specificity of 0.97 on the
 29 PD@Home data. On the PPP data we found a comparable performance in the group with
 30 clinically observed rest tremor in the device-sided arm. Furthermore, the amount of tremor

1 detected in non-PD controls of PPP aligned with the specificity found in non-tremor PD
2 participants and non-PD controls of PD@Home. The generalizability of published tremor
3 detection algorithms is often unknown. One study found a decrease in specificity from 94.2%
4 to 89.8% when applying a previously published tremor detection algorithm to their own
5 collected dataset ²¹, showing the importance of external validation. We found a slightly
6 decreased sensitivity and increased specificity in the group without clinically observed rest
7 tremor in the device-sided arm in PPP. A possible explanation is that our annotations based
8 on visual inspection of gyroscope signals were more sensitive for tremor than the video-based
9 annotations. Missed subtle tremor on the video recordings from PD@Home could also explain
10 the decreased specificity in the tremor PD participants compared to the non-tremor PD
11 participants and non-PD controls (i.e., some of the false-positives in the tremor PD group
12 might have been actual tremor which was missed). Comparison to the performance of other
13 tremor detection algorithms described in the literature is difficult due to variability in data
14 collection across studies ¹⁴⁻²⁶. In addition, our main aim was to obtain generalizable
15 performance across different studies, rather than maximizing performance on a single
16 dataset.

17 For its application in continuous, real-life monitoring of tremor, we prioritized high
18 specificity, even at the expense of a slight decrease in sensitivity. A high specificity will limit
19 the amount of false positives, which is important for continuous monitoring given the
20 significant amount of time without tremor. Although our trained classifier combined with
21 filtering by a clinical criterion achieved high specificity, it may not be feasible to eliminate all
22 false positives using gyroscope signals, because we observed several rhythmic activities that
23 closely resembled tremor (examples included e.g. brushing teeth, stirring in a cup of tea,
24 nodding). Besides, the wrist-worn gyroscope sensor may detect subtle physiologic tremor in
25 the forearm. Physiologic tremor is often described as a 8-12 Hz tremor, but its frequency
26 depends on the mechanical characteristics of the joint ³⁶. The frequency of the unexplained
27 rhythmic, oscillatory signals that were detected as tremor in the control group of PD@Home
28 corresponded to normal forearm tremor at 3-5 Hz ³⁶.

29 Our weekly tremor measures showed good to excellent test-retest reliability and were
30 able to distinguish between non-PD controls and PD groups with different clinical rest
31 tremor severity. Tremor time observed in PD subjects with a clinical rest tremor severity
32 score of 0 (in both ON and OFF conditions) in their device-sided arm was larger than in
33 controls, suggesting that remote monitoring may offer improved sensitivity for measuring
34 PD tremor compared to clinical assessments in the device-sided arm. After excluding
35 subjects with clinically observed rest tremor in the device-sided leg or postural tremor in
36 the device-sided arm, the subgroup did not differ significantly from the control group
37 regarding tremor time. We believe that a likely explanation is that participants with rest
38 tremor in their device-sided leg have rest tremor in their device-sided arm as well

1 (perhaps subtle and subclinical), which was not observed during the brief clinical
2 assessment. Another possible explanation is that rest tremor in other limbs was
3 detected at the wrist, for example when the device-sided arm was resting on the device-
4 sided leg. The association between the clinical evaluation of postural tremor and the
5 measured tremor time can likely be explained by the presence of re-emergent tremor in
6 the device-sided arm, which is a continuation of rest tremor during stable posturing with
7 a similar frequency³⁵.

8 The moderate correlation of weekly tremor measures with MDS-UPDRS part III scores is in
9 line with previous studies that continuously monitored PD tremor for more than 24 hours
10^{14,22,23}. Similar findings have been obtained in essential tremor³⁷. At first glance, this may be
11 viewed as a limitation of our algorithm, but in fact it has been shown that clinical assessments
12 of PD tremor are very variable⁸, in line with the waxing and waning nature of PD tremor³⁸,
13 as well as its strong sensitivity to stress^{39,40}. Hence, the primary goal of tremor detection
14 algorithms should not be to maximize the correlation with an imperfect measure (i.e., the
15 MDS-UPDRS, which provides only a snapshot in a very specific setting that is prone to observer
16 bias). In fact, this raises the question which approach should become the gold standard
17 against which to compare future tremor outcomes: the subjective and episodic clinically
18 based score, or the objective score obtained at home from a digital device.

19 Our weekly tremor measures also moderately correlated with the patient-reported tremor
20 score (2.10), with the highest correlation for tremor time. Understanding how sensor-derived
21 tremor measures relate to experienced tremor burden by patients is an important
22 prerequisite for implementing remote tremor monitoring in clinical care, as well as for
23 obtaining regulatory approval to use the sensor-derived tremor measures as primary or
24 secondary endpoint in clinical trials²⁸.

25 There are some limitations to this study. First, the number of PD@Home participants with
26 tremor was limited (n=8), causing the estimated sensitivity of the tremor detector to be less
27 precise. This was partially offset by the large number of participants in the second dataset
28 where we replicated the findings. Second, the PPP dataset did not contain gold-standard
29 tremor labels based on video-recordings. However, visual inspection of gyroscope signals was
30 considered adequate to show that the performance on PD@Home was generalizable to PPP.
31 Future research could elaborate on this by validating our algorithm on other external datasets
32 and sensor devices. Third, by using only a single wrist-worn sensor, we focused on measuring
33 tremor in the arm, although tremor in the legs and jaw may also occur. Besides, a wrist-worn
34 sensor might not be able to detect subtle tremor in the fingers, limiting the sensitivity for
35 tremor. Nevertheless, including more sensors comes at the cost of decreasing compliance,
36 and tremor in the arm is the most prevalent type of tremor in PD⁴¹. Our prior work in the PPP
37 cohort showed that long-term compliance (over a 2-year time frame) with using a wrist-worn

1 device is excellent, with a median wear time of 22 hours per day, and a drop-out rate of just
2 5.4%⁸.

3 Real-life monitoring of PD rest tremor with our developed tremor detector could benefit
4 both patient care and clinical trials. By averaging over larger time periods in patients' own
5 environments, sensor-based tremor quantification provides a solution to the problem that
6 hospital-based clinical tremor ratings are often not representative of the actual PD tremor
7 severity⁸. Clinical trials may benefit from the potentially improved sensitivity of objective
8 tremor measures compared to current clinical scales, because this will reduce the required
9 sample size and costs²⁸. However, before objective tremor measures can be implemented as
10 outcome measures in randomized controlled trials, it will be necessary to determine their
11 sensitivity to detect a clinically relevant change⁴². Besides, sensor-based tremor assessments
12 over longer periods of time may provide more insight into clinically relevant tremor
13 characteristics at the individual level, such as the effects of (dopaminergic) medication, the
14 effect of stress on tremor in daily life, and change in tremor severity over time. This could all
15 help in developing personalized treatments and ultimately improving the quality of life of
16 persons with PD.

17 **Methods**

18 **Datasets**

19 The Parkinson@Home Validation study was used to train the tremor detector. Details
20 concerning the data collection procedure are described elsewhere³³. Briefly, 25 persons with
21 PD and 25 age-matched non-PD controls were visited at their own homes and recorded on
22 video for at least one hour while they performed unscripted daily life activities, before and
23 after intake of dopaminergic medication for PD participants. Simultaneously, sensor data was
24 collected at several body locations. For this study, we used the gyroscope data collected with
25 the wrist-worn Physilog 4 device at a sampling rate of 200 Hz. The advantage of using
26 gyroscope data over accelerometer data is that filtering out the gravitational component is
27 not needed⁴³. For PD participants with tremor, the sensor at the side with most severe tremor
28 was chosen. We selected sides to match for hand dominance for the other participants. We
29 excluded two subjects due to technical issues with the sensor devices, yielding a training
30 dataset of 24 PD participants and 24 controls (for demographic and clinical characteristics see
31 Supplementary Table 3).

32 Based on the concurrent video recordings, the presence and severity of tremor during the
33 unscripted activities were annotated by trained research assistants, and checked by a
34 neurologist. The annotations included any type of tremor observed, because the video
35 recordings during daily life activities were not deemed suitable to reliably distinguish between
36 different tremor subtypes. The video annotation protocol can be found in Supplementary

1 Figure 4. In 8 PD patients, tremor was observed during the unscripted daily life activities,
2 whereas in the other 16 PD patients, no tremor was observed. Because of the low prevalence
3 of moderate and severe tremor in the dataset, we focused on detecting only the presence of
4 tremor. The median amount of annotated tremor used for training the tremor detector was
5 14 minutes (inter-quartile range (IQR) 7 – 36 minutes across the eight PD participants with
6 tremor in PD@Home), and the median amount of annotated non-tremor was 109 minutes
7 (IQR 86 – 145 minutes across all 48 participants in PD@Home). For tremulous PD participants,
8 significant upper limb activities and periodic activities that could resemble tremor were
9 annotated as well. The presence of the following activities was annotated for all participants:
10 sitting, standing, gait, postural transitions, running or exercising, cycling and driving a
11 motorized vehicle.

12 The trained tremor detector was subsequently applied to gyroscope data collected with
13 the Verily Study Watch during the Personalized Parkinson Project ³⁴. 520 early-stage PD
14 participants were asked to wear the watch preferably 24/7 for 2 years, except during charging
15 (approximately 1 hour per day), at their preferred side. In addition, 50 non-PD controls were
16 asked to wear the same watch for 1 year. The Verily Study Watch typically had a sampling
17 frequency 100 Hz, although a lower sampling frequency of 50 Hz occurred in some
18 participants at the beginning of the study. In this study, we used data collected during the
19 first 3 weeks of follow-up. For three patients, no sensor data was collected, yielding a dataset
20 of 517 patients and 50 controls. Demographic and clinical information, including MDS-UPDRS
21 part II and III scores in OFF and ON motor conditions for PD subjects, was collected at baseline
22 (see Supplementary Table 4).

23

24 **Tremor detection**

25 **Preprocessing, feature extraction and tremor detection**

26 Gyroscope signals were down-sampled to 50 or 100 Hz, depending on the closest original
27 sampling frequency. Based on 4-second windows, we obtained the power spectral density
28 (PSD) of each gyroscope axis signal (using Welch’s method). From the summed PSD across the
29 three axes, we extracted the peak frequency and 12 mel-frequency cepstral coefficients
30 (MFCCs). MFCCs were computed using 15 filterbanks in the range of 0-25 Hz with filter edges
31 defined by the adjusted mel-scale for inertial signals ⁴⁴. MFCCs have been popular in audio
32 signal analysis because they capture the overall shape of the spectral envelope with a small,
33 uncorrelated and scale-invariant feature set. The scale-invariance of MFCCs is important for
34 the generalizability of the tremor detection algorithm and its robustness to gyroscope drift.
35 Earlier work on tremor detection algorithms based on MFCCs showed promising results ²⁶.

36 Based on the video annotations, windows extracted from the PD@Home dataset were
37 labeled as tremor if this was present for at least 50% of window time. The most prevalent
38 type of activity was chosen as the activity label. A logistic regression classifier was trained to

1 detect tremor based on the MFCCs, with the threshold set at 95% specificity. During training,
2 windows labeled as cycling were oversampled by a factor of 100 to penalize the algorithm for
3 detecting cycling as tremor.

4 We aimed to detect rest and re-emergent tremor with high specificity. Therefore, an
5 additional clinical criterion was set for detecting tremor: the peak frequency must be within
6 the rest tremor frequency range of 3-7 Hz ⁷. Furthermore, we used a rule-based non-tremor
7 arm movements detector to filter out windows with simultaneous non-tremor arm
8 movements. The aim of this was 1) to further improve the specificity for rest tremor, and 2)
9 to improve the accuracy of measured tremor power. If rest tremor and non-tremor arm
10 movements are simultaneously present (e.g. tremor during gait with arm swing), the tremor
11 peak in the power spectral density could be distorted by the higher harmonics of the slower
12 movement (Supplementary Figure 5). Gyroscope power within the 0.5-3 Hz band was used to
13 distinguish slow voluntary movements from higher frequency tremor and lower frequency
14 gyroscope drift. We visualized the distribution of this feature across all windows of the
15 PD@Home dataset. The threshold for non-tremor arm movements was selected based on K-
16 means clustering with two clusters (see Supplementary Figure 6).

18 **Evaluation of tremor detection performance**

19 First, the tremor detection algorithm was evaluated on the PD@Home dataset using
20 Leave-One-Subject-Out Cross-Validation (LOSO-CV), i.e., we evaluated its sensitivity and
21 specificity on one subject, after training on all other subjects. In addition to the overall
22 performance, we compared the specificity stratified for different annotated activities. Video
23 recordings of detected tremor windows in the control group were visually inspected to
24 identify other activities that resemble tremor.

25 Next, we assessed the generalizability of the developed tremor detection algorithm to PPP
26 as the external dataset. We visually inspected and annotated a sample of 7160 windows,
27 sampled from the third week of data collection in 179 PD participants. Based on an earlier
28 version of our tremor detection algorithm, for each subject we randomly selected 20
29 predicted tremor and 20 predicted non-tremor windows between 7:00 am and 11:00 pm in
30 the third week of data collection. The ratio between predicted tremor and non-tremor
31 windows in the final dataset changed from 1:1 to approximately 1:3, because of further
32 improvements of the tremor detection algorithm to increase its specificity based on
33 PD@Home (i.e., the addition of an extra criterion for rest tremor and filtering out windows
34 with non-tremor arm movements). After visual inspection of the raw gyroscope signals and
35 PSDs, we annotated the windows as tremor, non-tremor or doubt (the annotation protocol is
36 given in Supplementary Figure 7). The first 1800 windows were annotated by the first author
37 of this paper and by a second annotator. Windows for which there was disagreement or doubt
38 were discussed. The inter-rater agreement was substantial (Cohen's kappa of 0.73), therefore

1 only the first author of this paper annotated the remaining windows. The signal-based tremor
2 labels were used to estimate the sensitivity and specificity for PPP, which were then
3 compared to the performance on PD@Home.

4 To evaluate the effect of tremor power on algorithm performance, we also stratified the
5 subjects across their clinical rest tremor severity score (MDS-UPDRS 3.17 OFF) at the device-
6 sided arm (60 subjects had a score of 0, 60 had a score of 1 and 59 had a score of ≥ 2). The
7 effect of other PD motor symptoms on algorithm performance was assessed by splitting the
8 subjects into two other subgroups based on the total UPDRS part III score minus the total
9 tremor score (for the first group this score was < 28 and for the second group ≥ 28).

11 **Weekly tremor measures**

12 After we obtained window-level tremor predictions for the first two weeks of PPP data, we
13 quantified tremor time and power based on data collected during daytime (08:00 am - 10:00
14 pm). These tremor measures were aggregated over one-week periods, to enable assessment
15 of longitudinal progression in future research. Only valid weeks were considered, which were
16 defined, based on previous research, as weeks with at least three valid days with ≥ 10 hours
17 of sensor data ⁴⁵.

18 For each valid week, tremor time was calculated as the number of detected tremor
19 windows divided by the number of windows without detected non-tremor arm movements
20 during valid days. Tremor power was first calculated for each detected tremor window by
21 $\log_{10}(P_T + 1)$, where P_T is the power within a bandwidth of 1.25 Hz around the dominant
22 frequency in the tremor frequency band of 3-7 Hz ⁴⁶. The logarithm was taken because of the
23 linear relationship between clinical tremor severity ratings and the log of tremor
24 displacement or angular velocity ⁴⁷. We added a constant of 1 before taking the logarithm to
25 make sure that P_T of 0 corresponds to a tremor power of 0. Then we derived three weekly
26 aggregated tremor power measures: the median, mode (i.e., the peak in the probability
27 density function of tremor power) and 90th percentile of tremor power.

28 In the first week, a median of 75.1 hours (IQR 65.4 – 78.0 hours) of gyroscope data
29 collected during daytime in 501/517 PD participants was used for analysis. For controls, this
30 was a median of 75.6 hours (IQR 65.5 – 80.4 hours) in 48/50 subjects. The median percentage
31 of time with non-tremor arm movements was 45.1% (IQR 37.4% - 53.1%) in the PD group and
32 50.2% (IQR 44.3% - 59.3%) in the control group (for more details see Supplementary Table 5).
33 Weekly tremor power measures were only derived for the PD group, and only if aggregated
34 tremor time exceeded 3.5%, which held for 207/517 PD participants in the first week. The
35 threshold of 3.5% was based on the 90th percentile of tremor time detected in non-PD
36 controls, and used to avoid inaccurate estimates due to a large false positive fraction of all
37 detected tremor windows.

1 **Construct validity and test-retest reliability**

2 The construct validity of weekly tremor measures was assessed in three ways. First, we
3 compared the weekly tremor measures between three groups of PD subjects with increasing
4 rest tremor severity scores (MDS-UPDRS 3.17 ON) of their device-sided arm. PD group 0 had
5 a score of 0 (no tremor; n=351), group 1 had a score of 1 (slight tremor, n=108), and group 2
6 had a score of ≥ 2 (mild to severe tremor; n=42). Group 2 contained a wider range of clinical
7 tremor severities (MDS-UPDRS scores of 2, 3, or 4), because the number of participants for
8 each of these scores was relatively low. We used the scores assessed in ON motor condition
9 for participants on anti-parkinsonian medication, because we expected that this best reflects
10 their condition in daily life. Weekly tremor time was also quantified in non-PD controls of PPP.

11 Second, we assessed the contribution of different tremor types in different body parts to
12 the weekly tremor measures by univariable correlations (Spearman's rank correlation) and by
13 multivariable linear regression, hypothesizing that the weekly tremor measures are
14 predominantly associated with the rest tremor severity in the device-sided arm. Third, we
15 further investigated the potential causes of tremor detected in PD subjects with an MDS-
16 UPDRS 3.17 ON score in the device-sided arm of 0, by incrementally excluding subjects with
17 other tremor scores of 1 or larger.

18 Finally, the test-retest reliability of all weekly tremor measures was assessed by the
19 Intraclass Correlation Coefficient (ICC) between the first two weeks in PD subjects of the PPP
20 dataset ⁴⁸.

21

22 **Data availability**

23 The data used to train the tremor detector on the Parkinson@Home dataset is publicly
24 available in the Radboud Data Repository: <https://doi.org/10.34973/2xxa-g520> (now
25 accessible by [https://data.ru.nl/login/reviewer-
26 3097040350/RZ4UID54SNSKTSTSX3M6KDEJKNHOVKC4PPLSYCI](https://data.ru.nl/login/reviewer-3097040350/RZ4UID54SNSKTSTSX3M6KDEJKNHOVKC4PPLSYCI)). Access to the full
27 PD@Home Validation study - with the exception of the raw video recordings - can be arranged
28 through a request to the Michael J Fox Foundation (www.michaeljfox.org). Data from the
29 Personalized Parkinson Project is available on reasonable request via:
30 <https://www.personalizedparkinsonproject.com/home/data/requesting>.

31

32 **Code availability**

33 The code used to train and validate the tremor detector on PD@Home is publicly available
34 in the Git repository: https://github.com/biomarkersParkinson/pdathome_tremor. An
35 implementation of the tremor detection algorithm, used to generate the results in this study

1 on the PPP dataset, is available as part of the ParaDigMa python toolbox
2 (<https://doi.org/10.5281/zenodo.14916072>).

3 **References**

- 4 1. Steinmetz, J. D. *et al.* Global, regional, and national burden of disorders affecting the
5 nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study
6 2021. *Lancet Neurol.* **23**, 344–381 (2024).
- 7 2. Gupta, D. K., Marano, M., Zweber, C., Boyd, J. T. & Kuo, S.-H. Prevalence and Relationship
8 of Rest Tremor and Action Tremor in Parkinson’s Disease. *Tremor Hyperkinetic Mov.* **10**,
9 58.
- 10 3. Zach, H., Dirx, M., Bloem, B. R. & Helmich, R. C. The Clinical Evaluation of Parkinson’s
11 Tremor. *J. Park. Dis.* **5**, 471–474 (2015).
- 12 4. Evers, L. J. W., Peeters, J. M., Bloem, B. R. & Meinders, M. J. Need for personalized
13 monitoring of Parkinson’s disease: the perspectives of patients and specialized healthcare
14 providers. *Front. Neurol.* **14**, 1150634 (2023).
- 15 5. Politis, M. *et al.* Parkinson’s disease symptoms: the patient’s perspective. *Mov. Disord.*
16 *Off. J. Mov. Disord. Soc.* **25**, 1646–1651 (2010).
- 17 6. Dirx, M. F. & Bologna, M. The pathophysiology of Parkinson’s disease tremor. *J. Neurol.*
18 *Sci.* **435**, 120196 (2022).
- 19 7. Chan, P. Y. *et al.* Motion characteristics of subclinical tremors in Parkinson’s disease and
20 normal subjects. *Sci. Rep.* **12**, 4021 (2022).
- 21 8. Burq, M. *et al.* Virtual exam for Parkinson’s disease enables frequent and reliable remote
22 measurements of motor function. *NPJ Digit. Med.* **5**, 65 (2022).
- 23 9. Helmich, R. C. The cerebral basis of Parkinsonian tremor: A network perspective. *Mov.*
24 *Disord.* **33**, 219–231 (2018).

- 1 10. Evers, L. J. W., Krijthe, J. H., Meinders, M. J., Bloem, B. R. & Heskes, T. M. Measuring
2 Parkinson's disease over time: The real-world within-subject reliability of the MDS-
3 UPDRS. *Mov. Disord.* **34**, 1480–1487 (2019).
- 4 11. Papapetropoulos, S. S. Patient diaries as a clinical endpoint in Parkinson's disease clinical
5 trials. *CNS Neurosci. Ther.* **18**, 380–387 (2012).
- 6 12. Del Din, S., Godfrey, A., Mazzà, C., Lord, S. & Rochester, L. Free-living monitoring of
7 Parkinson's disease: Lessons from the field. *Mov. Disord.* **31**, 1293–1313 (2016).
- 8 13. Espay, A. J. *et al.* A Roadmap for Implementation of Patient-Centered Digital Outcome
9 Measures in Parkinson's disease Obtained Using Mobile Health Technologies. *Mov.*
10 *Disord. Off. J. Mov. Disord. Soc.* **34**, 657–663 (2019).
- 11 14. Adams, J. L. *et al.* A real-world study of wearable sensors in Parkinson's disease. *NPJ Park.*
12 *Dis.* **7**, 106 (2021).
- 13 15. Braybrook, M. *et al.* An Ambulatory Tremor Score for Parkinson's Disease. *J. Park.*
14 *Dis.* **6**, 723–731 (2016).
- 15 16. Cole, B. T., Roy, S. H., De Luca, C. J. & Nawab, S. H. Dynamical learning and tracking of
16 tremor and dyskinesia from wearable sensors. *IEEE Trans. Neural Syst. Rehabil. Eng. Publ.*
17 *IEEE Eng. Med. Biol. Soc.* **22**, 982–991 (2014).
- 18 17. García-Magariño, I., Medrano, C., Plaza, I. & Oliván, B. A smartphone-based system for
19 detecting hand tremors in unconstrained environments. *Pers. Ubiquitous Comput.* **20**,
20 959–971 (2016).
- 21 18. Hoff, J. I., Wagemans, E. A. & van Hilten, B. J. Ambulatory objective assessment of tremor
22 in Parkinson's disease. *Clin. Neuropharmacol.* **24**, 280–283 (2001).

- 1 19. Hssayeni, M. D., Jimenez-Shahed, J., Burack, M. A. & Ghoraani, B. Wearable Sensors for
2 Estimation of Parkinsonian Tremor Severity during Free Body Movements. *Sensors* **19**,
3 4215 (2019).
- 4 20. Lang, M. *et al.* A Multi-Layer Gaussian Process for Motor Symptom Estimation in People
5 With Parkinson's Disease. *IEEE Trans. Biomed. Eng.* **66**, 3038–3049 (2019).
- 6 21. Mahadevan, N. *et al.* Development of digital biomarkers for resting tremor and
7 bradykinesia using a wrist-worn wearable device. *Npj Digit. Med.* **3**, 1–12 (2020).
- 8 22. McNames, J. *et al.* A Two-Stage Tremor Detection Algorithm for Wearable Inertial Sensors
9 During Normal Daily Activities. in *2019 41st Annual International Conference of the IEEE*
10 *Engineering in Medicine and Biology Society (EMBC)* 2535–2538 (2019).
11 doi:10.1109/EMBC.2019.8857133.
- 12 23. Powers, R. *et al.* Smartwatch inertial sensors continuously monitor real-world motor
13 fluctuations in Parkinson's disease. *Sci. Transl. Med.* **13**, eabd7865 (2021).
- 14 24. Rigas, G. *et al.* Tremor UPDRS estimation in home environment. *Annu. Int. Conf. IEEE Eng.*
15 *Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.* **2016**, 3642–3645 (2016).
- 16 25. Salarian, A. *et al.* Quantification of tremor and bradykinesia in Parkinson's disease using
17 a novel ambulatory monitoring system. *IEEE Trans. Biomed. Eng.* **54**, 313–322 (2007).
- 18 26. San-Segundo, R. *et al.* Parkinson's Disease Tremor Detection in the Wild Using Wearable
19 Accelerometers. *Sensors* **20**, 5817 (2020).
- 20 27. Moreau, C. *et al.* Overview on wearable sensors for the management of Parkinson's
21 disease. *NPJ Park. Dis.* **9**, 153 (2023).
- 22 28. Artusi, C. A. *et al.* Implementation of Mobile Health Technologies in Clinical Trials of
23 Movement Disorders: Underutilized Potential. *Neurotherapeutics* **17**, 1736–1746 (2020).

- 1 29. Zhang, A., De la Torre, F. & Hodgins, J. Comparing laboratory and in-the-wild data for
2 continuous Parkinson's Disease tremor detection. in *2020 42nd Annual International*
3 *Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* 5436–5441
4 (2020). doi:10.1109/EMBC44109.2020.9176255.
- 5 30. Sun, Y., Wang, Z., Liang, Y., Hao, C. & Shi, C. Digital biomarkers for precision diagnosis and
6 monitoring in Parkinson's disease. *Npj Digit. Med.* **7**, 1–10 (2024).
- 7 31. di Biase, L., Pecoraro, P. M., Pecoraro, G., Shah, S. A. & Di Lazzaro, V. Machine learning
8 and wearable sensors for automated Parkinson's disease diagnosis aid: a systematic
9 review. *J. Neurol.* **271**, 6452–6470 (2024).
- 10 32. De, A., Bhatia, K. P., Volkmann, J., Peach, R. & Schreglmann, S. R. Machine Learning in
11 Tremor Analysis: Critique and Directions. *Mov. Disord.* **38**, 717–731 (2023).
- 12 33. Evers, L. J. *et al.* Real-Life Gait Performance as a Digital Biomarker for Motor Fluctuations:
13 The Parkinson@Home Validation Study. *J. Med. Internet Res.* **22**, e19068 (2020).
- 14 34. Bloem, B. R. *et al.* The Personalized Parkinson Project: examining disease progression
15 through broad biomarkers in early Parkinson's disease. *BMC Neurol.* **19**, 160 (2019).
- 16 35. Dirkx, M. F., Zach, H., Bloem, B. R., Hallett, M. & Helmich, R. C. The nature of postural
17 tremor in Parkinson disease. *Neurology* **90**, e1095–e1103 (2018).
- 18 36. Deuschl, G. *et al.* The clinical and electrophysiological investigation of tremor. *Clin.*
19 *Neurophysiol.* **136**, 93–129 (2022).
- 20 37. McGurrin, P., McNames, J., Haubenberger, D. & Hallett, M. Continuous Monitoring of
21 Essential Tremor: Standards and Challenges. *Mov. Disord. Clin. Pract.* **9**, 1094–1098
22 (2022).
- 23 38. Dirkx, M. F., Shine, J. M. & Helmich, R. C. Integrative Brain States Facilitate the Expression
24 of Parkinson's Tremor. *Mov. Disord.* mds.29506 (2023) doi:10.1002/mds.29506.

- 1 39. Van Der Heide, A. *et al.* Stress and mindfulness in Parkinson's disease – a survey in 5000
2 patients. *Npj Park. Dis.* **7**, 7 (2021).
- 3 40. Zach, H., Dirkx, M. F., Pasman, J. W., Bloem, B. R. & Helmich, R. C. Cognitive Stress Reduces
4 the Effect of Levodopa on Parkinson's Resting Tremor. *CNS Neurosci. Ther.* **23**, 209–215
5 (2017).
- 6 41. Pasquini, J. *et al.* The Clinical Profile of Tremor in Parkinson's Disease. *Mov. Disord. Clin.*
7 *Pract.* **n/a**,.
- 8 42. Haubenberger, D. *et al.* Transducer-based evaluation of tremor. *Mov. Disord.* **31**, 1327–
9 1336 (2016).
- 10 43. Elble, R. J. Gravitational artifact in accelerometric measurements of tremor. *Clin.*
11 *Neurophysiol.* **116**, 1638–1643 (2005).
- 12 44. San-Segundo, R., Montero, J. M., Barra-Chicote, R., Fernández, F. & Pardo, J. M. Feature
13 extraction from smartphone inertial signals for human activity segmentation. *Signal*
14 *Process.* **120**, 359–372 (2016).
- 15 45. Chan, A., Chan, D., Lee, H., Ng, C. C. & Yeo, A. H. L. Reporting adherence, validity and
16 physical activity measures of wearable activity trackers in medical research: A systematic
17 review. *Int. J. Med. Inf.* **160**, 104696 (2022).
- 18 46. Dai, H., Zhang, P. & Lueth, T. C. Quantitative Assessment of Parkinsonian Tremor Based
19 on an Inertial Measurement Unit. *Sensors* **15**, 25055–25071 (2015).
- 20 47. Elble, R. J. Estimating Change in Tremor Amplitude Using Clinical Ratings:
21 Recommendations for Clinical Trials. *Tremor Hyperkinetic Mov.* **8**, 600 (2018).
- 22 48. Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation
23 Coefficients for Reliability Research. *J. Chiropr. Med.* **15**, 155–163 (2016).
- 24

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

Acknowledgements

The authors would like to thank Anna Gansen for her contributions to developing the tremor detection algorithm, and Kars Veldkamp for his help with annotating the selected windows from the PPP dataset. For providing input to the project we thank Max Little. Furthermore, we would like to acknowledge Peter Kok and Vedran Kasalica of the Netherlands eScience Center for their work on the ParaDigMa toolbox and their contributions to the codebase.

This work was financially supported by the Michael J Fox Foundation (grant #020425), the Dutch Research Council (grant #ASDI.2020.060 & grant #2023.010), the Dutch Research Council Long-Term Program (project #KICH3.LTP.20.006, financed by the Dutch Research Council, Verily, and the Dutch Ministry of Economic Affairs and Climate Policy), the Dutch Ministry of Economic Affairs (Topconsortium voor Kennis en Innovatie, Life Sciences & Health, grant #LSHM20090-H048). The Personalized Parkinson Project is co-funded by Verily Life Sciences LLC, Radboud university medical center and by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships. H.C. received research grants from the UK Medical Research Council (grant #MR/R020418/1 & grant #MR/X023141/1). I.G.B. was financially supported via the Radboud Healthy Data program. R.C.H. received research grants from the MJ Fox Foundation, the Netherlands Brain Foundation, ParkinsonNL, and the Netherlands Organization for Health Research and Development. All the funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Author contributions

N.A.T.: conceptualization, formal analysis, visualization, writing – original draft, writing – review and editing. R.T.: conceptualization, formal analysis, visualization, writing – original draft, writing – review and editing. D.C.S.: conceptualization, formal analysis, writing – review and editing. H.C.: writing – review and editing, funding acquisition. Y.P.R.: writing – review and editing. I.G.B.: supervision, writing – review and editing. B.R.B.: writing – review and editing, funding acquisition. R.C.H.: conceptualization, supervision, writing – review and editing. L.J.W.E.: conceptualization, supervision, writing – review and editing, funding acquisition. All authors read and approved the final manuscript.

Ethics declarations

Ethics approval

1 The Parkinson@Home Validation study and Personalized Parkinson Project were both
2 conducted in accordance with the Declaration of Helsinki and Good Clinical Practice
3 guidelines, and were approved by the local medical ethics committee (Commissie
4 Mensgebonden Onderzoek, regio Arnhem-Nijmegen, reference number 2015–1776 for the
5 Parkinson@Home Validation study and reference number 2016–2934 for the Personalized
6 Parkinson Project). All participants provided informed consent prior to enrollment.

7

8 **Competing interests**

9 N.A.T., R.T., D.C.S., H.C., Y.P.R, I.G.B, R.C.H. and L.J.W.E. declare no competing interests.
10 B.R.B. serves as the co-Editor in Chief for the Journal of Parkinson’s disease, serves on the
11 editorial board of Practical Neurology and Digital Biomarkers, has received fees from serving
12 on the scientific advisory board for the Critical Path Institute, Gyenno Science, MedRhythms,
13 UCB, Kyowa Kirin and Zambon (paid to the Institute), has received fees for speaking at
14 conferences from AbbVie, Bial, Biogen, GE Healthcare, Oruen, Roche, UCB and Zambon (paid
15 to the Institute), and has received research support from Biogen, Cure Parkinson’s, Davis
16 Phinney Foundation, Edmond J. Safra Foundation, Fred Foundation, Gatsby Foundation,
17 Hersenstichting Nederland, Horizon 2020, IRLAB Therapeutics, Maag Lever Darm Stichting,
18 Michael J Fox Foundation, Ministry of Agriculture, Ministry of Economic Affairs & Climate
19 Policy, Ministry of Health, Welfare and Sport, Netherlands Organization for Scientific Research
20 (ZonMw), Not Impossible, Parkinson Vereniging, Parkinson’s Foundation, Parkinson’s UK,
21 Stichting Alkemade-Keuls, Stichting Parkinson NL, Stichting Woelse Waard, Health Holland /
22 Topsector Life Sciences and Health, UCB, Verily Life Sciences, Roche and Zambon. B.R.B. does
23 not hold any stocks or stock options with any companies that are connected to Parkinson’s
24 disease or to any of his clinical or research activities.

25

Supplementary tables

Supplementary Table 1: Tremor detection performance on PD@Home (average across leave-one-subject out cross-validation folds, and standard deviation between brackets), without filtering out windows with detected non-tremor arm movements. The number of participants with data from the different sub-classes is indicated between brackets. Annotations for "Significant upper limb activity" and "Periodic" were only available for the 8 PD patients with tremor.

	Sensitivity	Specificity	Weighted specificity across sitting, standing, gait and postural transitions
Overall (n=48)	0.70 (0.18)	0.96 (0.05)	
Subgroup			
Tremor PD (n=8)	0.70 (0.18)	0.90 (0.04)	0.90 (0.05)
Non-tremor PD (n=16)	-	0.96 (0.06)	0.96 (0.05)
Non-PD controls (n=24)	-	0.98 (0.04)	0.98 (0.01)
Type of activity			
Sitting (n=8/n=48)	0.70 (0.25)	0.95 (0.07)	
Standing (n=7/n=48)	0.66 (0.31)	0.97 (0.03)	
Gait (n=7/n=48)	0.39 (0.27)	0.98 (0.04)	-
Postural transitions (n=48)	-	0.99 (0.03)	
Running/Exercising (n=5)	-	1 (0)	
Cycling (n=14)	-	0.97 (0.08)	-
Driving motorized vehicle (n=2)	-	0.92 (0.10)	
Significant upper limb activity (n=8)	-	0.95 (0.06)	-
Suspicious activity (n=5)	-	0.83 (0.18)	-

Supplementary Table 2: Performance of the tremor detection algorithm on PPP based on the visual inspection and annotation of a sample of windows, for different subgroups of PD participants. The number of subjects in each subgroup is given.

Subgroup	Sensitivity	Specificity
MDS-UPDRS 3.17 OFF = 0 (n=60)	0.25 – 0.48	0.99 - 1
MDS-UPDRS 3.17 OFF = 1 (n=60)	0.41 – 0.65	0.99
MDS-UPDRS 3.17 OFF ≥ 2 (n=59)	0.46 – 0.62	0.97 – 0.99
MDS-UPDRS part 3 OFF – tremor subscore < 28 (n=89)	0.34 – 0.57	0.99 - 1
MDS-UPDRS part 3 OFF – tremor subscore ≥ 28 (n=90)	0.40 – 0.59	0.98 – 0.99
Overall (n=179)	0.37 – 0.58	0.98 – 1

Supplementary Table 3: Demographic and clinical characteristics of PD participants and non-PD controls of the PD@Home dataset included in the analyses. IQR: inter-quartile range. MDS-UPDRS: Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale. Part 1: non-motor experiences of daily living. Part 2: motor experiences of daily living. Part 3: motor examination. Part 4: motor complications. *: 1 missing value.

	PD subjects with annotated tremor (n=8)	PD subjects with no annotated tremor (n=16)	Non-PD controls (n=24)
Age (years), median (IQR)	61.0 (58.3 - 69.0)	66.0 (61.0 - 70.5)	67.5 (55.0 - 70.0)
Gender (men), n (%)	4 (50%)	7 (44%)	13 (54%)
Time since diagnosis of PD (years), median (IQR)	7.0 (5.3 - 9.5)	7.0 (5.0 - 11.0)	-
Hoehn and Yahr stage in off state, n (%)			
Stage 1	1 (12.5%)	0 (0.0%)*	-
Stage 2	6 (75.0%)	10 (62.5%)	-
Stage 3	0 (0.0%)	4 (25.0%)	-
Stage 4	1 (12.5%)	1 (6.3%)	-
MDS-UPDRS, median (IQR)			
Part 1 (scale range: 0 to 52)	10.5 (7.0 - 17.3)	9.5 (8.0 - 15.0)	3.0 (0.3 - 4.0)
Part 2 (scale range: 0 to 52)	10.5 (8.3 - 14.5)	9.0 (7.3 - 13.0)	0.0 (0.0 - 0.0)*
Part 3 (off state) (scale range: 0 to 132)	50.4 (38.8 - 61.8)	35.0 (30.0 - 46.8)	6.5 (4.3 - 11.0)
Part 3 (on state) (scale range: 0 to 132)	32.0 (24.0 - 37.5)	25.5 (17.5 - 38.0)	-
Part 4 (scale range: 0 to 24)	6.0 (3.5 - 9.8)	6.0 (4.3 - 8.5)	-
Tremor sub-score of MDS-UPDRS part III, median (IQR)			
Off state (scale range: 0 to 40)	14.0 (10.3 - 18.8)	4.0 (2.3 - 8.5)	0.5 (0.0 - 1.8)
On state (scale range: 0 to 40)	9.5 (3.0 - 11.8)	2.0 (1.0 - 5.0)	-
Rest tremor severity (arm of most affected side), n (%)			
0: normal (off on)	0 (0.0%) 2 (25.0%)	14 (87.5%) 15 (93.8%)	24 (100%)
1: slight (off on)	1 (12.5%) 2 (25.0%)	0 (0.0%) 1 (6.3%)	0 (0.0%)
2: mild (off on)	3 (37.5%) 2 (25.0%)	1 (6.3%) 0 (0.0%)	0 (0.0%)
3: moderate (off on)	3 (37.5%) 2 (25.0%)	1 (6.3%) 0 (0.0%)	0 (0.0%)
4: severe (off on)	1 (12.5%) 0 (0.0%)	0 (0.0%) 0 (0.0%)	0 (0.0%)

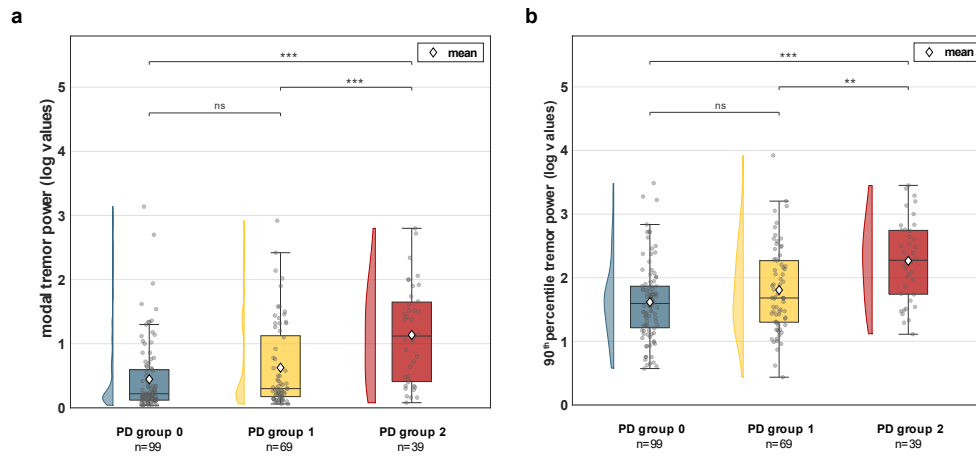
Supplementary Table 4: Demographic and clinical characteristics of PD participants and non-PD controls of the PPP dataset included in the analyses. IQR: inter-quartile range. MDS-UPDRS: Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale. Part 1: non-motor experiences of daily living. Part 2: motor experiences of daily living. Part 3: motor examination. Part 4: motor complications. *: 23 missing values. †: 21 missing values. ‡: 6 missing values. §: 32 missing values. ¶: 12 missing values. #: 26 missing values.

	PD subjects (n=517)	Non-PD controls (n=50)
Age (years), median (IQR)	62 (55 - 69)	70 (66 – 74)
Gender (men), n (%)	305 (59%)	18 (36%)
Time since diagnosis of PD (years), median (IQR)	2.7 (1.4 - 3.9)	-
Hoehn and Yahr stage in off state, n (%)		
Stage 1	46 (8.9%)	-
Stage 2	406 (78.5%)	-
Stage 3	59 (11.4%)	-
Stage 4	6 (1.2%)	-
MDS-UPDRS, median (IQR)		
Part 1 (scale range: 0 to 52)	9 (7 - 14)*	-
Part 2 (scale range: 0 to 52)	7 (4 - 12)†	-
Part 3 (off state) (scale range: 0 to 132)	32 (24 - 42)‡	-
Part 3 (on state) (scale range: 0 to 132)	27 (19 - 36)§	-
Part 4 (scale range: 0 to 24)	1 (0 - 5)¶	-
Tremor sub-score of MDS-UPDRS part III, median (IQR)		
Off state (scale range: 0 to 40)	4 (2 - 7)	-
On state (scale range: 0 to 40)	3 (1 – 5.5)#	-
Rest tremor severity (device-sided arm), n (%)		
0: normal (off on)	336 (65.0%) 353 (71.9%)	-
1: slight (off on)	121 (23.4%) 101 (20.6%)	-
2: mild (off on)	37 (7.2%) 27 (5.5%)	-
3: moderate (off on)	23 (4.5%) 10 (2.0%)	-
4: severe (off on)	0 (0.0%) 0 (0.0%)	-

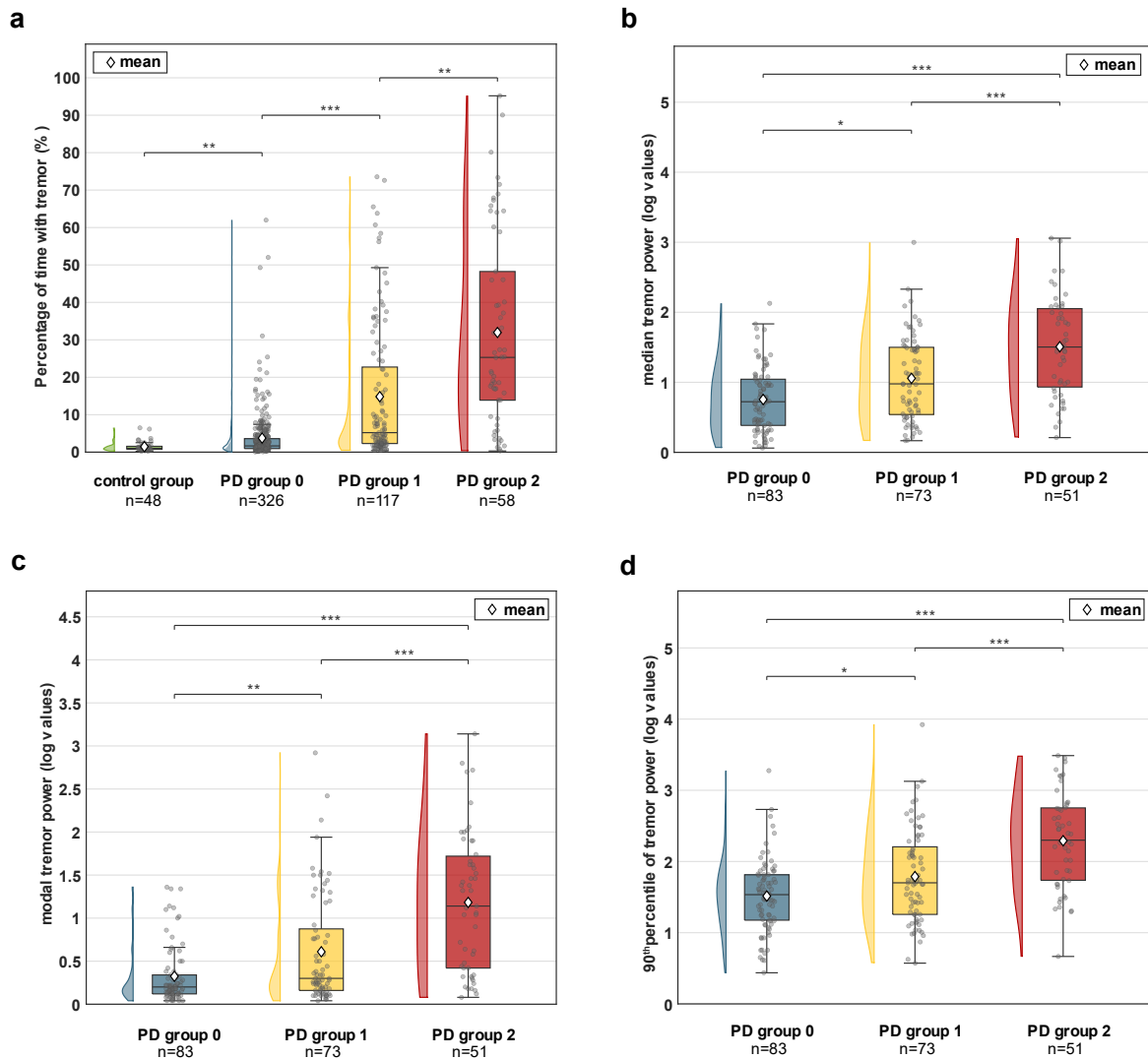
Supplementary Table 5: Median (interquartile range) percentage of daytime with non-tremor arm movements and percentage of detected tremor windows removed by filtering out windows with non-tremor arm movements in the first week of collected data of PPP. PD participants were split in three groups: PD group 0 had an MDS-UPDRS 3.17 score of 0 in the device-sided arm in ON motor state, PD group 1 had a score of 1 and PD group 2 a score of ≥2.

Subgroup	Percentage of daytime with detected arm movements (%)	Percentage of detected tremor removed by filtering out windows with detected arm movements (%)
Non-PD controls	50 (44 – 59)	38 (27 – 53)
PD group 0	47 (39 – 54)	28 (16 – 42)
PD group 1	41 (35 – 49)	17 (7 – 30)
PD group 2	39 (30 – 46)	18 (8 – 25)

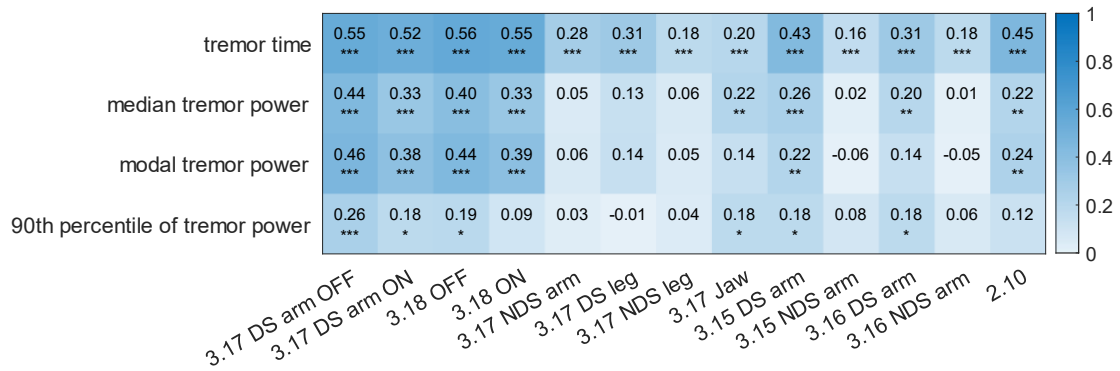
Supplementary figures



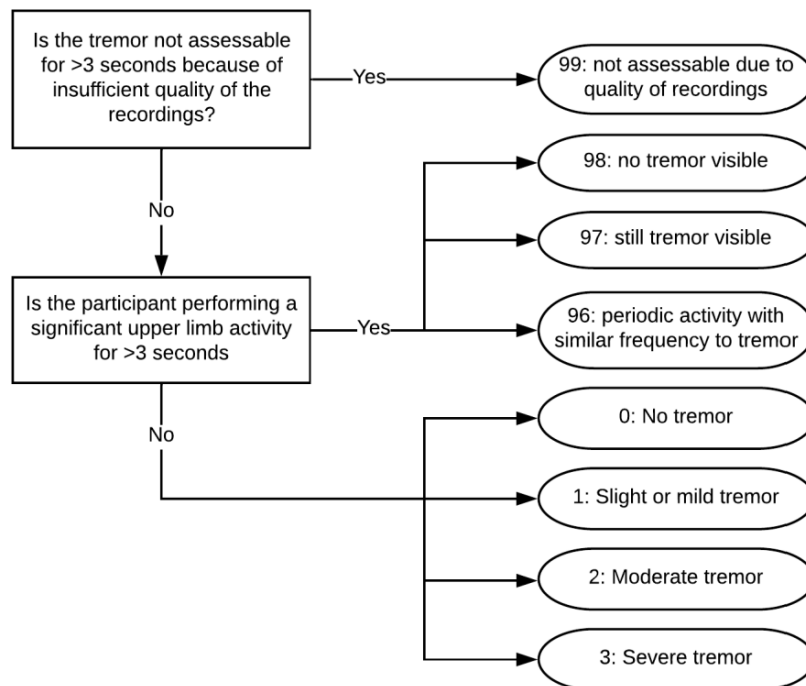
Supplementary Figure 1: Group differences in weekly tremor measures. **a:** Modal tremor power and **b:** 90th percentile of tremor power, measured in the first week of collected data of PPP in three groups of PD participants (groups 0, 1 and 2 with MDS-UPDRS 3.17 of 0, 1 and ≥ 2 assessed in ON motor state in the device-sided arm). Both measures were calculated across all detected tremor windows during daytime, but only assessed if the tremor time was $\geq 3.5\%$. The number of subjects in each subgroup is indicated. Significant differences (using Dunn's test with Bonferroni correction) between subsequent groups are shown (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).



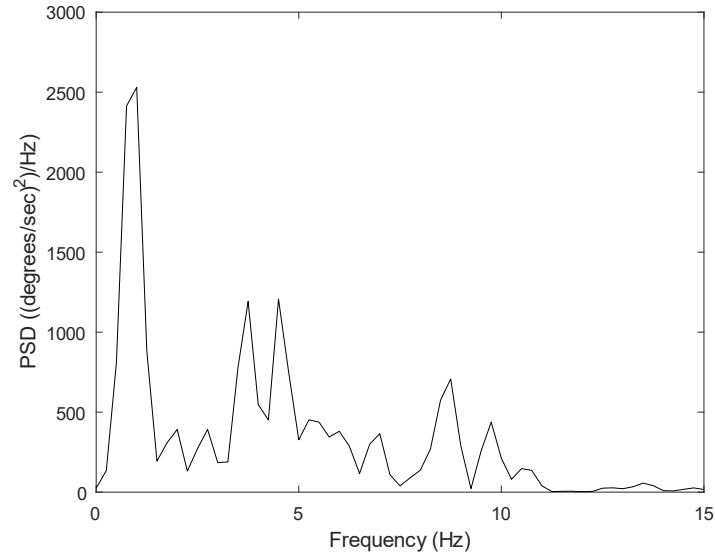
Supplementary Figure 2: Group differences in weekly tremor measures, based on clinical tremor severity scores assessed in OFF motor state (groups 0, 1 and 2 with MDS-UPDRS 3.17 of 0, 1 and ≥ 2 assessed in OFF motor state in the device-sided arm). a: Tremor time was calculated as the number of detected tremor windows divided by all windows without detected non-tremor arm movements during daytime (08:00 am – 10:00 pm), and expressed as percentage. **b:** Median tremor power, **c:** modal tremor power and **d:** 90th percentile of tremor power were calculated across all detected tremor windows during daytime, but only assessed if the tremor time was $\geq 3.5\%$. For all measures, the first week of collected data of PPP was used. The number of subjects in each subgroup is indicated. Significant differences (using Dunn's test with Bonferroni correction) between subsequent groups are shown (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).



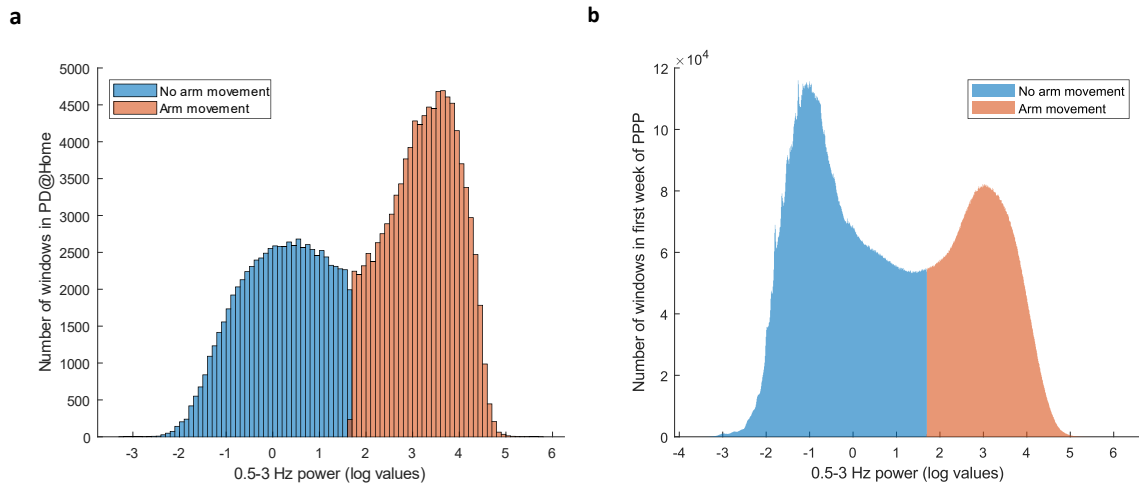
Supplementary Figure 3: Correlation of weekly tremor measures with clinical tremor scores, without filtering out the windows with detected non-tremor arm movements. Spearman's correlation coefficients are shown with their significance level (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, corrected using the false discovery rate method). Correlations with the rest tremor severity (3.17) and constancy (3.18) scores in the device-sided arm were assessed for ON and OFF scores separately. The other scores were averaged over ON and OFF motor states. 3.15 = postural tremor severity, 3.16 = kinetic tremor severity, 3.17 = rest tremor severity, 3.18 = rest tremor constancy, DS = device-sided, NDS = non-device-sided.



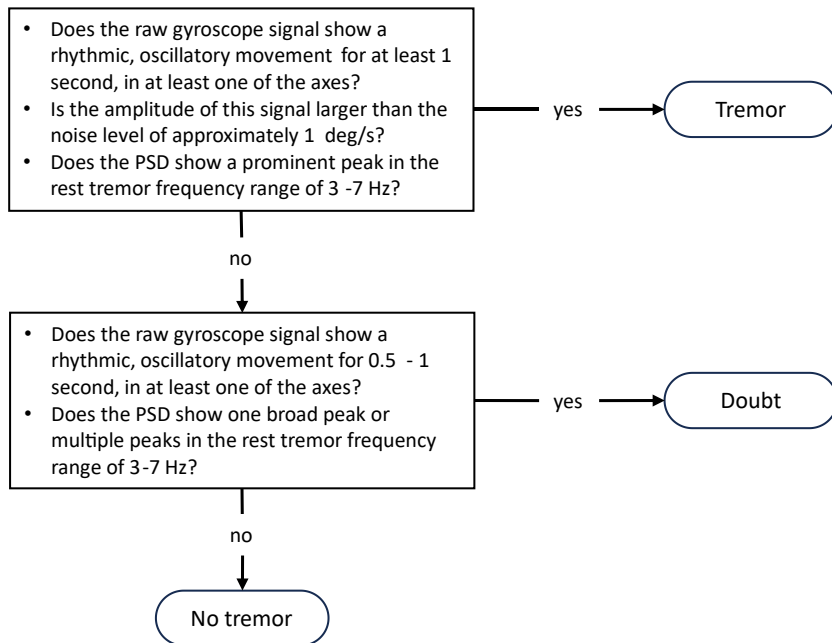
Supplementary Figure 4: Schematic overview of the video annotation protocol for the presence and severity of tremor used in the PD@Home study.



Supplementary Figure 5: Power spectral density (PSD) estimation of a 4-second gyroscope signal measured during tremor and gait in a subject of the Parkinson@Home dataset. The PSD was summed over all three gyroscope axes. The 4-5 Hz tremor peak is distorted by the higher harmonics of gait.



Supplementary Figure 6: Detection of non-tremor arm movements. a: Distribution of the power in the 0.5-3 Hz band (log values) across all windows of PD@Home. The threshold for non-tremor arm movements was determined by K-means clustering with 2 clusters, yielding a threshold of $50 \text{ deg}^2/\text{s}^2$. **b:** Distribution of the same feature across all windows collected during daytime in the first week of PPP. Using the same threshold, the proportion of windows with non-tremor arm movements is smaller than in PD@Home where people were encouraged to perform activities of daily living.



Supplementary Figure 7: Schematic overview of the annotation protocol for the presence of tremor in the PPP study, based on visual inspection of 4-second gyroscope signals and power spectral density (PSD) estimations.