

Original Paper

# Reliability of Average Daily Steps Measured Through a Consumer Smartwatch in Parkinson Disease Phenotypes, Stages, and Severities: Cross-Sectional Study

Edoardo Bianchini<sup>1,2</sup>, MD; Domiziana Rinaldi<sup>1</sup>, MD; Lanfranco De Carolis<sup>1</sup>, MD; Silvia Galli<sup>1</sup>, MD; Marika Alborghetti<sup>1</sup>, MD; Clint Hansen<sup>2,3</sup>, PhD; Antonio Suppa<sup>4,5</sup>, MD; Marco Salvetti<sup>1,5</sup>, Prof Dr Med; Francesco Ernesto Pontieri<sup>1</sup>, Prof Dr Med; Nicolas Vuillerme<sup>2,6</sup>, PhD

<sup>1</sup>Department of Neuroscience, Mental Health and Sensory Organs (NESMOS), Sapienza University of Rome, Rome, Italy

<sup>2</sup>AGEIS, Université Grenoble Alpes, Grenoble, France

<sup>3</sup>Department of Neurology, Kiel University, Kiel, Germany

<sup>4</sup>Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy

<sup>5</sup>IRCCS Neuromed Institute, Pozzilli, Italy

<sup>6</sup>Institut Universitaire de France, Paris, France

## Corresponding Author:

Edoardo Bianchini, MD

Department of Neuroscience, Mental Health and Sensory Organs (NESMOS)

Sapienza University of Rome

Via di Grottarossa, 1035-1039

Rome, 00189

Italy

Phone: 390633775579

Fax: 390633775900

Email: [edoardo.bianchini@uniroma1.it](mailto:edoardo.bianchini@uniroma1.it)

## Abstract

**Background:** Average daily steps (avDS) could be a valuable indicator of real-world ambulation in people with Parkinson disease (PD), and previous studies have reported the validity and reliability of this measure. Nonetheless, no study has considered disease phenotype, stage, and severity when assessing the reliability of consumer wrist-worn devices to estimate daily step count in unsupervised, free-living conditions in PD.

**Objective:** This study aims to assess and compare the reliability of a consumer wrist-worn smartwatch (Garmin Vivosmart 4) in counting avDS in people with PD in unsupervised, free-living conditions among disease phenotypes, stages, and severity groups.

**Methods:** A total of 104 people with PD were monitored through Garmin Vivosmart 4 for 5 consecutive days. Total daily steps were recorded and avDS were calculated. Participants were dichotomized into tremor dominant (TD; n=39) or postural instability and gait disorder (PIGD; n=65), presence (n=57) or absence (n=47) of tremor, and mild (n=65) or moderate (n=39) disease severity. Based on the modified Hoehn and Yahr scale (mHY), participants were further dichotomized into earlier (mHY 1-2; n=68) or intermediate (mHY 2.5-3; n=36) disease stages. Intraclass correlation coefficient (ICC; 3,k), standard error of measurement (SEM), and minimal detectable change (MDC) were used to evaluate the reliability of avDS for each subgroup. The threshold for acceptability was set at an ICC  $\geq 0.8$  with a lower bound of 95% CI  $\geq 0.75$ . The 2-tailed Student *t* tests for independent groups and analysis of 83.4% CI overlap were used to compare ICC between each group pair.

**Results:** Reliability of avDS measured through Garmin Vivosmart 4 for 5 consecutive days in unsupervised, free-living conditions was acceptable in the overall population with an ICC of 0.89 (95% CI 0.85-0.92), SEM below 10%, and an MDC of 1580 steps per day (27% of criterion). In all investigated subgroups, the reliability of avDS was also acceptable (ICC range 0.84-0.94). However, ICCs were significantly lower in participants with tremor ( $P=.03$ ), with mild severity ( $P=.04$ ), and earlier stage ( $P=.003$ ). Moreover, SEM was below 10% in participants with PIGD phenotype, without tremor, moderate disease severity, and intermediate disease stage, with an MDC ranging from 1148 to 1687 steps per day (18%-25% of criterion). Conversely, in participants with TD phenotype, tremor, mild disease severity, and earlier disease stage, SEM was  $>10\%$  of the criterion and MDC values ranged from 1401 to 2263 steps per day (30%-33% of the criterion).

**Conclusions:** In mild-to-moderate PD, avDS measured through a consumer smartwatch in unsupervised, free-living conditions for 5 consecutive days are reliable irrespective of disease phenotype, stage, and severity. However, in individuals with TD phenotype, tremor, mild disease severity, and earlier disease stages, reliability could be lower. These findings could facilitate a broader and informed implementation of avDS as an index of ambulatory activity in PD.

*JMIR Form Res* 2025;9:e63153; doi: [10.2196/63153](https://doi.org/10.2196/63153)

**Keywords:** gait; Parkinson disease; phenotype; wearable sensors; smartwatch; step count; reliability; activity monitor; digital health technology; digital outcome measures; wearable; mHealth; motor; quality of life; fall; posture; mobile health

## Introduction

Walking is a fundamental motor ability, and it is pivotal for functional independence and social well-being [1]. Gait disturbances are common in people with Parkinson disease (PD), and include shuffling gait, shortened step length, altered automaticity, decreased arm swing, and freezing of gait [2]. These represent a particularly disabling group of symptoms, significantly hampering the quality of life of people with PD [3] and increasing the risk of falls [4].

Daily steps are an easy-to-collect and useful measure of ambulatory activity and mobility [5]. Although this parameter could not provide details regarding subtle gait features, several evidence linked a reduced daily step count to overall mortality risk [6-9], as well as to a range of health conditions, such as dementia [10] and oncological and cardiovascular diseases [11,12]. Previous studies have also reported a negative correlation between daily steps and disease severity in PD [13] and proposed a minimum daily step goal of 4200 to match with physical activity recommendations in the early stages of the disease [14].

Wearable devices, including smartwatches, could represent a useful option to estimate daily steps in an unobtrusive, ecological way [15]. Moreover, wearables are widely available on the market, easily used by the general population, and enable unobtrusive continuous long-term data collection [16]. However, since these devices are usually tested in healthy populations, knowledge of the validity and reliability of collected data is generally limited when applied to different groups of patients. Therefore, a growing body of literature focused on the validity and reliability of consumer wearable devices for step counting, with generally positive results [17-19].

In PD, motor and gait manifestations could render step detection and step count challenging and hence significantly diminish the validity and reliability of device algorithms [20-22]. Nevertheless, a prior study by Ginis et al [23] demonstrated a good criterion validity in estimating average daily steps (avDS) of 2 Fitbit devices (Fitbit Alta and Fitbit Inspire 3) in 28 people with PD in a real-life setting, compared to a research-grade device (Dynaport Movemonitor, McRoberts, NL).

Similarly, a prior study from our group involving 47 individuals with PD demonstrated a good criterion validity in step counting using a consumer smartwatch (Garmin Vivosmart 4), when worn on the side least affected by the

disease and under well-controlled pharmacological conditions in a supervised, in-clinic setting [20].

Besides criterion validity, we recently demonstrated that optimal reliability of avDS recorded by Garmin Vivosmart 4 in real-life conditions could be achieved if the smartwatch is worn for a minimum of 4 days [24].

However, the clinical heterogeneity of PD when evaluating metrological characteristics of step-counting devices has been neglected so far. Indeed, the clinical presentation of PD is highly variable among individuals, thus, significant efforts have been made to identify distinct clusters and subtypes [25]. Several classifications have been proposed over the years from clinical-based [26-28] to more recent biomarker-based classifications [29-31]. In this regard, one of the most used classifications distinguishes PD with predominant features of tremor (ie, tremor dominant [TD]) or gait, posture, and balance issues (ie, postural instability and gait disorder [PIGD]) [26-28] based on subitems scores of the Unified Parkinson's Disease Rating Scale (UPDRS) and its revision by the Movement Disorder Society (MDS-UPDRS) [32]. This classification could be relevant when measuring avDS in PD since tremors could increase the noise-to-signal ratio, making step detection more challenging [21]. Similarly, in PIGD individuals, the higher degree of gait alterations could alter the performance of step-detection algorithms [20,22].

Symptom severity and disease stage could also represent other relevant parameters to be considered when assessing the reliability of any wearable devices in counting avDS. Indeed, with disease progression, gait features increasingly deviate from normality [2,33], and tremor and bradykinesia could further alter the spatiotemporal and kinematic characteristics of walking, dampening, in turn, algorithm performance in step detection [20-22].

Nevertheless, no study to date considered disease phenotype, stage, and symptom severity when assessing the reliability of consumer wrist-worn devices for step counting in unsupervised, free-living conditions in mild-to-moderate people with PD. This study was hence specifically designed to address this issue. We hypothesized that reduced reliability might be observed in TD individuals and people with PD with more severe symptoms and in more advanced disease stages due to the aforementioned increased signal noise due to tremor, the higher degree of motor symptoms, and the more marked gait alterations. In addition, since no previous study specifically investigated the criterion validity of Garmin Vivosmart 4 in estimating avDS, we performed a proof-of-concept experiment in a subgroup of participants. Details of

methods and results of this latter experiment are presented in [Multimedia Appendix 1](#).

## Methods

### Population

Participants were consecutively screened and recruited during scheduled visits at the Movement Disorder Outpatient Service of the Sant'Andrea University Hospital (Rome, Italy) in the period between March 2023 and March 2024. Inclusion criteria were: (1) diagnosis of idiopathic PD according to the MDS criteria (Postuma et al [34]); (2) aged 18 years or older; (3) disease stage <4 according to the modified Hoehn and Yahr scale (mHY) [35]; (4) classification as TD or PIGD according to Stebbins et al [27]; and (5) stable medication in the 4 weeks before the experimental procedure. Exclusion criteria were: (1) cognitive impairment, defined by Montreal Cognitive Assessment [36] score <21; and (2) orthopedic, rheumatologic, or systemic conditions affecting mobility as judged by the assessor.

### Ethical Considerations

This cross-sectional study was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. Approval was granted by the local Ethical Committee of Sapienza, University of Rome, Italy (0372/2022). Data collection and processing followed the current European regulations for data protection. All participants provided written informed consent before the beginning of measurements. All data were deidentified. Participants did not receive any form of compensation.

### Demographic and Clinical Data

Participants were evaluated during scheduled visits. Demographics and anthropometric measures (including age, sex, weight, height, and BMI) were collected. Disease duration and disease stage according to mHY and levodopa equivalent daily dose [37] were also collected. MDS-UPDRS [32] part III was used to assess motor symptoms severity.

Participants were divided into 4 subgroup pairs. Based on MDS-UPDRS part II and III scores, participants were classified into TD or PIGD disease subtypes according to Stebbins et al [27]. To evaluate the effect of tremor presence on device reliability, participants were also classified as those with and without tremor based on a score of  $\geq 1$  at item 2.10 of MDS-UPDRS part II. Concerning disease severity, participants were grouped into those with mild or moderate disease severity based on the MDS-UPDRS score as proposed by Martínez-Martín et al [38]. Similarly, participants were dichotomized in earlier (mHY 1-2) or intermediate (mHY 2.5-3) stages, based on mHY score.

### Experimental Procedure

Participants received the smartwatch Garmin Vivosmart 4 after the visit and were instructed to wear it at home for a minimum of 5 consecutive days, including at least 1 weekend day, on the wrist of the body side least affected

by the disease [20]. No reminders or further instructions were provided to participants during registration. We chose a 5-day period since we demonstrated previously that a minimum of 4 days of monitoring is needed to reliably estimate daily step count in PD [24]. Each smartwatch was configured according to the producer's recommendations and participants were asked to perform daily activities as usual. After 5 days, participants returned the smartwatch. The total daily number of steps for each day was recorded and avDS were calculated [24]. Compliance was assessed based on the participants' dashboard data. The device recognizes that it is worn through heart rate and inertial motion unit signal. We considered all recording days with >80% wear time while awake and no interruption of device use greater than 3 hours to be valid. Details of methods of the proof-of-concept experiment regarding criterion validity of Garmin Vivosmart 4 are presented in [Multimedia Appendix 2](#).

### Data and Statistical Analysis

The statistical analyses were performed using JASP (version 0.18.3.0; JASP Team), R (version 4.0.3; R Core Team), and RStudio (version 2023.12.0+369; R Foundation for Statistical Computing) for Windows. Descriptive statistics were calculated for the examined variables. The normality of distributions was assessed by histogram and residual plots inspection.

To evaluate the relative reliability for the 5-day monitoring period in the overall population and each subgroup, a 2-way intraclass correlation coefficient (ICC) with a fixed set of raters and averaged ratings was used (ICC (3,k), where k was the number of days of measurement), together with a custom R script. The following reference cut-off values for ICC interpretation were used [39]: excellent: >0.90; good: 0.75-0.90; moderate: 0.50-0.75; and poor: <0.50. The a priori threshold for acceptable ICC was set at a point estimate  $\geq 0.80$  with a lower bound of 95% CI  $\geq 0.75$  in accordance with a previous study from our group [24].

To compare ICCs between the 4 subgroup pairs, 2 methods were applied. First, standard errors and point estimates of ICCs were used to compute *t* statistics and perform 2-tailed independent groups Student *t* tests. Second, the CI overlap between each group pair was graphically and numerically assessed. Nonoverlapping CIs were considered indicators of significantly different ICCs [40]. Previous evidence underscored that a 95% CI overlap assessment could inflate the risk of type II error and suggested that an 83.4% CI could be a more powerful option [40-42]. Therefore, we adopted this method for CI overlap evaluation.

Standard error of measurement (SEM) and minimal detectable change (MDC) with a CI of 95% were used to compute the absolute reliability for the 5-day recordings in the overall population and each subgroup [43]. SEM and MDC were reported as absolute value and percentage of criterion measure (SEM% and MDC%, respectively). The criterion was the avDS count derived from the 5 days. For all analyses, the significance threshold was set at  $\alpha < .05$ . All data were reported as mean (SD) or median (IQR) for numerical data and n (%) for categorical variables.

## Results

### Overview

A total of 104 people with PD were included in the study. All participants were monitored through Garmin Vivosmart 4 at home for a period of 5 consecutive days. No participants or days were excluded based on the prespecified compliance criteria. Participants took on average 5923 (SD 3014) daily

steps, ranging from 357 to 12,620. Details of demographic, anthropometric, and clinical variables of the study population are shown in [Table 1](#). The overlap between the subgroups is shown in [Multimedia Appendix 1](#).

Overall, the results of the proof-of-concept experiment suggest that the smartwatch Garmin Vivosmart 4 is valid and decently accurate in estimating avDS in PD. Details are presented in the [Multimedia Appendix 2](#).

**Table 1.** Demographic, anthropometric, and clinical characteristics of the study population.

	Overall (N=104)	TD <sup>a</sup> (n=39)	PIGD <sup>b</sup> (n=65)	Tremor (n=57)	No tremor (n=47)	Mild (n=65)	Moderate (n=39)	mHY <sup>c</sup> 1-2 (n=68)	mHY 2.5-3 (n=36)
Age (years), mean (SD)	68.0 (8.4)	66.4 (9.0)	68.9 (8.0)	69.5 (7.9)	66.7 (8.8)	66.8 (8.7)	69.9 (7.7)	65.3 (7.8)	73.0 (7.3)
Height (cm), mean (SD)	171 (9.0)	173 (7.6)	170 (9.6)	170 (9.6)	172 (8.4)	173 (9.3)	168 (7.6)	173 (8.1)	168 (9.8)
Weight (kg), mean (SD)	75.7 (13.1)	76.7 (13.0)	75.1 (13.2)	75.0 (12.1)	76.3 (13.9)	77.2 (13.7)	73.2 (11.8)	77.7 (13.8)	71.9 (10.8)
BMI (kg/m <sup>2</sup> ), mean (SD)	25.7 (3.4)	25.5 (3.6)	25.8 (3.4)	25.8 (3.0)	25.7 (3.8)	25.7 (3.5)	25.8 (3.4)	25.9 (3.7)	25.4 (3.0)
Sex (female), n (%)	34 (33)	11 (28)	23 (35)	19 (33)	15 (32)	22 (34)	12 (31)	22 (32)	12 (33)
Disease duration (years), mean (SD)	6.4 (4.4)	5.2 (4.4)	7.0 (4.4)	7.2 (3.9)	5.7 (4.8)	5.7 (4.5)	7.5 (4.2)	5.3 (4.3)	8.4 (4.1)
LEDD <sup>d</sup> (mg), mean (SD)	553 (302)	418 (247)	634 (304)	623 (271)	495 (316)	489 (289)	659 (296)	453 (248)	741 (307)
mHY, median (IQR)	2 (2-2.5)	2 (1-2)	2 (2-2.5)	2 (2-2.5)	2 (2-2)	2 (2-2)	2.5 (2-3)	2 (2-2)	2.5 (2.5-3)
MDS-UPDRS-III <sup>e</sup> , median (IQR)	27 (21-32)	26 (18-31)	29 (22-33)	29 (22-33)	26 (21-32)	23 (19-29)	33 (29-37)	23 (18-29)	33 (29-36)
avDS <sup>f</sup> , mean (SD)	5923 (3014)	6654 (2733)	5485 (3109)	4594 (2612)	7020 (2898)	6512 (2857)	4942 (3049)	6838 (2908)	4195 (2419)

<sup>a</sup>TD: tremor dominant.

<sup>b</sup>PIGD: postural instability and gait disorder.

<sup>c</sup>mHY: modified Hoehn and Yahr scale.

<sup>d</sup>LEDD: levodopa equivalent daily dose.

<sup>e</sup>MDS-UPDRS III: Movement Disorder Society Unified Parkinson's Disease Rating Scale part III.

<sup>f</sup>avDS: average daily steps.

### Reliability of avDS in People With PD

#### Relative Reliability

AvDS collected during 5 consecutive days showed a level of relative reliability above the threshold of acceptability, as

indicated by an ICC point estimate of  $\geq 0.80$  and a lower 95% CI limit of  $\geq 0.75$ , in the overall population and all subgroups. Moreover, daily step count showed excellent reliability in PD in the intermediate disease stage. Details of ICC and CI limits are shown in [Table 2](#).

**Table 2.** ICC<sup>a</sup> (3,k) values with 95% CI for the overall population and each subgroup<sup>b</sup>.

	Overall (N=104)	TD <sup>c</sup> (n=39)	PIGD <sup>d</sup> (n=65)	Tremor (n=57)	No tremor (n=47)	Mild (n=65)	Moderate (n=39)	mHY <sup>e</sup> 1-2 (n=68)	mHY 2.5-3 (n=36)
ICC (3,k)	0.888	0.854	0.899	0.838	0.914	0.856	0.919	0.839	0.939
Lower 95% CI	0.850	0.767	0.855	0.760	0.868	0.793	0.871	0.769	0.900
Upper 95% CI	0.919	0.916	0.933	0.896	0.947	0.905	0.953	0.892	0.966
Lower 83.4% CI	N/A <sup>f</sup>	0.797	0.869	0.786	0.883	0.814	0.887	0.792	0.914
Upper 83.4% CI	N/A	0.901	0.925	0.881	0.939	0.893	0.945	0.878	0.959

<sup>a</sup>ICC: intraclass correlation coefficient.

<sup>b</sup>For subgroups, 83.4% CI to assess intervals overlap are also reported.

<sup>c</sup>TD: tremor dominant.

<sup>d</sup>PIGD: postural instability and gait disorder.

<sup>e</sup>mHY: modified Hoehn and Yahr scale.

<sup>f</sup>Not applicable.

## Absolute Reliability

AvDS showed an SEM below 10% in the overall population with an MDC of 1580 (26.7% of the criterion). AvDS also showed an SEM below 10% in the PIGD disease subtype, in participants without tremor, with a moderate disease severity, and in an intermediate disease stage with an MDC ranging

from 1148 to 1687 steps per day (18% to 25% of criterion). Conversely, in the TD disease subtype, in participants with tremor, with mild disease severity and in an early disease stage, SEM was >10% of criterion and MDC values ranged from 1401 to 2263 steps per day (30% to 33% of the criterion). Details of SEM and MDC are shown in [Table 3](#).

**Table 3.** Absolute and percentage values of SEM<sup>a</sup> and MDC<sup>b</sup> for the overall population and each subgroup.

	Overall (N=104)	TD <sup>c</sup> (n=39)	PIGD <sup>d</sup> (n=65)	Tremor (n=57)	No tremor (n=47)	Mild (n=65)	Moderate (n=39)	mHY <sup>e</sup> 1-2 (n=68)	mHY 2.5-3 (n=36)
SEM	570	742	495	506	609	701	414	817	267
SEM%	9.6	11.1	9.0	11.0	8.7	10.8	8.4	11.9	6.4
MDC	1580	2056	1372	1401	1687	1944	1148	2263	741
MDC%	26.7	30.9	25.0	30.5	24.0	29.8	23.2	33.1	17.7

<sup>a</sup>SEM: standard error of measurement.

<sup>b</sup>MDC: minimal detectable change.

<sup>c</sup>TD: tremor dominant.

<sup>d</sup>PIGD: postural instability and gait disorder.

<sup>e</sup>mHY: modified Hoehn and Yahr scale.

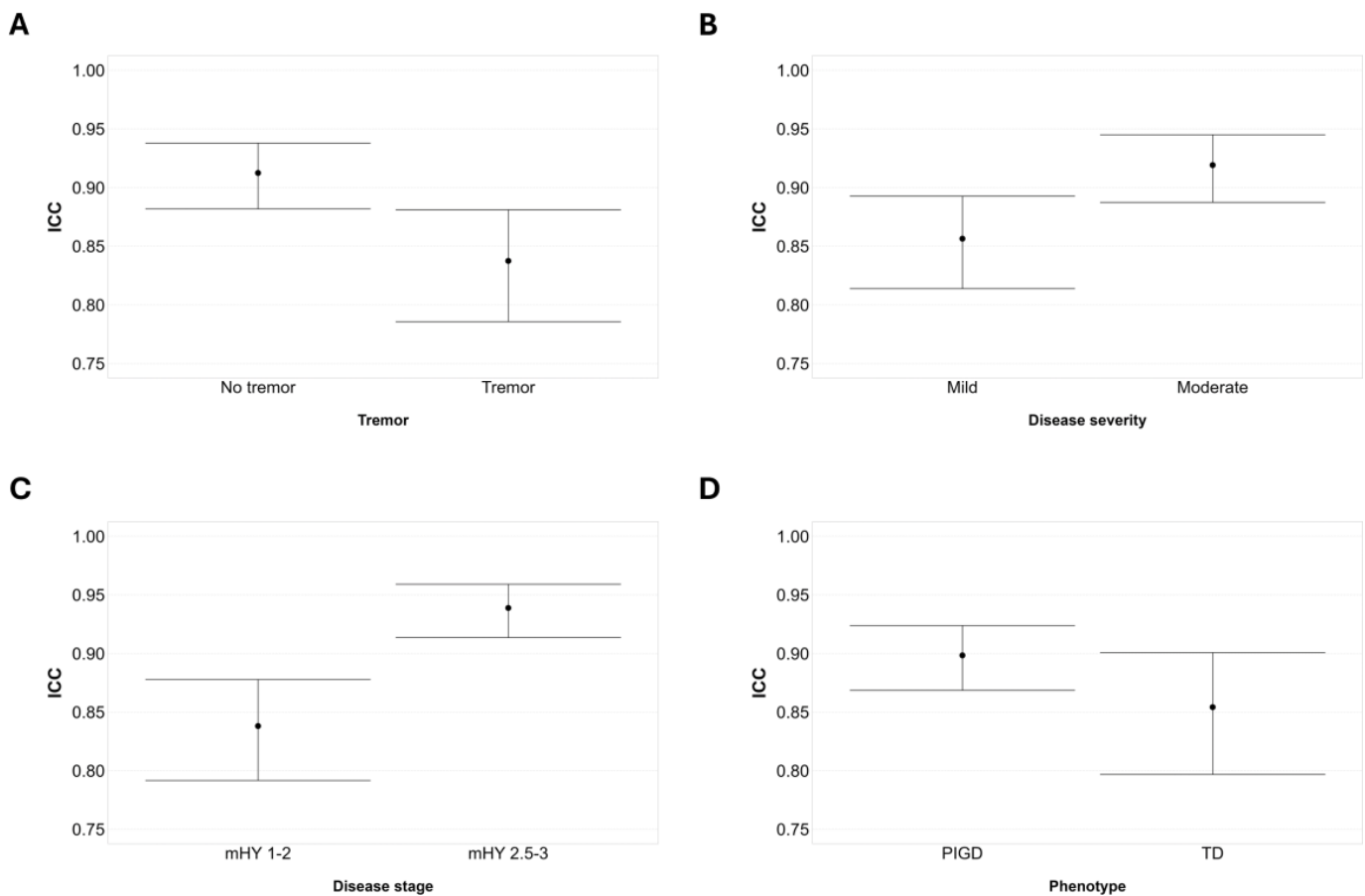
## Reliability Comparison Between Subgroups

When comparing ICCs between subgroups pairs, 2-tailed Student *t* test for independent groups showed a significant difference between participants with and without tremor ( $t_{102}=1.897$ ;  $P=.03$ ), between PD with mild and moderate disease severity ( $t_{102}=1.765$ ;  $P=.04$ ), and between individuals in early and intermediate disease stage ( $t_{102}=2.817$ ;  $P=.003$ ). Conversely, no significant difference was found between TD and PIGD participants ( $t_{102}=1.048$ ;  $P=.15$ ).

The analysis of 83.4% CI showed no overlap between interval limits between participants with and without tremor ([Figure 1A](#)), in early and intermediate disease stage ([Figure 1C](#)), and only a negligible overlap between individuals with mild and moderate disease severity ([Figure 1B](#)). Conversely, a degree of overlap of the two 83.4% CIs was observed between TD and PIGD participants ([Figure 1D](#)). Details of 83.4% CI limits are shown in [Table 2](#).



**Figure 1.** ICC and 83.4% CI comparison for each subgroup pair. (A) Participants with reported presence or absence of tremor. (B) Participants with mild and moderate disease severity as indicated by MDS-UPDRS. (C) Participants with earlier (mHY 1-2) and intermediate (mHY 2.5-3) disease stage. (D) PIGD and TD phenotypes. ICC (3,k) point estimate is indicated by the black dot, CI limits are represented by the vertical bars. ICC: intraclass correlation coefficient; MDS-UPDRS: International Movement Disorders Society Unified Parkinson's Disease Rating Scale; mHY: modified Hoehn and Yahr stage; PIGD: postural instability and gait disorder; TD: tremor dominant.



## Discussion

### Principal Findings

This cross-sectional study aimed at assessing and comparing the reliability of a consumer wrist-worn smartwatch (Garmin Vivosmart 4) in counting avDS in PD in unsupervised, free-living conditions for 5 consecutive days between disease phenotypes, stages, and severity groups.

Overall, our results showed, for the first time, that avDS were acceptably reliable in mild-to-moderate PD irrespective of disease severity, stage, or phenotype. Nevertheless, our results further indicated lower reliability in people with PD with TD phenotype, tremor, lower disease severity, and earlier stage.

### Reliability of avDS in the Overall PD Population

We found that the relative reliability of avDS, measured in unsupervised, free-living conditions, by Garmin Vivosmart 4 for 5 consecutive days was within the a priori criteria for acceptability in the overall PD population. We identified only 2 studies investigating the reliability of wearable devices in measuring avDS in people with PD [24,44]. Paul et al [44] reported that 2 consecutive days of monitoring were sufficient

to obtain an ICC of  $>0.9$ , using a research-grade, ankle-mounted, step counter (Step Activity Monitor) in 92 people with PD. One study from our group investigated the reliability of a wrist-worn consumer device in measuring avDS in PD before. In that study, we found an ICC (3,k) of 0.88 (0.82-0.93) for 4 days of monitoring in 56 mild-to-moderate people with PD using Garmin Vivosmart 4 [24]. The present results are consistent with our previous study, but also with other studies investigating the reliability of avDS through wrist-worn wearables in healthy older people [44-47], and in people with various neurological conditions, such as multiple sclerosis [48] and stroke [49]. Regarding absolute reliability, we found that SEM% was below 10% (9.6%) with an MDC of 1580 steps per day (27% of criterion). Only 1 study, from our group, investigated SEM and MDC for avDS in PD [24]. The present results are consistent with our previous study in which we reported an SEM% of 9.2% and an MDC of 1495 steps per day (or 26% of criterion) [24]. Taken together, these results confirm the reliability of avDS measured through Garmin Vivosmart 4 in unsupervised, free-living conditions in mild-to-moderate PD for 5 consecutive days.

## Reliability of avDS in People With PD Subgroups

To our knowledge, this is the first study to investigate and compare the reliability of a consumer smartwatch in measuring avDS in different subgroups of PD based on disease phenotype, severity, and stage. We found that in all investigated subgroups (ie, TD vs PIGD, mild vs moderate, tremor vs nontremor, and earlier vs intermediate), ICC values were within the a priori criteria for acceptability (ICC range 0.84-0.94). However, a significantly lower ICC was observed in people with PD with tremor, mild disease severity, and earlier disease stage. Moreover, SEM% was below 10% in participants with PIGD phenotype, moderate disease severity, intermediate disease stage, and without tremor, with an MDC ranging from 1148 to 1687 steps per day (18%-25% of criterion). Conversely, in individuals with TD phenotype, tremor, mild disease severity, and earlier disease stage, SEM was >10% of criterion and MDC values ranged from 1401 to 2263 steps per day (30%-33% of the criterion).

MDC, defined as the minimal change that falls outside the measurement error of an instrument, is extremely relevant in study design since it allows to calculate the sample size of studies aiming to assess the effectiveness of interventions [50]. MDC could be also crucial to define the appropriateness and feasibility of using a determinate device to measure a given construct. A prior work from Handlery et al [14] reported an increase in 1250 steps per day following a high-intensity physical activity intervention in PD measured through a research-grade wrist-worn device (Actigraph GT3X). In this study, we found that in individuals with TD phenotype, mild disease severity, and earlier disease stage, MDC was ~2000 steps per day. Although a direct comparison with the metrics reported in the work from Handlery et al [14] could not be performed due to the different devices, we could hypothesize that only large modifications in avDS could be reliably measured through Garmin Vivosmart 4 in the aforementioned PD subgroups. To this end, future studies will be needed to define the minimal clinically important difference for avDS measured through consumer wrist-worn devices and to assess the attainability of avDS modifications sufficiently large to be reliably detected by these devices.

The reduced reliability in participants with tremor is in line with our hypothesis that tremor could reduce the performance of the step-detection algorithm. In fact, tremor could increase the noise-to-signal ratio in the accelerometer signal, and in turn, render step detection more challenging [21]. Indeed, a previous study highlighted that tremor and dyskinesia together contributed to more than 19% of the variation in daily step counts when comparing measurements from waist-worn and wrist-worn devices in 46 people with PD with similar characteristics to those included in this study [21]. In this regard, our study further supports the assumption that tremor could reduce the step-detection performance of wrist-worn devices in PD.

On the other hand, our hypothesis that a reduced reliability might be observed in participants with higher symptom severity and more advanced stages was not supported by

our results. Indeed, the reduced reliability of avDS observed in individuals with mild disease severity and earlier disease stage is somehow counterintuitive. In fact, previous evidence highlighted that step count was less accurate in people walking at slower gait speed and with shorter step length in several neurological and musculoskeletal conditions, including PD [20,22,51-54]. Reduced step length and slower walking speed are typical features of Parkinsonian gait, with a higher prevalence along the disease course [2,33]. Moreover, another typical characteristic of walking in PD is the reduced automaticity that leads to a more discontinuous and irregular gait pattern that can further reduce device accuracy in step detection [2,22].

Despite these considerations, our results showed that avDS estimation was more reliable in people with PD with moderate disease severity and intermediate disease stage, compared with individuals with mild disease severity and earlier disease stage. In this regard, it must be considered that reliability is a measure of consistency and reproducibility of measurement and not a measure of accuracy [55]. Therefore, a reduced accuracy could not directly translate into reduced reliability. We could hypothesize that in more advanced stages of PD, the variability of clinical presentation could be lower. In an earlier stage, indeed, symptom heterogeneity, both in terms of motor and nonmotor features, could be extremely high [56-58]. However, this variability could decrease with disease progression since motor symptoms tend to consistently worsen along the disease course and motor features such as gait and balance impairment become increasingly prevalent [59,60]. In addition, the phenotype is dynamic along the disease course and some researchers have proposed that the classification into PIGD/TD evolves over time [61,62]. One study, indeed, reported that over a period of 8 years, approximately 70% of TD individuals transitioned to PIGD, whereas only 4% of PIGD individuals transitioned to TD [61]. Another study reported that 45% of TD participants at baseline had a subtype shift along a 2-year follow-up while 85% of PIGD participants remained as PIGD [62]. This is mirrored in our study cohort, where 32 out of 65 (49%) individuals with mild disease severity were classified as TD, whereas only 7 out of 39 (18%) were in the moderate group. Similarly, 33 out of 68 (49%) participants with earlier disease stage were classified as TD, whereas only 6 out of 36 (17%) were in the intermediate group. Therefore, a regression toward a more uniform motor impairment along the disease course might be considered. Moreover, since we found that tremor could be a relevant factor in reducing avDS reliability, the different prevalence of TD phenotype could also contribute to explaining our results. However, it must be underlined that no study, to our knowledge, systematically compared the heterogeneity of PD features across early, intermediate, and advanced disease stages. Therefore, our hypothesis should be taken with caution and future studies are needed to confirm it.

In conclusion, our findings highlight that, although avDS were reliable across the examined subgroups, clinicians and researchers should consider disease phenotype, stage, and severity when implementing wrist-worn wearables and

interpreting mobility data collected through these devices in PD.

## Limitations

We acknowledge that this study has some limitations. First, the participants included in our study displayed relatively preserved cognitive functions, due to our exclusion of participants with a Montreal Cognitive Assessment of <21. Additionally, those with more advanced disease stages or requiring walking aids were not included. This potentially limits the generalizability of our findings. However, it should be considered that the sample in our study can be seen as representative of the typical target for interventions using consumer-grade wearable technology. Moreover, including individuals with a disease stage >3, using walking aids, or with more severe cognitive impairments poses significant challenges in the utilization of consumer technology and is beyond our scope. Nevertheless, future research incorporating PD with lower functional scores, higher disease severity, and more impactful cognitive impairments would be valuable. Furthermore, we used only PD subtyping based on clinical features, yet other classification methods and clustering techniques have been proposed incorporating instrumental and biological data. Future studies are warranted to investigate the reliability of consumer smartwatches in PD subgroups defined using multimodal biomarkers. Finally, we did not account for antiparkinsonian treatment to control tremor. All included participants were under dopaminergic medical treatment and stable medical treatment in the 4 weeks before data collection, and we did not enroll drug-naïve individuals.

In addition, we did not account for different drug classes since it would have required a much higher sample size to cover all the possible combinations. Future studies might investigate the metrological features of commercial smartwatches in drug-naïve PD to avoid the influencing factor of dopaminergic treatment. However, since almost all patients with a diagnosis of PD are medically treated, we considered it relevant to assess the measurement properties of commercial devices in a setting more representative of the real-world experience.

## Conclusions

In mild-to-moderate PD, avDS measured through a consumer smartwatch in unsupervised, free-living conditions for 5 consecutive days are reliable irrespective of disease phenotype, stage, and severity. Researchers and clinicians who want to implement these instruments should consider that in individuals with TD phenotype, tremor, mild disease severity, and earlier disease stage, reliability could be lower and MDC could be higher. Future studies are needed to define the minimal clinically important difference for avDS measured through consumer wrist-worn devices and to assess the attainability of avDS modifications sufficiently large to be reliably detected by wrist-worn consumer devices. Taken together, we believe that these results could facilitate a broader implementation and an informed application of avDS as an index of ambulatory activity in PD and could be highly relevant to developing monitoring, preventive, educational, and rehabilitation strategies for PD.

## Acknowledgments

The authors sincerely thank all people who were involved in the data collection, from study participants to students, who aided and assisted the assessments. This research was performed as part of EB's PhD (Sapienza University of Rome, Italy and Université Grenoble Alpes, France) and funded through a Doctoral Bursary from Sapienza University of Rome (student number 1472455). The research was supported by Sapienza University of Rome as part of the program Sapienza Research Call 2020 (grant AR120172B97AF626) and Research Call 2022 (grant AR222181683A15AA), by the French National Research Agency (ANR) in the framework of the Investissements d'avenir program (ANR-10- AIRT-05 and ANR-15-IDEX-02), and the MIAI Cluster (ANR-23-IACL-0006). This work also forms part of a broader translational and interdisciplinary GaitAlps research program (NV)

## Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Overlap between the groups of participants with and without self-reported tremor and TD or PIGD phenotypes as well as between the groups with mild or moderate disease severity and early or intermediate disease stage. mHY: modified Hoehn and Yahr scale; PIGD: postural instability and gait disorder; TD: tremor dominant.

[\[DOCX File \(Microsoft Word File\), 22 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Proof-of-concept experiment to test criterion validity of Garmin Vivosmart 4 in estimating average daily steps in free-living conditions in people with Parkinson disease.

[\[DOCX File \(Microsoft Word File\), 26 KB-Multimedia Appendix 2\]](#)



## References

1. MacKinnon CD. Sensorimotor anatomy of gait, balance, and falls. *Handb Clin Neurol*. 2018;159(3–26):3-26. [doi: [10.1016/B978-0-444-63916-5.00001-X](https://doi.org/10.1016/B978-0-444-63916-5.00001-X)] [Medline: [30482322](https://pubmed.ncbi.nlm.nih.gov/30482322/)]
2. Mirelman A, Bonato P, Camicioli R, et al. Gait impairments in Parkinson's disease. *Lancet Neurol*. Jul 2019;18(7):697-708. [doi: [10.1016/S1474-4422\(19\)30044-4](https://doi.org/10.1016/S1474-4422(19)30044-4)] [Medline: [30975519](https://pubmed.ncbi.nlm.nih.gov/30975519/)]
3. Lee M, Noh Y, Youm C, et al. Estimating health-related quality of life based on demographic characteristics, questionnaires, gait ability, and physical fitness in Korean elderly adults. *IJERPH*. Nov 11, 2021;18(22):11816. [doi: [10.3390/ijerph182211816](https://doi.org/10.3390/ijerph182211816)]
4. Creaby MW, Cole MH. Gait characteristics and falls in Parkinson's disease: a systematic review and meta-analysis. *Parkinsonism Relat Disord*. Dec 2018;57:1-8. [doi: [10.1016/j.parkreldis.2018.07.008](https://doi.org/10.1016/j.parkreldis.2018.07.008)] [Medline: [30041848](https://pubmed.ncbi.nlm.nih.gov/30041848/)]
5. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. *Sports Med*. Jul 2017;47(7):1303-1315. [doi: [10.1007/s40279-016-0663-1](https://doi.org/10.1007/s40279-016-0663-1)] [Medline: [28005190](https://pubmed.ncbi.nlm.nih.gov/28005190/)]
6. Inoue K, Tsugawa Y, Mayeda ER, Ritz B. Association of daily step patterns with mortality in US adults. *JAMA Netw Open*. Mar 1, 2023;6(3):e235174. [doi: [10.1001/jamanetworkopen.2023.5174](https://doi.org/10.1001/jamanetworkopen.2023.5174)] [Medline: [36976556](https://pubmed.ncbi.nlm.nih.gov/36976556/)]
7. Lee IM, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. *JAMA Intern Med*. Aug 1, 2019;179(8):1105-1112. [doi: [10.1001/jamainternmed.2019.0899](https://doi.org/10.1001/jamainternmed.2019.0899)] [Medline: [31141585](https://pubmed.ncbi.nlm.nih.gov/31141585/)]
8. Paluch AE, Bajpai S, Bassett DR, et al. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. *Lancet Public Health*. Mar 2022;7(3):e219-e228. [doi: [10.1016/S2468-2667\(21\)00302-9](https://doi.org/10.1016/S2468-2667(21)00302-9)] [Medline: [35247352](https://pubmed.ncbi.nlm.nih.gov/35247352/)]
9. Saint-Maurice PF, Troiano RP, Bassett DR Jr, et al. Association of daily step count and step intensity with mortality among US adults. *JAMA*. Mar 24, 2020;323(12):1151-1160. [doi: [10.1001/jama.2020.1382](https://doi.org/10.1001/jama.2020.1382)] [Medline: [32207799](https://pubmed.ncbi.nlm.nih.gov/32207799/)]
10. Del Pozo Cruz B, Ahmadi M, Naismith SL, Stamatakis E. Association of daily step count and intensity with incident dementia in 78 430 adults living in the UK. *JAMA Neurol*. Oct 1, 2022;79(10):1059-1063. [doi: [10.1001/jamaneurol.2022.2672](https://doi.org/10.1001/jamaneurol.2022.2672)] [Medline: [36066874](https://pubmed.ncbi.nlm.nih.gov/36066874/)]
11. Del Pozo Cruz B, Ahmadi MN, Lee IM, Stamatakis E. Prospective associations of daily step counts and intensity with cancer and cardiovascular disease incidence and mortality and all-cause mortality. *JAMA Intern Med*. Nov 1, 2022;182(11):1139-1148. [doi: [10.1001/jamainternmed.2022.4000](https://doi.org/10.1001/jamainternmed.2022.4000)] [Medline: [36094529](https://pubmed.ncbi.nlm.nih.gov/36094529/)]
12. Banach M, Lewek J, Surma S, et al. The association between daily step count and all-cause and cardiovascular mortality: a meta-analysis. *Eur J Prev Cardiol*. Dec 21, 2023;30(18):1975-1985. [doi: [10.1093/eurjpc/zwad229](https://doi.org/10.1093/eurjpc/zwad229)] [Medline: [37555441](https://pubmed.ncbi.nlm.nih.gov/37555441/)]
13. Skidmore FM, Mackman CA, Pav B, et al. Daily ambulatory activity levels in idiopathic Parkinson disease. *J Rehabil Res Dev*. 2008;45(9):1343-1348. [doi: [10.1682/JRRD.2008.01.0002](https://doi.org/10.1682/JRRD.2008.01.0002)] [Medline: [19319758](https://pubmed.ncbi.nlm.nih.gov/19319758/)]
14. Handlery R, Stewart JC, Pellegrini C, et al. Physical activity in De Novo Parkinson disease: daily step recommendation and effects of treadmill exercise on physical activity. *Phys Ther*. Oct 1, 2021;101(10):34244805. [doi: [10.1093/ptj/pzab174](https://doi.org/10.1093/ptj/pzab174)] [Medline: [34244805](https://pubmed.ncbi.nlm.nih.gov/34244805/)]
15. Mattison G, Canfell O, Forrester D, et al. The influence of wearables on health care outcomes in chronic disease: systematic review. *J Med Internet Res*. Jul 1, 2022;24(7):e36690. [doi: [10.2196/36690](https://doi.org/10.2196/36690)] [Medline: [35776492](https://pubmed.ncbi.nlm.nih.gov/35776492/)]
16. Bianchini E, Maetzler W. Chapter five: wearable systems in movement disorders. In: *International Review of Movement Disorders*. Vol 5. Academic Press; 2023. [doi: [10.1016/bs.irmvd.2023.04.002](https://doi.org/10.1016/bs.irmvd.2023.04.002)]
17. Chevance G, Golaszewski NM, Tipton E, et al. Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits against reference measures: systematic review and meta-analysis. *JMIR Mhealth Uhealth*. Apr 13, 2022;10(4):e35626. [doi: [10.2196/35626](https://doi.org/10.2196/35626)] [Medline: [35416777](https://pubmed.ncbi.nlm.nih.gov/35416777/)]
18. Fuller D, Colwell E, Low J, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. *JMIR Mhealth Uhealth*. Sep 8, 2020;8(9):e18694. [doi: [10.2196/18694](https://doi.org/10.2196/18694)] [Medline: [32897239](https://pubmed.ncbi.nlm.nih.gov/32897239/)]
19. Nakagata T, Murakami H, Kawakami R, et al. Step-count outcomes of 13 different activity trackers: results from laboratory and free-living experiments. *Gait Posture*. Oct 2022;98:24-33. [doi: [10.1016/j.gaitpost.2022.08.004](https://doi.org/10.1016/j.gaitpost.2022.08.004)] [Medline: [36030707](https://pubmed.ncbi.nlm.nih.gov/36030707/)]
20. Bianchini E, Caliò B, Alborghetti M, et al. Step-counting accuracy of a commercial smartwatch in mild-to-moderate PD patients and effect of spatiotemporal gait parameters, laterality of symptoms, pharmacological state, and clinical variables. *Sensors (Basel)*. Dec 25, 2022;23(1):214. [doi: [10.3390/s23010214](https://doi.org/10.3390/s23010214)] [Medline: [36616812](https://pubmed.ncbi.nlm.nih.gov/36616812/)]
21. Kim DW, Hassett LM, Nguy V, Allen NE. A comparison of activity monitor data from devices worn on the wrist and the waist in people with Parkinson's disease. *Mov Disord Clin Pract*. Nov 2019;6(8):693-699. [doi: [10.1002/mdc3.12850](https://doi.org/10.1002/mdc3.12850)] [Medline: [31745480](https://pubmed.ncbi.nlm.nih.gov/31745480/)]

22. Wendel N, Macpherson CE, Webber K, et al. Accuracy of activity trackers in Parkinson disease: should we prescribe them? *Phys Ther*. Aug 1, 2018;98(8):705-714. [doi: [10.1093/ptj/pzy054](https://doi.org/10.1093/ptj/pzy054)] [Medline: [29718452](https://pubmed.ncbi.nlm.nih.gov/29718452/)]
23. Ginis P, Goris M, De Groef A, et al. Validation of commercial activity trackers in everyday life of people with Parkinson's disease. *Sensors (Basel)*. Apr 21, 2023;23(8):4156. [doi: [10.3390/s23084156](https://doi.org/10.3390/s23084156)] [Medline: [37112496](https://pubmed.ncbi.nlm.nih.gov/37112496/)]
24. Bianchini E, Galli S, Alborghetti M, et al. Four days are enough to provide a reliable daily step count in mild to moderate Parkinson's disease through a commercial smartwatch. *Sensors (Basel)*. Nov 4, 2023;23(21):8971. [doi: [10.3390/s23218971](https://doi.org/10.3390/s23218971)] [Medline: [37960670](https://pubmed.ncbi.nlm.nih.gov/37960670/)]
25. Outeiro TF, Alcalay RN, Antonini A, et al. Defining the riddle in order to solve it: there is more than one "Parkinson's disease". *Mov Disord*. Jul 2023;38(7):1127-1142. [doi: [10.1002/mds.29419](https://doi.org/10.1002/mds.29419)] [Medline: [37156737](https://pubmed.ncbi.nlm.nih.gov/37156737/)]
26. Jankovic J, McDermott M, Carter J, et al. Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. The Parkinson Study Group. *Neurology (ECronicon)*. Oct 1990;40(10):1529-1534. [doi: [10.1212/wnl.40.10.1529](https://doi.org/10.1212/wnl.40.10.1529)] [Medline: [2215943](https://pubmed.ncbi.nlm.nih.gov/2215943/)]
27. Stebbins GT, Goetz CG, Burn DJ, Jankovic J, Khoo TK, Tilley BC. 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*. May 2013;28(5):668-670. [doi: [10.1002/mds.25383](https://doi.org/10.1002/mds.25383)] [Medline: [23408503](https://pubmed.ncbi.nlm.nih.gov/23408503/)]
28. Zetuský WJ, Jankovic J, Pirozzolo FJ. The heterogeneity of Parkinson's disease: clinical and prognostic implications. *Neurology (ECronicon)*. Apr 1985;35(4):522-526. [doi: [10.1212/wnl.35.4.522](https://doi.org/10.1212/wnl.35.4.522)] [Medline: [3982637](https://pubmed.ncbi.nlm.nih.gov/3982637/)]
29. Cardoso F, Goetz CG, Mestre TA, et al. A statement of the MDS on biological definition, staging, and classification of Parkinson's disease. *Mov Disord*. Feb 2024;39(2):259-266. [doi: [10.1002/mds.29683](https://doi.org/10.1002/mds.29683)] [Medline: [38093469](https://pubmed.ncbi.nlm.nih.gov/38093469/)]
30. Höglinger GU, Adler CH, Berg D, et al. A biological classification of Parkinson's disease: the SynNeurGe research diagnostic criteria. *Lancet Neurol*. Feb 2024;23(2):191-204. [doi: [10.1016/S1474-4422\(23\)00404-0](https://doi.org/10.1016/S1474-4422(23)00404-0)] [Medline: [38267191](https://pubmed.ncbi.nlm.nih.gov/38267191/)]
31. Horsager J, Borghammer P. Brain-first vs. body-first Parkinson's disease: an update on recent evidence. *Parkinsonism Relat Disord*. May 2024;122:106101. [doi: [10.1016/j.parkreldis.2024.106101](https://doi.org/10.1016/j.parkreldis.2024.106101)] [Medline: [38519273](https://pubmed.ncbi.nlm.nih.gov/38519273/)]
32. 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*. Nov 15, 2008;23(15):2129-2170. [doi: [10.1002/mds.22340](https://doi.org/10.1002/mds.22340)]
33. Schlachetzki JCM, Barth J, Marxreiter F, et al. Wearable sensors objectively measure gait parameters in Parkinson's disease. *PLoS ONE*. 2017;12(10):e0183989. [doi: [10.1371/journal.pone.0183989](https://doi.org/10.1371/journal.pone.0183989)] [Medline: [29020012](https://pubmed.ncbi.nlm.nih.gov/29020012/)]
34. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord*. Oct 2015;30(12):1591-1601. [doi: [10.1002/mds.26424](https://doi.org/10.1002/mds.26424)]
35. Goetz CG, Poewe W, Rascol O, et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov Disord*. Sep 2004;19(9):1020-1028. [doi: [10.1002/mds.20213](https://doi.org/10.1002/mds.20213)] [Medline: [15372591](https://pubmed.ncbi.nlm.nih.gov/15372591/)]
36. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. Apr 2005;53(4):695-699. [doi: [10.1111/j.1532-5415.2005.53221.x](https://doi.org/10.1111/j.1532-5415.2005.53221.x)] [Medline: [15817019](https://pubmed.ncbi.nlm.nih.gov/15817019/)]
37. Jost ST, Kaldenbach MA, Antonini A, et al. Levodopa dose equivalency in Parkinson's disease: updated systematic review and proposals. *Mov Disord*. Jul 2023;38(7):1236-1252. [doi: [10.1002/mds.29410](https://doi.org/10.1002/mds.29410)] [Medline: [37147135](https://pubmed.ncbi.nlm.nih.gov/37147135/)]
38. Martínez-Martín P, Rodríguez-Blázquez C, et al. Parkinson's disease severity levels and MDS-Unified Parkinson's Disease Rating Scale. *Parkinsonism Relat Disord*. Jan 2015;21(1):50-54. [doi: [10.1016/j.parkreldis.2014.10.026](https://doi.org/10.1016/j.parkreldis.2014.10.026)] [Medline: [25466406](https://pubmed.ncbi.nlm.nih.gov/25466406/)]
39. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. Jun 2016;15(2):155-163. [doi: [10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012)] [Medline: [27330520](https://pubmed.ncbi.nlm.nih.gov/27330520/)]
40. Knol MJ, Pestman WR, Grobbee DE. The (mis)use of overlap of confidence intervals to assess effect modification. *Eur J Epidemiol*. Apr 2011;26(4):253-254. [doi: [10.1007/s10654-011-9563-8](https://doi.org/10.1007/s10654-011-9563-8)] [Medline: [21424218](https://pubmed.ncbi.nlm.nih.gov/21424218/)]
41. Austin PC, Hux JE. A brief note on overlapping confidence intervals. *J Vasc Surg*. Jul 2002;36(1):194-195. [doi: [10.1067/mva.2002.125015](https://doi.org/10.1067/mva.2002.125015)] [Medline: [12096281](https://pubmed.ncbi.nlm.nih.gov/12096281/)]
42. Cho J, Seo DM, Uh Y. Clinical application of overlapping confidence intervals for monitoring changes in serial clinical chemistry test results. *Ann Lab Med*. May 2020;40(3):201-208. [doi: [10.3343/alm.2020.40.3.201](https://doi.org/10.3343/alm.2020.40.3.201)] [Medline: [31858759](https://pubmed.ncbi.nlm.nih.gov/31858759/)]
43. Balaguier R, Madeleine P, Vuillerme N. Intra-session absolute and relative reliability of pressure pain thresholds in the low back region of vine-workers: effect of the number of trials. *BMC Musculoskelet Disord*. Aug 18, 2016;17(1):350. [doi: [10.1186/s12891-016-1212-7](https://doi.org/10.1186/s12891-016-1212-7)] [Medline: [27538914](https://pubmed.ncbi.nlm.nih.gov/27538914/)]

44. Paul SS, Ellis TD, Dibble LE, et al. Obtaining reliable estimates of ambulatory physical activity in people with Parkinson's disease. *J Parkinsons Dis.* May 5, 2016;6(2):301-305. [doi: [10.3233/JPD-160791](https://doi.org/10.3233/JPD-160791)] [Medline: [27164042](https://pubmed.ncbi.nlm.nih.gov/27164042/)]
45. Yao J, Tan CS, Lim N, Tan J, Chen C, Müller-Riemenschneider F. Number of daily measurements needed to estimate habitual step count levels using wrist-worn trackers and smartphones in 212,048 adults. *Sci Rep.* May 5, 2021;11(1):9633. [doi: [10.1038/s41598-021-89141-3](https://doi.org/10.1038/s41598-021-89141-3)] [Medline: [33953288](https://pubmed.ncbi.nlm.nih.gov/33953288/)]
46. Clemes SA, Griffiths PL. How many days of pedometer monitoring predict monthly ambulatory activity in adults? *Med Sci Sports Exerc.* Sep 2008;40(9):1589-1595. [doi: [10.1249/MSS.0b013e318177eb96](https://doi.org/10.1249/MSS.0b013e318177eb96)] [Medline: [18685533](https://pubmed.ncbi.nlm.nih.gov/18685533/)]
47. Sigmundová D, Vašíčková J, Stelzer J, Repka E. The influence of monitoring interval on data measurement: an analysis of step counts of university students. *Int J Environ Res Public Health.* Jan 28, 2013;10(2):515-527. [doi: [10.3390/ijerph10020515](https://doi.org/10.3390/ijerph10020515)] [Medline: [23358235](https://pubmed.ncbi.nlm.nih.gov/23358235/)]
48. Norris M, Anderson R, Motl RW, Hayes S, Coote S. Minimum number of days required for a reliable estimate of daily step count and energy expenditure, in people with MS who walk unaided. *Gait Posture.* Mar 2017;53:201-206. [doi: [10.1016/j.gaitpost.2017.02.005](https://doi.org/10.1016/j.gaitpost.2017.02.005)] [Medline: [28199925](https://pubmed.ncbi.nlm.nih.gov/28199925/)]
49. Fini NA, Burge AT, Bernhardt J, Holland AE. Two days of measurement provides reliable estimates of physical activity poststroke: an observational study. *Arch Phys Med Rehabil.* May 2019;100(5):883-890. [doi: [10.1016/j.apmr.2018.10.006](https://doi.org/10.1016/j.apmr.2018.10.006)] [Medline: [31030730](https://pubmed.ncbi.nlm.nih.gov/31030730/)]
50. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. *Am J Phys Med Rehabil.* Sep 2005;84(9):719-723. [doi: [10.1097/01.phm.0000176452.17771.20](https://doi.org/10.1097/01.phm.0000176452.17771.20)] [Medline: [16141752](https://pubmed.ncbi.nlm.nih.gov/16141752/)]
51. Treacy D, Hassett L, Schurr K, Chagpar S, Paul SS, Sherrington C. Validity of different activity monitors to count steps in an inpatient rehabilitation setting. *Phys Ther.* May 1, 2017;97(5):581-588. [doi: [10.1093/ptj/pzx010](https://doi.org/10.1093/ptj/pzx010)] [Medline: [28339904](https://pubmed.ncbi.nlm.nih.gov/28339904/)]
52. Fokkema T, Kooiman TJM, Krijnen WP, Van Der Schans CP, De Groot M. Reliability and validity of ten consumer activity trackers depend on walking speed. *Med Sci Sports Exerc.* Apr 2017;49(4):793-800. [doi: [10.1249/MSS.0000000000001146](https://doi.org/10.1249/MSS.0000000000001146)] [Medline: [28319983](https://pubmed.ncbi.nlm.nih.gov/28319983/)]
53. Chow JJ, Thom JM, Wewege MA, Ward RE, Parmenter BJ. Accuracy of step count measured by physical activity monitors: the effect of gait speed and anatomical placement site. *Gait Posture.* Sep 2017;57:199-203. [doi: [10.1016/j.gaitpost.2017.06.012](https://doi.org/10.1016/j.gaitpost.2017.06.012)] [Medline: [28666177](https://pubmed.ncbi.nlm.nih.gov/28666177/)]
54. Lamont RM, Daniel HL, Payne CL, Brauer SG. Accuracy of wearable physical activity trackers in people with Parkinson's disease. *Gait Posture.* Jun 2018;63:104-108. [doi: [10.1016/j.gaitpost.2018.04.034](https://doi.org/10.1016/j.gaitpost.2018.04.034)] [Medline: [29729611](https://pubmed.ncbi.nlm.nih.gov/29729611/)]
55. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* May 2010;19(4):539-549. [doi: [10.1007/s11136-010-9606-8](https://doi.org/10.1007/s11136-010-9606-8)] [Medline: [20169472](https://pubmed.ncbi.nlm.nih.gov/20169472/)]
56. Erro R, Vitale C, Amboni M, et al. The heterogeneity of early Parkinson's disease: a cluster analysis on newly diagnosed untreated patients. *PLoS ONE.* 2013;8(8):e70244. [doi: [10.1371/journal.pone.0070244](https://doi.org/10.1371/journal.pone.0070244)] [Medline: [23936396](https://pubmed.ncbi.nlm.nih.gov/23936396/)]
57. Wüllner U, Borghammer P, Choe CU, et al. The heterogeneity of Parkinson's disease. *J Neural Transm (Vienna).* Jun 2023;130(6):827-838. [doi: [10.1007/s00702-023-02635-4](https://doi.org/10.1007/s00702-023-02635-4)] [Medline: [37169935](https://pubmed.ncbi.nlm.nih.gov/37169935/)]
58. Berg D, Borghammer P, Fereshtehnejad SM, et al. Prodromal Parkinson disease subtypes—key to understanding heterogeneity. *Nat Rev Neurol.* Jun 2021;17(6):349-361. [doi: [10.1038/s41582-021-00486-9](https://doi.org/10.1038/s41582-021-00486-9)] [Medline: [33879872](https://pubmed.ncbi.nlm.nih.gov/33879872/)]
59. Fasano A, Fung VSC, Lopiano L, et al. Characterizing advanced Parkinson's disease: OBSERVE-PD observational study results of 2615 patients. *BMC Neurol.* Apr 2, 2019;19(1):50. [doi: [10.1186/s12883-019-1276-8](https://doi.org/10.1186/s12883-019-1276-8)] [Medline: [30940119](https://pubmed.ncbi.nlm.nih.gov/30940119/)]
60. Stefani A, Tessitore A, Tambasco N, et al. Criteria for identification of advanced Parkinson's disease: the results of the Italian subgroup of OBSERVE-PD observational study. *BMC Neurol.* Jan 28, 2022;22(1):41. [doi: [10.1186/s12883-022-02554-z](https://doi.org/10.1186/s12883-022-02554-z)] [Medline: [35090406](https://pubmed.ncbi.nlm.nih.gov/35090406/)]
61. Alves G, Larsen JP, Emre M, Wentzel-Larsen T, Aarsland D. Changes in motor subtype and risk for incident dementia in Parkinson's disease. *Mov Disord.* Aug 2006;21(8):1123-1130. [doi: [10.1002/mds.20897](https://doi.org/10.1002/mds.20897)] [Medline: [16637023](https://pubmed.ncbi.nlm.nih.gov/16637023/)]
62. Che N, Ou R, Li C, et al. Plasma GFAP as a prognostic biomarker of motor subtype in early Parkinson's disease. *NPJ Parkinsons Dis.* Mar 1, 2024;10(1):48. [doi: [10.1038/s41531-024-00664-8](https://doi.org/10.1038/s41531-024-00664-8)] [Medline: [38429295](https://pubmed.ncbi.nlm.nih.gov/38429295/)]

## Abbreviations

- avDS:** average daily steps
- ICC:** intraclass correlation coefficient
- MDC:** minimal detectable change
- MDS-UPDRS:** International Movement Disorders Society Unified Parkinson's Disease Rating Scale
- mHY:** modified Hoehn and Yahr scale
- PD:** Parkinson disease

**PIGD:** postural instability and gait disorder

**SEM:** Standard error of measurement

**TD:** tremor dominant

*Edited by Amaryllis Mavragani; peer-reviewed by Erin Rainaldi, Jari Ruokolainen; submitted 16.06.2024; final revised version received 03.02.2025; accepted 21.02.2025; published 18.03.2025*

*Please cite as:*

*Bianchini E, Rinaldi D, De Carolis L, Galli S, Alborghetti M, Hansen C, Suppa A, Salvetti M, Pontieri FE, Vuillerme N  
Reliability of Average Daily Steps Measured Through a Consumer Smartwatch in Parkinson Disease Phenotypes, Stages,  
and Severities: Cross-Sectional Study*

*JMIR Form Res 2025;9:e63153*

*URL: <https://formative.jmir.org/2025/1/e63153>*

*doi: [10.2196/63153](https://doi.org/10.2196/63153)*

© Edoardo Bianchini, Domiziana Rinaldi, Lanfranco De Carolis, Silvia Galli, Marika Alborghetti, Clint Hansen, Antonio Suppa, Marco Salvetti, Francesco Ernesto Pontieri, Nicolas Vuillerme. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.