

Automated Real-World Video Analysis of Sit-to-Stand Transitions Predicts Parkinson's Disease Severity

Catherine Morgan^{a,b} Alessandro Masullo^c Majid Mirmehdi^c
Hanna Kristiina Isotalus^c Ferdian Jovan^c Ryan McConville^c
Emma L. Tonkin^c Alan Whone^{a,b} Ian Craddock^c

^aTranslational Health Sciences, University of Bristol, Bristol, UK; ^bMovement Disorders Group, Bristol Brain Centre, North Bristol NHS Trust, Southmead Hospital, Bristol, UK; ^cFaculty of Engineering, University of Bristol, Digital Health Offices, 1 Cathedral Square, Bristol, UK

Keywords

Parkinson's disease-related motor symptoms · Influence and/or predict health-related outcomes · Objective data · Home environment · Mobility · Video recording

Abstract

Introduction: Technology holds the potential to track disease progression and response to neuroprotective therapies in Parkinson's disease (PD). The sit-to-stand (STS) transition is a frequently occurring event which is important to people with PD. The aim of this study was to demonstrate an automatic approach to quantify STS duration and speed using a real-world free-living dataset and look at clinical correlations of the outcomes, including whether STS parameters change when someone withholds PD medications. **Methods:** Eighty-five hours of video data were collected from 24 participants staying in pairs for 5-day periods in a naturalistic setting. Skeleton joints were extracted from the video data; the head trajectory was estimated and used to estimate the STS parameters of duration and speed. **Results:** 3.14 STS

transitions were seen per hour per person on average. Significant correlations were seen between automatic and manual STS duration (Pearson rho = 0.419, $p = 0.042$) and between automatic STS speed and manual STS duration (Pearson rho = 0.780, $p < 0.001$). Significant and strong correlations were seen between the gold-standard clinical rating scale scores and both STS duration and STS speed; these correlations were not seen in the STS transitions when the participants were carrying something in their hand(s). Significant differences were seen at the cohort level between control and PD participants' ON medications' STS duration ($U = 6,263$, $p = 0.018$) and speed ($U = 9,965$, $p < 0.001$). At an individual level, only two participants with PD became significantly slower to STS when they were OFF medications; withholding medications did not significantly change STS duration at an individual level in any participant. **Conclusion:** We demonstrate a novel approach to automatically quantify and ecologically validate two STS parameters which correlate with gold-standard clinical tools measuring disease severity in PD.

© 2023 The Author(s).
Published by S. Karger AG, Basel

Introduction

Currently, there are no licensed disease-modifying therapies (DMT) for Parkinson's disease (PD) because PD progression measurement in clinical trials uses infrequent snapshot subjective [1] clinical rating scales, meaning there is insufficient power to see disease slowing against placebo [2]. An unmet need in DMT clinical trials is having biomarkers of disease progression which are objective, reproducible, reliable, ecologically valid from real-world data, affordable, and scalable to large populations of patients [3, 4].

Digital biomarkers show promise for the objective and frequent measurement of symptoms in PD, and in particular, the speed and flexibility of movements, including sit-to-stand (STS) speed, have been identified as being highly useful to the assessment of disease progression in PD [5]. STS transitions are altered in PD compared to healthy controls with longer transition times [6] and issues with the speed of torque production [7]. Previous work has indicated that STS ascent speed is minimally affected by whether someone is dual-tasking [8] (e.g., holding something in their hand while standing up); however, other work suggests that dual-tasking alters a number of gait and mobility parameters in PD [9, 10]. No work has explored how dual-tasking changes STS parameters from real-world free-living data.

A continuously collected in-home digital biomarker could overcome the issues of symptoms changing in front of a clinician [11] and during clinical assessments [12]. Single wearables [13] and smartphones [14] have been used to quantify STS parameters, but difficulties in algorithm accuracy with dyskinesia [15] and in differentiating the ON versus OFF medication state [16] have been noted. They are also limited to evaluating the movement from the body location they are attached to. Multiple inertial motor units have been used to detect STS and quantify angular velocities of hip, knee, and ankle during STS [17]. However, the use of multiple wearables longitudinally in real-world settings has issues relating to acceptability [18] and usability [19].

Video data could capture posture, gait, and movement of all limbs, overcoming issues experienced by wearable/app developers. PD patients have positive perceptions about video at home if the type of data collected was shown to the participants beforehand and no audio was collected [20]. The participants of this study quickly acclimatized to the cameras in communal rooms [21]. Any privacy concerns could be tackled by video data being processed to minimize such risks to participants [22, 23, 24]. STS transitions have been detected from

laboratory data [25, 26] and in home [27] settings using video data from structured and observed activities, but no free-living (unstructured naturalistic behaviour) STS parameter quantification from real-world video data has been attempted in PD. In the laboratory, video data can predict two sub-scores of the UPDRS (bradykinesia and postural instability gait disturbance) with better results than two clinician video raters [28], indicating the potential of cameras passively tracking STS transitions in PD to produce useful endpoints with which to measure disease progression.

To date, there are no truly free-living in-home video datasets in PD without researcher presence. Furthermore, this is the first work to develop and test an automatic approach to quantification of STS parameters (duration and speed) using video data in PD in any setting. Finally, to the best of the authors' knowledge, this is the first work which looks to understand whether real-world duration of STS and speed of STS ascent could be used as potential digital biomarkers of PD progression. These parameters were carefully chosen for this work due to their promise as digital biomarkers [5], the availability of human rater ground truth (STS duration), and the validated algorithms developed already by our group which could then be applied to PD (STS speed) [8].

The objectives of this study are firstly to describe and test a computational approach to automatically quantify how someone gets up to stand from a seated position. The second objective is to explore whether the STS transition duration and speed from free-living real-world data could be used (a) to assess PD severity, (b) to distinguish between a PD patient's ON and OFF medication states, and (c) to distinguish between PD and control.

Methods

Participants

Twenty-four participants were recruited, 12 of whom had a clinical diagnosis of PD according to the UK Brain Bank Criteria with a Modified Hoehn and Yahr Scale score of 3 or less in the practically defined OFF medication state. Exclusion criteria included the presence of (other) neurodegenerative or significant musculoskeletal diseases in any participant.

Data Collection

The participants stayed in a two-storey instrumented "test-bed" house in pairs (a PD participant and a healthy control spouse/friend/family member) for five days/four nights continuously, during which time they largely lived freely, apart from two researcher visits to the setting to conduct clinical assessments [29]. PD participants withheld their long-acting dopaminergic medication for 24 h and short-acting agents for

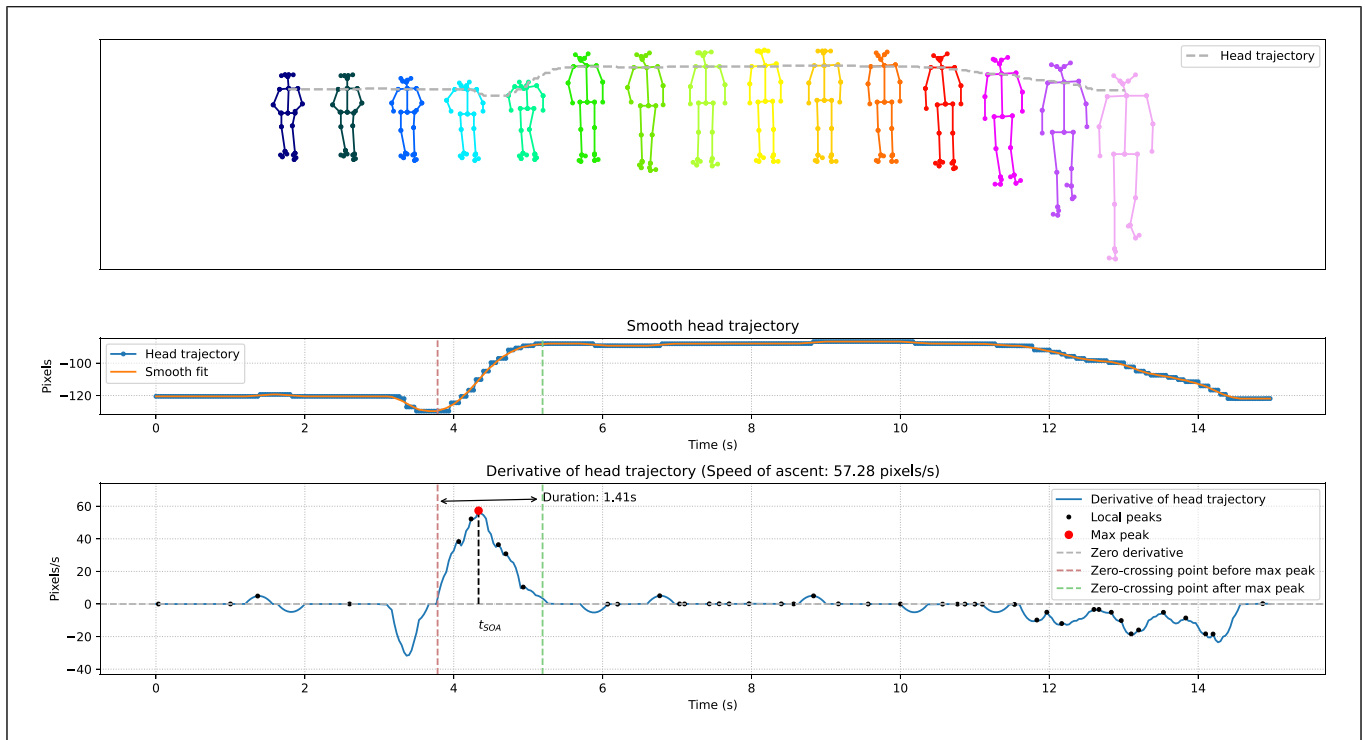


Fig. 1. Illustration of how the parameters for sit-to-stand were measured for this work.

12 h before (participants with deep brain stimulators switched stimulation off 1 h prior) the practically defined OFF data capture which lasted between 2- and 4-h mid-study. Wall-mounted video cameras in the kitchen, hall, dining, and living rooms recorded video for around 2 h a day at varying times based on participant preference. Each camera recorded around 8 h of footage, of which around 2 h were OFF medications for the PD participant. Resolution was 640×480 pixels with 30 frames/s.

Automatic Detection and Quantification Methodology

Videos were first analysed using OpenPose software [24] to detect human bodies and extract their skeleton joints in 15-second video-clips trimmed around each STS transition (first row of Fig. 1). The evolution in time of the head, or head trajectory, was estimated by averaging the vertical components of the five joints that compose the head plus the three joints that form the shoulders. The so-defined head trajectory was then used to estimate the STS parameters after being smoothed with a Savitzky-Golay filter [30] (a digital filter that can be applied to a set of digital datapoints for the purpose of smoothing the data; second row of Fig. 1). Two STS parameters were quantified: firstly, the speed of ascent (SOA) was evaluated similarly to our previous work [8] by looking at the maximum derivative of the head trajectory within each video clip; secondly, we estimated the STS “final attempt duration.” This process follows the previous estimation of the speed of ascent and the moment in time, namely t_{SOA} , at which the measure is detected. Once t_{SOA} is found, we identify the closest local minimum (when the person’s head is at its lowest point in the

video frame) before t_{SOA} , the closest the local maximum (highest point someone reaches in the frame after standing up) after t_{SOA} , and then we evaluate the time elapsed between the two points. The latter are easily identified in Figure 1 (third row) by looking at the points where the derivative of the head trajectory crosses the zero line.

It is important to note that the STS final attempt duration detected in this way only considers the time elapsed from when the person starts moving upwards until the movement is terminated; therefore, in case the participants require several attempts at standing up, only the last attempt will be timed. While this behaviour can create some large discrepancies with times measured by clinicians in a laboratory setting (which instead starts timing from when the participant first moves towards standing up), this helps improve robustness and consistency of the algorithm in a free-living environment.

Although the analysis performed in this work used the human rater label timestamps to identify STS episodes, the entire process (including STS detection) can be fully automated as we showed in our previous work to detect STS transitions in the real world [8]. The focus of this work is to demonstrate automatic quantification of STS parameters where detection has already taken place, show how these automatic parameters correlate with human annotations, and explore clinical correlations.

Implementation Details

An important step in measuring STS parameters is their conversion to physical measurements. While the STS duration can be easily converted from frames to seconds, knowing the frame

rate of the camera, the conversion to STS speed requires additional processing. In fact, skeleton coordinates are measured in pixels; therefore, the STS speed is expressed in pixels/s. These values must be converted into physical quantities (metres/second, m/s) before different measurements can be compared; otherwise, the STS speed will depend on the distance of the participant from the camera. To override this issue without having to perform lengthy calibration procedures, we simply normalize the STS speed by the height of the skeleton, in pixels, at the moment they complete the transition (i.e., zero-crossing point after max peak). This value, multiplied by the height of the participant in metres, allows us to estimate the speed of each participant in m/s.

The impact of the filter size for the Savitzky-Golay filter was also studied. This is defined by the window size and the polynomial order used to fit the head trajectory data and estimate the derivatives. In our experiment, we noticed that a smaller window size is beneficial for estimating STS speed but produces higher errors in the STS duration, while a larger window size has the opposite effect. This phenomenon is to be expected since larger window sizes increase smoothing and promote a reduction of false positives in the peaks used to estimate the STS duration. However, higher smoothing also means reduced gradients, which has a negative impact on the estimation of the STS speed. A compromise between the two measurements must therefore be made to obtain optimal STS parameters.

Human-Rater Annotations

The videos were watched post hoc by a number of different clinician and non-clinician human raters [31]. A widely available software called ELAN [32] was used. All STS episodes were identified and labelled (their location in the data was recorded), producing a start-to-end label of the “whole episode duration,” from when the person first moved their head forward, indicating they were about to stand up to when they were fully upright.

All the STS episodes were then rewatched by a clinician rater (neurology speciality registrar with training in MDS-UPDRS scoring including evaluation of STS and its subcomponents). This rater checked and, where they felt necessary (for 5% of the labels), adjusted the whole episode’s duration labels. They then provided the “final attempt duration” label in milliseconds, comprising their impression of the duration between the lowest point of the head (start) and when the person was fully upright/the maximum vertical position of the vertex of the head (end). For the purposes of this paper, this manually labelled “final attempt duration” is hereafter called “manual STS duration.” The rater also noted whether the person was dual-tasking – in this case, whether they were carrying something in their hand(s) as they stood up. 10% of the STS episodes were then randomly chosen to evaluate intra-rater reliability, with this subset scored several months after the first rating. The intra-rater agreement on labels was 98% for “final attempt duration.”

Statistical Approach

To investigate the correlations between the gold-standard clinical rating scales – the Movement Disorders Society-sponsored revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) [2] total score, motor subscore (III), and the postural instability gait dysfunction (PIGD) sub-score of the MDS-UPDRS [33] – and the STS parameters (duration and speed), Pearson or Spearman rank correlation coefficients were used.

Fisher’s *r*-to-*z* scores compared correlations from independent samples and Spearman’s partial correlations used to adjust for age and height as a potential confounder. To evaluate mean group difference between PD and control, Wilcoxon signed rank [34] tests were utilized, and to evaluate the statistical difference in parameter results between the ON and OFF medication states, the student paired *t* test [35] or Mann-Whitney U test [36] (depending on Shapiro-Wilk test of normality and tests of variance) was used.

The sample size was chosen as a convenience sample for this pilot study. The aim was to explore the hypothesis that certain activities such as STS show promise as digital biomarkers through correlations with gold-standard clinical rating scales and in their sensitivity to medication response.

Results

Participants

Table 1 shows the demographics for our study participants who had mean ages of 61.25 (PD) and 59.25 (control). The male:female sex ratios were 7:5 in the PD cohort and 3:9 in the control cohort, and years since PD diagnosis ranged from 0–2 to 18–20 years.

From 85 h of video data, there were 813 STS episodes seen by the human rater (383 control, 430 PD, of which 148 were OFF medications). This represents an average of 3.14 STS episodes per hour per person from daytime data.

Manual and Automatic Results Correlations

A significant correlation between the manual and automatic quantification of STS duration in seconds (Pearson $\rho = 0.419$, $p = 0.042$) is shown in Figure 2 (a), where each of the variables is averaged per participant. The speed of ascent in metres/s also correlates strongly with the manual STS duration in seconds (Pearson $\rho = -0.780$, $p < 0.001$), shown in Figure 2 (b). This negative correlation between speed and duration is expected since a faster speed of ascent leads to a shorter time to complete the STS transition. A comparison between the manual and automatic measurements also allowed us to calculate a bias and random error for our automatic STS duration, producing a bias of 0.10 s and a random error of 0.19 s.

Clinical Relevance of STS Speed and Duration

In Figure 3, we present an analysis where the automatic and manual measurements are grouped alongside each other by the medication state in the participant cohorts. Figure 3 (a) shows a good visual agreement in the spread of measurements between manual and automatic measurements for both of our participant cohorts. Mean automatic STS durations for the control group and PD group ON medications were 1.243 ($n = 200$) and 1.287 s

Table 1. Cohort-level participant demographic and clinical rating scale scores

	Participants with PD	Control participants
Age, years	61.25 (8.5)	59.25 (13.4)
Men, <i>n</i> %	7 (58)	3 (25)
Years since diagnosis	8.2 (6.5)	
Levodopa-equivalent daily dose, mg	517.5 (395.7)	
MDS-UPDRS total score	ON = 44.8 (16.1) OFF = 61.7 (29.9)	6.8 (4.8)
MDS-UPDRS III subscore	ON = 19.1 (10.4) OFF = 36.8 (23.0)	2.8 (1.9)
PIGD subscore	ON = 2.8 (1.9) OFF = 4.3 (4.8)	0.1 (0.3)

Values are means with standard deviation in parentheses unless indicated otherwise. MDS-UPDRS, Movement Disorders Society-Sponsored Revision of the Unified Parkinson’s Disease Rating Scale; PIGD, postural instability gait dysfunction.

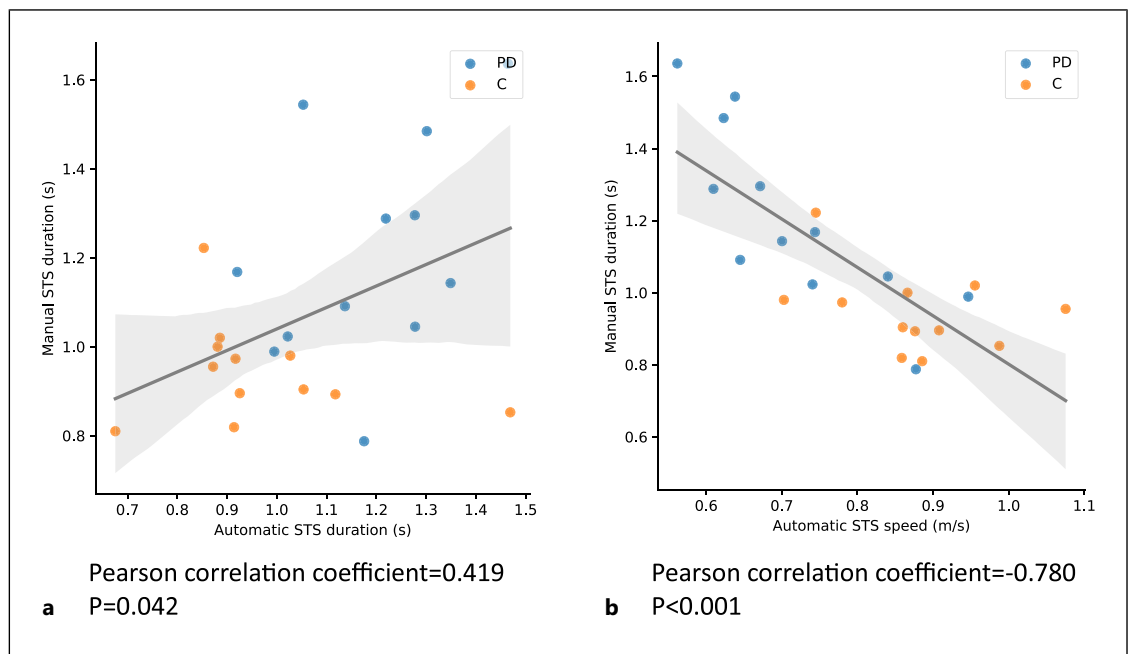


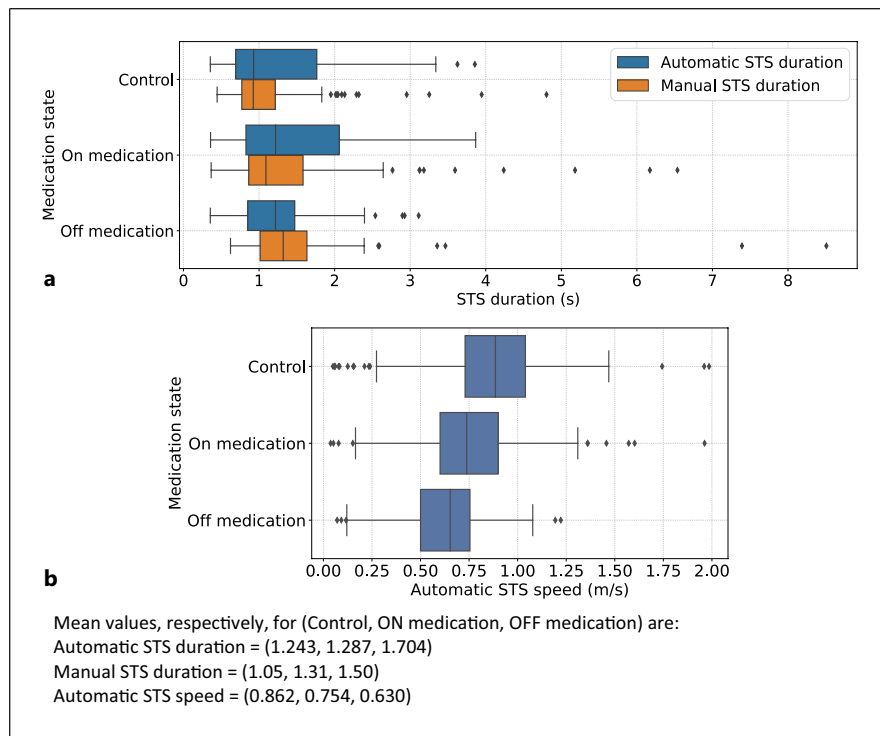
Fig. 2. Pearson correlations between manual STS duration and (a) automatic STS duration and (b) automatic STS speed.

(*n* = 75), respectively; these two means differed significantly (Mann-Whitney *U* = 6,263, *p* = 0.018). PD participants ON medications took significantly shorter mean automatic STS duration than when they were OFF medications (1.704 s), Wilcoxon *Z* = -1.835, *p* = 0.034. Mean automatic STS speed for the control group (0.862 m/s, *n* = 200) was significantly faster than the PD group ON medications (0.754 m/s, *n* = 75), Mann-Whitney

U = 9,965, *p* < 0.001. The PD cohort also had significantly slower automatic STS speed OFF medications (0.630 m/s, *n* = 43) than ON medications, Student *t* = 2.211, *p* = 0.016.

We next analysed whether the STS measurements could be used to differentiate between ON and OFF medication states in individual participants (Table 2). Using the automated measures, the speed of STS ascent

Fig. 3. Boxplots illustrating the differences between mean values of STS duration for control participants compared to PD participants' ON and OFF medications data, showing automatic and manual label outcomes (a), and the differences between mean speed of ascent values between control participants and PD participants ON and OFF medications (b).



was significantly reduced in two participants (1,028 and 1,034) while OFF medication. None of the participants showed significant differences in mean automatic STS duration between the ON and OFF medication states (the relatively small number of datapoints for individuals will have negatively influenced the statistical power).

To investigate how the STS automatic parameters correlate with clinical rating scales, Figure 4 displays the correlations with the MDS-UPDRS III scores, showing a positive correlation with STS duration (Spearman $\rho = 0.464, p = 0.022$) and a negative correlation with the STS speed of ascent (Spearman $\rho = -0.723, p < 0.001$). This information supports the clinical meaningfulness of the automatically quantified STS parameters in relation to the current gold-standard way of measuring disease severity (and progression) in PD. Of note and supporting the correlations relating to the automatic parameters, there is a significant correlation between the manual STS duration and the gold-standard clinical rating scale score, the MDS-UPDRS III score ON medications (Spearman $\rho = 0.746, p = 0.003$).

Further correlations are displayed in Table 3 with the whole dataset of STS labels; these are then split into subgroups of whether the participants are dual-tasking or not (and correlation data from all labels). Overall, the

positive correlations between scale scores commonly used as primary endpoints in clinical trials in PD (MDS-UPDRS total score, MDS-UPDRS III, and postural instability and gait dysfunction sub-score) and both STS speed and duration remained after adjusting for age and height as potential confounders. The only exception to this was a correlation between automatic STS duration and the MDS-UPDRS III score ($\rho = 0.365, p = 0.094$) when it was adjusted for height as well as age. The difference in participant heights in the PD cohort (mean = 171 cm) and control cohort (mean = 166 cm) was not significant ($t(22) = -1.578, p = 0.129$). The correlations from the dataset of STS while participants are dual-tasking do not reach significance, while those correlations with non-dual-tasking labels remain strong and significant. For speed, using Fisher's r-to-z statistical test, there is a significant difference between carrying something and not in the correlations with the MDS-UPDRS total score ON ($z = 3.351, p < 0.001$), MDS-UPDRS III ON ($z = 3.594, p < 0.001$), PIGD score ON ($z = 4.969, p < 0.001$). For duration, there is also a significant difference between carrying something and not in the correlations with the MDS-UPDRS total score ON ($z = -2.658, p = 0.004$), MDS-UPDRS III ON ($z = -3.569, p < 0.001$), PIGD score ON ($z = -1.995, p = 0.023$).

Table 2. Change in mean STS parameters (automatic duration, automatic speed) in individual participants with PD, comparing the ON with the OFF medication states, with paired samples Student's *t* tests *p* values to evaluate significance of the difference in STS duration between these two states

Participant ID	Years since diagnosis	Automatic STS speed mean, m/s			Automatic STS duration mean, s		
		medication status		<i>p</i> value	medication status		<i>p</i> value
		ON	OFF		ON	OFF	
1010	3–5	0.732 <i>n</i> = 15	0.778 <i>n</i> = 11	0.373	1.019 <i>n</i> = 15	1.215 <i>n</i> = 11	0.325
1012	6–8	0.635 <i>n</i> = 29	0.566 <i>n</i> = 5	0.222	1.597 <i>n</i> = 29	1.281 <i>n</i> = 5	0.812
1021	15–17	0.801 <i>n</i> = 13	0.387 <i>n</i> = 4	0.224	1.820 <i>n</i> = 13	1.210 <i>n</i> = 4	0.692
1023	15–17	0.919 <i>n</i> = 12	0.923 <i>n</i> = 6	0.381	1.484 <i>n</i> = 12	0.813 <i>n</i> = 6	0.929
1026	9–11	0.546 <i>n</i> = 17	0.768 <i>n</i> = 3	0.950	1.340 <i>n</i> = 17	1.000 <i>n</i> = 3	0.798
1028	18–20	0.853 <i>n</i> = 20	0.367 <i>n</i> = 9	0.004**	1.604 <i>n</i> = 20	1.454 <i>n</i> = 9	0.309
1030	0–2	0.827 <i>n</i> = 13	0.658 <i>n</i> = 5	0.460	1.343 <i>n</i> = 13	1.092 <i>n</i> = 5	0.582
1032	0–2	0.765 <i>n</i> = 15	0.684 <i>n</i> = 4	0.222	1.180 <i>n</i> = 15	1.282 <i>n</i> = 4	0.116
1034	0–2	0.656 <i>n</i> = 28	0.349 <i>n</i> = 8	0.005**	1.604 <i>n</i> = 28	1.397 <i>n</i> = 8	0.626
1036	3–5	0.683 <i>n</i> = 8	0.712 <i>n</i> = 10	0.602	1.840 <i>n</i> = 8	1.483 <i>n</i> = 10	0.664
1038	6–8	0.854 <i>n</i> = 18	0.917 <i>n</i> = 5	0.351	1.438 <i>n</i> = 18	1.063 <i>n</i> = 5	0.696
1040	9–11	0.857 <i>n</i> = 7	0.665 <i>n</i> = 9	0.083	1.897 <i>n</i> = 7	1.285 <i>n</i> = 9	0.885

ID, identification; STS, sit-to-stand; *n*, number of values; *s*, seconds; m/s, metres/second. Significant differences highlighted in bold. * *p* < 0.05; ** *p* < 0.005; *** *p* < 0.001.

Discussion

We present this state-of-the-art approach to automatically quantify how people sit-to-stand at home in PD built on a truly unique real-world dataset. This work illustrates novel ways to automatically quantify STS duration and ascent speed and validates that the two parameters correlate strongly with a human rater's impressions of STS duration. Furthermore, the clinical relevance of the STS transition outcomes is demonstrated through significant Spearman correlations between the automatic transition outcomes and the gold-standard clinical rating scales. Furthermore, there are significant differences in the two automatic STS parameters between the control cohort and the PD cohort ON medications (when their symptoms should be controlled).

We have collected many (85) hours of videos of people with PD (and control participants) in a real-world setting going about their daily living with no scripts. This video has all been watched and painstakingly annotated (second by second) by human raters. The amount of video of each participant has enabled many hundreds of STS transitions to be captured – producing a rich dataset of this functionally relevant activity. Since PD is a disease which is different in every person, this work prioritizes demonstrating and validating our approach to free-living, showing how people with PD differ from control participants and capturing data in PD ON and OFF medications. We believe our sample size is ample to show proof of concept (with strong correlations to clinical rating scales, for example), and we are sharing our novel approach with

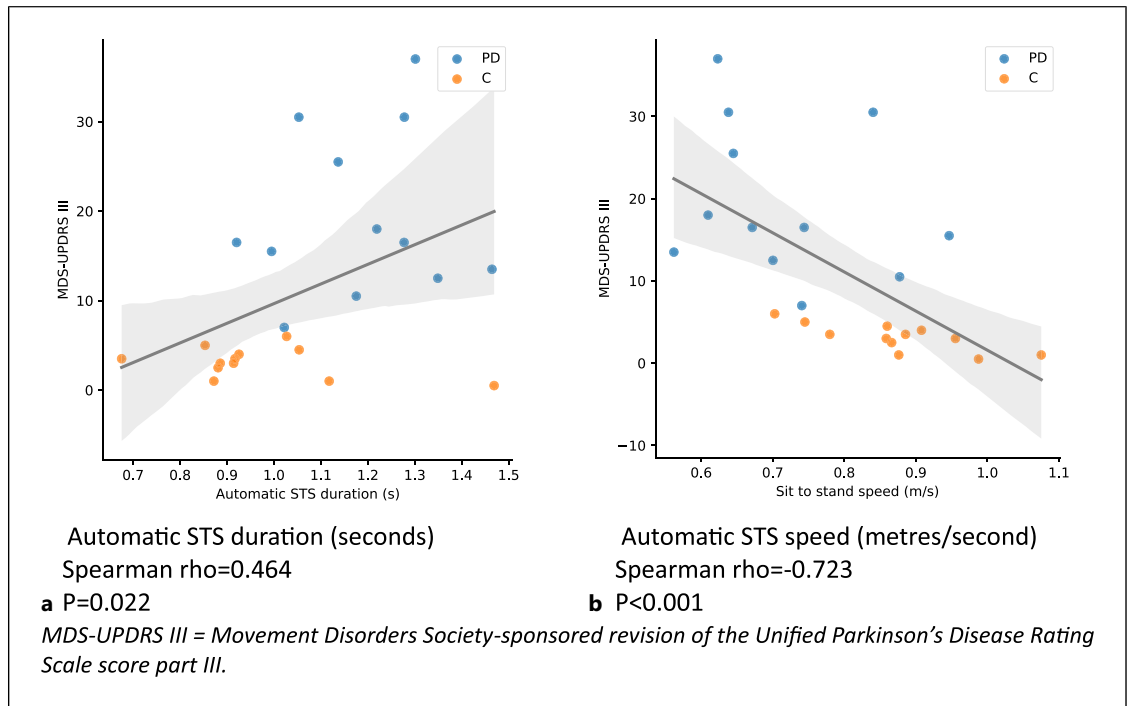


Fig. 4. Scatter plot illustrating the Spearman correlations between the MDS-UPDRS III scores (from PD participants ON medications and control participants) and automatic STS duration (seconds) (**a**), automatic STS speed (metres/second) (**b**).

Table 3. Spearman correlations between automatic STS parameters (speed and duration) and clinical rating scale scores, subdivided into groups of dual-tasking, not dual-tasking and all labels, adjusted for participant age and height

Variable	Test	Automatic STS speed correlations			Automatic STS duration correlations		
		dual-tasking	not dual-tasking	all	dual-tasking	not dual-tasking	all
MDS-UPDRS total score	Spearman's rho	-0.465	-0.705***	-0.648***	0.065	0.506*	0.479*
	p value	0.094	<0.001	0.001	0.824	0.016	0.024
MDS-UPDRS III	Spearman's rho	-0.381	-0.691***	-0.593**	-0.075	0.457*	0.365
	p value	0.179	<0.001	0.004	0.798	0.033	0.094
PIGD score	Spearman's rho	-0.290	-0.576**	-0.496*	-0.130	0.488*	0.439*
	p value	0.315	0.005	0.019	0.657	0.021	0.041

Results adjusted for participant age and height. MDS-UPDRS, Movement Disorders Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale; MDS-UPDRS III, motor sub-scale part 3 of the MDS-UPDRS; PIGD, postural instability gait dysfunction (a sub-score of MDS-UPDRS III comprising clinical evaluation of posture, gait, and postural instability); STS, sit-to-stand. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

other researchers so that they can potentially evaluate and build upon it. There are very few groups who have used cameras in real-world settings in PD, and this is the first such work with unobtrusive ground truth validation from wall-mounted cameras.

From indoor free-living, each participant did an average of 3.14 STS transitions per hour. If captured longitudinally, this could build a rich and informative dataset of how someone mobilizes at home. For example, if measuring for 16 waking hours per day over 18 months, around 27,500

episodes would be captured. Disease progression measurement would be enhanced with frequent such stereotyped outcomes, supplementing unreliable patient diaries [37] and the infrequent clinical rating scales used commonly in clinical trials [38, 39]. A particular strength of this study is in the ecological validation, through a second-by-second video labelling approach. The burden of time and resources to obtain this ground truth needed for algorithm validation is a challenge in this area [4].

We have shown significant correlations between the human-rater annotations and the automatically quantified parameters. The stronger group-level correlations between manual STS duration and automatic STS speed (compared to automatic STS duration) indicate that automatically-derived speed of ascent in STS may have greater ecological validity in mapping to the clinician-rater ground truth labels. Neither of the automatic parameters is a like-for-like comparison with the manual STS duration: the automatic STS duration has variability introduced by the fact that it evaluates the person's sternum and/or head movement, whereas the manual approach looks only at the position of the vertex of the head. When calculating the speed of ascent, the participant height (which is not factored into the automatic STS speed quantification) introduces an additional source of variability.

We have shown that STS transition duration and speed of ascent can differentiate between PD and control at a group level in our cohort of participants with early-stage to moderate PD, giving our approach face validity. Furthermore, there were significant differences between ON and OFF dopaminergic medication states demonstrated by both duration and speed of STS at a group level. At an individual level, two participants showed a significant slowing in their speed of STS ascent while OFF medications compared to being ON medications. This indicates an early promise of the automatic parameters to detect symptom fluctuations relating to dopaminergic medications in an individual. The lack of difference in automatic STS mean duration between the ON and OFF medication states in any participant is interesting and reflects the need for larger numbers of labels to see whether this is also true in bigger datasets – especially since previous reports have indicated that posture and gait symptoms are relatively refractory to standard dopaminergic therapies [40].

Strong correlations were seen between the gold-standard clinical rating scale used as primary outcome measures of disease progression in most clinical trials in PD [41] and both the manually quantified STS duration and automatically quantified STS duration and speed of ascent. Although previous correlations between laboratory-

derived STS parameters and the MDS-UPDRS III scores/PD stage have been shown [5, 42], this is the first work looking at these correlations from free-living real-world STS transitions in PD, with adjustment for age and height increasing the robustness of results. We did not adjust for participant sex since there is evidence that it does not alter significantly sit-to-stand outcomes [43, 44]. We propose that the automatically-derived STS results could provide a continuous and more granular output than the clinical rating scales, with larger numbers of datapoints that in turn could reduce the time taken to show the minimum clinically important difference in clinical trials of therapeutics.

There are many ways in which someone could dual-task while standing up from sitting. This work showed that the correlations between STS parameters and clinical rating scale scores remained significant for the data where the person was not carrying something, but lost their significance for the data where the participants were carrying something. The Fisher's *r*-to-*z* statistic proved that there was a true difference between the two datasets (carrying something or not) in terms of their correlation strength. The number of "carrying something" datapoints was limited at 54, so this needs further investigation to explore the change in transition quality resulting from adding the extra cognitive burden/motor task of carrying something, including looking at what they were carrying (hot drink, newspaper, etc.). Furthermore, the STS duration and speed, as measured in this paper, measure the final component of the transition – other features such as whether a person shifts forward in the chair, or whether they take multiple attempts before finally successfully standing, are not quantified. While the analysis of the full movement including the aforementioned components adds information clinically, their detection does not translate well into free-living environments, where variation is added as people may stand up from a variety of positions (e.g., lying) and while doing other activities (e.g., picking up something from a table). Therefore, only looking at the time taken for the final ascent could remove confounders and noise from the data experienced by other groups [15] when analysed longitudinally in a single participant. For future work, developing automatic detection of STS transition subcomponents could add useful clinical information.

The automatic approach in this paper performed well in a variety of conditions, indicating high reliability and robustness. At times, the participants had the lower halves of their bodies occluded (from the camera) by a table, for example. Nevertheless, the use of only the head and shoulder skeleton joints in this approach meant that

these occlusions did not affect its ability to measure STS speed or duration. The RGB data were relatively low resolution (640×480 pixels), but since our approach relies on skeleton data, which were produced satisfactorily from the RGB, it was robust to this moderately low RGB resolution. The number of frames of RGB data collected per second varies according to factors such as the complexity of the image. However, we find that this does not affect the use of the data for our purposes since the data collection rate is adequate to capture the characteristics of STS episodes. The camera hardware was wall-mounted in 4 different rooms, each with different lighting conditions, e.g., the kitchen had a large skylight while the hallway had minimal daylight. The data were also collected both in daylight with curtains open, and in the evenings/early mornings with curtains shut and room lights on. The subjects were allowed to walk wherever they pleased, use any chair, and move the chairs if they wished. This meant that the STS episodes occurred at various distances and angles from the cameras. However, skeleton pose data were successfully extracted, and our approach was robust to these variations.

There are factors which could influence STS duration and ascent speed, including the type of chair someone sits in and the physical room setting (e.g., how narrow the space is around the chair). If someone does not frequently change their furniture or room layout, the infinite variety of different peoples' homes should not preclude the aim to repeatedly measure specific activities so that one person's performance can be measured over long periods of time. However, work is still needed to identify and control for these sources of variability during mobility assessments for optimal accuracy in interpretation of results.

A limitation experienced in the development of algorithms used to generate the automatic parameters was the compromise between optimizing the system to be most accurate relating to automatic STS speed (e.g., accounting for when someone not only stood up but also then walked towards the camera, so appearing to continue moving "up" in frame even though their STS transition was finished) and trying not to exclude STS episodes because they did not conform to the "perfect" easily quantifiable transition. A further issue encountered was the occlusion of a person in frame by something/someone else. One way to solve this problem in future work is using 3D skeletons needing less calibration and more advanced pose extraction methods which should reduce the occlusion issue.

Another limitation in this work is sample size (discussed in methods). Future work towards evaluating longitudinal disease progression would require large (more generalizable) datasets stratified for a disease

sub-group (e.g., akinetic-rigid, tremor-dominant), over long periods of time, comparing changes in neuroimaging and selected wet biomarkers such as alpha-synuclein protein [45]. Because participant cohorts are likely to be more weighted towards sub-clinical and early-stage disease in clinical trials of putative disease-modifying agents [4], we propose that future work should include more participants in these stages of PD. Three of our participants had had PD more than 15+ years, although they maintained a H&Y score of 3 in the OFF medication state. Their results would not be representative of early-stage disease.

The cameras we have used are low cost and scalable to multiple homes over long periods of time [46–48], and vitally, they are acceptable in free-living for people with PD [20, 21]. This work shows that an automatic framework for the quantification of STS duration and speed of ascent in a free-living environment is possible and that these two parameters show promise in terms of predicting disease severity and can differentiate between medication states in some individuals. These are good first steps, but regulatory approval for sensor-derived outcomes is a further hurdle if a new digital tool is going to be adopted as a primary outcome measure in clinical trials in PD [4]. Close liaison with people living with PD, regulators, end-users (including pharmaceutical companies), and other stakeholders continues to be necessary moving forward. The end goal should be to validate a functionally relevant, patient-centred digital outcome measure which is reproducible, reliable, and can complement the existing clinical practices in PD symptom evaluation.

Acknowledgments

We gratefully acknowledge the study participants for their time and efforts in participating in this research. We also acknowledge the local Parkinson's and Other Movement Disorders Health Integration Team (Patient and Public Involvement Group) for their assistance at each step of study design.

Statement of Ethics

This study protocol was reviewed and approved by the National Health Service Wales Research Ethics Committee 6 on December 17, 2019, and approval from the Health Research Authority and Health and Care Research Wales was confirmed on January 14, 2020 (reference 19/WA/0051). All participants gave written informed consent to participate. This research was conducted ethically in accordance with the World Medical Association Declaration of Helsinki.

Conflict of Interest Statement

The authors declare no competing interests.

Funding Sources

This work was supported by the SPHERE Next Steps Project funded by the UK Engineering and Physical Sciences Research Council (EPSRC) [Grant EP/R005273/1], the Elizabeth Blackwell Institute for Health Research, and the Wellcome Trust Institutional Strategic Support Fund [Grant code: 204813/Z/16/Z]; by Cure Parkinson's [Grant code AW021]; and by IXICO [Grant code R101507-101]. Dr. Jonathan de Pass and Mrs. Georgina de Pass made a charitable donation to the University of Bristol through the Development and Alumni Relations Office; the funding pays for the salary of CM, but they have no input into her work.

Author Contributions

C.M. contributed to research project design, organization, execution; statistical analysis design, execution, and review/critique; writing of the first draft of manuscript, and manuscript

review/critique. A.M. contributed to research project execution; statistical analysis design, execution, and review/critique; writing of the first draft of manuscript, and manuscript review/critique. C.M. and A.M. contributed equally to this work. M.M. contributed to research project conception; statistical analysis design, execution, and review/critique; manuscript review/critique. H.K.I. contributed to statistical analysis design, execution, and review/critique; manuscript review/critique. F.J. and R.M. contributed to statistical analysis review/critique and manuscript review/critique. E.L.T. contributed to research project organization and execution; manuscript review/critique. I.C. and A.W. contributed to research project design, organization, execution; statistical analysis review/critique, and manuscript review/critique.

Data Availability Statement

The algorithm will be made available in due course at <https://github.com/ale152/SitToStandPD>. Restrictions apply to the availability of the data according to the permissions granted by the study participants, and so these data are not yet publicly available. Further enquiries can be directed to the corresponding author. Please address correspondence to Dr. Catherine Morgan.

References

- 1 Post B, Merkus MP, de Bie RMA, de Haan RJ, Speelman JD. Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? *Mov Disord*. 2005 Dec;20(12):1577–84.
- 2 Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord*. 2008 Nov;23(15):2129–70.
- 3 Artusi CA, Mishra M, Latimer P, Vizcarra JA, Lopiano L, Maetzler W, et al. Integration of technology-based outcome measures in clinical trials of Parkinson and other neurodegenerative diseases. *Parkinsonism Relat Disord*. 2018 Jan;46(Suppl 1):S53–S56.
- 4 Espay AJ, Hausdorff JM, Sánchez-Ferro Á, Klucken J, Merola A, Bonato P, et al. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. *Mov Disord*. 2019 May;34(5):657–63.
- 5 Toosizadeh N, Mohler J, Lei H, Parvaneh S, Sherman S, Najafi B. Motor performance assessment in Parkinson's disease: association between objective in-clinic, objective in-home, and subjective/semi-objective measures. *PLoS One*. 2015 Apr;10(4):e0124763.
- 6 da Cunha CP, Rao PT, Karthikbabu S. Clinical features contributing to the sit-to-stand transfer in people with Parkinson's disease: a systematic review. *Egypt J Neurol Psychiatr Neurosurg*. 2021 Dec;57(1):143–15.
- 7 Mak MKY, Hui-Chan CWY. The speed of sit-to-stand can be modulated in Parkinson's disease. *Clin Neurophysiol*. 2005 Apr;116(4):780–9.
- 8 Masullo A, Burghardt T, Perrett T, Damen D, Mirmehdi M. Sit-to-stand analysis in the wild using silhouettes for longitudinal health monitoring. *Lect Notes Computer Sci*. 2019 Oct;11663 LNCS:175–85.
- 9 Raffegau T, Krehbiel L, Kang N, Thijs F, Altmann L, Cauraugh J, et al. A meta-analysis: Parkinson's disease and dual-task walking. *Parkinsonism Relat Disord*. 2019;62:28–35.
- 10 Penko AL, Streicher MC, Koop MM, Dey T, Rosenfeldt AB, Bazyk AS, et al. Dual-task interference disrupts Parkinson's gait across multiple cognitive domains. *Neuroscience*. 2018 May;379:375–82.
- 11 Robles-García V, Corral-Bergantiños Y, Espinosa N, Jácome MA, García-Sancho C, Cudeiro J, et al. Spatiotemporal gait patterns during overt and covert evaluation in patients with Parkinson's disease and healthy subjects: is there a hawthorne effect? *J Appl Biomech*. 2015 Jun;31(3):189–94.
- 12 Morgan C, Jameson J, Craddock I, Tonkin EL, Oikonomou G, Isotalus HK, et al. Understanding how people with Parkinson's disease turn in gait from a real-world in-home dataset. *Parkinsonism Relat Disord*. 2022 Dec;105:114–22.
- 13 Pinto C, Schuch CP, Balbinot G, Salazar AP, Hennig EM, Kleiner AFR, et al. Movement smoothness during a functional mobility task in subjects with Parkinson's disease and freezing of gait - an analysis using inertial measurement units. *J NeuroEng Rehabil*. 2019 Sep;16(1):110.
- 14 González Rojas HA, Cuevas PC, Zayas Figueras EE, Foix SC, Sánchez Egea AJ. Time measurement characterization of stand-to-sit and sit-to-stand transitions by using a smartphone. *Med Biol Eng Comput*. 2018 May;56(5):879–88.
- 15 Pham MH, Warmerdam E, Elshehabi M, Schlenstedt C, Bergeest LM, Heller M, et al. Validation of a lower back "wearable"-based sit-to-stand and stand-to-sit algorithm for patients with Parkinson's disease and older adults in a home-like environment. *Front Neurol*. 2018 Aug;9(AUG):652.
- 16 Miller Koop M, Ozinga SJ, Rosenfeldt AB, Alberts JL. Quantifying turning behavior and gait in Parkinson's disease using mobile technology. *IBRO Rep*. 2018 Dec;5:10–6.
- 17 Memar S, Delrobaei M, Pieterman M, McIsaac K, Jog M. Quantification of whole-body bradykinesia in Parkinson's disease participants using multiple inertial sensors. *J Neurol Sci*. 2018 Apr;387:157–65.
- 18 AlMahadin G, Lotfi A, Zysk E, Siena FL, Carthy MM, Breedon P. Parkinson's disease: current assessment methods and wearable devices for evaluation of movement disorder motor symptoms: a patient and healthcare professional perspective. *BMC Neurol*. 2020 Dec;20(1):419.

- 19 Silva de Lima AL, Hahn T, Evers LJW, de Vries NM, Cohen E, Afek M, et al. Feasibility of large-scale deployment of multiple wearable sensors in Parkinson's disease. *PLoS One*. 2017;12(12):e0189161.
- 20 McNaney R, Morgan C, Kulkarni P, Vega J, Heidarivincheh F, McConville R, et al. Exploring perceptions of cross-sectoral data sharing with people with Parkinson's. CHI conference on human factors in computing systems. 2022 Apr;1–14.
- 21 Morgan C, Tonkin EL, Craddock I, Whone AL. Acceptability of an in-home multimodal sensor platform for Parkinson disease: non-randomized qualitative study. *JMIR Hum Factors*. 2022;9(3):e36370.
- 22 Li Y, Vishwamitra N, Knijnenburg B, Hu H, Caine K. Effectiveness and users' experience of obfuscation as a privacy-enhancing technology for sharing photos. *Proc ACM Hum Comput Interact*. 2017 Nov;1(CSCW):1–24.
- 23 Agrawal R, Srikant R, Thomas D. Privacy preserving OLAP. Proceedings of the ACM SIGMOD International Conference on Management of Data. 2005:251–62.
- 24 Cao Z, Hidalgo G, Simon T, Wei S-E, Sheikh Y. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE Trans Pattern Anal Mach Intell*. 2021;43(1):172–86.
- 25 Galna B, Barry G, Jackson D, Mhiripiri D, Olivier P, Rochester L. Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. *Gait Posture*. 2014 Apr;39(4):1062–8.
- 26 Li T, Chen J, Hu C, Ma Y, Wu Z, Wan W, et al. Automatic timed up-and-go sub-task segmentation for Parkinson's disease patients using video-based activity classification. *IEEE Trans Neural Syst Rehabil Eng*. 2018;26(11):2189–99.
- 27 Ejupi A, Brodie M, Gschwind YJ, Lord SR, Zagler WL, Delbaere K. Kinect-based five-times-sit-to-stand test for clinical and in-home assessment of fall risk in older people. *Gerontology*. 2015 Dec;62(1):118–24.
- 28 Mehta D, Asif U, Hao T, Bilal E, von Cavallar S, Harrer S, et al. Towards automated and marker-less Parkinson disease assessment: predicting UPDRS scores using sit-stand videos. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2021;3836–44.
- 29 Morgan C, Craddock I, Tonkin EL, Kinnunen KM, Mcnaney R, Whitehouse S, et al. Protocol for PD SENSORS: Parkinson's Disease Symptom Evaluation in a Naturalistic Setting producing Outcome measuRes using SPHERE technology. An observational feasibility study of multi-modal multi-sensor technology to measure symptoms and activities of daily living in Parkinson's disease. *BMJ Open*. 2020 Nov;10(11):e041303.
- 30 `scipy.signal.savgol_filter` — SciPy v1.9.0 Manual [Internet]. [cited 2022 Aug 3]. Available from: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.savgol_filter.html.
- 31 Morgan C, Heidarivincheh F, Craddock I, McConville R, Nieto MP, Tonkin EL, et al. Data labelling in the wild: annotating free-living activities and Parkinson's disease symptoms. 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events, PerCom Workshops 2021. 2021 Mar;471–4.
- 32 ELAN (Version 6.2). *The Language Archive*. Nijmegen. 2021. Available from: <https://archive.mpi.nl/tla/elan>.
- 33 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*. 2013 May;28(5):668–70.
- 34 Wilcoxon F. Individual comparisons by ranking methods. *Biometrics Bull*. 1945 Dec;1(6):80.
- 35 Student. The probable error of a mean. *Biometrika*. 1908;6:1–25. [cited 2022 Aug 4]
- 36 Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Statist*. 1947 Mar [cited 2022 Nov 30;18(1):50–60.
- 37 Erb MK, Karlin DR, Ho BK, Thomas KC, Parisi F, Vergara-Diaz GP, et al. mHealth and wearable technology should replace motor diaries to track motor fluctuations in Parkinson's disease. *NPJ Digit Med*. 2020 Jan;3(1):6–10.
- 38 Hauser R, Friedlander J, Zesiewicz TA, Adler CH, Seeberger LC, O'Brien CF, et al. 2000 A home diary to assess functional status in patients with Parkinson's disease with motor fluctuations and dyskinesia. *Clin Neuropharmacol*. 2000;23(2):75–81.
- 39 Stacy M, Bowron A, Guttman M, Hauser R, Hughes K, Larsen JP, et al. Identification of motor and nonmotor wearing-off in Parkinson's disease: comparison of a patient questionnaire versus a clinician assessment. *Mov Disord*. 2005 Jun;20(6):726–33.
- 40 Sethi K. Levodopa unresponsive symptoms in Parkinson disease. *Mov Disord*. 2008 Jan;23(Suppl 3):S521–33.
- 41 Van Uem JMT, Isaacs T, Lewin A, Bresolin E, Salkovic D, Espay AJ, et al. A viewpoint on wearable technology-enabled measurement of wellbeing and health-related quality of life in Parkinson's disease. *J Parkinsons Dis*. 2016;6(2):279–87.
- 42 di Lazzaro G, Ricci M, Saggio G, Costantini G, Schirinzi T, Alwardat M, et al. Technology-based therapy-response and prognostic biomarkers in a prospective study of a de novo Parkinson's disease cohort. *NPJ Parkinsons Dis*. 2021 Dec;7(1):82.
- 43 Pua YH, Tay L, Clark RA, Thumboo J, Tay EL, Mah SM, et al. Reference ranges for gait speed and sit-to-stand performance in a cohort of mobility-intact community-dwelling older adults from Singapore. *J Am Med Dir Assoc*. 2022 Sep;23(9):1579–84. e1.
- 44 Bohannon RW, Bubela DJ, Magasi SR, Wang YC, Gershon RC. Sit-to-stand test: performance and determinants across the age-span. *Isokinet Exerc Sci*. 2010 Jan;18(4):235–40.
- 45 Ganguly U, Singh S, Pal S, Prasad S, Agrawal BK, Saini RV, et al. Alpha-synuclein as a biomarker of Parkinson's disease: good, but not good enough. *Front Aging Neurosci*. 2021;13:702639.
- 46 Elsts A, Burghardt T, Byrne D, Camplani M, Damen D, Fafoutis X, et al. *A guide to the SPHERE 100 homes study dataset*; 2018 May. [cited 2018 Nov 15]. Available from: <http://arxiv.org/abs/1805.11907>.
- 47 Zhu N, Diethe T, Camplani M, Tao L, Burrows A, Twomey N, et al. Bridging e-health and the internet of things: the SPHERE project. *IEEE Intell Syst*. 2015 Jul;30(4):39–46.
- 48 Woznowski P, Burrows A, Diethe T, Fafoutis X, Hall J, Hannuna S, et al. SPHERE: a sensor platform for healthcare in a residential environment. *Designing, developing, and facilitating smart cities*. Cham: Springer International Publishing; 2017. p. 315–33.