

Data Descriptor Template

Title

WearGait-PD: An Open-Access Wearables Dataset for Gait in Parkinson's Disease and Age-Matched Controls

Authors

Anthony J. Anderson^{1*}, David Eguren^{1*}, Michael A. Gonzalez^{1*}, Naima Khan¹, Sophia Watkinson¹, Michael Caiola¹, Siegfried S Hirczy², Cyrus P. Zabetian^{2,3}, Kelly Mills⁴, Emile Moukheiber⁴, Laureano Moro-Velazquez⁵, Najim Dehak⁵, Chelsie Motely⁴, Brittney C. Muir², Ankur Butala⁴, and Kimberly Kontson¹

*These authors contributed equally to the manuscript.

Affiliations

1. US Food and Drug Administration, Center for Devices and Radiological Health, Office of Science and Engineering Laboratories, Silver Spring, MD, USA
2. VA Puget Sound Healthcare System, Seattle, WA, USA
3. Department of Neurology, University of Washington, Seattle, WA, USA
4. Johns Hopkins University School of Medicine, Baltimore, MD USA
5. Johns Hopkins University Whiting School of Engineering, Baltimore, MD, USA

corresponding author: Kimberly Kontson (Kimberly.Kontson@fda.hhs.gov)

Abstract

Wearable movement sensors are powerful tools for objectively characterizing and quantifying movement. They enhance the precise characterization of gait, balance, and motor symptoms in Parkinson's disease and related disorders, facilitating in-clinic and remote assessments, disease management, and therapeutic intervention development. Access to high-quality data from these sensors can accelerate discoveries in this clinical population. The WearGait-PD open-access dataset contains raw inertial measurement unit (IMU) and sensorized insole data from individuals with PD and age-matched controls, synchronized to a gait walkway reference system. IMU data include 3-degree of freedom (DOF) acceleration, rotational velocity, magnetic field strength, and orientation for each of 13 sensors on the participant's body. Sensor insole data include absolute pressure from 16 sensors in each insole and 3-DOF acceleration and rotational velocity. Walkway data include 2D position and relative pressure for each active sensor during every footfall. Frame-by-frame annotation of participant actions during gait and balance tasks was incorporated using synchronized video cameras. All data were associated with demographic information and clinical evaluations (e.g., medications, DBS-status, MDS-UPDRS scores).

Background & Summary

Parkinson's disease (PD) is a progressive neurodegenerative disorder that often impairs gait and balance, resulting in reduced mobility and independence, and diminished quality of life¹⁻³. Wearable movement sensors have emerged as powerful tools for objectively characterizing and quantifying movement. These offer the potential to enhance the precise characterization of gait, balance, and motor symptoms in Parkinson's disease (PD) and related disorders, facilitating in-clinic and remote assessments, disease management, and therapeutic intervention development⁴⁻⁷. In contrast, current clinical assessments of gait and balance are semi-quantitative but remain largely qualitative, relying on subjective observations and patient reports. Wearable sensors promise to improve care by enabling quantitative

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

52 assessments as part of routine clinical practice¹¹, remote patient monitoring^{12,13}, and objective
53 endpoint measures for the regulatory evaluation of new drugs and medical devices¹⁴. By
54 providing continuous, high-resolution data on mobility-related health, these sensors and their
55 associated algorithms can identify motor signs of PD, track disease progression, and offer a
56 more complete picture of an individual's functional status¹⁵⁻¹⁷. This data-driven approach can
57 potentially transform PD treatment, ultimately improving patient outcomes and quality of life.

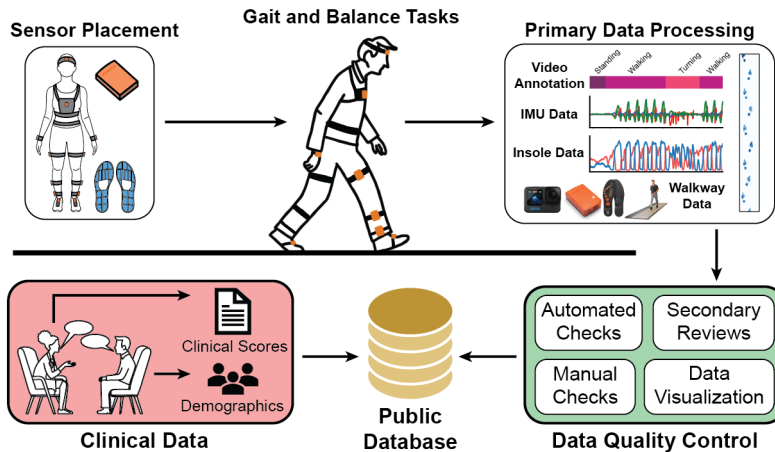
58
59 Despite the potential of wearable movement sensors to improve the treatment of PD, their
60 translation and widespread clinical adoption have been limited due to various scientific,
61 technical, economic, and regulatory issues¹⁸⁻²⁰. However, efforts are underway across
62 academia, the medical device industry, and regulatory bodies to overcome these limitations.
63 For instance, academic-led consortiums, such as Mobilise-D, are working to establish, verify,
64 and validate digital mobility outcomes, creating a framework for their use in clinical settings²¹.
65 Medical device developers have created wrist-worn devices that monitor the presence of
66 upper limb symptoms of PD and incorporate these measurements into clinician-facing
67 dashboards in an aim to improve treatment decisions^{22,23}. Additionally, the FDA has issued
68 guidance on using digital health technologies for remote data acquisition in clinical
69 investigations²⁴ and participates in the Digital Health Measurement Collaborative Community
70 (DATAcc) by the Digital Medicine Society, while the European Medicines Agency recently
71 approved a digitally-derived measure for use in Duchenne Muscular Dystrophy clinical trials¹⁴.
72 These activities indicate a growing recognition by regulatory organizations of the
73 transformative potential of the technology.

74
75 The need for high-quality, diverse, and well-annotated data remains a significant obstacle to
76 the development, validation, and clinical translation of wearable sensor digital health
77 technologies. As digital health technologies rely on a combination of computing platforms,
78 sensors, and algorithms²⁴, data are a fundamental requirement at every stage of development,
79 from prototyping through clinical trials²⁵. However, obtaining high-quality movement data
80 from people with PD is challenging due to the high costs associated with data collection, the
81 fragility of the population, and the necessity for skilled clinicians to score and annotate the
82 data. These barriers create a development landscape where only well-funded organizations,
83 public or private, have the resources to participate in the innovation process, limiting the
84 diversity of ideas and approaches explored.

85
86 Open-access datasets offer one solution to the challenge posed by the lack of data in the
87 development of wearable technologies for PD. By providing high-value clinical datasets to the
88 research and development communities, open-access data efforts can democratize the
89 innovation process, enabling a wider range of researchers and developers to contribute to
90 advancing these technologies. Open-access datasets also promote reproducibility by allowing
91 researchers and developers to validate their findings independently and with diverse data.
92 While several open-access wearable sensor and smartphone datasets for PD already exist²⁶⁻³³,
93 their overall utility may be limited by: (1) a primary focus on wrist-based monitoring, which
94 may not capture the full spectrum of gait and balance impairments, (2) a lack of associated
95 clinical information, such as disease severity scales and medication status, (3) insufficient
96 contextual information surrounding the data, such as time-series descriptions of specific
97 locomotor activities, and (4) a lack of reference measurements for the analytical validation of
98 algorithms²⁵. Addressing these limitations is crucial for streamlining the integration of
99 wearable sensors for clinical use in PD.

100
101 In response to these limitations, we present a large, multi-site, multi-modal open dataset
102 (WearGait-PD) of older adults with and without PD, engaged in both structured and semi-
103 structured gait and balance tasks while wearing a full-body suite of sensors (Figure 1). This

104 dataset fills several vital gaps in the field, as it: (1) provides data from a full-body sensor suite
 105 of inertial measurement units and sensorized insoles, (2) includes comprehensive clinical
 106 information, such as Movement Disorder Society – Unified Parkinson’s Disease Rating Scale
 107 (MDS-UPDRS) scores and demographic details for each participant, (3) offers frame-by-frame,
 108 human-readable annotations applied by two expert reviewers based on video assessment,
 109 with clinician annotation of PD-specific symptoms such as freezing of gait, and (4) incorporates
 110 ground truth measurements from a pressure-sensing walkway for independent validation of
 111 digitally-derived metrics. By making this dataset publicly available to the research and medical
 112 product development communities, we aim to accelerate innovation and ultimately access to
 113 effective digital health technologies for assessing gait and balance in PD.
 114



115
 116 *Figure 1: Overview of data workflow for the WearGait-PD dataset. Sensor data were collected*
 117 *via a set of wearable sensors (IMUs and sensorized insoles) and reference systems (cameras*
 118 *and a sensorized walkway). The sensor data were processed, time synchronized, visualized,*
 119 *and reviewed before uploaded to the public repository. Clinical data included demographic*
 120 *information and results from clinical tests when possible.*
 121
 122

123 **Methods**

124 Participants

125 At the time of submission, 126 participants were included in the dataset. Sixty-one PD
 126 participants (25F/36M) with an average age of 66 ± 9 years and modified Hoehn & Yahr score
 127 of 2.15 ± 0.48 were included. The total MDS-UPDRS Part III score for these participants ranged
 128 from 6-46, with a mean and standard deviation of 24 ± 10 . The protocol did not require
 129 participants to perform tasks in a specific medication state (i.e., ON or OFF). Instead, the time
 130 elapsed since their last PD medication dose was recorded as a proxy. A total of 41 freezing of
 131 gait events were captured across five unique PD participants. Within the PD group, 158 stagger
 132 events were captured. Sixty-five control participants (40F/25M) with an average age of 76 ± 9
 133 years were also included. Within this control group, 163 staggering events were captured.
 134

135 Inclusion/Exclusion Criteria

136 The study included adults (≥ 18 years) with physician-diagnosed PD and controls without PD.
 137 Exclusion criteria encompassed conditions that could prevent safe task performance, including
 138 non-neurological disorders significantly affecting speech, vision, or balance, cognitive or
 139 psychological impairments, or pregnancy. These criteria allowed for the inclusion of
 140 participants with a wide variety of typical and mildly pathological movement characteristics
 141 within the control group (e.g., individuals with essential tremor were not excluded, as we
 142 aimed to avoid spectrum bias by having only controls).
 143

144 Consent and Ethics

145 Written informed consent was obtained from all study participants, including permission to
146 publish de-identified data openly. These data were collected under the Johns Hopkins School
147 of Medicine (JHM) IRB00234370 with an IRB Authorization Agreement (95-CDRH-2021-12-18)
148 and VA IRB01702255 with an IRB Authorization Agreement (2023-CDRH-095), for which FDA
149 is the relying institution.

150

151 Sensor Systems

152 The following sensor systems were used during the collection of data for this study. All study
153 sites used the same acquisition software versions to maintain consistency.

154

155 *Movella/Xsens MTw Awinda Inertial Measurement Units (IMUs):* MTw Awinda is a wireless
156 human motion tracking sensor suitable for real-time applications. Each sensor was 47 x 30 x
157 13 mm. Each sensor communicated wirelessly with one of two base stations. Each base station
158 was connected to a computer running the MT Manager 2019.1.1 acquisition software.

159

160 *Moticon OpenGo Sensor Insoles:* Moticon OpenGo is a wireless in-shoe system with
161 Bluetooth™- and Wi-Fi-enabled electronics. Each sensor insole is designed to be used inside a
162 shoe and measures plantar pressure distribution at the sole of foot with sixteen distributed
163 pressure sensors. An on-board sensor at the center of the insole measures acceleration and
164 rotation of the foot along three axes. The insoles slip into the shoes of the participant and a
165 mobile Android application (OpenGo Mobile App Version 03.11.00) was used to control data
166 acquisition.

167

168 *ProtoKinetics Pressure Walkway:* The 16-foot x 2-foot ProtoKinetics Zeno™ Walkway Gait
169 Analysis System detects pressure data and uses that data to define foot contacts. The walkway
170 was equipped with over 18,000 individual sensors that each registered 16 different activation
171 levels. Each sensor was 0.5" x 0.5" (1.27 cm x 1.27 cm). The ProtoKinetics Movement Analysis
172 Software (PKMAS) 6.00c3 was used for data acquisition and raw data export.

173

174 *GoPro HERO10 Black Video Camera:* Two GoPro cameras were used in this study to film the
175 participants from the frontal and sagittal planes as they completed each task. Each camera
176 recorded video at 50 fps with a resolution of 1080p using either the linear or wide camera
177 lens. The video compression was set to H.264 + HEVC. The cameras were triggered wirelessly
178 via remote control by experimenters during gait and balance tasks.

179

180 Experimental Protocol

181

182 *Study Sites*

183 Data were collected across four study sites: FDA Office of Science and Engineering Labs (FDA),
184 Johns Hopkins Outpatient Center (JHOC), Johns Hopkins Bayview Campus (Bayview), and the
185 VA Puget Sound Health Care System Center for Limb Loss and Mobility (VA-Seattle). Subject
186 identifiers specific to each study site were used to denote the location at which data were
187 collected. The prefix 'FHC' is associated with the FDA site, 'NLS' is associated with the JHOC
188 site, prefix 'HC' is associated with the Bayview site, and prefixes 'WPD' and 'WHC' are
189 associated with the VA-Seattle site. Staff at each study site were trained on the same
190 equipment and the same experimental protocol was used across all sites.

191

192 *Participant Preparation*

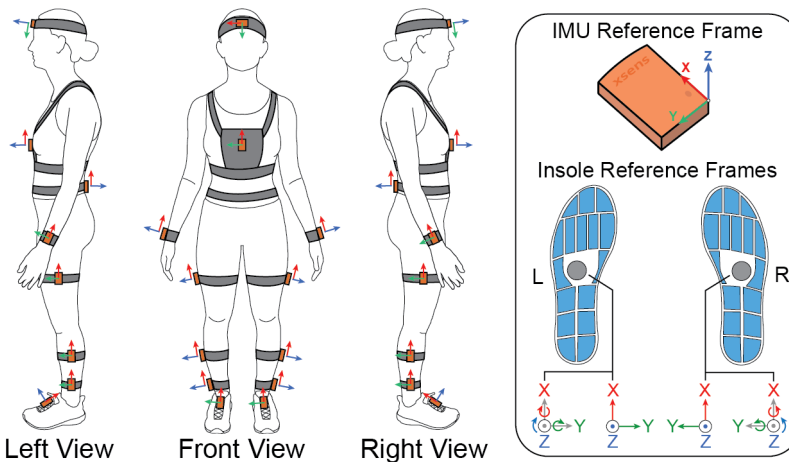
193 IMU sensors were attached to the participant's body at the following locations (see Figure 2):

194

- Mid forehead
- Xiphoid process of sternum

195

- 196 • L4/L5 of lower back
- 197 • Right/left wrist halfway between the ulnar and radial styloid processes
- 198 • Right/left lateral thigh, midway between greater trochanter of femur and head of
- 199 fibula
- 200 • Right/left lateral shank, midway between head of fibula and center of lateral malleolus
- 201 • Right/left ankle, just above the lateral malleolus
- 202 • Right/left dorsum of feet 50% of distance between 1st and 5th metatarsal bases and
- 203 50% of distance between distal end of 3rd metatarsal and center of anterior aspect of
- 204 talus.
- 205
- 206



207 Left View Front View Right View
 208 Figure 2: (Left) Locations and orientations of IMU sensors on the body and (Right) the
 209 reference frames for both the IMUs and insoles.

210
 211 The sensor locations and orientations were standardized and consistently applied across all
 212 participants and study sites. Each participant was fitted with Moticon OpenGo sensor insoles
 213 placed directly in the shoes.

214
 215 *Tasks*

216 Once prepped with all study equipment, participants were video-recorded and asked to
 217 perform specific tasks. Most tasks required the participants to make several passes along the
 218 16-foot pressure walkway, where a pass constituted walking one length of the walkway. A list
 219 of the tasks performed is below, with the abbreviations used in data file names in parentheses.
 220 Throughout these task descriptions, the term ‘mat’ is synonymous with ‘walkway’.

- 222 • SelfPace (SP): Participants started approximately 5 feet off the mat, walked at a self-
 223 selected pace across the mat and approximately 5 feet off the other side of the mat.
 224 Participants turned around off the mat and repeated this pass until a total of 4
 225 passes were made.
- 226 • HurriedPace (HP): Participants started approximately 5 feet off the mat, walked at a
 227 hurried pace across the mat and approximately 5 feet off the other side of the mat.
 228 Participants turned around off the mat and repeated this pass until a total of 4
 229 passes were made.
- 230 • SelfPace_mat (SPm): Participants repeated the SelfPace walking task as above but
 231 remained on the mat the entire time. Participants started on the mat, walked at a
 232 self-selected pace across the mat, stopped, and turned around at the other end of
 233 the mat. Participants repeated this until a total of 4 passes were made.

- 234 • HurriedPace_mat (HPm): Participants repeated the HurriedPace walking task as
235 above but remained on the mat the entire time. Participants started on the mat,
236 walked at a hurried pace across the mat, stopped, and turned around at the other
237 end of the mat. Participants repeated this until a total of 4 passes were made.
- 238 • SelfPace_matTURN (SPmT): Participants repeated the SelfPace_mat walking task as
239 above but *were instructed to alternate between left and right turns*. Participants
240 started on the mat, walked at a self-selected pace across the mat, stopped, and
241 turned around at the other end of the mat. Participants repeated this until a total of
242 5 passes were made so that two turns in each direction were captured.
- 243 • TandemGait (TG): Participants made 2 full passes on the mat while doing a tandem
244 gait walk (walking heel to toe). Participants remained on the mat the entire time and
245 turned around naturally after the first pass.
- 246 • Timed Up and Go (TUG): Participants started seated in a chair with armrests with
247 their back against the chair. The participant rose from the chair, walked to the other
248 end of the mat at a comfortable, self-selected pace, and turned at a line taped on
249 the mat 3m (9ft 10in) from the chair front. From there, they walked back to the
250 chair, and sat down. Participants were instructed to only use the armrests as
251 needed. Participants did NOT need to cross their arms when standing and sitting.
252 The *front* legs of the chair were placed on the starting line on the mat and turns
253 were made on the mat. The participant completed 3 trials of the TUG test within one
254 continuous recording.
- 255 • Balance (B): Participants started this task off the mat, typically near the midpoint of
256 the mat on the side. For each of 6 subtasks, the participant was asked to step onto
257 the mat, complete the balance subtask for 10 seconds, and then step off the mat.
- 258 ○ Standing with eyes open and feet shoulder width apart
259 ○ Standing with eyes closed and feet shoulder width apart
260 ○ Standing with eyes open and feet together
261 ○ Standing with eyes closed and feet together
262 ○ Standing with eyes open and right foot heel touching the left foot toe
263 ○ Standing with eyes open and left foot heel touching the right foot toe
264
- 265 • SelfPace_doorpat (SPdoorpat): A mock door frame approximately 3 ft x 8 ft was
266 placed across the center of the mat over a pattern of lines perpendicular to the
267 direction of travel. Participants started approximately 5 feet off the mat, walked at a
268 self-selected pace across the mat through the mock door frame and approximately 5
269 feet off the other side of the mat. Participants turned around off the mat and
270 repeated this pass until a total of 4 passes were made.
- 271
- 272 • FreeWalk (FW): Participants started approximately 5 feet off the mat, walked at a
273 self-selected pace across the mat and out of the session room into the hallway.
274 Participants proceeded to walk through a defined path within the research facility as
275 they navigated hallways, a small staircase (if available), and sitting down/rising from
276 a chair before returning to the session room. Each study site had a unique free walk
277 path, which is described in greater detail in the Synapse repository.
278

279 Tasks were completed sequentially in the order listed above.

280

281 Clinical Information and Assessments

282 In coordination with the clinical collaborator at each study site, information about medication
283 usage and clinical evaluations was obtained from the most recent patient visit when available.
284 If medical records were not available, medication usage and any other relevant medical history
285 were obtained by participant self-report. The MDS-UPDRS was used as the main
286 clinician/patient reported outcome measure. When possible, the MDS-UPDRS evaluation was
287 completed by a trained specialist the day of the session to ensure the evaluation corresponded
288 with the data collected. For instances where the MDS-UPDRS was not captured the day of data
289 collection, the most recent MDS-UPDRS evaluation was used with the date captured denoted
290 in the clinical spreadsheet.

291 Additional clinical information collected included:

- 292 • Time since first diagnosis
- 293 • Current medication type and dose
- 294 • Time since last medication dose
- 295 • Current physical/occupational therapy status and frequency of visits
- 296 • DBS (if patient has it); Bilateral vs unilateral; location of electrodes; date of surgery
- 297 • Basic demographics (age, gender, sex, height, weight, race)
- 298 • Modified Hoehn and Yahr scale score

299

300 Data Processing

301

302 *Movella/Xsens MTw Awinda Inertial Measurement Units (IMUs)*: During data collection, each
303 MTw IMU internally sampled data at 1000 Hz, applied factory calibration, and low pass filtered
304 at 184 Hz. An onboard step-down integration algorithm processed these samples, computing
305 orientation and velocity increments that were transmitted wirelessly at 100 Hz. The MT
306 Manager software was used for post-processing, applying the manufacturer's XKF3-hm
307 Kalman filter to generate 100 Hz outputs of acceleration, angular velocity, orientation, and
308 gravity-compensated (free) acceleration. For a detailed description of the system's signal
309 processing architecture, refer to the MTw Awinda whitepaper³⁴.

310

311 *ProtoKinetics Pressure Walkway*: After data collection, the ProtoKinetics software
312 automatically labels groups of active sensor cells as belonging to either the left or right foot.
313 Additional manual preprocessing of the walkway data was completed to ensure left and right
314 footfalls were correctly identified. Partial footfalls on the walkway were labeled as 'Other'
315 rather than left or right. After the footfall labeling was completed, a research staff member
316 conducted a review of the labeling to ensure no mistakes or mislabeled footfalls exist.

317

318 *Moticon OpenGo Sensor Insoles*: Sensor insole waveforms were reviewed visually. Although
319 infrequent, instances of desynchronization occurred during wireless communication between
320 the sensor insoles and the Moticon acquisition App. This resulted in asynchronous data
321 capture or prolonged recording from one insole even after the 'Stop' command was issued.
322 When such desynchronization was detected, the data from the insoles were adjusted—shifted
323 and/or trimmed—within the OpenGo Software to ensure proper alignment between both
324 insoles and to accurately synchronize the cessation of recording with the conclusion of the
325 data stream.

326

327 *GoPro HERO10 Black Video Camera*: Manual annotation of the participant's activity using the
328 video data was also completed. After the data collection session ended, the videos recorded
329 for each trial from the side and front views were combined and deidentified. The result was a
330 single video with both views shown side-by-side, where each view had the participant's face
331 cropped out. On some occasions, only a single video camera view was available. The primary
332 annotation process of the session videos used the Video Labeler application in MATLAB to
333 provide a frame-by-frame annotation of general and clinical events for the duration of each
334 task video, with clinical events being labeled concurrently with general events.

335 After the primary annotations were completed, a research staff member conducted a review
336 of the existing annotations to ensure no mistakes or mislabeled events existed. If any clinical
337 events were identified by the research staff during primary annotation or review, those events
338 were reviewed by at least one neurologist collaborator at Johns Hopkins or the VA to confirm
339 or reject the annotation of the clinical event.

340 Detailed definitions of the general and clinical gait events used in this dataset can be found in
341 the Data [Description – Video Annotation](#) section of the Wiki.. These definitions were provided
342 to ensure consistency across annotators.

343 Data Synchronization and Alignment

344 Each task included four main data sources: IMUs, sensor insoles, walkway, and video cameras.
345 The walkway served as the primary clock for all system synchronization and alignment. Upon
346 starting a recording in the PKMAS walkway software, a 3.3 V TTL square wave signal was sent
347 from the ProtoKinetics walkway interface box to the Sync In port of each MTw Awinda base
348 station. The MT Manager software for the MTw Awinda IMUs was configured to start
349 recording upon receipt of the TTL signal from the walkway. Synchronization between the
350 sensor insoles and walkway was achieved by setting the Moticon OpenGo App to produce an
351 audible beep during recording start/stop. The audio signal from the phone was converted to
352 a 3.3 V square wave via custom electronics from Moticon, which was read into the
353 ProtoKinetics walkway interface box and subsequently the walkway acquisition software.
354 Walkway, sensor insole, and IMU systems were all set to record data at 100 Hz.

355 Two video cameras set up perpendicular and parallel to the direction of participant travel
356 collected data to facilitate frame-by-frame annotation of participant movement. A green light
357 connected to the ProtoKinetics walkway interface box was programmed to turned on for five
358 seconds when recording was initiated. The video cameras started recording prior to the
359 walkway to ensure each video camera captured the green light turning on. The green light was
360 then used to align the two video streams to each other as well as the walkway. For each task,
361 the general order of operations for collecting synchronized data were as follows:

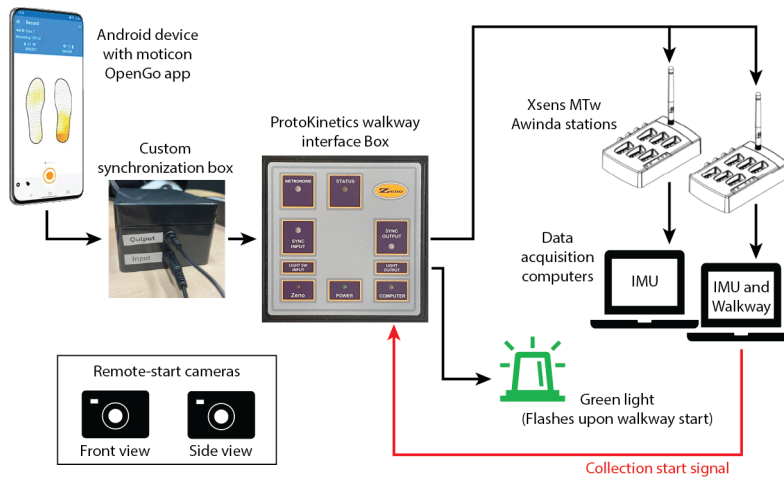
- 362 1. Prime IMU acquisition software on computers to accept a sync signal.
- 363 2. Start the video cameras using a remote control.
- 364 3. Start the trial in walkway acquisition software (which simultaneously starts
365 recording IMU data and triggers green light).
- 366 4. Start recording sensor insole data in phone app.

367 After a participant completed a task, the opposite order of operations was taken to stop
368 recording data:

- 369 1. Stop recording sensor insole data in phone app.
- 370 2. Stop recording the trial walkway acquisition software (which simultaneously
371 stops the IMU recording).

372 3. Stop the video cameras using the remote control.

373 Figure 3 shows a schematic of the overall hardware synchronization approach.



374

375 Figure 3: Synchronization approach across the walkway, IMU, sensor insole, and camera
376 systems.

377 MATLAB scripts were written to read in the raw data from each sensor modality, align the data
378 in time, and export to the MAT and CSV file formats. A brief description of the subroutines
379 implemented to generate these aligned MAT and CSV files follows.

380 *Align walkway and IMU data* → Since the IMU system received a hardwired sync signal from
381 the walkway system (primary clock) at the beginning of recording, the data from the IMU
382 sensors and walkway were concatenated without further alignment. Occasionally, the IMU
383 sensor data contained more or fewer time points than the walkway data. When shorter,
384 padding of NaN was added to the end of the IMU data. When longer, the data that extended
385 past the end of the length of the walkway data were discarded. Benchtop characterization
386 confirmed the walkway and IMU data were aligned within 2 frames at 100 Hz.

387 *Align walkway and sensor insole data* → Custom electronics enabled a square-wave sync signal
388 to be sent from the sensor insole system to the walkway system that denoted the start and
389 stop of sensor insole recording. The end of the square wave (denoting sensor insole recording
390 stop) was used as the alignment point as it was more consistent. The last frame of the sensor
391 insole data was aligned with the last transition point of the sync square-wave in the walkway
392 data. Padding of the beginning and end of sensor insole data with NaN was added to keep
393 arrays the same size.

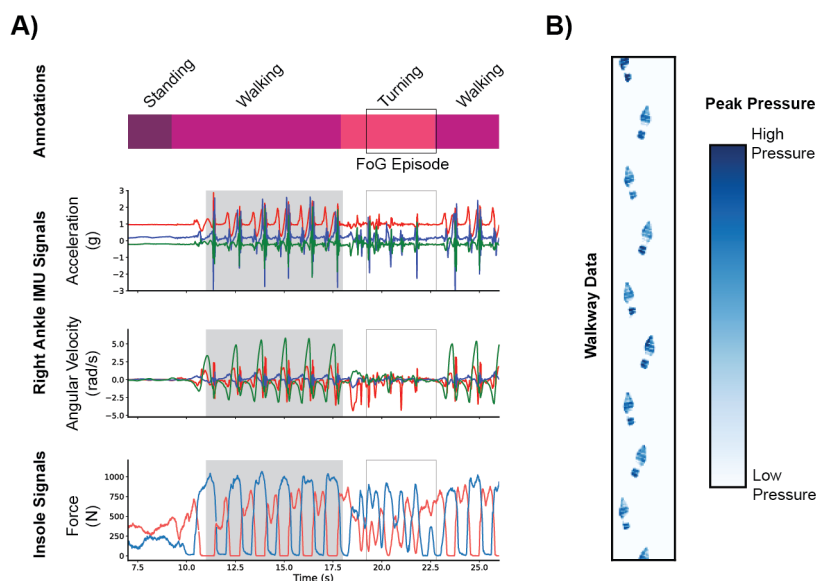
394 During pilot work, visual inspection of the aligned walkway and sensor insole data revealed a
395 consistent time delay in the sensor insole signals relative to the walkway signals. We attributed
396 this delay to the non-deterministic nature of the Android operating system running the
397 Moticon Open Go App, which led to lag between stopping the sensor insole recording and the
398 sync signal being sent to the walkway acquisition software. Benchtop experiments were
399 conducted to characterize this delay across all study sites, sensor insole sizes, and data
400 collection sessions. Results indicated site-specific, consistent delays that were independent of
401 insole size and data collection session. These site-specific corrections were applied to the
402 sensor insole data during initial import and alignment processing to address the constant bias
403 component of the delay. However, cross-correlation analyses indicated that the delay

404 between walkway and insole waveforms has both a constant component and small random
405 components. Dataset users should proceed with caution when interpreting data that depends
406 on precise alignment of the sensor insoles with the walkway and IMU systems. For relevant
407 analyses, users may consider first aligning the insole total force waveforms to the walkway
408 foot pressure waveforms using cross-correlation or other methods. More details on the
409 system characterization can be found in the [Data Description – Sensor insole sync
410 characterization](#) section of the Wiki.

411

412 *Align walkway and video annotation data* → The annotation file exported from MATLAB's
413 Video Labeler App was formatted such that the string variable for each annotation event at a
414 given time point was listed under a 'GeneralEvent' or 'ClinicalEvent' column. The annotation
415 data were resampled from 50 Hz to 100 Hz. Alignment between the walkway and video
416 annotation data was achieved using the green light connected to the walkway system. A
417 MATLAB script automatically detected the first video frame in which the green light turned on
418 and used this to align the video annotation data with the start of the walkway data. All video
419 data before the green light turned on were discarded.

420 The output of these processes is a set of aligned time-series data across each data collection
421 source (Figure 4A), as well as the spatial information collected by the walkway (Figure 4B).



422

423 Figure 4: (A) Time series data from a participant with Parkinson's disease, during which a
424 freezing of gait (FoG) episode occurred. The data include annotation of activities captured via
425 camera, 3-DOF accelerometer and gyroscope data from an IMU, and force data from a pair
426 of sensorized insoles. The portion of the data with a gray background corresponds to the
427 spatial information in part b. (B) Spatial data from a participant with Parkinson's disease,
428 during one pass across a sensorized walkway. Blue regions indicate areas of walkway
429 activation, with darker colors corresponding to higher forces.

430

431 Data Records

432 Detailed descriptions of the experimental set-up, data collection approach, processing, and
433 data quality control for the WearGait-PD dataset are provided in the [main Wiki page](#) on the
434 Synapse platform (SAGE Bionetworks) The [Data Access](#) tab within the Wiki provides further
435 instructions on how to gain access to the data.

436

437 There are two main parts to the data: the clinical/demographic information and the sensor
438 data. All clinical and demographic information is available for all PD patients and control
439 participants and can be found in CSV spreadsheets. In these spreadsheets, each row
440 constitutes one participant, with the first column listing the participant ID. All participant IDs
441 are alphanumeric, with the letters denoting the study site at which the participant was
442 recorded. All columns are labelled with the variable name. The clinical and demographic
443 information contained within these spreadsheets was briefly described in the Methods –
444 Clinical Information and Assessments section of this paper. When available, the response to
445 each question of the MDS-UPDRS was included, allowing for maximum flexibility in use of the
446 scores.

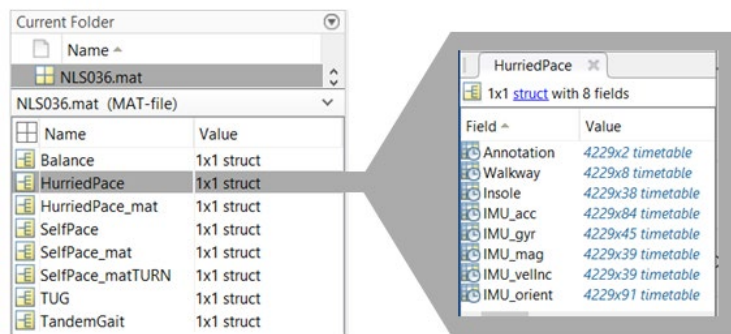
447

448 Formatting scripts were written to align all wearables and reference sensor data and save data
449 as MAT files and comma-separated value (CSV) files. For a given participant, there are 8 CSV
450 files (1 file/task) that contain the annotation events, walkway data, insole data, and IMU data.
451 The CSV files follow the naming convention of <SubjectID>_<Task Name>.csv (see Methods –
452 Tasks section of this paper for task descriptions). These data can be found in a folder labelled
453 'CSV files' in the Synapse repository. Of note, the use of the word 'mat' in the task names
454 refers to participants performing a task entirely on the walkway mat.

455

456 The same data for each participant is also represented as one MATLAB MAT file that contains
457 8 structure variables for each task, with each task containing 8 fields (Figure 5).

458



459

460 Figure 5: Sample MAT file structure for one participant.

461

462 A brief description of the fields within each task structure variable is below:

- 463 • Annotation: tx2 timetable containing video annotation information
- 464 • Walkway: tx8 timetable containing sensor and contact walkway data
- 465 • Insole: tx38 timetable containing insole data
- 466 • IMU_acc: tx84 timetable containing Acc and FreeAcc variables for each IMU
467 and insole
- 468 • IMU_gyr: tx45 timetable containing gyroscope data for each IMU and insole
- 469 • IMU_mag: tx39 timetable containing magnetometer data for each IMU
- 470 • IMU_yellnc: tx39 timetable containing Δv data for each IMU
- 471 • IMU_orient: tx91 timetable containing Δq and roll, pitch, yaw data for each IMU

472

473 For missing numeric data due to connectivity issues, sensor malfunctions, or lack of
474 instrumentation for a given participant/task, NaN values were used. The data alignment
475 process also used NaN values at the beginning and/or end to pad the time series, so NaN
476 values in these locations are expected. Sensors not available during a given data collection
477 were excluded from CSV output files but represented as NaN columns in the MAT files. For the

478 GeneralEvent and ClinicalEvent variables containing frame-by-frame annotations, the string
479 “unlabeled” was used to represent portions of a trial that did not match any annotation
480 category (e.g., shuffling at the end of a trial instead of standing). In total, there were 346
481 variables associated with each task. A description of all the variables as well as reference
482 coordinate systems for the IMUs, walkway, and sensor insoles is provided in the [Data](#)
483 [Description](#) section of the Wiki.
484

485 **Technical Validation**

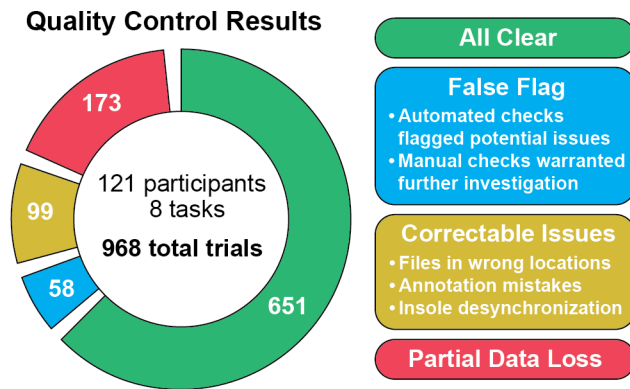
486 Data quality is of the utmost importance in the curation of an open-access dataset. A quality
487 assurance process was implemented to ensure consistent and accurate data collection.
488 Specifically, a core experimental protocol was developed and reviewed by each researcher,
489 including cross-site review of how the protocol was executed by various researchers. This
490 protocol detailed the steps involved in system set-up, participant preparation, data collection,
491 and data processing and export. All files were renamed to include the participant ID and task
492 abbreviation immediately after data collection to minimize the risk of file loss due to incorrect
493 folder allocation. For those modalities that required more manual processing (i.e., walkway
494 data and video annotation data), a second researcher reviewed the primary processing
495 completed by the first researcher to ensure proper footfall identification and conformance
496 with the agreed upon annotation event definitions. An Excel spreadsheet accessible to all
497 researchers was also used to track the progression of data processing, from initial collection
498 to MAT and CSV file generation.
499

500 After MAT and CSV files were created, a data quality control process on those files was
501 implemented through a series of automatic and manual data checks that involved visual
502 inspection of data streams by a researcher to identify data cleanliness and validity issues.
503

504 Specifically, this process identified issues related to:

- 505 • Processing and inclusion of all tasks in the MAT and CSV files
- 506 • Integrity of individual IMU sensor data, including the identification of sensor values
507 outside of expected ranges
- 508 • Unexpected missing columns of data
- 509 • Unexpected large gaps in data
- 510 • Expectations around the variable type for a given data variable (e.g., annotations
511 contain no numeric data, IMU data does not contain any errant non-numeric data)
- 512 • Video annotation events and expectations surrounding the inclusion of specific events
513 for specific tasks
- 514 • Alignment of all data, with a particular focus on walkway pressure and sensor insole
515 force data
516

517 If an issue was identified, a second researcher was assigned to review and address the issue.
518 Once an issue was addressed, the relevant files were put through the data quality control
519 process again to confirm that all issues were resolved before the final MAT file and final set of
520 CSVs files were generated and included in the dataset.
521



522

523

Figure 6: Quality control results for each trial, applied prior to inclusion of data to the WearGait-PD dataset. A trial is defined by a single task performed by a single participant. Category values add up to greater than 968 as the 'False Flag', 'Correctable Issues', and 'Partial Data Loss' categories were not mutually exclusive.

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

Of the 968 total trials that comprised the initial release of data, the majority of files passed all quality checks (651 trials) and another portion were flagged as potential issues but required no corrective action (58 trials) (Figure 6). Of the remaining trials with confirmed issues, 99 trials contained fixable issues such as the placement of files in the wrong location, missed annotation frames, or insole desynchronization. Each of these passed subsequent quality checks. The remaining 173 files had significant data loss (data dropout of 7 frames or greater, or loss of a sensor), either from connectivity or battery issues. While these data could not be recovered, the remaining data present in these trials were checked to confirm they were clean and valid. Overall, 82% of trials were considered complete, while the remaining 18% of trials passed all quality control checks with the exception that they were missing some data from either the insoles or IMUs.

540

Usage Notes

541

542

543

544

545

546

547

548

549

550

551

552

553

554

To access the data on Synapse, users must first register with Synapse.org. Users register by providing a name and valid e-mail address and agreeing to the Synapse Terms and Conditions of Use and reviewing the Privacy Policy and Code of Conduct. As part of the registration process, users must agree to each of the terms of the Synapse Pledge, recapitulating the themes of the Synapse Terms and Conditions of Use, Privacy Policy, and Code of Conduct. Specifically, this pledge requires users to confirm they will (1) adhere to the community standards of inclusion and respect, (2) adhere to all conditions and data use limitations, (3) act ethically and responsibly, (4) use appropriate physical, technical, and administrative measures to keep data secure and protect participant's privacy, (5) support open access best practices, (6) credit research participants and all data sources, (7) confirm the data will not be used for marketing and/or advertising, and (8) report suspected data breaches and/or misuse to the Synapse team.

553

554

555

Code Availability

556

557

558

559

Using MATLAB 2023a, we developed code to implement the quality control process detailed in the Technical Validation section. This code is available in the 'Code' folder under 'Files' on the Synapse repository.

559

560

Acknowledgements

561

562

The authors would like to thank all study participants for their enthusiastic participation. We'd also like to thank the FDA research fellows (Ms. Isabella Zuccaroli, Mr. Bryan Sabogal, Mr. Brian

563 Nogh) who have assisted in data collection and processing, as well as former FDA staff
564 members (Dr. Edward Nyman) who contributed to the development of the experimental
565 protocol. The authors would also like to thank research coordinators from Johns Hopkins
566 Bayview Medical Center (Ms. Jackie Langdon, Ms. Meredith Dobrosielski, and Ms. Crystal
567 Szczesny) for their assistance with subject recruitment and consent, and Dr. Peter Abadir for
568 his thoughtful insight into the experimental protocol. In this work, researchers at the Johns
569 Hopkins University were partially funded by the Consolidated Anti-Aging Foundation.
570

571 **Author contributions**

572 Conceptualization (KK, MC, AB), Methodology (All authors), Software (MC, MG, AA, KK),
573 Validation (AA, MG, KK, MC), Formal Analysis (KK, AA, MG, DE), Investigation (KK, AA, DE, MG,
574 NK, SW, MC, CM, BM, SZ, KM, AB, EM), Resources (KK, BM, AB), Data Curation (KK, AA, DE,
575 MG, MC, NK, SW, SZ, AB, KM, EM), Writing – Original Draft (AA, DE, MG, KK), Writing – Review
576 and Editing (All authors), Visualization (KK, AA, MG), Supervision (KK, BM, AB), Project
577 Administration (KK, BM, AB, CM), Funding Acquisition (KK, BM, AA, LM, AB)
578

579 **Competing interests**

580 The authors declare no competing interests.
581

582 **Disclaimer:**

583 The mention of commercial products, their sources, or their use in connection with material
584 reported herein is not to be construed as either an actual or implied endorsement of such
585 products by the Department of Health and Human Services.
586

587 **References**

- 588 1 DeMaagd, G. & Philip, A. Parkinson's disease and its management: part 1: disease
589 entity, risk factors, pathophysiology, clinical presentation, and diagnosis. *Pharmacy
590 and therapeutics* **40**, 504 (2015).
- 591 2 Duncan, R. P. & Earhart, G. M. Measuring participation in individuals with Parkinson
592 disease: relationships with disease severity, quality of life, and mobility. *Disability
593 and rehabilitation* **33**, 1440-1446 (2011).
- 594 3 Santos García, D. *et al.* Non-motor symptoms burden, mood, and gait problems are
595 the most significant factors contributing to a poor quality of life in non-demented
596 Parkinson's disease patients: Results from the COPPADIS Study Cohort. *Parkinsonism
597 Relat Disord* **66**, 151-157 (2019). <https://doi.org/10.1016/j.parkreldis.2019.07.031>
- 598 4 Schlachetzki, J. C. *et al.* Wearable sensors objectively measure gait parameters in
599 Parkinson's disease. *PloS one* **12**, e0183989 (2017).
- 600 5 Morris, R. *et al.* Validity of Mobility Lab (version 2) for gait assessment in young
601 adults, older adults and Parkinson's disease. *Physiol Meas* **40**, 095003 (2019).
602 <https://doi.org/10.1088/1361-6579/ab4023>
- 603 6 Salis, F. *et al.* A multi-sensor wearable system for the assessment of diseased gait in
604 real-world conditions. *Front Bioeng Biotechnol* **11**, 1143248 (2023).
605 <https://doi.org/10.3389/fbioe.2023.1143248>
- 606 7 Sotirakis, C. *et al.* Identification of motor progression in Parkinson's disease using
607 wearable sensors and machine learning. *npj Parkinson's Disease* **9**, 142 (2023).
- 608 8 Goetz, C. G. *et al.* Movement Disorder Society-sponsored revision of the Unified
609 Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric
610 testing results. *Mov Disord* **23**, 2129-2170 (2008).
611 <https://doi.org/10.1002/mds.22340>
- 612 9 Evers, L. J. W., Krijthe, J. H., Meinders, M. J., Bloem, B. R. & Heskes, T. M. Measuring
613 Parkinson's disease over time: The real-world within-subject reliability of the MDS-
614 UPDRS. *Mov Disord* **34**, 1480-1487 (2019). <https://doi.org/10.1002/mds.27790>

- 615 10 Regnault, A. *et al.* Does the MDS-UPDRS provide the precision to assess progression
616 in early Parkinson's disease? Learnings from the Parkinson's progression marker
617 initiative cohort. *J Neurol* **266**, 1927-1936 (2019). [https://doi.org:10.1007/s00415-](https://doi.org:10.1007/s00415-019-09348-3)
618 [019-09348-3](https://doi.org:10.1007/s00415-019-09348-3)
- 619 11 Guerra, A., D'Onofrio, V., Ferreri, F., Bologna, M. & Antonini, A. Objective
620 measurement versus clinician-based assessment for Parkinson's disease. *Expert Rev*
621 *Neurother* **23**, 689-702 (2023). <https://doi.org:10.1080/14737175.2023.2229954>
- 622 12 Shah, V. V. *et al.* Digital biomarkers of mobility in Parkinson's disease during daily
623 living. *Journal of Parkinson's disease* **10**, 1099-1111 (2020).
- 624 13 Shah, V. V. *et al.* Gait and turning characteristics from daily life increase ability to
625 predict future falls in people with Parkinson's disease. *Frontiers in neurology* **14**,
626 1096401 (2023).
- 627 14 Servais, L. *et al.* Stride Velocity 95th Centile: Insights into Gaining Regulatory
628 Qualification of the First Wearable-Derived Digital Endpoint for use in Duchenne
629 Muscular Dystrophy Trials. *J Neuromuscul Dis* **9**, 335-346 (2022).
630 <https://doi.org:10.3233/JND-210743>
- 631 15 Klucken, J., Krüger, R., Schmidt, P. & Bloem, B. R. Management of Parkinson's disease
632 20 years from now: towards digital health pathways. *Journal of Parkinson's disease* **8**,
633 S85-S94 (2018).
- 634 16 Mughal, H., Javed, A. R., Rizwan, M., Almadhor, A. S. & Kryvinska, N. Parkinson's
635 disease management via wearable sensors: a systematic review. *IEEE Access* **10**,
636 35219-35237 (2022).
- 637 17 Rovini, E., Maremmani, C. & Cavallo, F. How wearable sensors can support
638 Parkinson's disease diagnosis and treatment: a systematic review. *Frontiers in*
639 *neuroscience* **11**, 288959 (2017).
- 640 18 Del Din, S., Kirk, C., Yarnall, A. J., Rochester, L. & Hausdorff, J. M. Body-worn sensors
641 for remote monitoring of Parkinson's disease motor symptoms: vision, state of the
642 art, and challenges ahead. *Journal of Parkinson's disease* **11**, S35-S47 (2021).
- 643 19 Fasano, A. & Mancini, M. Wearable-based mobility monitoring: the long road ahead.
644 *Lancet Neurol* **19**, 378-379 (2020). [https://doi.org:10.1016/S1474-4422\(20\)30033-8](https://doi.org:10.1016/S1474-4422(20)30033-8)
- 645 20 Espay, A. J. *et al.* A roadmap for implementation of patient-centered digital outcome
646 measures in Parkinson's disease obtained using mobile health technologies. *Mov*
647 *Disord* **34**, 657-663 (2019). <https://doi.org:10.1002/mds.27671>
- 648 21 Rochester, L. *et al.* A Roadmap to Inform Development, Validation and Approval of
649 Digital Mobility Outcomes: The Mobilise-D Approach. *Digit Biomark* **4**, 13-27 (2020).
650 <https://doi.org:10.1159/000512513>
- 651 22 US Food and Drug Administration - Center for Devices and Radiological Health.
652 *NeuroRPM K221772 Approval Letter*,
653 <https://www.accessdata.fda.gov/cdrh_docs/pdf22/K221772.pdf> (2023).
- 654 23 US Food and Drug Administration - Center for Devices and Radiological Health. *Rune*
655 *Labs Kinematics System K213519 Approval Letter*,
656 <https://www.accessdata.fda.gov/cdrh_docs/pdf21/K213519.pdf> (2022).
- 657 24 US Food and Drug Administration. *Digital Health Technologies for Remote Data*
658 *Acquisition in Clinical Investigations - Guidance for Industry, Investigators, and Other*
659 *Stakeholders*, <<https://www.fda.gov/media/155022/download>> (2023).
- 660 25 Goldsack, J. C. *et al.* Verification, analytical validation, and clinical validation (V3): the
661 foundation of determining fit-for-purpose for Biometric Monitoring Technologies
662 (BioMeTs). *NPJ Digit Med* **3**, 55 (2020). <https://doi.org:10.1038/s41746-020-0260-4>
- 663 26 Bot, B. M. *et al.* The mPower study, Parkinson disease mobile data collected using
664 ResearchKit. *Scientific data* **3**, 1-9 (2016).

665 27 de Oliveira, C. E. N. *et al.* A public data set with ground reaction forces of human
666 balance in individuals with Parkinson's disease. *Frontiers in neuroscience* **16**, 865882
667 (2022).

668 28 Gilmore, G., Gouelle, A., Adamson, M. B., Pieterman, M. & Jog, M. Forward and
669 backward walking in Parkinson disease: a factor analysis. *Gait & posture* **74**, 14-19
670 (2019).

671 29 Ribeiro De Souza, C. *et al.* A public data set of videos, inertial measurement unit, and
672 clinical scales of freezing of gait in individuals with parkinson's disease during a
673 turning-in-place task. *Frontiers in Neuroscience* **16**, 832463 (2022).

674 30 Vergara-Diaz, G. *et al.* Limb and trunk accelerometer data collected with wearable
675 sensors from subjects with Parkinson's disease. *Scientific Data* **8**, 47 (2021).

676 31 Zhang, W. *et al.* Multimodal data for the detection of freezing of gait in Parkinson's
677 disease. *Scientific data* **9**, 606 (2022).

678 32 Varghese, J. *et al.* Machine Learning in the Parkinson's disease smartwatch (PADS)
679 dataset. *NPJ Parkinsons Dis* **10**, 9 (2024). [https://doi.org/10.1038/s41531-023-00625-](https://doi.org/10.1038/s41531-023-00625-7)
680 [7](https://doi.org/10.1038/s41531-023-00625-7)

681 33 Morgan, C. *et al.* A multimodal dataset of real world mobility activities in Parkinson's
682 disease. *Sci Data* **10**, 918 (2023). <https://doi.org/10.1038/s41597-023-02663-5>

683 34 Paulich, M., Schepers, M., Rudigkeit, N. & Bellusci, G. Xsens MTw Awinda: Miniature
684 Wireless Inertial-Magnetic Motion Tracker for Highly Accurate 3D Kinematic
685 Applications. *Xsens Technologies White Paper*

686