Wearable devices can identify Parkinson's disease up to 7 years before clinical diagnosis

Ann-Kathrin Schalkamp[1], Kathryn J Peall[2], Neil A Harrison[2,3], Cynthia Sandor[1*]

[1] Psychological Medicine and Clinical Neuroscience, UK Dementia Research Institute, Cardiff University, United Kingdom

[2] Neuroscience and Mental Health Innovation Institute, Division of Psychological Medicine and Clinical Neurosciences, Cardiff, United Kingdom

[3] Cardiff University Brain Research Imaging Centre (CUBRIC), Cardiff, United Kingdom

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative movement disorder with a long latent phase and no currently existing disease-modifying treatments. Reliable predictive biomarkers that could transform efforts to develop neuroprotective treatments remain to be identified. Using UK Biobank, we investigated the predictive value of accelerometry in identifying prodromal PD in the general population and compared this digital biomarker to models based on genetics, lifestyle, blood biochemistry, and prodromal symptoms data. Machine learning models trained using accelerometry data achieved high test performance in distinguishing both clinically diagnosed PD (AUROC: 0.83$\pm$0.03, AUPRC: 0.51$\pm$0.06) and prodromal PD up to seven years pre-diagnosis (AUROC: 0.81$\pm$0.03, AUPRC: 0.37$\pm$0.05) from the general population, and outperformed all other modalities tested. Accelerometry is a potentially important, low-cost screening tool for identifying people at risk of developing PD and identifying subjects for clinical trials of neuroprotective treatments.

Keywords

Accelerometry, Parkinson's disease, Prodromal, Sleep, General Population

Introduction

Most patients diagnosed with Parkinson's disease (PD) 50-70% of nigral dopaminergic neurons will have degenerated by the time the hallmark motor symptoms manifest and a clinical diagnosis is made (Fearnley & Lees, 1991). There remains a need to identify cheap, reliable, easily accessible, and sensitive biomarkers to detect early pathological changes, with success in this field likely to be transformative in identifying suitable participants for involvement in clinical trials of potential neuroprotective therapeutics.

It is well recognised that at the point of a clinical motor diagnosis of PD, multiple prodromal symptoms could have been present for several years including Rapid-Eye-Movement Sleep Behaviour Disorder (RBD), constipation, hyposmia, depression, anxiety, and excessive daytime somnolence with urinary dysfunction, orthostatic hypotension, sub-threshold motor symptoms, and abnormal dopaminergic molecular brain imaging more recently added to the criteria for prodromal PD (Heinzel et al., 2019; Postuma & Berg, 2016, 2019). Multiple previous studies have examined these symptoms, together with fluid, tissue, and imaging biomarkers in determining their sensitivity in identifying prodromal PD (Fereshtehnejad et al., 2017; Hustad & Aasly, 2020). However, the absence of multi-modal models, which combine the predictive power of multiple data sources, has limited this work (Postuma & Berg, 2019). Furthermore, most studies have tended to compare those with prodromal PD to control cohorts lacking any comorbidity limiting the translational validity and real-world applicability of the findings. More research is thus needed to understand the specificity and effective role of prodromal markers in the general population.

To facilitate the identification of clinical change, a key aim would be continuous or regular monitoring for sustained periods to obtain robust estimates and detect subtle changes at the earliest possible opportunity. However, such monitoring cannot be achieved through clinical assessments given the limitations of time, cost, accessibility, and sensitivity (Dorsey et al., 2020). By contrast digital sensors passively collect data continuously in real-world settings

without added cost or effort (Brognara et al., 2019). Preliminary analyses of such digital sensors collecting acceleration and heart rate data have demonstrated the potential for distinguishing those with a clinical diagnosis of PD from those without, as well as added capabilities of monitoring motor progression, and describing sleeping behaviour (Johansson et al., 2018; Schlachetzki et al., 2017; Shah et al., 2020). However, these quantitative motor measures remain largely understudied (Heinzel et al., 2019) with studies often limited by small sample sizes (Del Din et al., 2019), or restricted to analysis only after clinical PD diagnosis (Mirelman et al., 2016; Williamson et al., 2021). Detecting early movement alterations using digital sensors to identify diseases before the clinical diagnosis is a largely unexplored field with much potential for application in the general population.

This study uses the large, prospective population-based cohort recruited to the UK Biobank (UKBB) to investigate the specificity and real-world applicability of accelerometry as a prodromal marker for PD. Since 2006 data has been collected for >500,000 individuals aged 40-69 years with ongoing passive follow-up of clinical status through routinely collected clinical data, including primary care, hospital admissions, and death registry details (Bycroft et al., 2018). Accelerometry data were collected for a subset of this cohort (N = 103,712, collected 2013-2015) (Doherty et al., 2017). Using these data, we have sought to determine whether acceleration data can serve as a prodromal marker for PD, examining its specificity by comparing data from those diagnosed with PD to both unaffected controls as well as related disorders, namely neurodegenerative disorders, movement disorders, and comorbid clinical disorders. We compared the performance of the accelerometry model to models using other modalities, such as genetics, blood, lifestyle, prodromal symptoms and accelerometry to determine the best data sources to identify prodromal PD.

Results

The UK Biobank provides a large and increasing cohort of individuals diagnosed with Parkinson's disease

Clinical diagnoses within the UKBB currently derive from multiple sources including self-reported symptoms and diagnoses, hospital records, death records, and primary care data; however, there is no clinical diagnostic validation. To ensure we did not over- or underestimate the true number of PD cases, we compared the prevalence in our cohort to that expected based on 2015 UK population statistics (Parkinson's UK, 2017).

At baseline data collection 967 subjects had previously received a diagnosis of PD with an additional 2869 subjects diagnosed by March 2021, meaning a total of around 0.76% of the UKBB participants had been diagnosed with PD at that time point.

Annual incidence of PD increased (Supplementary Figure 1A) as the cohort aged, with the highest number of diagnoses made amongst those aged 70-80 years (Supplementary Figure 1B). Overall, the observed number of PD cases was slightly lower than that expected based on 2015 UK population statistics (2015 expected: 2252.61, 2015 observed: 1984; Supplementary Figure 1C and Figure 1), with 5255 expected by 2030 assuming no further deaths were to occur. Although the PD diagnosis in the UKBB is not curated by specialists, as is often done in disease-specific, deeply phenotyped cohorts, and the date of diagnosis may lack accuracy due to incomplete coverage of the electronic health records in terms of time and clinics, in general we found the prevalence and incidence in the UKBB cohort to closely resemble that expected from such a population.

Sensor-measured acceleration is reduced several years prior to Parkinson's disease diagnosis

We compared average acceleration over a 7-day period for each hour of the day between the diagnostic groups. At the time of or within two years after accelerometry data collection, 273 subjects were diagnosed with PD (mean years since diagnosis: 5.04±6.37; Supplementary Table 1, Figure 2A). An additional 196 individuals received a new PD diagnosis more than two years after accelerometer data collection (mean years to diagnosis: 4.33±1.30, Figure 2A). The prodromal group was significantly older than the diagnosed group (t-statistic = 3.18, p-value = 1.6×10^{-3}) and were therefore not directly compared. We randomly sampled age- and sex-matched unaffected controls for each subject with PD. Prodromal and diagnosed PD cases both showed a significantly smaller acceleration profile over all waking hours (from 7am: p-value = 3.84×10^{-5}, p-value = 6.12×10^{-8} to 11pm: p-value = 5.92×10^{-6}, p-value = 1.27×10^{-8}) than age- and sex-matched unaffected controls (Figure 2B, Supplementary Table 3). No differences in average acceleration between the groups was found at night (from 11pm: p-value = 0.001, p-value = 0.056 to 7am: p-value = 0.016, p-value = 0.022).

Reduced sensor-detected acceleration at the prodromal stage is unique to Parkinson's disease

As physical activity varies between individuals irrespective of health status, we explored whether the observed reduction in acceleration was unique to PD or whether it could also be observed in other clinical disorders, in particular other neurodegenerative and movement disorders (Supplementary Table 1). We calculated residual average acceleration corrected for age, sex, and BMI as these showed significant association with average acceleration (Supplementary Table 4). Secondly, as anticipated, several subjects were diagnosed with multiple comorbidities (Supplementary Figure 2): for example, 4.7% of those diagnosed with 'AllCauseDementia' were also diagnosed with PD and 63% with depression. In this setting, those with a comorbid PD diagnosis were excluded. As depression is a potential prodromal

marker of PD (Heinzel et al., 2019), individuals with a co-diagnosis of depression were also removed. Analyses without these adjustments can be found in Supplementary Figure 3.

There was a significant reduction in average acceleration in diagnosed PD and prodromal PD compared to unaffected controls (t-statistic = 10.194, p-value = 2.42×10^{-22} and t-statistic = 10.492, p-value = 2.65×10^{-23}, respectively) (Figure 2C), but no significant difference between individuals with prodromal and diagnosed PD (p-value = 0.87). The reduction of acceleration in prodromal PD, consistent with the level observed in the diagnosed PD group, suggests that impaired locomotion is detectable several years prior to a clinical diagnosis of PD. Of those investigated, 'Depression' was the only other disorder to show a reduction in acceleration following diagnosis, and none of the disorders investigated were found to have a reduction in acceleration prior to diagnosis that persisted following clinical diagnosis, as was observed for PD (Figure 2C). In addition, subjects included in the 'AllCauseParkinsonism' group demonstrated the same between-group differences as PD, but this effect was lost with exclusion of those diagnosed with PD (Figure 2C, Supplementary Figure 3). Following removal of the PD cases, the 'OtherParkinsonism' group consisted of those diagnosed with Multiple System Atrophy (MSA) and Progressive Supranuclear Palsy (PSP), for which no group differences were observed, likely due to small sample sizes (prodromal group n = 11, diagnosed group n = 23). Overall, the reduction in acceleration prior to, and following diagnosis, was unique to PD, suggesting this measure to be specific with potential for use in early identification in those likely to be diagnosed with PD.

Reduced sleep duration and quality only become apparent after Parkinson's disease diagnosis

We downloaded and extracted sleep features from raw accelerometry data for individuals with the following diagnoses: 'AllCauseParkinsonism', 'AllCauseDementia', 'Dystonia', and their respective unaffected control matches; a total of 4472 different individuals were processed.

We labelled physical activity classes using a pretrained Random Forest (Walmsley et al., 2021). Night-time was defined as 10pm to 9:59am and daytime as 10am to 9:59pm. Sleep features derived from acceleration data indicated reduced quality and duration of sleep in those diagnosed with PD, but not in individuals in the prodromal phase (Figure 3). Individuals with a diagnosis of PD compared to both unaffected controls and the prodromal PD group overall slept for fewer hours (p-value = 2.89×10^{-8}, p-value = 8.9×10^{-7}), had fewer consecutive hours of sleep (p-value = 8.41×10^{-22}, p-value = 8.76×10^{-8}), and woke up more frequently overnight (p-value = 2.14×10^{-12}, p-value = 7.27×10^{-6}). Prodromal PD cases and those diagnosed with PD slept more frequently during the day compared to unaffected controls (p-value = 0.0002, p-value = 0.001), with no significant difference between prodromal and diagnosed PD cohorts (p-value = 0.6). Individuals with prodromal PD showed less variation in their acceleration while asleep (t-statistic = 3.16, p-value = 0.002) and slept longer than unaffected controls (t-statistic = 4.37, p-value = 1.3×10^{-5}) and diagnosed PD cases (t-statistic = 5.8, p-value = 1.33×10^{-8}). Supplementary Figure 4 shows the same analyses without exclusion of cases diagnosed with comorbid depression. Most accelerometry-derived sleep measures deteriorate only after PD diagnosis.

Examination of the other diagnostic cohorts identified less deterioration in the sleep measures than found in PD (Figure 3). Prodromal dementia and diagnosed 'OtherParkinsonism' cases showed a reduction in uninterrupted sleep compared to unaffected controls. Diagnosed dementia cases slept more and napped more during daytime than unaffected controls. Wake-ups during night-time were not increased in any disorder other than PD, rendering it a potential specific diagnostic marker.

Accelerometry also has the capacity to identify those with prodromal Parkinson's disease from the general population

We next explored the predictive power of accelerometry data in terms of area under receiver operator curve (AUROC) and area under precision recall curve (AUPRC) at an individual level using Lasso logistic regression models with average acceleration, age, and sex as features (Supplementary Table 5). Average acceleration distinguished those diagnosed with PD from unaffected controls with a mean AUROC of 0.76±0.06 and could identify prodromal PD from unaffected controls with a mean AUROC of 0.79±0.04. Prodromal PD and diagnosed PD could also be identified from a general population cohort including unaffected controls, prodromal and diagnosed cases of 'Dystonia', 'OtherParkinsonism', and 'AllCauseDementia' using only average acceleration with a respective AUROC of 0.79±0.02 and 0.72±0.02. However, the AUPRC, which is better suited for such imbalanced datasets (Davis & Goadrich, 2006), showed values of 0.28±0.04 and 0.28±0.03, respectively, indicating the lack of sensitivity of these models to identify the smaller class. Adding the derived physical activity and sleep features increased the performance of the model to identify individuals diagnosed with PD from the general population to 0.83±0.03 AUROC and 0.51±0.06 AUPRC. There was no significant improvement in identifying prodromal PD cases from the general population when these features were added (Supplementary Table 6). The most robustly selected feature is mean movement during epochs classified as light physical activity (Supplementary Figure 5).

We then sought to evaluate whether other disorders might be misclassified by the models, and if so, which ones (Figure 4). Neither the model trained to identify prodromal PD nor the one trained to identify diagnosed PD misclassified any of the disorders examined as PD (mean probability <0.5) demonstrating the specificity of the models. We additionally evaluated the models on held-out data that was not involved in the nested cross-validation. For the model trained to identify diagnosed PD, we additionally tested it on data from prodromal PD cases. Conversely, the model trained to identify prodromal PD was tested on data from diagnosed PD cases. The model trained to identify diagnosed PD also assigns high probabilities to prodromal

PD cases to have PD and the model trained to identify prodromal PD, in turn, also assigns high probabilities of developing PD to the diagnosed PD group. This highlights the similarity between the prodromal and diagnosed PD group and their respective models.

Accelerometry identifies prodromal Parkinson's disease more accurately than any other risk or prodromal factor

Several modalities have been previously explored for their value in identifying prodromal PD; however, these were often investigated in isolation and in clinically refined cohorts, rather than the general population. Here, we inspected genetics, lifestyle, blood biochemistry, recognised prodromal symptoms for PD, as well as accelerometry (Table 1). We trained Lasso logistic regression models on these different modalities to identify diagnosed PD or prodromal PD from matched unaffected controls or the general population, the latter group including individuals diagnosed with other neurodegenerative or movement disorders. Models trained on lifestyle, serum biochemical blood markers, recognised prodromal symptoms, or genetic factors showed lower AUROC and AUPRC scores than models trained on accelerometry features (Figure 5A-D, Supplementary Table 5). Models trained using the general population as controls showed lower AUPRCs reflecting the class imbalance and resulting bias to assign individuals to the larger class. We therefore used the AUPRC values to compare the modalities in terms of their performance. Across all but one task, the accelerometry modality significantly outperformed any other single-modality model significantly, for example, the genetic modality in identifying prodromal PD from the general population (p-value = 2.35×10^{-6}) (Supplementary Table 6). The only task where the accelerometry modality did not outperform the other modalities was for models trained to identify prodromal PD from matched unaffected controls; the models showed a high variability between cross-validation folds leading to insignificant differences in performance. Combining modalities did not significantly improve performance for any task (Supplementary Table 6).

We next evaluated which factors of each modality are the most relevant (Supplementary Figure 6A-E) and provided a measure of how many features from a modality were significant in the combined model. Features from the accelerometry modality made up the largest portion (67% for the model identifying prodromal PD and 70% for the model identifying diagnosed PD) of the most important features in the combined model (Supplementary Figure 5). The second modality in order of importance to the model was the serum biochemical markers, with values of 25% and 20%, respectively. The prodromal model identified one prodromal symptom as important, and the diagnosis model identified one genetic feature as important. Mean acceleration during epochs that are classified as light activity was the strongest protective factor for both models. The most important features in the combined model resembled those found in the modality-specific models (Supplementary Figure 6A-E). As the accelerometry and the blood modalities were selected as the most relevant but also have the highest number of features (Table 1), one could argue that this importance is purely due to their dimensionality. We investigated this through a stacked model where the predicted probabilities of the modality-specific models serve as input to a final logistic regression model. This approach also identified the same modalities (accelerometry and blood biochemistry) as the most important ones (Supplementary Figure 7).

Acceleration-derived phenotypes can predict the time to diagnosis

An estimate for the time to diagnosis would not only have potential clinical utility but would be central in clinical trials evaluating the efficacy of disease-modifying or curative therapies. As such, we next explored which modality would be most beneficial in predicting the time to clinical diagnosis of PD. A simple linear association was not found between residual average acceleration (age- and sex- corrected) and time to diagnosis for the prodromal cases ($r = 0.11$, p-value = 0.13). A survival random forest model trained on accelerometry features achieved a mean AUROC (from the five-folds time-dependent AUROCs) of 0.76 ± 0.06 when restricted to unaffected healthy controls and 0.84 ± 0.02 when trained on the general population (Figure 5E-F, Supplementary Table 7). This model can predict the probability of not receiving a PD

diagnosis over time significantly better than any other single modality model. This finding highlights that acceleration data not only allows us to predict who will develop PD but also when this diagnosis might be expected.

Discussion

We showed here the potential of accelerometry as a prodromal biomarker for PD. We found that reduced acceleration manifests years prior to clinical PD diagnosis. This reduction before diagnosis was unique to PD, it was not observed for any other disorder examined. By comparing the predictive value of accelerometry with other modalities including genetics, lifestyle, blood biochemistry, and prodromal symptoms, we found that no other data type performed better. This high predictive value remained when assessing the models in a more real-world scenario where the control group contained individuals with related disorders as well as unaffected controls. Finally, we showed that accelerometry can further predict the time-point at which a PD diagnosis can be expected.

We demonstrated here for the first time the clinical value of accelerometry-based biomarkers for prodromal PD in the general population with a large sample size of individuals who convert to PD after data collection while assessing the specificity and drawing a comparison to other biomarkers. Previous work has explored the usage of digital gait markers for monitoring disease onset, severity, progression, and medication response in PD (Breasail et al., 2021; Liu et al., 2022). Using the UKBB data, Williamson et al. (2021) demonstrated high accuracy in detecting diagnosed PD cases from controls in a large sample size using free-world collected wrist-worn accelerometer data. However, they focused on prevalent PD cases and thus did not explore the possibility of these digital markers to identify PD years before clinical diagnosis. The only other study that investigated the prodromal phase of PD with gait-measuring sensors (Del Din et al., 2019) was limited to 16 subjects who converted to PD. Further, gait data in that study was only collected during a specified task in the clinic and not in free-living conditions. Other digital markers have been investigated for their use as prodromal biomarkers in equally small sample sizes (n = 12), for example, nocturnal breathing patterns predicted severity and progression in diagnosed PD cases but were also examined as risk assessment tools for prodromal PD (Yang et al., 2022). Other modalities have been explored considering specificity of the markers by examining other related disorders; for example, cognitive and functional

impairment prior to diagnosis has been assess in several neurodegenerative disorders using UKBB data (Swaddiwudhipong et al., 2022). Transferability to the general population and biomarker specificity have not been reported in previous work on accelerometry due to the data being collected in PD cohorts; however, the population-based data from the UKBB allows to investigate this because it contains data from subjects diagnosed with diverse disorders. Overall, we identified five major gaps in research that this work aims to address: studying the (1) prodromal phase of PD using passively collected (2) real-world gait-sensor-based data in a (3) large sample size (4) while comparing its performance to other established markers and (5) its generalizability to the general population. The transferability to the general population and the specificity of the biomarkers have not been reported in previous work due to the data being collected in PD cohorts but the population-based data from the UKBB allows us to investigate specificity as it contains subjects diagnosed with diverse disorders. We therefore showed here for the first time the actual clinical value of accelerometry based biomarkers for prodromal PD compared to other modalities.

As the data is easily accessible and low-cost, using accelerometry data in clinical research and practice is feasible. Currently, electronic health records do not include data from medical-grade wearables. Such data, however, is readily available as smart-devices capable of collecting accelerometer data are used daily by most people (Chandrasekaran et al., 2020). This resource could be linked to electronic health records so that a large retrospective study could be undertaken. A rapidly growing interest in leveraging such data for personalised medicine accelerates this research area with many new digital sensors and wearables being developed (Xu et al., 2022). Challenges to overcome include measurement validity and capability, data privacy, and liability concerns (Simon et al., 2022). If these limitations are addressed, wearables and other health-sensor devices hold the ability to transition medicine into a digital health era improving accessibility in remote areas, reducing cost, and improving healthcare (Xu et al., 2022).

This study has several limitations. One is the lack of replication although extensive cross-validation was performed. UKBB provides a unique resource with no equivalent in terms of scale and amount of data collected to study the prodromal phase of several disorders retrospectively. Similar resources could easily be obtained as about 30% of adults in the United States already use wearable health care devices (Chandrasekaran et al., 2020). Several restraints concerning data availability within the UKBB should be noted. Except for a small subset of individuals, accelerometry data was only collected over one seven-day period. Longitudinal data on acceleration would allow for an investigation of individual trajectories improving the sensitivity of detection deterioration in physical activity. Further, not all included features were available for all subjects and hence, we trained the models only on a subset of individuals who had complete information (Table 1). This artificially reduced our sample size but made the models more comparable. As the accelerometry study was only done in a subset of the individuals, this modality is the main limiting factor. This data availability in the UKBB, however, does not reflect the real-world availability of the modalities. Genetic data is much more sparsely available in the general population but was prioritised within the UKBB. Accelerometry data can, in general, be easily and cost-efficiently gathered. As the downloading of the raw accelerometry data and the processing pipeline to label the physical activity is time-consuming (30 seconds to download the raw data for one subject ~250MB, ~3 minutes to process one subject), we processed only the raw data for 'AllCauseDementia', 'AllCauseParkinsonsim', 'Dystonia', and their respective unaffected control matches. This led to an artificial overrepresentation of these diagnosis groups in the general population and thus limiting the transferability of our model to clinical practice as it was not trained on a perfect representation of the general population. Processing the data of more unaffected controls, could result in a more representative population.

In conclusion, our results suggest that accelerometry collected with wearable devices could be used to monitor those at elevated risk of PD such that they can be diagnosed earlier and, importantly, individuals who will likely convert within the next few years can be included in studies for neuroprotective treatments.

Material and Methods

Study Population

The UKBB holds in-depth information on ~500,000 participants and is approved by the Research Ethics Committee (reference 16/NW/0274). It was accessed under the application code 69610 with data released to Cardiff University. We explored PD and related disorders, namely 'AllCauseDementia', 'AllCauseParkinsonism', 'AlzheimerDisease', 'Dystonia', 'Osteoarthritis', and 'Depression'. We identified patient groups based on ICD10 and ICD9 codes in the hospital inpatient data (fields 41270 and 41271) and the death registry (fields 40001 and 40002) which were curated from UKBB provided tables and phecodes (Wu et al., 2018), as well as self-reported diagnoses (field 20002). Primary care data was also included (field 42040) using read codes (version 2 and 3) respective to the ICD10 codes as mapped through TRUD NHS Read browser (NHS). The respective codes for each diagnosis can be found in Supplementary Table 2. We distinguished prodromal (incident) and diagnosed (prevalent) cases at the date of accelerometer data collection (field 90003) based on the earliest reported date across all resources and allowed for a two-year margin of error, meaning that patients diagnosed before or within the two years after accelerometer data collection are classified as diagnosed/prevalent cases (Figure 2A) and individuals receiving a diagnosis >2 years later are labelled prodromal/incident cases. From the set of unaffected controls, which we defined as having no neurological or behavioural disorder across all included sources, we randomly sampled unique age- and sex-matched individuals to our patients (1:1). Only participants who passed quality control for the accelerometer data (field 90016) were included. Health-related outcome data is available up to March 2021. Using the same approach, we also identified subjects with recognised prodromal signs and symptoms, namely depression, anxiety, orthostatic hypotension, RBD, hyposmia, urinary incontinence, and constipation. We defined these as prodromal symptoms if they were reported before a PD diagnosis was made.

Accelerometer data

Participants wore an Axivity AX3 wrist-worn triaxial accelerometer on their dominant hand for a 7-day-period. UKBB provides summary statistics (category 1009) describing daily and hourly averages. We augmented the accelerometer data by pre-processing the raw data into time-series data and classifying 30 second intervals into physical activity categories, namely imputed, sleep, sedentary, light, or Moderate to Vigorous Physical Activity (MVPA), with a machine-learning model, using balanced random forests with Markov confusion matrices, using the accelerometer package provided by the Oxford Wearables Group (Willetts et al., 2018). The pre-processing steps we employed are the same as those used to derive the summary statistics from UKBB. These steps include device calibration, resampling to 100 Hz, and removal of noise and gravity (Doherty et al., 2017). From this machine learning labelled time-series data we derived measures of uninterrupted duration, mean movement, and number of interruptions for each physical activity category. For sleep this entailed measures of sleep quality, for example frequency of night-time waking and frequency of daytime napping.

Additional data

We merged the accelerometer statistics (category 1009), blood biochemistry measures at baseline (category 17518), physical health measures at baseline (category 100006), Polygenic Risk Scores (PRS) (category 301), and our derived physical activity phenotypes. We further included age at baseline visit, age at accelerometer data collection, and sex.

Statistical Analyses

All statistical analyses were carried out in python 3.9 using scipy 1.6.1, scikit-learn 0.23.2 (Pedregosa et al., 2011) and sksurv 0.14.0 (Pölsterl, 2020) packages and figures were generated with seaborn 0.11.1 (Waskom, 2021) and matplotlib.

Prevalence

We validated our established cohort of PD cases by comparing the observed to the expected prevalence. For each year between 1950 and 2021 we identified the number of diagnosed and undiagnosed cases in each age-group. Based on the date of death (field 40007) subjects were excluded from the statistics from their year of death onwards. We calculated the estimated number of PD cases for each year based on the number of people alive in each age group and the prevalence rates for individual age groups from a population-based study from 2015 (Parkinson's UK, 2017). We extrapolated this expected prevalence until 2030, making the assumption of no deaths taking place.

Identifying and Adjusting for Covariates

Age and sex are known covariates of acceleration. To address this, we subsampled the unaffected controls in an age- and sex-matched manner. However, prodromal and clinically diagnosed groups differed in age (t-statistic = 3.18, p-value = 1.6×10^{-3}). BMI is also a covariate for acceleration (Supplementary Table 4). To address this, we calculated the residuals of average acceleration using coefficients for age, BMI and sex learned from the unaffected control group (N = 39932) with a linear regression model. When examining the other (non-PD) diagnostic groups, we removed any cases in which co-morbid PD was observed to attempt to maintain cohort homogeneity (Supplementary Figure 2). We also removed subjects with a diagnosis of depression from the other diagnosis groups as this diagnostic group was found to have a significantly reduced acceleration and is a prodromal marker for PD. We compared the residual average acceleration measure between the healthy, prodromal, and diagnosed groups for each included diagnosis class with two-sided T-tests and Bonferroni-correction. We also computed the residual sleep features, which were age-, BMI- and sex-corrected, using the same method based on a cohort of 2144 unaffected controls for which we derived the sleep features.

Prediction Models

To quantify the predictive power of the acceleration data on an individual level and to compare it to other modalities, we fitted logistic regression models. We identified five modalities: genetics, lifestyle, blood biochemistry, prodromal symptoms, and accelerometry (Table 1). We restricted the dataset to subjects with information available for all five modalities. We estimated the predictive performance of each modality with logistic regression with fitted least absolute shrinkage and selection operator (LASSO) penalty in a nested cross-validation. Three different model types were trained: 1) diagnostic biomarker: identifying diagnosed PD from control, 2) prodromal marker: identifying prodromal PD from control, 3) screening: identifying diagnosed and prodromal PD from control. The control group was either matched unaffected controls or a representation of the general population, which included unaffected controls and subjects diagnosed with other disorders such as dementia, dystonia, and other forms of parkinsonism.

We trained models on different modalities, always including the covariates age and sex: genetics, lifestyle, blood, prodromal symptoms, all acceleration features, all modalities combined (Table 1). We trained the models in a nested 5-fold cross-validation using a stratified split of 33% for testing and 67% for training and validation with grid search for the best Lasso penalty hyperparameter (10 equidistant values between 10^{-5} and 10^{4}). Parameter selection was applied independently to each training fold. Real valued predictors were standardised based on the training data of the outer split to have a standard deviation of one and a mean of zero. Binary data was encoded as 0/1. Balanced class weighting was applied to the models trained on the general population to adjust for class imbalances. We report the mean and 95% confidence interval (CI) of the area under the receiver operator curve (AUROC) and the area under the precision and recall curve (AUPRC) on the outer cross-validation splits to compare models.

Performance of the classifiers was compared using two-sided T-tests with multiple testing accounted for using Bonferroni correction at 0.05. We determined which disorders were most

likely to be misidentified by the models as PD by assessing the mean predicted probability of having PD as assigned by the model on the outer folds of the test data. We further investigated feature importance by calculating the mean effect of each predictor over the five outer cross-validation splits. We validated their stability through checking their effect size in each outer cross-validation split. A feature was labelled as important and significant if the mean effect size across folds was significantly different from zero (95% Bonferroni-corrected CI does not cross zero). In addition to the combined model which used the union across all modalities, we further trained a stacked model which took the predicted probabilities from each modality-specific model and integrated it in a final lasso logistic regression model (alpha = 0.1) that predicts the overall probability of having or getting a PD diagnosis. Each modality was thus assigned a coefficient with which its prediction contributed to the final prediction. This model was trained on the same outer cross-validation splits as the other models using their predictions on the training data to train the final model and the respective test data for testing. The performance was evaluated in a similar fashion with AUROC and AUPRC across folds and the assigned coefficients across folds were evaluated for their stability across folds by examining the mean and the Bonferroni-adjusted 95% Confidence Interval.

Survival Models

We explored the value of each modality to predict the time to diagnosis for the prodromal PD cohort. We did so by first calculating the Pearson correlation of the residual average acceleration (age- and sex-corrected) and time to diagnosis in the prodromal PD cohort. We then used survival modelling on the prodromal and matched unaffected controls to predict when each individual would be diagnosed. To this end, we used survival random forests with a five-fold cross-validation trained on a stratified split of 67%. The survival random forest is made up of 1000 trees which requires at least 10 samples for a split and 15 samples per leaf. The matched unaffected controls were modelled as right censored, as we did not know whether or when they would receive a diagnosis. We modelled the time from accelerometer data collection to PD diagnosis, thus defining the time of accelerometer collection as time 0.

For the prodromal symptoms modality, we hence restricted the time of diagnosis of prodromal symptoms to the date of accelerometer data collection and removed all subsequent diagnoses of prodromal symptoms. We reported the mean time-dependent AUROC on the five cross-validation 33% test sets.

Data Availability

Data from the UK Biobank (ukbiobank.ac.uk/) are available on application to the UK Biobank.

Funding

A.-K. Schalkamp is supported by a PhD studentship funded by Health and Care Research Wales.
C. Sandor is supported by the UK Dementia Research Institute (UK DRI) funded by the Medical Research Council (MRC), Alzheimer's Society and Alzheimer's Research UK (AR-UK) and by the Ser Cymru II programme which is part-funded by Cardiff University and the European Regional Development Fund through the Welsh Government.
N. Harrison has nothing to declare.
K. Peall is funded by an MRC Clinician-Scientist Fellowship (MR/P008593/1).

Acknowledgements

The work described here is part-funded by the European Regional Development Fund, administered through the Welsh Government. We are also grateful for the Advanced Research Computing at Cardiff.

Author Contribution

A-K.S. and C.S. participated in designing the study, topic definition, and review of relevant studies. Machine learning models and statistical analyses were designed and implemented by A-K.S.. Figures and tables were done by A-K.S. with the support of C.S.. A-K.S. wrote the first draft. C.S., N.A.H., and K.J.P. contributed to subsequent versions of the manuscript. All authors critically reviewed the paper, all authors have a clear understanding of the content, results, and conclusions of the study and agree to submit this manuscript for publication. The corresponding author (C.S.) declares that all authors listed meet the authorship criteria and

that no other authors involved in this study are omitted. C.S. is ultimately responsible for this article.

Competing interests

The authors declare no competing interests.

References

Breasail, M. O., Biswas, B., Smith, M. D., Mazhar, M. K. A., Tenison, E., Cullen, A., Lithander, F. E., Roudaut, A., & Henderson, E. J. (2021). Wearable GPS and Accelerometer Technologies for Monitoring Mobility and Physical Activity in Neurodegenerative Disorders: A Systematic Review. *Sensors (Basel)*, *21*(24). https://doi.org/10.3390/s21248261

Brognara, L., Palumbo, P., Grimm, B., & Palmerini, L. (2019). Assessing Gait in Parkinson's Disease Using Wearable Motion Sensors: A Systematic Review. *Diseases*, *7*(1). https://doi.org/10.3390/diseases7010018

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203-209. https://doi.org/10.1038/s41586-018-0579-z

Chandrasekaran, R., Katthula, V., & Moustakas, E. (2020). Patterns of Use and Key Predictors for the Use of Wearable Health Care Devices by US Adults: Insights from a National Survey. *J Med Internet Res*, *22*(10), e22443. https://doi.org/10.2196/22443

Del Din, S., Elshehabi, M., Galna, B., Hobert, M. A., Warmerdam, E., Suenkel, U., Brockmann, K., Metzger, F., Hansen, C., Berg, D., Rochester, L., & Maetzler, W. (2019). Gait analysis with wearables predicts conversion to parkinson disease. *Ann Neurol*, *86*(3), 357-367. https://doi.org/10.1002/ana.25548

Doherty, A., Jackson, D., Hammerla, N., Plotz, T., Olivier, P., Granat, M. H., White, T., van Hees, V. T., Trenell, M. I., Owen, C. G., Preece, S. J., Gillions, R., Sheard, S., Peakman, T., Brage, S., & Wareham, N. J. (2017). Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. *PLoS One*, *12*(2), e0169649. https://doi.org/10.1371/journal.pone.0169649

Dorsey, E. R., Omberg, L., Waddell, E., Adams, J. L., Adams, R., Ali, M. R., Amodeo, K., Arky, A., Augustine, E. F., Dinesh, K., Hoque, M. E., Glidden, A. M., Jensen-Roberts, S., Kabelac, Z., Katabi, D., Kieburtz, K., Kinel, D. R., Little, M. A., Lizarraga, K. J., Myers, T., Riggare, S., Rosero, S. Z., Saria, S., Schifitto, G., Schneider, R. B., Sharma, G., Shoulson, I., Stevenson, E. A., Tarolli, C. G., Luo, J., & McDermott, M. P. (2020). Deep Phenotyping of Parkinson's Disease. *J Parkinsons Dis*, *10*(3), 855-873. https://doi.org/10.3233/JPD-202006

Fearnley, J. M., & Lees, A. J. (1991). Ageing and Parkinson's disease: substantia nigra regional selectivity. *Brain*, *114 (Pt 5)*, 2283-2301. https://doi.org/10.1093/brain/114.5.2283

Fereshtehnejad, S. M., Montplaisir, J. Y., Pelletier, A., Gagnon, J. F., Berg, D., & Postuma, R. B. (2017). Validation of the MDS research criteria for prodromal Parkinson's disease: Longitudinal assessment in a REM sleep behavior disorder (RBD) cohort. *Mov Disord*, *32*(6), 865-873. https://doi.org/10.1002/mds.26989

Heinzel, S., Berg, D., Gasser, T., Chen, H., Yao, C., Postuma, R. B., & Disease, M. D. S. T. F. o. t. D. o. P. s. (2019). Update of the MDS research criteria for prodromal Parkinson's disease. *Mov Disord*, *34*(10), 1464-1470. https://doi.org/10.1002/mds.27802

Hustad, E., & Aasly, J. O. (2020). Clinical and Imaging Markers of Prodromal Parkinson's Disease. *Front Neurol*, *11*, 395. https://doi.org/10.3389/fneur.2020.00395

Johansson, D., Malmgren, K., & Alt Murphy, M. (2018). Wearable sensors for clinical applications in epilepsy, Parkinson's disease, and stroke: a mixed-methods systematic review. *J Neurol*, *265*(8), 1740-1752. https://doi.org/10.1007/s00415-018-8786-y

Liu, Y., Zhang, G., Tarolli, C. G., Hristov, R., Jensen-Roberts, S., Waddell, E. M., Myers, T. L., Pawlik, M. E., Soto, J. M., Wilson, R. M., Yang, Y., Nordahl, T., Lizarraga, K. J., Adams, J. L., Schneider, R. B., Kieburtz, K., Ellis, T., Dorsey, E. R., & Katabi, D. (2022). Monitoring gait at home with radio waves in Parkinson's disease: A marker of severity, progression, and medication response. *Sci Transl Med*, *14*(663), eadc9669. https://doi.org/10.1126/scitranslmed.adc9669

Mirelman, A., Bernad-Elazari, H., Thaler, A., Giladi-Yacobi, E., Gurevich, T., Gana-Weisz, M., Saunders-Pullman, R., Raymond, D., Doan, N., Bressman, S. B., Marder, K. S., Alcalay, R. N., Rao, A. K., Berg, D., Brockmann, K., Aasly, J., Waro, B. J., Tolosa, E., Vilas, D., Pont-Sunyer, C., Orr-Urtreger, A., Hausdorff, J. M., & Giladi, N. (2016). Arm swing as a potential new prodromal marker of Parkinson's disease. *Mov Disord*, *31*(10), 1527-1534. https://doi.org/10.1002/mds.26720

NHS, D. T. *NHS Read Browser*. Retrieved March 28, 2022 from https://isd.digital.nhs.uk/trud/users/guest/filters/0/categories/9/items/8/releases

Parkinson's UK. (2017). *The incidence and prevalence of Parkinson's in the UK*. Retrieved November 21, 2022 from https://www.parkinsons.org.uk/sites/default/files/2018-01/Prevalence%20%20Incidence%20Report%20Latest_Public_2.pdf

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825—2830. http://jmlr.org/papers/v12/pedregosa11a.html

Pölsterl, S. (2020). scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn. *Journal of Machine Learning Research*, *21*(212), 1–6. http://jmlr.org/papers/v21/20-729.html

Postuma, R. B., & Berg, D. (2016). Advances in markers of prodromal Parkinson disease. *Nat Rev Neurol*, *12*(11), 622-634. https://doi.org/10.1038/nrneurol.2016.152

Postuma, R. B., & Berg, D. (2019). Prodromal Parkinson's Disease: The Decade Past, the Decade to Come. *Mov Disord*, *34*(5), 665-675. https://doi.org/10.1002/mds.27670

Schlachetzki, J. C. M., Barth, J., Marxreiter, F., Gossler, J., Kohl, Z., Reinfelder, S., Gassner, H., Aminian, K., Eskofier, B. M., Winkler, J., & Klucken, J. (2017). Wearable sensors objectively measure gait parameters in Parkinson's disease. *PLoS One*, *12*(10), e0183989. https://doi.org/10.1371/journal.pone.0183989

Shah, V. V., McNames, J., Mancini, M., Carlson-Kuhta, P., Nutt, J. G., El-Gohary, M., Lapidus, J. A., Horak, F. B., & Curtze, C. (2020). Digital Biomarkers of Mobility in Parkinson's Disease During Daily Living. *J Parkinsons Dis*, *10*(3), 1099-1111. https://doi.org/10.3233/JPD-201914

Swaddiwudhipong, N., Whiteside, D. J., Hezemans, F. H., Street, D., Rowe, J. B., & Rittman, T. (2022). Pre-diagnostic cognitive and functional impairment in multiple sporadic neurodegenerative diseases. *Alzheimers Dement*. https://doi.org/10.1002/alz.12802

Walmsley, R., Chan, S., Smith-Byrne, K., Ramakrishnan, R., Woodward, M., Rahimi, K., Dwyer, T., Bennett, D., & Doherty, A. (2021). Reallocation of time between device-measured movement behaviours and risk of incident cardiovascular disease. *Br J Sports Med*, *56*(18), 1008-1017. https://doi.org/10.1136/bjsports-2021-104050

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, *6*, 3021. https://doi.org/10.21105/joss.03021

Willetts, M., Hollowell, S., Aslett, L., Holmes, C., & Doherty, A. (2018). Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. *Sci Rep*, *8*(1), 7961. https://doi.org/10.1038/s41598-018-26174-1

Williamson, J. R., Telfer, B., Mullany, R., & Friedl, K. E. (2021). Detecting Parkinson's Disease from Wrist-Worn Accelerometry in the U.K. Biobank. *Sensors (Basel)*, *21*(6). https://doi.org/10.3390/s21062047

Yang, Y., Yuan, Y., Zhang, G., Wang, H., Chen, Y. C., Liu, Y., Tarolli, C. G., Crepeau, D., Bukartyk, J., Junna, M. R., Videnovic, A., Ellis, T. D., Lipford, M. C., Dorsey, R., & Katabi, D. (2022). Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. *Nat Med*, *28*(10), 2207-2215. https://doi.org/10.1038/s41591-022-01932-x

modality	cases	controls	features	N features	Color code
Genetics & Family history	165	2320	Polygenic risk scores of 34 traits Family history of: Stroke, Diabetes, Severe Depression, Alzheimer's disease Dementia & Parkinson's disease	41	
Blood biochemistry	114	1751	Albumin, Alkalinephosphatase, Alanineaminotransferase, ApolipoproteinA, ApolipoproteinB, Aspartateaminotransferase, Urea, Calcium, Cholesterol, Creatinine, C reactiveprotein, CystatinC, Gammaglutamyltransferase, Glucose, Glycatedhaemoglobin HbA1c, HDLcholesterol, IGF 1, LDLdirect, Phosphate, SHBG, Totalbilirubin, Testosterone, Totalprotein, Triglycerides, Urate, Vitamin D	28	
Lifestyle	168	1598	AlcoholStatus Current, AlcoholStatus Previous, SmokeStatus Current, SmokeStatus Previous, Daytime Sleepiness Often, AlcoholFrequency LessThanWeekly, BMI, Waist Circumference, Hip Circumference, Diastolic BloodPressure, PulseRate, Body-Fat Percentage, TownsendDeprivationIndex	15	
Prodromal Symptoms	196	2990	UrinaryIncontinence, Constipation, ErectileDysfunction, Anxiety, REM Behavioral Sleep Disorder (RBD), Hyposmia, OrthostaticHypotension, Depression	10	
All Accelerometry features	196	2990	UKBB provided averages, weartime-bias corrected value, and standard deviations for days and hours, self-derived features for physical activity epochs (sleep, sedentary, light, MVPA, imputed)	101	
Combined	113	1589	union of above	187	
Stacked	113	1589	predicted probabilities from single-modality models (genetics + family, blood, lifestyle, prodromal symptoms, all accelerometry)	5	

Table 1: Feature set for each of the modalities.

For each modality we show which and how many predictors were included. Every modality includes the covariates (age at accelerometry data collection, sex). We further indicate the class imbalance and the

total number of available data for each modality in the population-based model identifying prodromal PD.

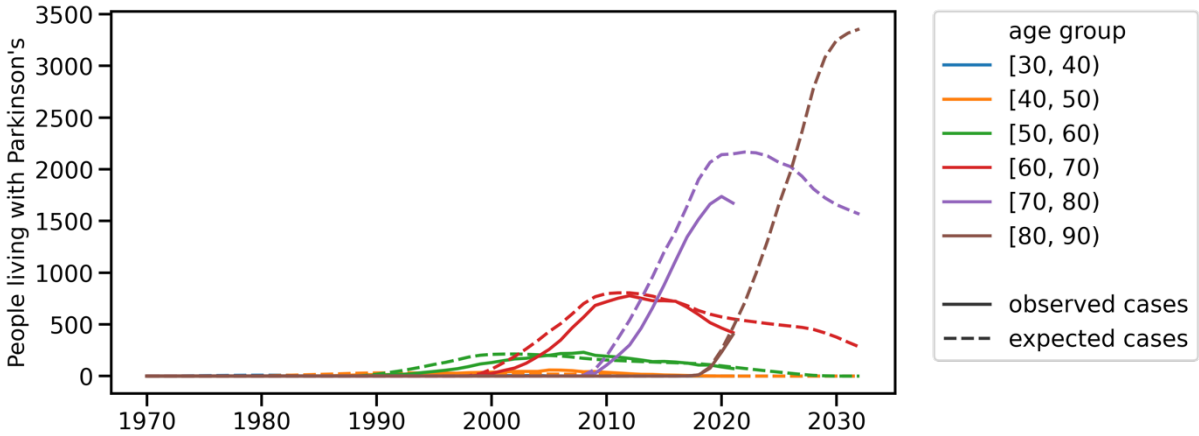

Figure 1: Estimated and Observed Prevalence of Parkinson's disease in UK Biobank.

Estimated (dashed) and observed (solid) number of people living with PD in the UK Biobank over time within age groups is shown. Estimated number of cases uses the population-based UK statistics from 2015 (Parkinson's UK, 2017)

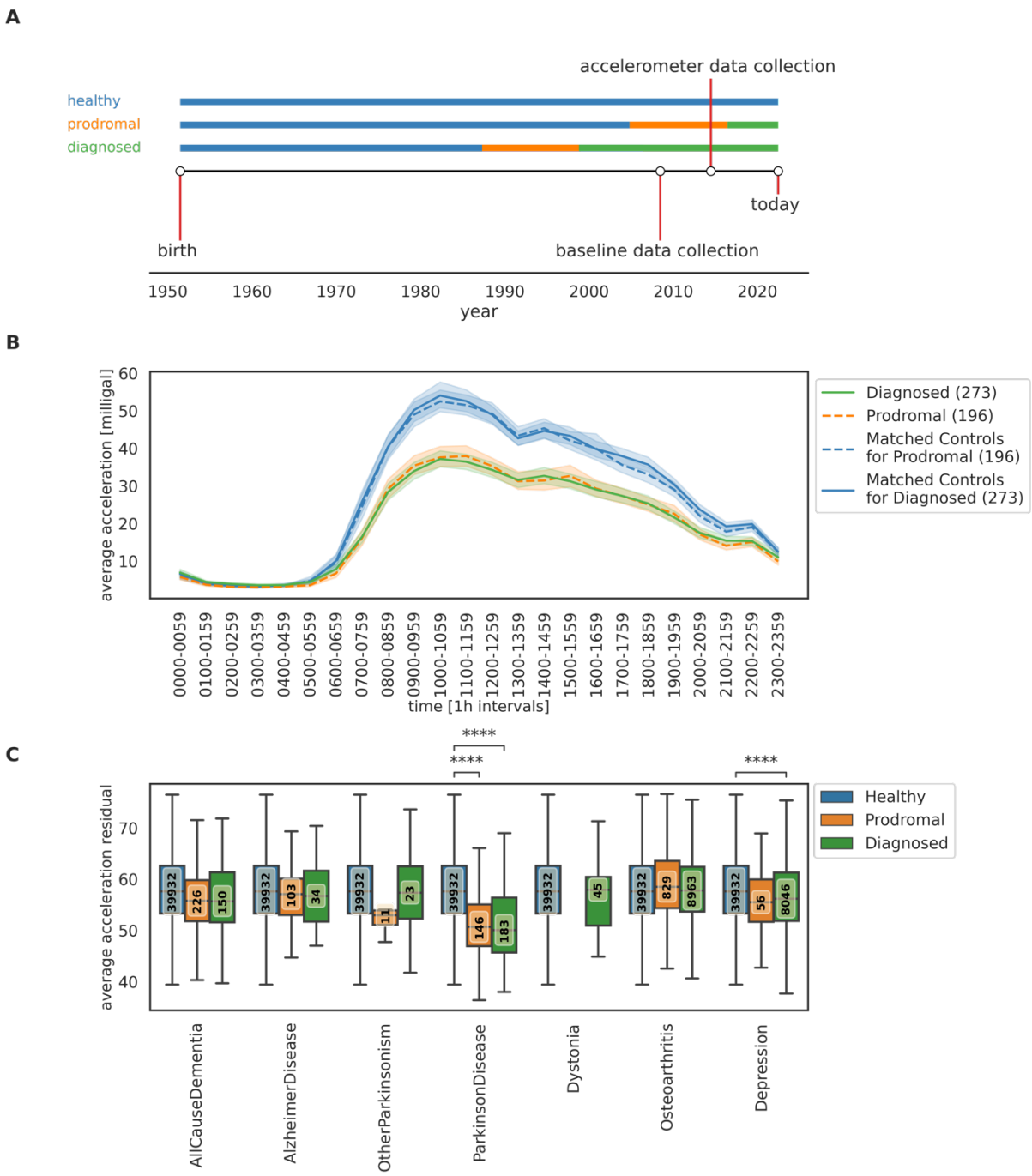

Figure 2: Unique reduction in acceleration in prodromal Parkinson's disease

[A] Baseline data were collected between 2006 and 2010; accelerometry data was gathered for a subset between 2013 and 2015. Diagnosed cases (green) were diagnosed prior to or within the subsequent two years of accelerometry data collection. Prodromal cases (orange) were diagnosed two or more years after accelerometry data collection.

[B] Average acceleration in milligal (0.01 mm/s^2) is shown in one-hour intervals over the course of one day. Group means for prodromal subjects (n = 196, orange, dashed), unaffected controls matched to

the prodromal ones (n = 196, blue, dashed), diagnosed subjects (n = 273, green, solid), and unaffected controls matched to the diagnosed ones (n = 273, blue, solid) is plotted with the respective 95% confidence interval.

[C] Boxplots for residual (age-, BMI-, and sex-corrected through unaffected control cohort) no-wear time bias corrected average acceleration after removal of cases diagnosed with comorbid depression or PD are shown for seven disease groups and unaffected controls. For each disease group we differentiate between diagnosed (green), prodromal (orange), and healthy (blue). The number of individuals in each group is indicated in the central box. Significance of group differences (two sided T-test) are indicated with star symbols, where significance is reached with a 0.05 Bonferroni-corrected threshold of 2.38×10^{-3} (ns: $2.38 \times 10^{-3} < p <= 1$, *: $2 \times 10^{-4} < p <= 2.38 \times 10^{-3}$, **: $2 \times 10^{-5} < p <= 2 \times 10^{-4}$, ***: $2 \times 10^{-6} < p <= 2 \times 10^{-5}$, ****: $p <= 2 \times 10^{-6}$).

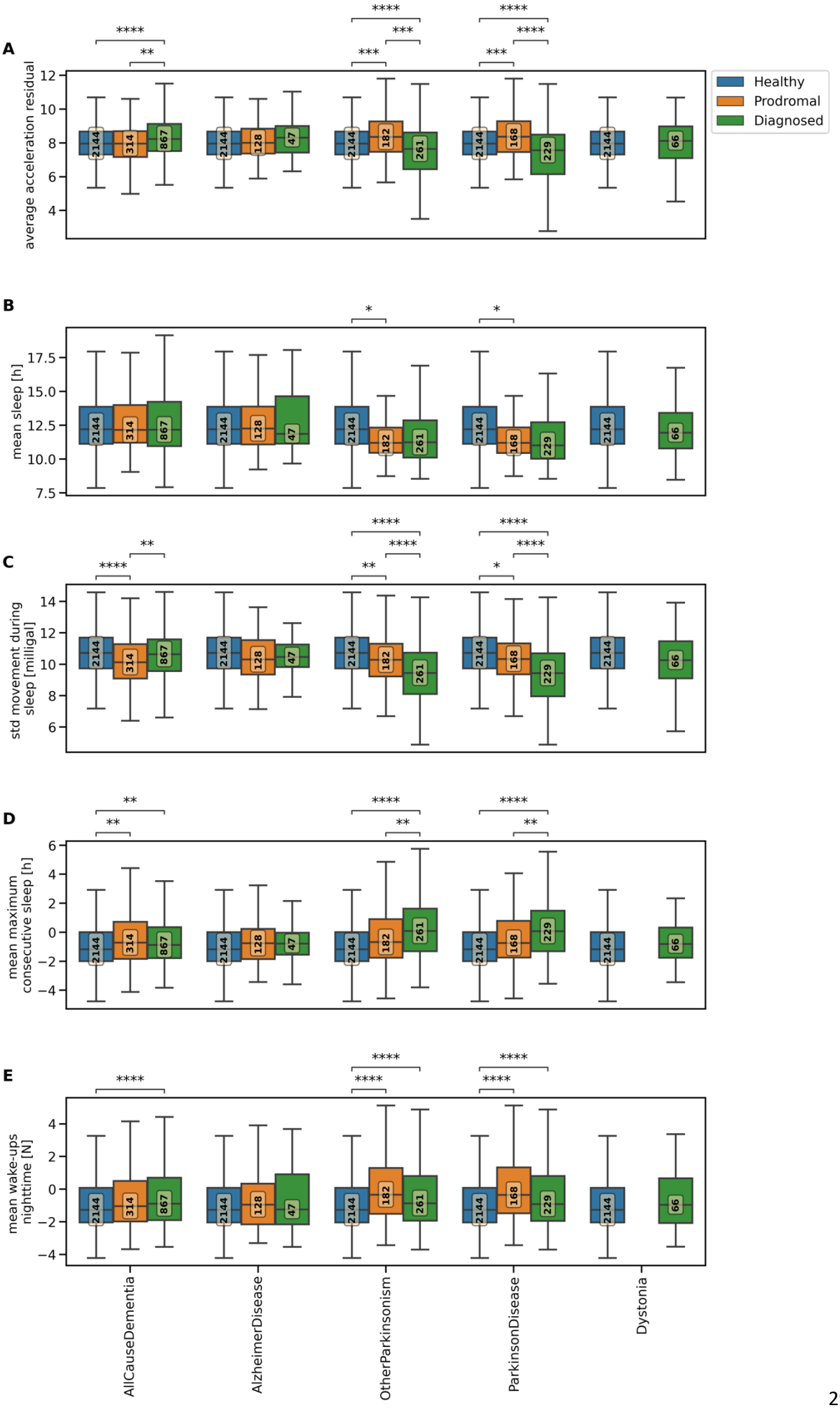
A
average acceleration residual
B
mean sleep [h]
C
std movement during sleep [milligal]
D
mean maximum consecutive sleep [h]
E
mean wake-ups nighttime [N]
Healthy
Prodromal
Diagnosed
AllCauseDementia
AlzheimerDisease
OtherParkinsonism
ParkinsonDisease
Dystonia

Figure 3: Quality and duration of sleep are reduced in diagnosed but not prodromal Parkinson's or any other disorder

Boxplots for five residual (age-, BMI-, and sex-corrected through unaffected control cohort) accelerometry derived sleep features after removal of cases diagnosed with comorbid depression or PD are shown for seven disease groups and unaffected controls. For each disease group we differentiate between diagnosed (green), prodromal (orange), and healthy (blue). The number of individuals in each group is indicated in the central box. Significance of group differences (two sided T-test) are indicated with star symbols, where significance is reached with a 0.05 Bonferroni-corrected threshold of 2.38×10^{-3} (ns: $2.38 \times 10^{-3} < p <= 1$, *: $2 \times 10^{-4} < p <= 2.38 \times 10^{-3}$, **: $2 \times 10^{-5} < p <= 2 \times 10^{-4}$, ***: $2 \times 10^{-6} < p <= 2 \times 10^{-5}$, ****: $p <= 2 \times 10^{-6}$).

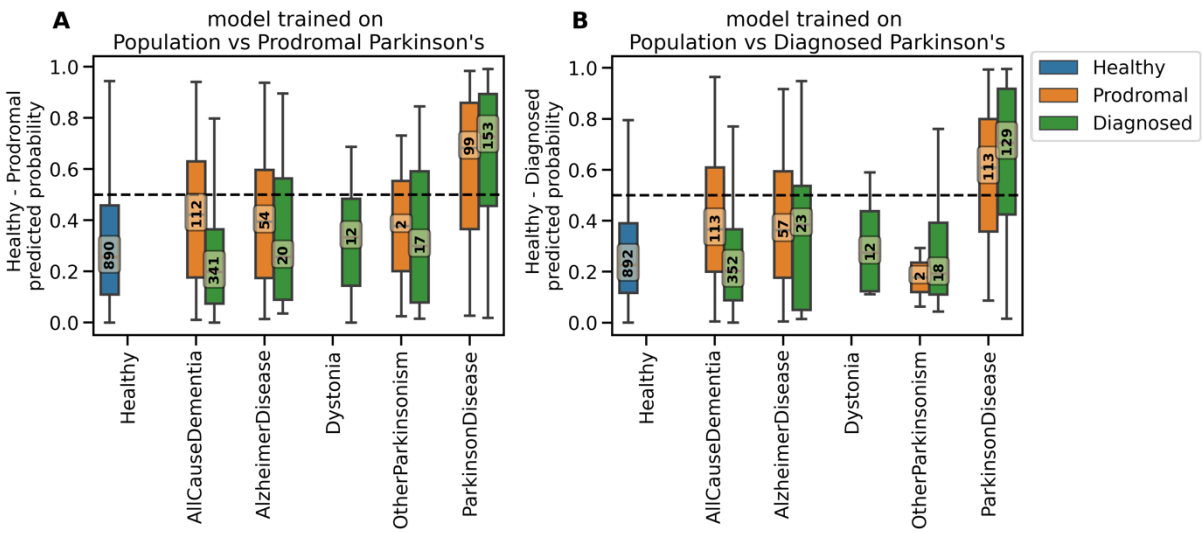

Figure 4: Misclassification of other disorders as Parkinson's disease.

The probability of having prodromal or diagnosed Parkinson's Disease (PD) as assigned by the best performing model on the test data is shown for each diagnostic group. Black, dashed line indicates the 0.5 cut-off for classification as case vs control. [A] Model trained to identify prodromal PD from the general population. We show the predicted probability on the test set and the held-out diagnosed PD population. [B] Model trained to identify diagnosed PD from the general population. We show the predicted probability on the test set and the held-out prodromal PD population.

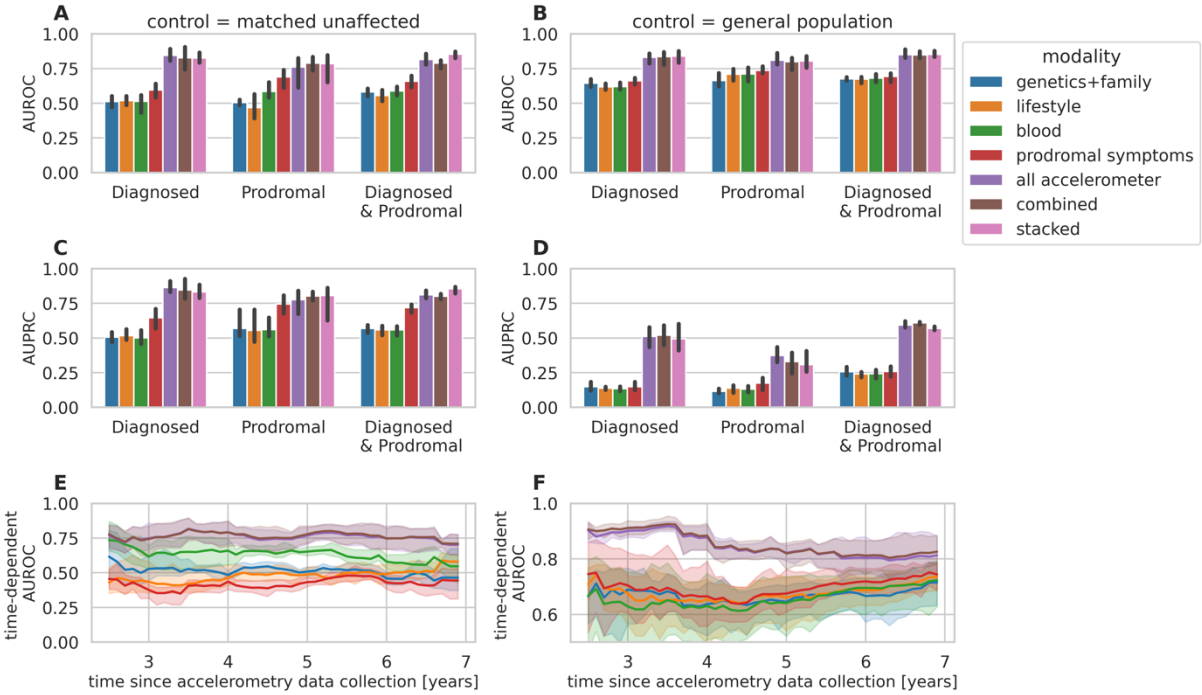

Figure 5: Accelerometry identifies Parkinson's disease and predicts time to diagnosis better than any other risk factors.

[A-D] Bar plots indicate the performance of each logistic regression model using different feature sets (Table 1). The mean area under receiver operator curve (AUROC) [A and B] and the mean area under precision recall curve (AUPRC) [C and D] are plotted across the five outer cross-validation folds with the error bars indicating the 95% confidence interval. We show this for five single modality models, one combined, and one stacked model for three different tasks with two different control groups, [A, C] matched unaffected controls and [B, D] general population. [E-F] Time-dependent AUROC of the random survival forests is plotted together with Bonferroni-adjusted 95% confidence interval for seven years since accelerometry data collection. We show this for [E] a control group made up of unaffected controls and [F] a control group representing the general population.