# Longitudinal analysis of step counts in Parkinson's disease patients: insights from a web-based application and generalized additive model

Yuan Gu[1] and Yishu Gong[2]

[1] School of Medicine, Stanford University, Palo Alto, CA, United States of America
[2] The Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, United States of America

## ABSTRACT

**Background**. Parkinson's disease (PD) is a chronic neurological disorder that affects millions of people worldwide. A common motor symptom associated with PD is gait impairment, leading to reduced step count and mobility.

**Methods**. Monitoring and analyzing step count data can provide valuable insights into the progression of the disease and the effectiveness of various treatments. In our study, the generalized additive model (GAM) was used to identify statistically significant variables for step counts. Additionally, a web application was developed as an interactive visualization tool.

**Results**. The GAM model shows that the following variables are statistically significant for daily step counts: sex ($p = 0.03$), handedness ($p = 0.015$), PD status of father ($p = 0.056$), COVID-19 status (Yes *vs.* No, $p = 0.008$), cohort (PD *vs.* Healthy, $p < 0.0001$), the cubic regression spline with three basis functions of age by cohorts ($p < 0.0001$), and the random effect of individual age trajectories ($p = 0.0001$).

**Conclusions**. Based on the PPMI data, we find that sex, handedness, PD status of father, COVID-19 status, cohort, and the smoothing functions of age are all statistically significant for step counts. Additionally, a web application tailored specifically for step count analysis in PD patients was developed. This tool provides a user-friendly interface for patients, caregivers, and healthcare professionals to track and analyze step count data, facilitating personalized treatment plans and enhancing the management of PD.

## INTRODUCTION

Parkinson's disease (PD) has emerged as the second most prevalent age-related neurodegenerative disorder worldwide, resulting in a substantial burden of disability and a significantly increased risk of developing dementia and mortality. The overall prevalence rate of PD is 572 cases per 100,000 among individuals aged 45 and above, and over 93 cases per 100,000 among individuals aged 65 and older (*Willis et al., 2022*). Despite significant progress in PD medications over the past half-century, none can be classified as a cure. Although these drugs have improved in effectiveness and duration, they still do not fully alleviate PD symptoms or offer a long-term solution. Therefore, treatment options must be

customized to meet the specific needs of each individual patient (*Jankovic & Aguilar, 2008*; *Ellis & Fell, 2017*; *Stoker & Barker, 2020*). Additionally, the cause of PD remains unknown, and there are no established methods to prevent the disease. While several risk factors, such as pesticide exposure, have been identified, confirmed causes of PD are genetic in nature. When PD is not linked to genetics, it is categorized as idiopathic (iPD), indicating that the precise cause remains unclear (*Mahlknecht et al., 2022*; *Puigròs et al., 2022*).

Today, the emergence of wearable solutions—such as smartwatches, fitness trackers, and health monitoring devices—has become increasingly popular due to their ability to provide low-cost, instant, convenient, and accurate measurements. These wearables offer a wide range of functions, including tracking physical activity, monitoring vital signs, and even detecting certain health conditions. With their affordability, real-time data availability, and ease of use, wearable solutions have empowered individuals to take charge of their health and well-being in an accessible and accurate manner (*Channa, Popescu & Ciobanu, 2020*).

Thanks to the reliability of wearable solutions, monitoring physical activity, especially during mild to moderate PD stages, holds particular significance. Previous research has extensively investigated clinically significant daily physical activities, providing evidence related to physical performance metrics such as gait, muscle strength, and step counts. Individuals with PD typically exhibit lower levels of physical activity compared to individuals of similar age without the condition. Even those newly diagnosed and not yet on anti-Parkinson medications typically take around 5,000 steps per day if they are community-dwelling and able to walk independently. This falls short of the recommended 7,000 steps per day for older adults to maintain good health, highlighting the sedentary lifestyles often observed in people with PD (*von Rosen et al., 2021*).

Therefore, step counts over time serve as a crucial indicator of movement in telemonitoring, with previous studies consistently demonstrating that continuous monitoring of step counts may be an important way to quantify declining ambulatory behavior due to disease progression or improved ambulatory behavior resulting from rehabilitation, medical, or surgical interventions in persons with PD (*Cavanaugh et al., 2012*; *Master et al., 2022*). These findings provide valuable insights for informing public health recommendations on physical activity, aiming to enhance prevention strategies for better overall health outcomes (*Gong et al., 2023*).

## RELATED WORK

As discussed in the introduction, PD has a high prevalence, particularly among the elderly, with no current medications that offer a cure. Step counts serve as a valuable indicator of health and well-being in PD patients. Additionally, there is a strong correlation between reduced step counts and increased mortality rates among PD patients, as demonstrated in multiple studies. One study (*Stens et al., 2023*) analyzed 111,309 individuals across 12 studies and found that significant risk reductions in all-cause mortality were observed at 2,517 steps per day. Another study (*Inoue et al., 2023*) of 3,101 adults found a curvilinear dose–response relationship, showing that taking 8,000+ steps on multiple days per week lowered the 10-year risk of all-cause and cardiovascular mortality. Other studies, such

as *Banach et al. (2023)*, *Paluch et al. (2022)* and *Ahmadi et al. (2024)*, also indicated that a similar daily step count can reduce mortality risk, even in individuals with high sedentary time (*Stens et al., 2023*).

Several studies have focused on PD patients. For example, the WATCH-PD study assessed whether a combination of a smartwatch and smartphone could accurately measure early PD-related features (*Adams et al., 2023*). The Personalized Parkinson's Project (PPP), a single-center study of patients with early-stage PD in the Netherlands co-sponsored by Verily (Google Life Sciences), aimed to more precisely measure aspects of patient function and well-being (*Diao et al., 2022*). Another study investigated predictors of sustained physical activity in PD patients during the COVID-19 pandemic in Sweden, using step count data from the Actigraph GT3x, although it included only 63 subjects with a one-year follow-up (*Moulaee Conradsson et al., 2024*). In 2024, the WATCH-PD team published their 12-month findings in *npj Parkinson's Disease* (*Adams et al., 2023*). This study explored gait and tremor metrics from commercially available smartwatches and smartphones to assess treatment efficacy for early-stage PD. The study enrolled eighty-two individuals with early, untreated PD and fifty age-matched controls across seventeen US research sites. The authors emphasized the importance of step count as a critical indicator of patient burden and mortality. Over 12 months, they observed a mean daily step count decrease among PD patients not taking dopaminergic medication. However, these studies face limitations due to small sample sizes and brief follow-up durations.

A landmark study is the Parkinson Progression Marker Initiative (PPMI), an ongoing longitudinal study conducted by The Michael J. Fox Foundation. It aims to identify biomarkers for PD progression, improving disease understanding and enhancing therapeutic trials targeting PD modification (*Marek et al., 2011*). In 2018, Verily partnered with PPMI to use a multi-sensor investigational wearable device for continuous data collection on movement, physiological measures, and environmental factors (*Atri et al., 2022*).

Other studies, such as the BETA-PD project (*von Rosen et al., 2021*), aim to delineate physical activity (PA) profiles—namely "Sedentary," "Light Movers," and "Steady Movers"—of 301 individuals with PD to yield valuable insights for customizing PA interventions. *De Carvalho Lana et al. (2021)* investigated the validity of various mHealth devices (Google Fit, Health, STEPZ, Pacer, Fitbit) in estimating step counts among 34 individuals with idiopathic PD. Despite the growing interest in PD research, few studies have analyzed step count trends over extended periods across various age groups for both healthy individuals and PD patients. Although some studies using PPMI data have examined features such as brain changes (*Koros et al., 2023*), blood markers of inflammation (*Bartl et al., 2023*), and MDS-UPDRS Scores (*Holden et al., 2018*), minimal research has focused on physical activity. Studies (*Tsukita, Sakamaki-Tsukita & Takahashi, 2022*; *Amara et al., 2019*) examined self-reported activity using the Physical Activity Scale for the Elderly, but neither analyzed objective step count data from PPMI, possibly due to the large dataset from years of passive monitoring.

Moreover, large sample sizes in this area of research are lacking. The absence of an open-source web application also limits clinicians' ability to examine step count trends

over time among different groups, based on factors like sex, COVID-19 status, and hand dominance. Without such a tool, the exploration of potential correlations and insights that could contribute to personalized treatment approaches and further understanding of the disease is restricted.

Therefore, it is essential to conduct a comprehensive longitudinal analysis of step counts, incorporating potential risk factors, using extensive, robust, long-term follow-up data. The PPMI study, for example, offers a rich data source well-suited to such analyses, enabling more definitive conclusions about the implications of step count variations over time.

Wearable technology has ranked as the top fitness trend for the past two years, and forecasted financial trends suggest its use will continue to grow annually in the foreseeable future. One study (*Montes et al., 2020*) evaluated step-count functionality in wearable devices, providing strong evidence of their reliability and validity for tracking steps during free movement. To date, to the best of our knowledge, features (*e.g.*, gait, speech) were extracted using the same methods described in the baseline manuscript, with the exception of step count and a composite speech measure, which had not been previously analyzed (*Adams et al., 2023*).

Walking is the most common form of human movement, with a safe and efficient gait essential for maintaining independence throughout life. Gait is a key indicator of overall health, highlighting the importance of regular gait assessments in healthcare settings. However, gait is a complex, multidimensional activity that cannot be fully captured by a single parameter. Methods such as video motion capture systems, force platforms, and instrumented walkways are considered the gold standards for quantitative gait analysis. Although highly accurate, these methods are expensive, resource-intensive, and limited to stationary use in laboratory environments, which restricts their accessibility and practicality for widespread clinical use (*Werner et al., 2023*). Step count analysis using wearable devices offers a more affordable, user-friendly alternative to traditional gait analysis, enabling out-of-lab monitoring without the need for specialized equipment, trained personnel, or controlled conditions. These devices support unobtrusive, continuous tracking over long periods, capturing a more comprehensive view of natural gait and activity in daily life (*Werner et al., 2023*; *Gauthier-Beaupré & Grosjean, 2023*).

Although a few Shiny apps have been published online in our previous studies (*Ahmadi et al., 2024*; *Adams et al., 2023*; *Gu et al., 2023a*; *Gu et al., 2023b*), focusing on different diseases, such as cancer and kidney failure, no accessible online platform currently exists for describing baseline characteristics of PD patients or for modeling time-varying trends in step counts between PD patients and healthy participants. An online app offers greater convenience for clinicians and researchers, allowing real-time step count analysis from their own devices. This study addresses this gap by constructing a longitudinal model based on PPMI data and providing an online tool for clinical investigators to visualize these trends in detail. Our online tool is available at the following link: https://baran-shad.shinyapps.io/Rshiny_PPMIstepcounts/. Portions of this text were previously published as part of a preprint (*Gong et al., 2023*).

## METHODS

We requested access to the full breadth of individual-level PPMI longitudinal data from the Parkinson's Progression Markers Initiative. Details of the PPMI study can be found at the following link: https://www.ppmi-info.org/. A total of 353 subjects recorded their step counts, with a follow-up period exceeding 2.11 years. The cumulative number of steps taken by each subject was recorded on an hourly basis. Our analysis includes additional demographic and characteristic factors, such as sex, handedness (right, left/both), presence of Parkinson's disease in parents (yes or no), COVID-19 status, and cohort (PD or healthy control). After incorporating these variables and excluding subjects who were not successfully enrolled or resided outside the United States during the study, our final sample consists of 172 subjects.

### Matching

To retain as many observations as possible, we employed a 1:4 (32 control *vs.* 126 PD) matching approach to select subjects from the control and PD groups, utilizing the "matchit" package in R. The matching model was specifically designed using the nearest neighbors approach, matching the two groups based on factors such as age, sex, handedness, and COVID-19 status to align their propensity scores. The aim of this 1:4 matching was to pair each treated unit with the four closest control units based on their respective propensity scores. By implementing this matching methodology, we sought to enhance comparability between the treated and control groups, thereby improving the validity of our analysis and mitigating potential confounding factors. After merging with data on the presence of Parkinson's disease in parents, the final analysis dataset contains 32 healthy controls and 126 PD patients.

### Time binning

To account for the varying decimal values of ages represented in days, we employed the binning method in our study. This approach allowed for the creation of equally spaced time bins, specifically using intervals of 3 months. For example, age groups ranging from 66 to 67 were binned into intervals of 0.25. This resulted in age bin labels such as 66.25, 66.50, 66.75, and 67. By grouping the data in this manner, we were able to increase the sample size within each binned age interval and aggregate and cumulate the step counts within each age bin. Using binning in conjunction with cumulative step counts enabled us to capture and analyze overall activity patterns across different age intervals in a more manageable and informative way. This approach also facilitated the examination of age-related patterns and trends with greater resolution, providing valuable insights into the relationship between age and other variables under investigation.

### Baseline characteristics and descriptive plots

There are a total of 32 healthy control subjects and 126 PD patients (Table 1). Among these subjects, 61 are females and 97 are males. In previous studies (*Yust-Katz et al., 2008*; *Van der Hoorn et al., 2012*), left-handedness has been considered a potential risk factor for PD, so we incorporated handedness as a factor in our analysis. Our dataset includes 131

**Table 1  Baseline characteristics by healthy and PD groups.**

|  | Healthy control (*n* = 32) | Parkinson's disease (*n* = 126) | *P*-value |
|---|---|---|---|
| COVID-19 status |  |  | 0.14 |
|    No | 23 (72%) | 105 (83%) |  |
|    Yes | 9 (28%) | 21 (17%) |  |
| Sex |  |  | 0.8854 |
|    Female | 12 (37%) | 49 (39%) |  |
|    Male | 20 (63%) | 77 (61%) |  |
| Handedness |  |  | 0.1831 |
|    Right | 24 (75%) | 107 (85%) |  |
|    Left/Both | 8 (25%) | 19 (15%) |  |
| PD status of father |  |  | 0.0278 |
|    No | 32 (100%) | 109 (87%) |  |
|    Yes | 0 (0%) | 17 (13%) |  |
| PD status of mother |  |  | 0.0831 |
|    No | 32 (100%) | 115 (91%) |  |
|    Yes | 0 (0%) | 11 (9%) |  |

right-handed individuals and 27 left-handed individuals or those exhibiting ambidextrous tendencies. The study coincided with the outbreak of the COVID-19 pandemic, which poses significant risks, particularly for the elderly population, including individuals already diagnosed with Parkinson's disease. COVID-19 status was also collected, and there are 30 subjects who had COVID-19. Another unconventional risk factor considered is the PD status of both parents. A total of 17 patients has fathers with Parkinson's disease, and 11 patients have mothers with Parkinson's disease.

The *p*-values for the chi-square test are presented in Table 1 to determine whether there is a statistically significant difference between the two groups (healthy and PD) across different variables. Notably, most of these *p*-values are greater than 0.05, indicating no statistical significance. However, the variable "PD status of Father" has a *p*-value of 0.0278, indicating a significant difference in the PD status of fathers between the healthy and PD groups. For other baseline characteristics, all the *p*-values are > 0.05, suggesting that there is no significant difference in these characteristics between the PD and healthy groups, and we can assume that these variables are evenly distributed between the two groups.

In Fig. 1, the step counts over time are depicted for each subject, with the left side representing the healthy group and the right side illustrating the PD group. Moving on to Fig. 2, it showcases the change in step counts from baseline over time at a 3 months time intervals. The left side pertains to the healthy group (green box), while the right side corresponds to the PD group (gray box). In this figure, each pair of box plots represents the average daily steps over 3-month intervals for two groups. The seven pairs of box plots correspond to baseline (0 months) and subsequent time points at 3, 6, 9, 12, 15, and 18 months. Each box plot displays the distribution of step counts within each group at that
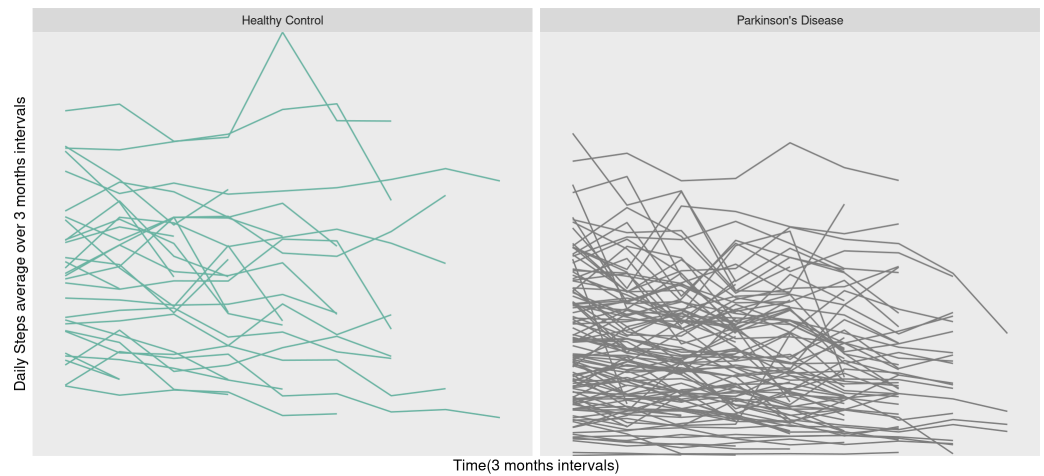
**Figure 1  Trajectories of daily steps over time.**

**Figure 2  Box plots of daily steps average over 3 months by different time (3 months intervals).**

specific interval, with the central line indicating the median step count and the height of the box representing the interquartile range, or the spread of middle values.

Overall, the healthy control group consistently shows higher median step counts and broader ranges across all time points compared to the PD group, indicating that healthy individuals maintain greater physical activity levels over time. Meanwhile, the PD group displays lower and relatively stable step counts, suggesting that mobility in PD patients is persistently limited, with minimal variation across the 18-month period. This trend highlights a clear and sustained difference in daily physical activity between the two groups, underscoring the impact of PD on long-term mobility. When comparing the two groups in Fig. 1, it becomes evident that individuals in the healthy group take more daily steps

compared to PD patients. Furthermore, we observe greater variability in step counts among PD patients over time, suggesting a potential decline in daily step counts as PD progresses.

Overall, in Fig. 2, these box plots clearly illustrate the difference in step counts between the two groups at every time point. The trend of higher step counts among individuals in the healthy group, as opposed to the PD group, is consistently depicted in both Figs. 1 and 2. Additional descriptive plots are available in our application for more comprehensive insights.

## GENERALIZED ADDITIVE MODEL RESULTS

The standard linear mixed-effects model is a common choice for longitudinal data analysis (*Hanff et al., 2024*). However, it may not effectively capture the curvature of time-varying trajectories, especially when quadratic relationships exist in repeated measurements. To address this limitation, we employed the generalized additive model (GAM) in our study. Unlike the standard linear mixed-effects model, which relies solely on linear feature contributions, GAM is constructed as a sum of smooth functions of features. This flexibility allows GAM to better accommodate and model complex, non-linear relationships within the data, enhancing the accuracy and interpretability of our analysis.

Here, we employ a GAM similar to the one used in *Adams et al. (2023)*. This choice addresses the challenges of handling step count data, which tends to be dense, irregularly sampled, and highly variable across individuals due to differing personal habits. While the WATCH-PD study collected limited step count data, they were unable to apply the framework to step count data and reported only empirical results. In our analysis of the richer PPMI step count data, we extend the model by incorporating an additional two years of step count tracking and integrating multiple covariates.

Our specific model is defined as follows:

$$\text{steps}_{i,j} = \beta_0 + \beta_1 * \text{Sex} + \beta_2 * \text{Handedness} + \beta_3 * \text{PD}_{\text{father}} + \beta_4 * \text{PD}_{\text{mother}} + \beta_5 * \text{Covid}$$
$$+ \beta_6 * \text{Cohort} + \beta_7 f_1(\text{age}_i, \text{cohort}) + \beta_8 f_{2j}(\text{age}_i, \text{subject}_j) + \varepsilon_{ij}.$$

In this model, step count serves as the response variable. The independent variables include sex, handedness, parental PD status (father and mother), COVID-19 status, and cohort (Parkinson's Disease *vs.* Healthy Control). We applied a cubic regression spline with three basis functions to model age within cohorts, as well as random effects to capture individual age trajectories. The temporal variable is modeled as a piecewise cubic Hermite interpolating polynomial (PCHIP) spline with four basis functions, allowing us to estimate distinct smooth functions across time (*Eilers & Marx, 1996*). For random effects, we introduced a random intercept and smooth temporal trajectories for each participant, accounting for participant-specific variation.

For complex models such as GAMs, standard goodness-of-fit metrics can be inadequate because the non-linear, smooth terms cannot be effectively summarized by simple linear correlations. Therefore, we employed residual diagnostics to assess the model's fit more
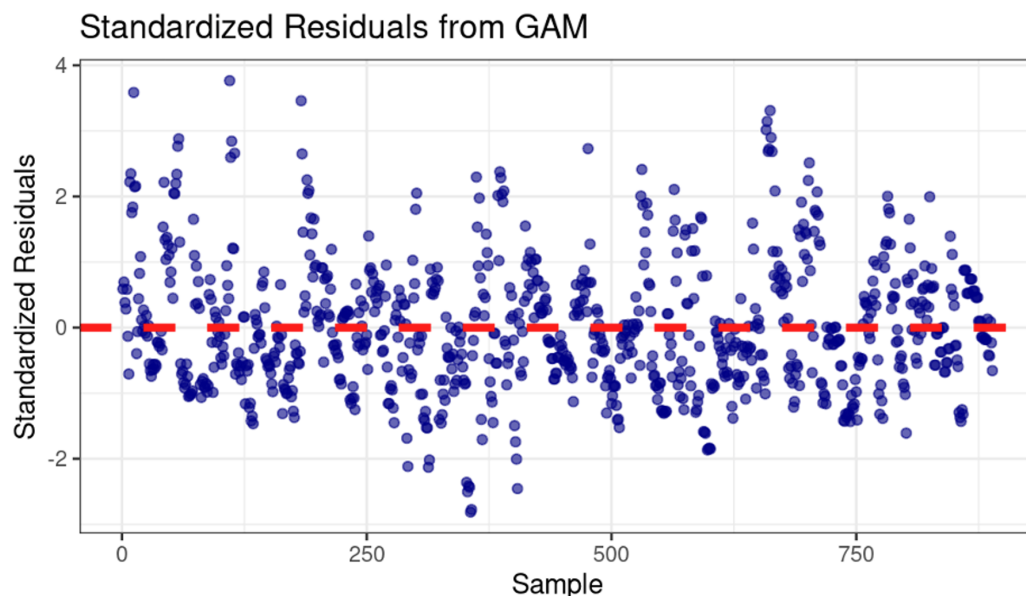
**Figure 3  Standardized residuals *vs.* sample for GAM model diagnostics.**
Full-size ☒ DOI: 10.7717/peerj.19519/fig-3

appropriately. As shown in Fig. 3, the residuals appear normally distributed and centered around zero, suggesting that the model captures the main structure of the data well.

The Q-Q plot of standardized residuals (Fig. 4) specifically examines the normality assumption. The plot shows that residuals near the center closely follow the theoretical quantile line, indicating that the model accurately represents central trends. However, a rightward deviation in the upper tail suggests a slight positive skew, likely influenced by extreme values from individual events, such as travel or other variations from normal activity.

The model result is displayed in Table 2. We observe that several variables, including sex ($p$-value = 0.03), handedness ($p$-value = 0.015), PD status of the father ($p$-value = 0.056, close to the significance level), COVID-19 status ($p$-value = 0.008), cohort ($p$-value < 0.0001), as well as the smoothing curve on age ($p$-value < 0.0001), and the random effect of the smoothing trajectory on age ($p$-value =0.0001), all demonstrate statistical significance or are close to significance in relation to step counts, except for the PD status of the mother ($p$-value = 0.47). Notably, among these features, handedness displays a positive coefficient, implying that right-handed individuals exhibit higher step counts compared to their left-handed counterparts or those who are ambidextrous. Furthermore, as the study extended beyond the COVID-19 period, it's noteworthy that COVID-19 status is another factor that statistically affects daily step counts.

Figure 5 below depicts the step counts estimated by our GAM model. These results are presented as smooth curves, with green curves representing the healthy group and gray curves for the PD group. Additionally, the original daily step data is overlaid as connected lines, providing background information to enhance the clarity and comprehensibility of the visualization. This visualization reinforces the observation that individuals within
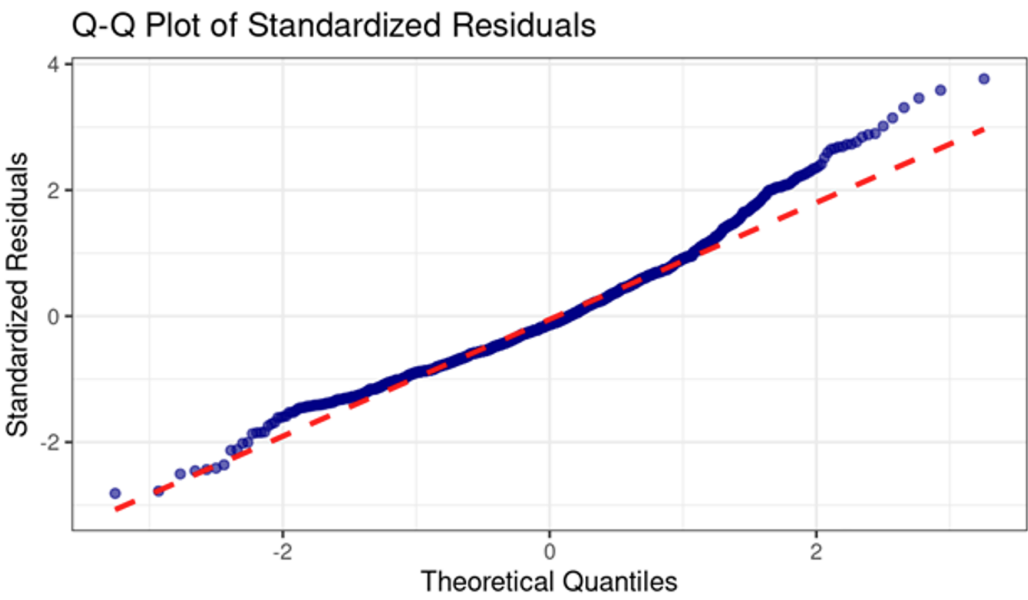
**Figure 4** Q-Q plot of standardized residuals for GAM model diagnostics.

Full-size ☑ DOI: 10.7717/peerj.19519/fig-4

**Table 2 GAM model results.**

| Variable | Coefficient | *p*-value |
|---|---|---|
| Sex | 388.8 | 0.03035 |
| Handedness (Yes: Right; No; Left/Both) | 564.5 | 0.01545 |
| PD status of father | 588.4 | 0.05633 |
| PD status of mother | −251.9 | 0.47156 |
| COVID-19 | 538.5 | 0.00871 |
| Cohort (PD: 1; Healthy control: 2) | 1,906.6 | <0.0001 |
| Smoothing on age | – | <0.0001 |
| Random effect of smoothing trajectory on age | – | 0.0001 |

the PD group exhibit significantly fewer step counts compared to the control group. Moreover, in both groups, step counts decrease with age, especially after age 70. The figures are displayed within the Shiny application, allowing clinicians to customize the presentation of the figures by making selections from a drop-down menu.

## DISCUSSION

Although numerous studies have concentrated on PD, there is a notable scarcity of investigations that delve into step counts of PD patients over extended periods. This limitation is primarily due to the large volume of daily data records required, making continuous follow-up challenging. Even in the extensive PPMI study, the number of participants in the healthy control group followed for up to two years remains limited compared to the PD groups, leading to imbalanced data between the two groups. This imbalance may constrain the model's ability to offer generalized conclusions. However,
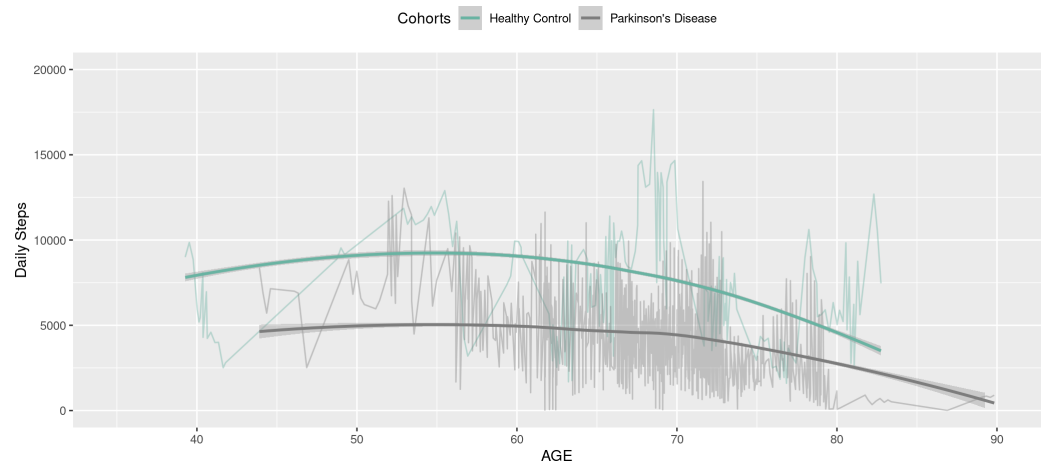
**Figure 5** Daily steps comparison between healthy (blue) and PD groups over age curves: GAM fitted curve; connected lines: original daily steps.

our analysis effectively handled the voluminous and complex daily step count data derived from the PPMI.

Throughout our study, several statistically significant factors affecting step counts were identified. These factors include gender, COVID-19 status, handedness, the PD status of the father, as well as the application of cubic spline regression for age and the inclusion of individual age as a random effect.

The GAM model revealed that, after adjusting for other variables, handedness carries a positive coefficient, suggesting that right-handed individuals record higher step counts than left-handed or ambidextrous individuals. This finding aligns with previous studies and may serve as a foundation for future research examining the imaging disparities between right- and left-handed individuals. Potential analysis methods could include exploring time-course differences in functional loci between these groups through advanced imaging techniques, such as image segmentation using location networks combined with 3D U-net or augmented pyramid networks like QAP-NET and RAR-U-Net, *etc* (*Wang et al., 2023*; *Wang & Voiculescu, 2021*; *Wang, Zhang & Voiculescu, 2021*; *Song, Gu & Kumar, 2023*; *Gu et al., 2024*).

Gender also plays a significant role in step counts, with men exhibiting higher daily step counts compared to women. The curvature of cubic regression over age is another significant factor, as daily step counts tend to decrease with age, particularly after the age of 70. Furthermore, COVID-19 status has proven to be a statistically significant factor impacting daily step counts. Another noteworthy finding pertains to the family history of PD: while the PD status of the father approached statistical significance ($p$-value = 0.056), the small subset of patients with a parental history of PD limits our ability to fully explore familial influences. Additional data collection is necessary to conduct a more comprehensive investigation of this factor.

The flexibility of the GAM model is one of its strengths, allowing for the integration of varying degrees of complexity within the data. By adjusting smoothness degrees, GAM

can effectively capture both linear and polynomial patterns. Moreover, GAMs inherently account for autocorrelation in longitudinal data, unlike linear mixed models, which rely on a predefined correlation matrix that may misrepresent the actual data structure. Additionally, the polynomial smoothing method in GAMs can automatically interpolate missing data, providing a more reliable approximation compared to traditional methods that assume specific missing data patterns (*e.g.*, missing completely at random or missing at random). However, the use of cubic regression splines as the basis for the GAM model adds complexity to model interpretation due to the flexibility of basis functions. While this enhances the model's ability to capture intricate patterns, it may make interpreting individual effects more challenging. Nevertheless, our goal is to identify statistically significant risk factors for step counts, and GAM proves to be an appropriate approach for this purpose.

To further enhance accessibility and facilitate deeper analysis, we developed an R Shiny web application. This tool allows investigators to interactively explore descriptive plots, visualize step count trajectories over time for different groups, and understand the model results directly through a user-friendly interface. The app provides clinicians and researchers with the flexibility to examine demographic insights and group key features for a more comprehensive understanding of the statistical significance of various risk factors on step counts in the context of Parkinson's Disease.

## CONCLUSION

Overall, our research builds upon our previously published study (*Marek et al., 2011*) and addresses a significant gap in the literature, as limited studies have focused on step counts—an essential metric for tracking PD patients. Step count data has been largely overlooked due to several challenges, including small sample sizes, short follow-up periods, complex study designs, and the practical difficulties associated with ensuring consistent daily use of digital wearables by PD patients.

We utilized the Parkinson's PPMI data, which includes longitudinal data on hundreds of patients followed for over two years, capturing a wide range of important features. By applying an advanced GAM, we were able to accommodate polynomial data structures and address missing data through smoothing interpolation. This approach is a significant improvement over the commonly used linear mixed models, which only account for linear relationships and rely on a predefined correlation matrix that often fails to accurately represent the data's true correlation structure.

Our findings are groundbreaking in identifying innovative features such as handedness and the PD status of the father as statistically significant factors affecting step counts. These factors have never been identified in similar studies before, adding valuable insights to the field of PD research. Additionally, we developed a free web application that allows users to visualize step count trajectories over time across different groups, offering a dynamic and user-friendly tool for enhanced data analysis and interpretation. This app enables investigators to conveniently access and interpret the model results directly *via* a shared link.

Despite the advancements in our study, several limitations should be acknowledged. The sample size of the control group is relatively modest, which could potentially constrain the robustness of comparative analyses. Additionally, the number of independent variables considered in our study is limited, and there may be other factors contributing to step counts that were not explored. To address these limitations and gain a more comprehensive understanding of the dynamics of PD progression, a larger and more diverse dataset, coupled with an extended follow-up period, is necessary. This would not only provide a broader context for the analysis but also offer the opportunity to explore the nuances of PD progression and its interactions with various factors in greater depth.

## REFERENCES

**Adams JL, Kangarloo T, Tracey B, O'Donnell P, Volfson D, Latzman RD, Zach N, Alexander R, Bergethon P, Cosman J, Anderson D, Best A, Severson J, Kostrzebski MA, Auinger P, Wilmot P, Pohlson Y, Waddell E, Jensen-Roberts S, Gong Y, Kilambi KP, Herrero TR, Ray Dorsey E, Parkinson Study Group Watch-PD Study Investigators and Collaborators. 2023.** Using a smartwatch and smartphone to assess early Parkinson's disease in the WATCH-PD study. *Npj Parkinson's Disease* **9(1)**:64 DOI 10.1038/s41531-023-00497-x.

**Ahmadi MN, Rezende LFM, Ferrari G, del Pozo Cruz B, Lee IM, Stamatakis E. 2024.** Do the associations of daily steps with mortality and incident cardiovascular disease differ by sedentary time levels? A device-based cohort study. *British Journal of Sports Medicine* **58(5)**:261–268 DOI 10.1136/bjsports-2023-107221.

**Amara AW, Chahine L, Seedorff N, Caspell-Garcia CJ, Coffey C, Simuni T, The Parkinson's Progression Markers Initiative. 2019.** Self-reported physical activity levels and clinical progression in early Parkinson's disease. *Parkinsonism and Related Disorders* **61**:118–125 DOI 10.1016/j.parkreldis.2018.11.006.

**Atri R, Urban K, Marebwa B, Simuni T, Tanner C, Siderowf A, Frasier M, Haas M, Lancashire L. 2022.** Deep learning for daily monitoring of Parkinson's disease outside the clinic using wearable sensors. *Sensors* **22(18)**:6831 DOI 10.3390/s22186831.

**Banach M, Lewek J, Surma S, Penson PE, Sahebkar A, Martin SS, Bajraktari G, Henein MY, Reiner Ž, da Browa AB, Bytyçi I. 2023.** The association between daily step count and all-cause and cardiovascular mortality: a meta-analysis. *European Journal of Preventive Cardiology* **30(18)**:1975–1985 DOI 10.1093/eurjpc/zwad229.

**Bartl M, Dakna M, Schade S, Otte B, Wicke T, Lang E, Starke M, Ebentheuer J, Weber S, Toischer K, Schnelle M, Sixel-Döring F, Trenkwalder C, Mollenhauer B. 2023.** Blood markers of inflammation, neurodegeneration, and cardiovascular risk in early Parkinson's disease. *Movement Disorders* **38(1)**:68–81 DOI 10.1002/mds.29257.

**Cavanaugh JT, Ellis TD, Earhart GM, Ford MP, Foreman KB, Dibble LE. 2012.** Capturing ambulatory activity decline in Parkinson disease. *Journal of Neurologic Physical Therapy* **36(2)**:51–57 DOI 10.1097/NPT.0b013e318254ba7a.

**Channa A, Popescu N, Ciobanu V. 2020.** Wearable solutions for patients with Parkinson's disease and neurocognitive disorder: a systematic review. *Sensors* **20(9)**:2713 DOI 10.3390/s20092713.

**De Carvalho Lana R, De Paula AR, Silva AF, Costa PH, Polese JC. 2021.** Validity of mHealth devices for counting steps in individuals with Parkinson's disease. *Journal of Bodywork and Movement Therapies* **28**:496–501 DOI 10.1016/j.jbmt.2021.06.018.

**Diao JA, Raza MM, Venkatesh KP, Kvedar JC. 2022.** Watching Parkinson's disease with wrist-based sensors. *NPJ Digital Medicine* **5(1)**:73 DOI 10.1038/s41746-022-00619-4.

**Eilers PHC, Marx BD. 1996.** Flexible smoothing with B-Splines and penalties. *Statistical Science* **11(2)**:89–121 DOI 10.1214/ss/1038425655.

**Ellis JM, Fell MJ. 2017.** Current approaches to the treatment of Parkinson's disease. *Bioorganic & Medicinal Chemistry Letters* **27(18)**:4247–4255 DOI 10.1016/j.bmcl.2017.08.053.

**Gauthier-Beaupré A, Grosjean S. 2023.** Understanding acceptability of digital health technologies among francophone-speaking communities across the world: a meta-ethnographic study. *Frontiers in Communication* **8**:1230015 DOI 10.3389/fcomm.2023.1230015.

**Gong Y, Wang Y, Wang Z, Li X, Gu Y. 2023.** Longitudinal analysis of step counts in Parkinson's disease patients: insights from a web-based application. *MedRxiv* DOI 10.1101/2023.11.22.23298898.

**Gu Y, Gong Y, Wang M, Jiang S, Li C, Yuan Z. 2023a.** Enhancing kidney failure analysis: web application development for longitudinal trajectory clustering. *MedRxiv* DOI 10.1101/2023.05.31.23290804.

**Gu Y, Wang M, Gong Y, Li X, Wang Z, Wang Y, Jiang S, Zhang D, Li C. 2023b.** Unveiling breast cancer risk profiles: a survival clustering analysis empowered by an online web application. *Future Oncology* **19(40)**:2651–2667 DOI 10.2217/fon-2023-0736.

**Gu Y, Wang Z, Wang Y, Gong Y, Li C. 2024.** Exploring longitudinal MRI-based deep learning analysis in Parkinson's patients - a short survey focus on handedness. *Cancer Insight* **3(1)**:99–116 DOI 10.58567/ci03010006.

**Hanff AM, Krüger R, McCrum C, Ley C, Acharya G, Aguayo G, Alexandre M, Ali M, Ammerlann W, Arena G, Bassis M, Batutu R, Beaumont K, Béchet S, Berchem G, Bisdorff A, Boussaad I, Bouvier D, Castillo L, Contesotto G, Bremaeker Nd, Dewitt B, Diederich N, Dondelinger R, Ramia NE, Ferrari A, Frauenknecht K, Fritz J, Gamio C, Gantenbein M, Gawron P, Georges L, Ghosh S, Giraitis M, Glaab E, Goergen M, Lope EGd, Graas J, Graziano M, Groues V, Grünewald A, Hammot G, Hansen L, Heneka M, Henry E, Henry M, Herbrink S, Herzinger S, Hundt A, Jacoby N, Jónsdóttir S, Klucken J, Kofanova O, Lambert P, Landoulsi Z, Lentz R, Longhino L, Lopes AF, Lorentz V, Marques TM, Marques G, Conde PM, May P, McIntyre D, Mediouni C, Meisch F, Mendibide A, Menster M, Minelli M, Mittelbronn M, Mtimet S, Munsch M, Nati R, Nehrbass U, Nickels S, Nicolai B, Nicolay JP, Noor F, Gomes CPC, Pachchek S, Pauly C, Pauly L, Pavelka L, Perquin M, Pexaras A, Rauschenberger A, Rawal R, Bobbili DR, Remark L, Richard I, Roland O, Roomp K, Rosales E, Sapienza S, Satagopam V, Schmitz S, Schneider R, Schwamborn J, Severino R, Sharify A, Soare R, Soboleva E, Sokolowska K, Theresine M, Thien H, Thiry E, Loo RTJ, Trouet J, Tsurkalenko O, Vaillant M, Vega C, Boas LV, Wilmes P, Wollscheid-Lengeling E, Zelimkhanov G. 2024.** Mixed effects models but not t-tests or linear regression detect progression of apathy in Parkinson's disease over seven years in a cohort: a comparative analysis. *BMC Medical Research Methodology* **24(1)**:23 DOI 10.1186/s12874-024-02301-7.

**Holden SK, Finseth T, Sillau SH, Berman BD. 2018.** Progression of MDS-UPDRS scores over five years in de novo Parkinson disease from the parkinson's progression markers initiative cohort. *Movement Disorders Clinical Practice* **5(1)**:47–53 DOI 10.1002/mdc3.12553.

**Inoue K, Tsugawa Y, Mayeda ER, Ritz B. 2023.** Association of daily step patterns with mortality in US adults. *JAMA Network Open* **6**(**3**):e235174 DOI 10.1001/jamanetworkopen.2023.5174.

**Jankovic J, Aguilar LG. 2008.** Current approaches to the treatment of Parkinson's disease. *Neuropsychiatric Disease and Treatment* **4**(**4**):743–757 DOI 10.2147/ndt.s2006.

**Koros C, Simitsi AM, Papagiannakis N, Bougea A, Prentakis A, Papadimitriou D, Pachi I, Beratis I, Stanitsa E, Angelopoulou E, Antonelou R, Bregianni M, Lourentzos K, Papageorgiou SG, Bonakis A, Trapali XG, Stamelou M, Stefanis L. 2023.** Serum uric acid as a putative biomarker in prodromal Parkinson's disease: longitudinal data from the PPMI study. *Journal of Parkinson's Disease* **13**(**5**):811–818 DOI 10.3233/JPD-230007.

**Mahlknecht P, Marini K, Werkmann M, Poewe W, Seppi K. 2022.** Prodromal Parkinson's disease: hype or hope for disease-modifying trials? *Translational Neurodegeneration* **11**(**1**):1–3.

**Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, Simuni T, Coffey C, Kieburtz K, Flagg E, Chowdhury S, Poewe W. 2011.** The Parkinson progression marker initiative (PPMI). *Progress in Neurobiology* **95**(**4**):629–635 DOI 10.1016/j.pneurobio.2011.09.004.

**Master H, Annis J, Huang S, Beckman JA, Ratsimbazafy F, Marginean K, Carroll R, Natarajan K, Harrell FE, Roden DM, Harris P, Brittain EL. 2022.** Association of step counts over time with the risk of chronic disease in the all of US research program. *Nature Medicine* **28**(**11**):2301–2308 DOI 10.1038/s41591-022-02012-w.

**Montes J, Tandy R, Young J, Lee S-P, Navalta J. 2020.** Step count reliability and validity of five wearable technology devices while walking and jogging in both a free motion setting and on a treadmill. *International Journal of Exercise Science* **13**(**7**):410–426 DOI 10.70252/IZMD1804.

**Moulaee Conradsson D, Leavy B, Hagströmer M, Franzén E. 2024.** Predictors of sustained physical activity during the COVID-19 pandemic in people with Parkinson disease in Sweden. *Journal of Neurologic Physical Therapy* **48**(**2**):75–82 DOI 10.1097/NPT.0000000000000455.

**Paluch AE, Bajpai S, Bassett DR, Carnethon MR, Ekelund U, Evenson KR, Galuska DA, Jefferis BJ, Kraus WE, Lee IM, Matthews CE, Omura JD, Patel AV, Pieper CF, Rees-Punia E, Dallmeier D, Klenk J, Whincup PH, Dooley EE, Gabriel KP, Palta P, Pompeii LA, Chernofsky A, Larson MG, Vasan RS, Spartano N, Ballin M, Nordström P, Nordström A, Anderssen SA, Hansen BH, Cochrane JA, Dwyer T, Wang J, Ferrucci L, Liu F, Schrack J, Urbanek J, Saint-Maurice PF, Yamamoto N, Yoshitake Y, Newton Jr RL, Yang S, Shiroma EJ, Steps for Health Collaborative. 2022.** Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. *The Lancet Public Health* **7**(**3**):e219-e228 DOI 10.1016/S2468-2667(21)00302-9.

**Puigròs M, Calderon A, Perez-Soriano A, De Dios C, Fernandez M, Colell A, Marti MJ, Tolosa E, Trullas R. 2022.** Cell-free mitochondrial DNA deletions in idiopathic, but not LRRK2, Parkinson's disease. *Neurobiology of Disease* **174**:105885 DOI 10.1016/j.nbd.2022.105885.

**Song J, Gu Y, Kumar E. 2023.** Chest disease image classification based on spectral clustering algorithm. *Research Reports on Computer Science* **2(1)**:77–90 DOI 10.37256/rrcs.2120232742.

**Stens NA, Bakker EA, Mañas A, Buffart LM, Ortega FB, Lee DC, Thompson PD, Thijssen DHJ, Eijsvogels TMH. 2023.** Relationship of daily step counts to all-cause mortality and cardiovascular events. *Journal of the American College of Cardiology* **82(15)**:1483–1494 DOI 10.1016/j.jacc.2023.07.029.

**Stoker TB, Barker RA. 2020.** Recent developments in the treatment of Parkinson's disease. *F1000Research* **9**:862 DOI 10.12688/f1000research.25011.2.

**Tsukita K, Sakamaki-Tsukita H, Takahashi R. 2022.** Long-term effect of regular physical activity and exercise habits in patients with early Parkinson disease. *Neurology* **98(8)**:e859–e871 DOI 10.1212/WNL.0000000000201035.

**Van der Hoorn A, Burger H, Leenders KL, De Jong BM. 2012.** Handedness correlates with the dominant Parkinson side: a systematic review and meta-analysis. *Movement Disorders* **27(2)**:206–210 DOI 10.1002/mds.24051.

**von Rosen P, Hagströmer M, Franzén E, Leavy B. 2021.** Physical activity profiles in Parkinson's disease. *BMC Neurology* **21**:71 DOI 10.1186/s12883-021-02101-2.

**Wang Y, Feng A, Xue Y, Shao M, Blitz AM, Luciano MD, Carass A, Prince JL. 2023.** Investigation of probability maps in deep-learning-based brain ventricle parcellation. In: *Proceedings of SPIE - The International Society for Optical Engineering 12464:96.* Bellingham: SPIE, 565–570 DOI 10.1117/12.2653999.

**Wang Z, Voiculescu I. 2021.** Quadruple augmented pyramid network for multi-class COVID-19 segmentation via CT. In: *2021 43rd annual international conference of the IEEE engineering in medicine & biology society (EMBC), 2956–59.* Piscataway: IEEE DOI 10.1109/EMBC46164.2021.9629904.

**Wang Z, Zhang Z, Voiculescu I. 2021.** RAR-U-Net: a residual encoder to attention decoder by residual connections framework for spine segmentation under noisy labels. In: *2021 IEEE international conference on image processing (ICIP), 21–25.* Piscataway: IEEE DOI 10.1109/ICIP42928.2021.9506085.

**Werner C, Hezel N, Dongus F, Spielmann J, Mayer J, Becker C, Bauer JM. 2023.** Validity and reliability of the Apple Health app on iPhone for measuring gait parameters in children, adults, and seniors. *Scientific Reports* **13(1)**:5350 DOI 10.1038/s41598-023-32550-3.

**Willis AW, Roberts E, Beck JC, Fiske B, Ross W, Savica R, Van Den Eeden SK, Tanner CM, Marras C, Alcalay R, Schwarzschild M, Racette B, Chen H, Church T, Wilson B, Doria JM. 2022.** Incidence of Parkinson disease in North America. *Npj Parkinson's Disease* **8(1)**:170 DOI 10.1038/s41531-022-00410-y.

**Yust-Katz S, Tesler D, Treves TA, Melamed E, Djaldetti R. 2008.** Handedness as a predictor of side of onset of Parkinson's disease. *Parkinsonism & Related Disorders* **14(8)**:633–635 DOI 10.1016/j.parkreldis.2008.04.009.