CHANGES IN BODY COMPOSITION BEFORE AND AFTER PARKINSON'S DISEASE DIAGNOSIS - A LONGITUDINAL ANALYSIS

By

Shengfang (Eleanor) Song

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biostatistics – Master of Science

2020

ABSTRACT

CHANGES IN BODY COMPOSITION BEFORE AND AFTER PARKINSON'S DISEASE DIAGNOSIS - A LONGITUDINAL ANALYSIS

By

Shengfang (Eleanor) Song

Background: Weight loss is common in Parkinson's disease (PD). However, little is known when it starts, how it changes as PD progresses, and whether there is a differential loss of lean or fat mass.

Methods: In the Health ABC study (n=3075, age range 70-79), body composition was assessed by dual-energy X-ray absorptiometry from baseline (year 1) to year 10. For each PD patient at each year, we calculated the difference between body composition measures and their expected values had they not developed PD, and then realigned the time scale in reference to the year of PD diagnosis. Using both inverse variance-weighted fixed-effects linear model and linear mixed model with cross random effects, we examined the trend of change for each body composition measure before and after PD diagnosis.

Results: During an average of 7.8 years of follow-up, a total of 81 PD patients were identified. PD patients were more likely to be men (59.3% vs. 48.2%) and white (74.1% vs. 57.9%) than non-PD participants. Compared with their expected weight, PD cases began to lose weight and fat mass several years before PD diagnosis(p-values for trend were 0.003 and <0.001 for total body mass and fat mass, respectively).

Conclusions: In this longitudinal analysis, we found persistent weight loss, predominantly in fat mass, in PD patients starting a few years before diagnosis.

Keywords: Parkinson's disease; Body composition; Lean-mass; Fat-mass; Weight.

Dedicated to my parents and my husband Jisheng for their love, understanding, and continuing support.

ACKNOWLEDGEMENTS

I would like to first express my deep and sincere gratitude to my thesis advisor, Dr. Honglei Chen, for organizing and planning the ideas and designs of this study and providing invaluable guidance throughout the research. He really opened the door of neurodegenerative disorder and longitudinal data analysis to me and provided me valuable opportunities to learn and work with public health data. I am also very grateful to the support and instruction from my primary advisor Dr. Dorothy Pathak, who has been providing lots of useful advice and encouragement during my graduate study. I would like to thank Dr. Zhehui Luo and Dr. Chenxi Li for providing strong statistical expertise and precious feedback on the methodology applied in my thesis. It was a great privilege and honor to work with and learn from them.

I would also like to thank the faculty and staff in the Department of Epidemiology and Biostatistics for their invaluable advice and support during these two years of study.

Many thanks to my lab members Zichun Cao, Aiwen Yang, Yaqun Yuan, Keran Wang, and Pooja Skukla for their support in data analysis, study conducting, and thesis writing. Thank to my cohort of the biostatistics program, I cannot imagine how I can overcome all the difficulties without working together with them.

Finally, I would like to acknowledge my friends and family, I cannot accomplish this without their love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder that is clinically diagnosed by the presence of motor dysfunction such as bradykinesia, rigidity, tremor, and postural instability. PD affects approximately 1 million individuals 65 years or older in the US and is projected to increase 25% by 2030, and thus presents an increasing burden on healthcare systems.[1] Sporadic PD often takes decades to develop and is typically diagnosed when its cardinal motor signs become clinically evident late in life. In the past two decades, we learned that a wide range of nonmotor symptoms may develop many years prior to PD diagnosis. A thorough understanding of these nonmotor symptoms may be critical both to early disease recognition and a better understanding of disease development. To date, the best-studied nonmotor symptoms are olfaction loss, REM sleep behavior disorder, depression, cognitive changes, and constipation[2]. In comparison, other symptoms have received less attention.

Weight loss is common in PD cases[3–7] and is often associated with severe motor dysfunction, higher comorbidity, poor physical and mental function, frailty, and increased mortality[5,8–12]. Interestingly, two recent prospective studies reported that lower body mass index was associated with a higher risk of PD, suggesting that weight loss may start prior to PD diagnosis[13,14]. To the best of our knowledge, only one previous population-based study has used repeatedly obtained body weights to assess when and how body weight changes before and after PD diagnosis. In the Health Professionals Follow-up Study, Chen et al. reported that PD cases tend to begin weight loss 2-4 years prior to PD diagnosis and the trend persisted after PD diagnosis. On an average, compared to

individuals who did not have PD, cases lost 3.86 kg more of body weight throughout the follow-up[4]. However, this study used self-reported body weight, and did not differentiate loss of lean versus fat mass which by itself may have implications for the health and survival of older adults. We therefore utilized objective measures of body composition, including total, lean, and fat mass, repeatedly assessed in a community-based biracial cohort to examine changes in body composition in PD cases up to a decade before and after PD diagnosis.

METHODS AND MATERIALS

Study Design and Study Population

The Health, Aging, and Body Composition (Health ABC) study was designed to study risk factors for functional declines in older adults, especially changes in body composition, behavioral and physiological conditions in the context of aging. The study recruited 3075 well-functioning older adults (age 70-79, 51.4% women, 41.6% blacks) in 1997-1998 living in Pittsburgh, Pennsylvania, and Memphis, Tennessee. Inclusion criteria included 1) no difficulty in walking 1/4 mile or climbing up 10 steps; 2) no mobility-related difficulty in performing every-day tasks; 3) no intention to move out of the study area in the next three years. Exclusion criteria included 1) active cancer treatment in the past three years; 2) current participation in a lifestyle intervention trial. Participants enrolled in the study by completing the Year 1 baseline clinic visit from April 1997 to June 1998 and their health and survival were monitored for up to 17 years with annual or biennial clinic visits, semiannual/quarterly phone calls, and hospitalization and death surveillance. Body compositions were measured, using dual-energy X-ray absorptiometry (DXA), at clinic visits on an annual basis from year 1 to year 6, and then biennially from year 6 to year 10, and then again year 16. In this study, we excluded data from the 16th year because of the long gap between year 10 and 16, and few PD cases were alive and participated in the year 16 clinic visit. We followed eligible participants from baseline until the date of death, last contact, or year 10 clinic visit, whichever came first, with an average of 7.8 years. All participants provided written informed consent and the study protocol was approved by the Institutional Review Boards at University of

Pittsburgh, University of Tennessee – Memphis, NIH[15]. This specific secondary data analysis was IRB exempted as non-human research by the Michigan State University.

Measurements

Body composition Body weight or total body mass, lean mass, and fat mass were acquired from total body scans by using fan-beam dual-energy X-ray absorptiometry (Hologic QDR 4500A version 8.20a, Hologic, Waltham, MA) with dual-energy X-ray absorptiometry software (Hologic). The percentage of fat mass was calculated by dividing the total fat mass by the total body mass. The validity and reproducibility of the DXA scanner have been previously reported.[16,17] The reliability was monitored by quality assurance measurements included the use of daily and cross-calibration phantoms at both study sites[18]. It has been determined that for both study sites, Hologic QDR 4500A overestimated fat-free mass by 5.4% compared to criterion methods,[19] and underestimated fat mass and percent fat consequently. Also, the Pittsburgh DXA scanner overestimated total mass by about 2% relative to scale weight[19]. In this study, DXA data for all years were corrected and recalculated accordingly. Identical scan protocols were employed for all participants from both sites[20].

PD ascertainment We identified potential PD cases from multiple data sources. At the clinic visits of years 1, 2, 3, 5, and 6, the study asked participants to show all medications that they used in the past two weeks and, for each medication, to report the name, dose and frequency of use, reason for use, and year of first use. In years 8, 10, and 11, the study asked participants to report medication use in the past thirty days, including name, current use, frequency of use, and duration of use (years 8 and 10). The study also

queried about physician diagnosis of PD at enrollment and in a sub-study at year 13. In addition, the study conducted comprehensive hospitalization and death surveillance. For each hospitalization, up to twenty diagnoses were summarized on the discharge form according to the International Classification of Diseases-9-Clinical Modification (ICD-9-CM). This discharge summary, along with records of medical history and physical exams, was subsequently reviewed by a local event adjudicator, and diseases presented at the hospitalization were adjudicated. For each death event, the cohort conducted an exit interview with a knowledgeable proxy who provided information on physical functioning of the study participant while alive and details of the event. These data, together with other relevant information such as recent hospitalizations, were centrally reviewed by a team of experts and the underlying cause of death was adjudicated by consensus. At the time of the present study, hospitalization and death surveillances were complete through August 2012[21].

In 2015, we retrospectively adjudicated a total of 81 PD cases by comprehensively reviewing the above-referenced health data. We first identified a total of 156 participants who had at least one of the following: 1) reported the use of antiparkinsonian medications (carbidopa/levodopa, dopamine agonists, monoamine oxidase B inhibitors, amantadine, or anticholinergic drugs) at any of the medication surveys; 2) self-reported PD diagnosis; 3) local adjudication of PD as the cause of hospitalization or ICD-9 code of PD (332.0) on the discharge form; 4) PD as the centrally adjudicated cause of death or reported on the proxy interview following death. For each potential case, two experienced movement disorder specialists independently reviewed their PD relevant data over the entire follow-up, accounting for the number of sources that indicated a PD

diagnosis, internal consistency within each source, and evidence against PD diagnosis. Internal consistency was met if the participants reported PD medications in multiple years, reported PD as the reason for taking the medication, or local adjudication of PD as the reason for multiple hospitalizations. Otherwise, if the participants reported neuroleptic use before PD medication, restless leg syndrome as the reason for taking PD medication, use of dementia medication, or hospitalization with dementia prior to or at the first evidence of PD, we treated them as contradictory evidence. The final PD cases were defined as following: 1) at least two independent sources of PD identification without any contradictory evidence (n = 58) or 2) one source of PD identification with clear internal consistency and no contradictory evidence (n = 13). As hospitalization and death became the only sources to identify potential PD cases after the last medication survey in year 11, we also included potential cases whose only source of information was from adjudicated hospitalization (n = 7) or death (n = 3) after the last available medication survey. We further defined the year of diagnosis as the first year that PD medication (n = 48) or diagnosis (n = 1) was reported. If PD was first identified by hospitalization (n = 29) or death (n = 3), we defined year of diagnosis as the middle point of first identification and the previous year of medical survey without reports of PD medication use.[21] The current analysis included all PD cases identified in the cohort, defined as a dichotomous variable with 1 for participants who developed PD (cases) and 0 for those who did not develop PD (non-PD participants) during the follow-up. For simplicity, we used cases and non-PD participants to refer these two groups.

Years from PD diagnosis In the analysis, we used the year of PD diagnosis as the reference time point. For each PD case at each measurement, we calculated the number of years in reference to the time of PD diagnosis by subtracting the calendar year of PD diagnosis from the year of measurement. For example, if a case got PD diagnosis in 2003 and he had body composition measured five times respectively in 1999, 2001, 2003, 2006, 2008, the corresponding time in reference to PD diagnosis will be 4 and 2 years before, the year at, and 3 and 5 years after PD diagnosis, and the corresponding value for the variable Timepoint will be -4, -2, 0, 3, and 5.

Covariates We considered the following covariates in this study - age, sex, race, CVYear (clinic visit year), study site, and standing height. Age was defined as the age at each clinic visit. Sex, race, and study site were reported at baseline. Race is a two-level variable with 1 for white and 2 for black. Study site has two levels - Memphis or Pittsburgh. CVYear represents the number of years starting from baseline, which has values from 1 to 10 representing the 1^{st} (baseline) to the 10^{th} year of clinic visit. Body height was measured to the nearest 0.1 cm by using a wall-mounted stadiometer at the clinic visits of years 1, 4, 6, 8, and 10. At each visit, standing height measurements were taken up to 4 times. The final body height was determined as the following: 1) the average if 2-3 measurements were obtained; or 2) the average of the last 2 measurements if all 4 measurements were made; or 3) single measurement if only one was taken. For study visits that height was not measured, we used measures from the nearest previous visit.

STATISTICAL ANALYSIS

Summary Statistics

We first conducted a multi-variate longitudinal analysis to explore the yearly change trajectory of the body composition responses (total body mass, total body lean mass, total body fat mass, and percentage of fat mass) among PD cases and non-PD participants. To further examine how body composition changed in PD cases before and after the diagnosis, we used a statistical approach similar to what we published previously.[4] The goal is to describe changes in body composition in cases relative to what the changes in body composition would have been had they not developed PD. Briefly, at each clinic visit year, we first fitted among non-PD participants a linear regression model of each body composition measure on age, sex, race, study site, and height and obtained the beta-coefficients. We then applied these coefficients to cases and calculated their expected body composition measures and interpret it as their counterfactual body composition measures had they not developed PD (thereafter referred as expected measures). We finally calculated residuals which, by definition, represent the difference between their actual body composition measures in reference to the expected values in that particular calendar year had they not developed PD. After conducting this analysis for each clinic visit that body compositions were measured, we compiled all available residuals for PD cases and realigned the time scale using the year of PD diagnosis as reference. We subsequently fitted both an inverse variance-weighted linear regression model and a mixed model with crossed random effects to examine trends of the above residuals across all years in reference to PD diagnosis. For all the analyses in the present study, we assumed missing completely at random

and excluded those with missing responses from the corresponding analysis. The longitudinal analysis, the linear regressions for measures at each clinic visit, and the inverse variance weighted modeling were conducted with SAS version 9.4 (SAS Systems Inc., Cary, NC) and the mixed model with crossed random effects was conducted with Stata version 16.0 (StataCorp, College Station, TX). A p-value of less than 0.05 was considered statistically significant.

Longitudinal Analysis on the Timescale of Clinic Visit Year

This is an exploratory analysis for the yearly change trajectories of the body compositions among participants who had (cases) and had not developed PD (non-PD participants) during the follow-up. We first fitted a mixed model with a separate mean for each of the 16 $PD_i * CVYearc_i$ combinations (thereafter referred as unstructured mean), together with separate unstructured variance-covariance matrices for cases and non-PD participants (thereafter referred as separate unstructured variance-covariance matrix) as:

$$BC_{mi} = \beta_{m0_1} + \beta_{m1_1} * PD_i + \beta_{m2_1} * CVYearc_i + \beta_{m3_1} * PD_i * CVYearc_i + \beta_{m4_1} * Age_i + \beta_{m5_1} * Age_i * CVYearl_i + \beta_{m6_1} * Sex_i + \beta_{m7_1} * Race_i + \beta_{m8_1} * Height_i + \beta_{m9_1} * Height_i * CVYearl_i + \beta_{m10_1} * Site_i + \varepsilon_{mi_1} \quad (1)$$

where BC_{mi} represents the vector of yearly repeated body composition measurement *m* (taking from total body mass, total body lean mass, total body fat mass, and percentage body fat) for participant *i*; $CVYearc_i$ and $CVYearl_i$ stand for the categorical and linear clinic visit year for participant *i*, respectively; $\varepsilon_{mi_1} \sim N(0, \Sigma_{m_1})$ represents the vector of error components for body composition *m* in equation (1) where Σ_{m_1} is the corresponding variance-covariance matrix.

We then tried to further reduce both the mean and variance-covariance structure to fit a more parsimonious model. Model comparisons were conducted using likelihood ratio tests and the unstructured mean and separate unstructured variance-covariance matrix model was selected to be the final model for all the body composition responses (Table 1).

Table 1 Longitudinal mixed model comparisons by likelihood ratio tests

	Total mass			Total lean mass			Total fat mass			Percentage of fat mass		
	Chi-Square Statistic*	DF	P-values	Chi-Square Statistic*	DF	P-values	Chi-Square Statistic*	DF	P-values	Chi-Square Statistic*	DF	P-values
Mean Structure												
Unstructured† vs. Different quadratic trends	42	10	<.01	66.8	10	<.01	57.3	10	<.01	79.6	10	<.01
Unstructured vs. Different cubic trends	32.3	8	<.01	42.7	8	<.01	56	8	<.01	79	8	<.01
Unstructured vs. Different quartic trends	27.7	6	<.01	23.8	6	<.01	16.8	6	0.01	10.6	6	0.10
Unstructured vs. Different quintic trends	23.5	4	<.01	22.3	4	<.01	11	4	0.03	7.7	4	0.10
Unstructured vs. Different sextic trends	23.4	2	<.01	20.9	2	<.01	8.9	2	0.01	1.9	2	0.39
Variance-Covariance Structure												
Unstructured vs Heterogeneous Toeplitz	2328.8	42	<.01	1216.1	42	<.01	2471.6	42	<.01	1407.8	42	<.01
Common vs. Separate‡ variance-covariance	63.8	36	<.01	54.7	36	0.02	62.9	36	<.01	49.9	36	0.06

* Calculated from the difference of the -2log likelihood for the nested model and the full model.

† Separate mean for each of the 16 $PD_i * CVYearc_i$ combinations.

‡ Whether the variance-covariance structure was the same for cases and non-PD participants.

Trend Analysis on the Timescale of Years from PD Diagnosis

Step 1: Linear Regression The purpose of this step is to obtain the residuals for PD cases for the subsequent analyses in step 2. At each clinic visit year, we first fitted a linear regression in non-PD participants:

$$BC_{mijn} = \beta_{m0_2} + \beta_{m1_2} * Age_{ij} + \beta_{m2_2} * Sex_i + \beta_{m3_2} * Race_i + \beta_{m4_2} * Height_{ij} + \beta_{m5_2} * Site_i + \varepsilon_{mij_2} \quad (2)$$

where BC_{mijn} represents body composition measurements of non-PD participant *i* for body composition *m* (taking from total body mass, total body lean mass, total body fat mass, and percentage body fat) measured at clinic visit year *j*; $\varepsilon_{mij_2} \sim N\ (0, \sigma^2_{mij_2})$, represents the corresponding error component for body composition *m* in equation (2).

Assuming the effects of age, sex, race, height, and site on body composition are the same for PD cases and non-PD participants within each clinic visit year, we then applied the beta-coefficients obtained from equation (2) to PD cases and calculated their predicted values if they had not developed PD:

$$\widehat{BC}_{mij} = \beta_{m0_2} + \beta_{m1_2} * Age_{ij} + \beta_{m2_2} * Sex_i + \beta_{m3_2} * Race_i + \beta_{m4_2} * Height_{ij} + \beta_{m5_2} * Site_i \quad (3)$$

where $\widehat{BC}_{mij}$ denotes the predicted response for PD case *i* calculated for body composition *m* at clinic visit year *j*; $\beta_{m0_2} \sim \beta_{m5_2}$ were the corresponding coefficients obtained from equation (2).

Residuals were calculated by subtracting the predicted body composition values from the observed measures, which represent the differences in body compositions in PD cases in reference to their expected values if they had not had PD:

$$\boldsymbol{Residual_{mij} = BC_{mij_{observed}} - \widehat{BC}_{mij}} \qquad (4)$$

where $Residual_{mij}$ denotes the residual of PD case *i* calculated for body composition *m* at clinic visit year *j*; $BC_{mij_{observed}}$ is the observed body composition *m* for PD case *i* measured at clinic visit year *j*.

Step 2: Trend Analysis To evaluate the change trajectory of body compositions across years from PD diagnosis in PD cases relative to if they had not had PD, both an inverse variance-weighted linear regression model and a mixed model with crossed random effects were fitted on the residuals calculated from step 1 ($\boldsymbol{Residual_{mij}}$ from equation (4)). Only PD cases were included in the analyses in step 2.

Inverse variance-weighted (IVW) fixed-effects linear model

The inverse variance-weighted fixed-effects model is predominantly used in multivariate statistics and meta-analyses for synthesizing data from different sources[22–25]. The idea is that assuming the predictions from larger sample size and smaller variance groups are more precise, and then allocating larger weights to these groups to make them have greater influences on the analyses than the groups with smaller sample size and/or higher variability[26].

In the present analysis, we were mainly interested in the average of residuals at each timepoint (years from PD diagnosis) and their change trajectory. Therefore, we fitted a

fixed-effects linear model to examine the trajectory of the averages of residuals with timepoint:

$$MResiduals_{mk} = \beta_{m0_5} + \beta_{m1_5} * Timepoint + \beta_{m2_5} * Timepoint^2 + \varepsilon_{m_5} \qquad (5)$$

where $\boldsymbol{MResiduals_{mk}}$ is the mean of residuals for body composition *m* at *k*th timepoint; Timepoint represents years from PD diagnosis; $\boldsymbol{\varepsilon_{m_5}} \sim \boldsymbol{N}\ (\mathbf{0}, \boldsymbol{\sigma^2_{m_5}})$, is the error component for body composition *m* in equation (5).

To reduce the bias that might be introduced by the unbalanced sample sizes between timepoint groups in our data (Figure 1), we first excluded those timepoints with a sample size less than or equal to 5 from the analyses (excluding timepoint -13, -12, 10, 11, 12, 13, and 15). The example dataset after the exclusion is showed in Table 2. We then borrowed the idea of inverse variance weighting to allocate a weight for each timepoint group based on the reciprocal of its standard error of the mean to allocate higher weights to those groups with smaller variances. The weights for each timepoint group were given by[26,27]:

$$\boldsymbol{w_{mk}} = \frac{\mathbf{1}}{\boldsymbol{SE^2_{mk}}} \qquad (6)$$

where $\boldsymbol{w_{mk}}$ and $\boldsymbol{SE_{mk}}$ denote the weight and the standard error of the mean at *k*th timepoint for body composition *m*, respectively.

Thus, the weighted least-squares estimates from equation (5) will achieve minimum variance. Under the assumption that $\boldsymbol{\varepsilon_{m_5}}$ is independent with each other and with timepoint and is normally distributed $\sim \boldsymbol{N}\ (\mathbf{0}, \boldsymbol{\sigma^2_{m_5}})$, the weighted least-squares estimates can be derived as both the maximum-likelihood estimator and the best linear unbiased

estimators (BLUE)[28–30]. The predicted means and their 95% confidence intervals were then obtained based on these estimators. The β coefficient is given by:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m_5}} = (\boldsymbol{X}'_{\boldsymbol{m}}\boldsymbol{W}_{\boldsymbol{m}}\boldsymbol{X}_{\boldsymbol{m}})^{-1}(\boldsymbol{X}'_{\boldsymbol{m}}\boldsymbol{W}_{\boldsymbol{m}}\boldsymbol{Y}_{\boldsymbol{m}}) \quad (7)$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{m_5}}$ is a 3x1 matrix of estimators of the regression parameters in the regression model in equation (5) for body composition *m*, $\boldsymbol{X}_{\boldsymbol{m}}$ is a 21x3 matrix of independent variables for body composition *m*; $\boldsymbol{W}_{\boldsymbol{m}}$ is a 21x21 diagonal matrix of w_i for body composition *m;* $\boldsymbol{Y}_{\boldsymbol{m}}$ is a 21x1 matrix of means of residuals for body composition *m*, taking the example dataset in table 2 for instance).

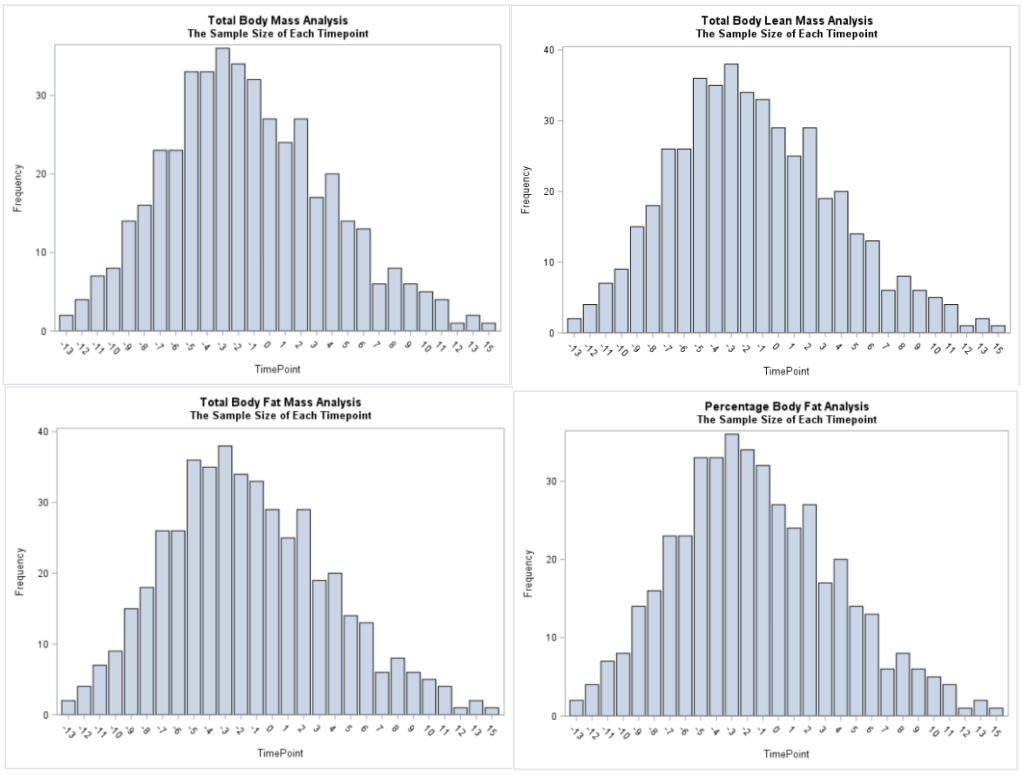


Figure 1 The sample size of timepoint groups for various body composition measures

Table 2 Example dataset for the IVW fixed-effect linear regression

TimePoint	Sample Size	Mean of Residuals	Standard Error	Weight
-11	7	4.17	6.11	0.03
-10	8	9.29	5.22	0.04
-9	14	3.11	3.56	0.08
-8	16	1.92	2.79	0.13
-7	23	0.67	1.94	0.26
-6	23	2.41	2.35	0.18
-5	33	0.51	1.46	0.47
-4	33	-2.13	1.83	0.30
-3	36	0.36	1.73	0.33
-2	34	-1.45	1.69	0.35
-1	32	2.39	2.10	0.23
0	27	-4.93	1.85	0.29
1	24	-0.44	2.10	0.23
2	27	-2.17	2.12	0.22
3	17	0.800	2.61	0.15
4	20	-4.39	2.77	0.13
5	14	-5.55	3.14	0.10
6	13	-0.05	3.62	0.08
7	6	-8.08	4.79	0.04
8	8	-5.35	4.54	0.05
9	6	-5.80	4.19	0.06

Timepoint, years from PD diagnosis.

Other assumptions should be satisfied for IVW fixed-effect linear regression including[24,26]: 1) $\boldsymbol{MResiduals_{mk}}$ are independent with each other; 2) $\boldsymbol{\varepsilon_{m_5}}$ are independent with each other and with timepoint; 3) the conditional variance of $\boldsymbol{MResiduals_{mk}}$ is $\boldsymbol{\sigma^2/w_{mk}}$.

However, the above assumptions might be violated in our data as the means of residuals at different timepoints were correlated since each PD case contributed measurements for multiple timepoints. Thus, $\boldsymbol{\varepsilon_{m_5}}$ were also dependence with each other. Such violations could bias the predicted means and return an incorrect standard error for $\boldsymbol{\widehat{\beta}_{m_5}}$. Thus, mixed models with crossed random effects were conducted to obtain more precise estimates.

Mixed model with crossed random effects

The mixed model can capture the heterogeneous variances and covariances in the data and is, therefore, more powerful compared to the repeated-measures ANOVA especially when the sphericity and homoscedasticity assumptions are violated[31]. Unlike the IVW fixed-effect linear model, the mixed model with random effects estimated the effects of timepoint on the residuals at an individual level instead of only on their summary statistics. The example dataset for mixed modeling is in Table 3.

Table 3 Example dataset for the mixed model with crossed random effects

Obs	ID	CVYear	TimePoint	PD	Residual
1	1	1	0	1	13.05
2	1	2	1	1	7.38
3	1	3	2	1	3.29
4	1	4	3	1	8.88
5	1	5	4	1	5.55
6	1	6	5	1	1.47
7	2	1	0	1	-4.84
8	2	2	1	1	-3.30
9	2	3	2	1	0.59
10	2	4	3	1	0.16
11	2	5	4	1	1.06
12	2	6	5	1	-3.87
13	2	8	7	1	-11.04
14	3	1	-2	1	-10.60
15	3	2	-1	1	-7.81
16	3	3	0	1	-8.66
17	3	4	1	1	-12.01
18	3	5	2	1	-15.85
19	3	6	3	1	-21.22
20	3	8	5	1	-25.56
21	3	10	7	1	-28.23

Timepoint, years from PD diagnosis

In traditional hierarchical or multilevel models, random effects are assumed to be nested so that the lower-level units are uncorrelated across higher-level units[31,32]. However, our

data did not show a typical hierarchical structure, in other words, none of a unit is nested in another in our data and the levels of units are not fixed. For easy explaining, we chose timepoint as the “highest level” followed by participant and CVYear as an example to illustrate how our data is different from a traditional hierarchical structure. To be specific, in our data, 1) the residuals from the same participant (“lower-level”) were correlated across timepoint (“higher-level”) (Figure 2, solid lines); and 2) the residuals calculated from the same clinic visit year were correlated across the “higher levels” (participant and timepoint) due to the application of the same β coefficients for all the observations within the same year (Figure 2, orange/blue dashed lines for correlation across participant/timepoint, respectively). Thus, we included both random effects of participants and CVYear simultaneously into a single analysis and treated the two random effects as crossed[31].

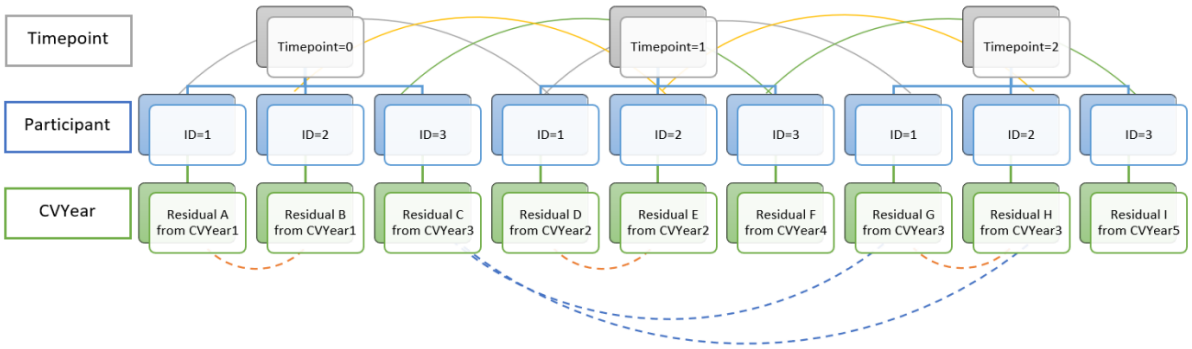


Figure 2 Example hierarchical structure extracted from Table 3.

Correlations are shown by the same colored lines that connected two blocks. The solid lines denote the correlations of multiple measurements from the same participant across timepoint; the orange and blue dashed lines denote the correlations of the values from the same clinic visit year across participant and timepoint, respectively.

The general mixed-effects model is given as:

$$Residual_{mi} = \beta_{m0_8} + \beta_{m1_8} * Timepoint_i + \beta_{m2_8} * Timepoint_i^2 + Z_{mi} b_{mi} + \varepsilon_{mi_8} \quad (8)$$

where $\boldsymbol{Residual_{mi}}$ is a vector of the residuals of PD case *i* in the analysis for body composition *m*; $\boldsymbol{Timepoint_i}$ is a vector of the years from PD diagnosis of the *i*th PD case; $\boldsymbol{Z_{mi}}$ denotes the design matrix for the random effects for the *i*th PD case in the analysis for body composition *m*; $\boldsymbol{b_{mi}}$ represents the corresponding vector of random effects; $\boldsymbol{\varepsilon_{mi_8}}$ represents the corresponding error vector in equation (8).

In the current analysis, the inference was based on the marginal distribution for the outcome variable residual.

$$\boldsymbol{Residual_{mi}} \sim \boldsymbol{N}\left(\boldsymbol{X_{mi}\beta_m},\ \boldsymbol{Z_{mi}D_mZ'_{mi}} + \boldsymbol{\Sigma_{mi}}\right)^{33}, \quad (9)$$

where $\boldsymbol{Residual_{mi}}$ is a vector of the residuals of PD case *i* in the analysis for body composition *m*; $\boldsymbol{X_{mi}}$ is a vector of the fixed effects in model (8) from the *i*th PD case in the analysis for body composition *m*; $\boldsymbol{D_m}$ is the covariance matrix of $\boldsymbol{b_{mi}}$ and $\boldsymbol{\Sigma_{mi}}$ is the covariance matrix of $\boldsymbol{\varepsilon_{mi_8}}$.

Let $\boldsymbol{\alpha}$ be a vector of variance components found in $\boldsymbol{Z_{mi}D_mZ'_{mi}} + \boldsymbol{\Sigma_{mi}}$ and be estimated by the restricted maximum likelihood (REML) estimation in the current analyses. The vector of $\boldsymbol{\beta}$ coefficient was estimated by[33]:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{N} \boldsymbol{X'_{mi}V_{mi}^{-1}}(\boldsymbol{\alpha})\boldsymbol{X_{mi}}\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X'_{mi}V_{mi}^{-1}}(\boldsymbol{\alpha})\boldsymbol{Residual_{mi}} \quad (10)$$

where N is the total number of PD cases in our dataset.

The main assumptions under the present mixed random-effects model are[33]: 1) $\boldsymbol{b_{mi}}$ is multivariate normal with mean vector 0 and with covariance matrix $\boldsymbol{D_m}$, 2) $\boldsymbol{\varepsilon_{mi_8}}$ is multivariate normal with mean vector 0 and with covariance matrix $\boldsymbol{\Sigma_{mi}}$, 3) all $\boldsymbol{b_{mi}}$ and $\boldsymbol{\varepsilon_{mi_8}}$ are independent of each other, and 4) $\boldsymbol{Z_{mi}D_mZ'_{mi}} + \boldsymbol{\Sigma_{mi}}$ is a positive (semi-)definite matrix.

The testing of the need for random effects was conducted by the likelihood ratio test (LRT) and the test statistics are exhibited in Table 4. The structure of random effects was decided based on the LRT results and the criteria of smaller AIC and BIC, and structure 1 was selected as the final random effects structure for all four body composition measures. Thus, the item $\boldsymbol{Z_{mi}b_{mi}}$ in equation (8) included the crossed random effects of both by-CVYear and by-participant adjustments to the intercept and by-participant adjustments to the effect of Timepoint[32,34]. The final model was given by:

$$\boldsymbol{Residual_{mi}} = \boldsymbol{\beta_{m0_{11}}} + \boldsymbol{\beta_{m1_{11}}} * \boldsymbol{Timepoint_i} + \boldsymbol{\beta_{m2_{11}}} * \boldsymbol{Timepoint_i^2}$$
$$+\boldsymbol{CV_{mj}b1_{mj}} + \boldsymbol{P_{mi}b2_{mi}} + \boldsymbol{\varepsilon_{mi_{11}}} \quad (11)$$

where $\boldsymbol{Residual_{mi}}$ is a vector of the residuals of PD case *i* in the analysis for body composition *m*; $\boldsymbol{Timepoint_i}$ is a vector of the years from PD diagnosis of the *i*th PD case; $\boldsymbol{CV_{mj}b1_{mj}}$ represents the CVYear random effects for *j*th clinic visit year in the analysis for body composition *m*; $\boldsymbol{P_{mi}b2_{mi}}$ is the participant random effects for *i*th participant in the analysis for body composition *m*; $\boldsymbol{\varepsilon_{mi_{11}}}$ is the error vector for participant *i* in the analysis for body composition *m.*

Table 4 Random effects structure selection for CRE model

	AIC	BIC	Log-likelihood	Ref struct-ure	Chi-Square Statistic*	DF	P†
Total mass							
Structure 1	2401.8	2430.1	-1193.9				
Structure 2	2459.1	2483.4	-1223.6	1	59.3	1	<.01
Total lean mass							
Structure 1	1886.3	1915.0	-936.2				
Structure 2	1937.6	1962.2	-962.8	1	53.3	1	<.01
Total fat mass							
Structure 1	2182.4	2211.1	-1084.2				
Structure 2	2247.7	2272.3	-1117.8	1	67.3	1	<.01
Percentage of fat mass							
Structure 1	1891.1	1919.4	-938.5				
Structure 2	1946.5	1970.7	-967.2	1	57.4	1	<.01

* Calculated from the difference of the -2log likelihood for the nested model and the full model.

† P-values.

Structure 1: Random intercept for both participants and CVYear; By-participant slope for timepoint.

Structure 2: Only random intercept for both participants and CVYear.

RESULTS

Participant flow chart

In all analyses, we excluded participants who were initially identified as possible PD cases but did not get confirmed in the case-adjudication effort. Participants with missing values in body composition measures were excluded from the corresponding year's analysis. There was no missing value on covariates after this exclusion. The numbers of each exclusion and final sample size for each clinic visit are provided in Figure 3.

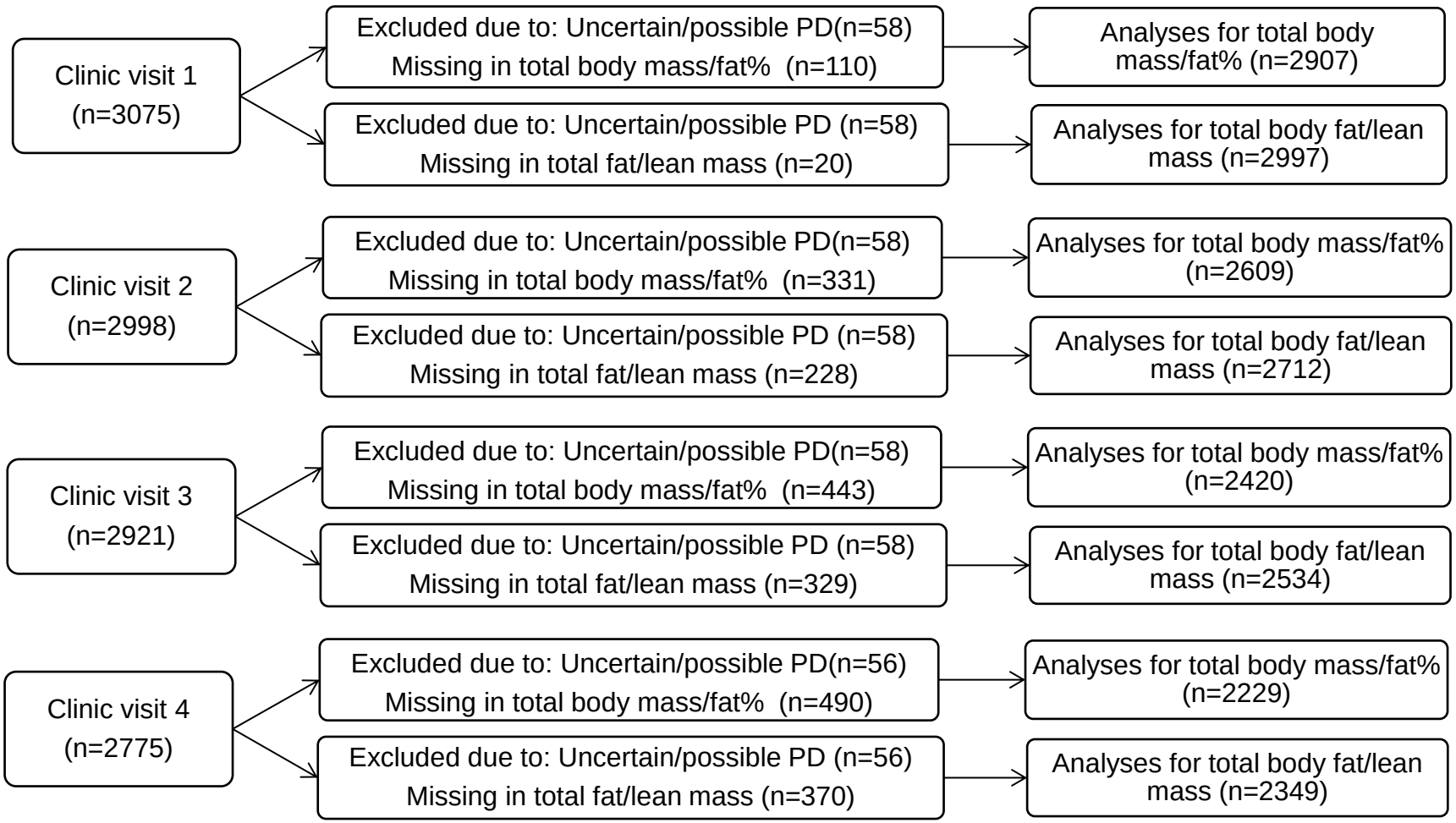
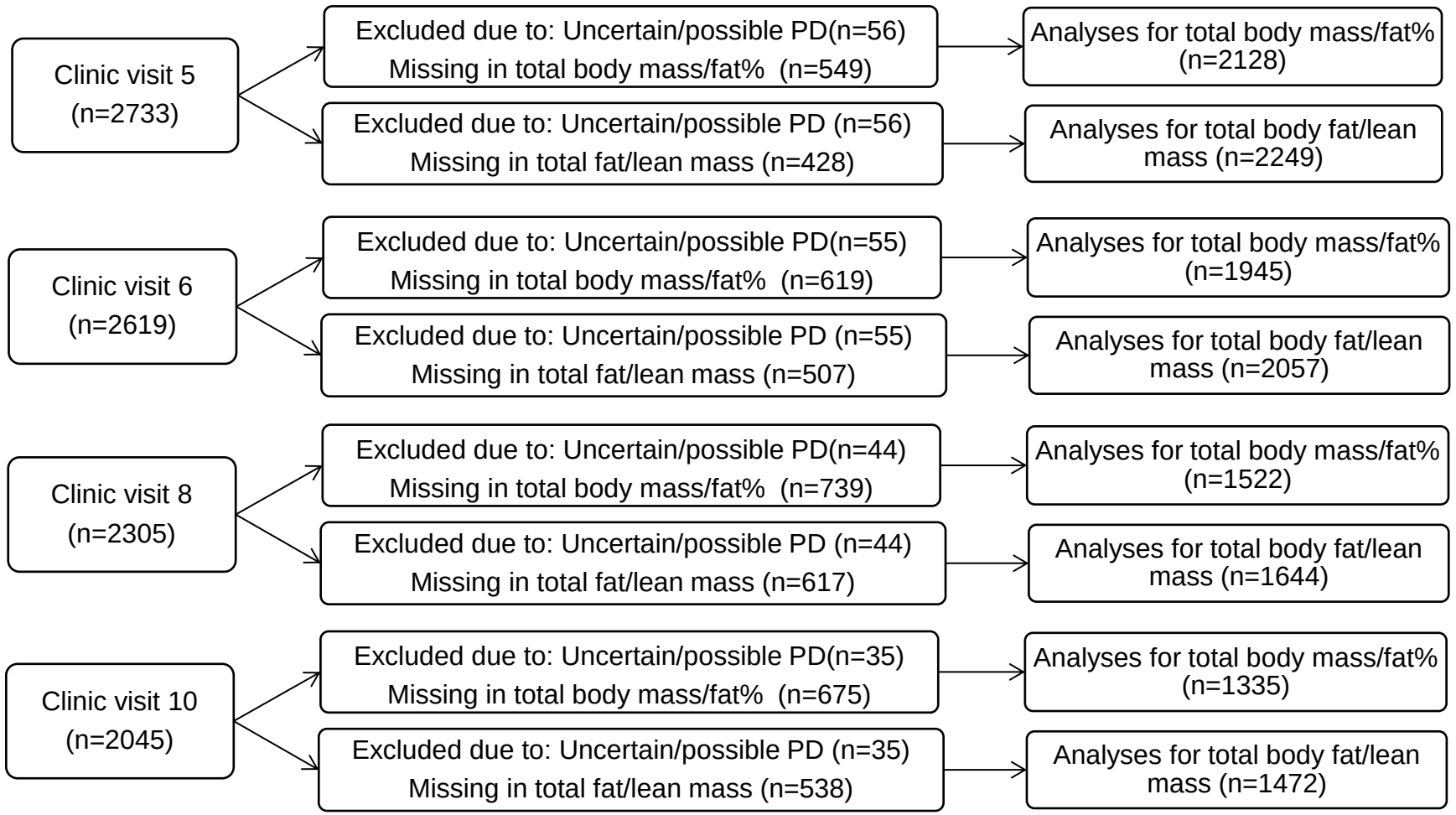


Figure 3 Participant flow chart

Figure 3 (cont'd):



Baseline characteristics

We present the baseline population characteristics of participants who had (cases) and had not developed PD (non-PD participants) during the follow-up in Table 5. For simplicity, we used cases and non-PD participants to refer these two groups. Because of the narrow age range of our study population, there was no significant age difference between PD cases and non-PD participants (74.0 ± 2.8 vs. 73.6 ± 2.9 years, p=0.18). Compared with non-PD participants, PD cases were more likely to be male (59.3% vs. 48.2%, p=0.049) and white (74.1% vs. 57.9%, p=0.004). The height and body composition measures did not significantly differ between PD cases and non-PD participants at baseline.

Table 5 Baseline Participant Characteristics of PD cases and non-PD participants

	PD Cases (n=81)	Non-PD Participants (n=2936)	P-values
Continuous Variables, Mean (SD)			
Age (years)	74.0 (2.8)	73.6 (2.9)	0.18
Standing height (m)	1.7 (0.1)	1.7 (0.1)	0.11
Total body mass (Kg)	75.9 (12.4)	75.5 (14.9)	0.62
Total body lean mass (Kg)	48.4 (9.1)	46.7 (10.0)	0.07
Total body fat mass (Kg)	25.6 (7.1)	26.8 (8.8)	0.21
Percentage of fat mass (%)	33.5 (7.0)	35.0 (7.9)	0.06
Categorical Variables, N (%)			
Sex			**0.049**
Female	33 (40.7)	1522 (51.8)	
Male	48 (59.3)	1414 (48.2)	
Race			**0.004**
White	60 (74.1)	1701 (57.9)	
Black	21 (25.9)	1235 (42.1)	
Site			0.67
Memphis	39 (48.15)	1483 (50.5)	
Pittsburgh	42 (51.85)	1453 (49.5)	

Due to missing data, the exact sample sizes of cases/controls were 77/2830 for total body mass and percentage of fat mass, and 80/2917 for total body lean and fat mass.
P-values were calculated through Mann Whitney U test and chi-square test for continuous and categorical variables, respectively.

Longitudinal Analysis on the Timescale of Clinic Visit Year

Table 6 shows the means of each body composition measure by clinic visit year. Overall, all four body composition measures slightly decreased over time in both PD cases and controls. Compare to non-PD participants, PD cases showed a lower total fat mass and percentage of fat mass from clinic visit year 6 to 10 ($p<0.05$), higher total lean mass at year 6 and 7 ($p<0.05$), and no difference in total body mass from year 1 to 10.

The comparison of fitted means from the linear mixed model (equation (1)) was showed in Table 7 and Figure 4. In this analysis, a significant downtrend in percentage of fat mass was observed among PD cases relative to non-PD participants (Figure 4, Table 7, $p = 0.03$). From baseline to clinic visit year 10, PD cases lose 1.4 kg and 1.8% more total fat mass compared to non-PD participants (Table 7, $p < 0.05$). In comparison, the total body mass and lean mass in PD cases were relatively stable over the entire observation period (Figure 4) and the change trajectories were not significantly different from non-PD participants (Table 7, p-values for total body mass and lean mass are 0.58 and 0.86, respectively).

Table 6 Compare main observed outcomes between PD cases and non-PD participants through clinic visit year 1 to 10

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 8	Year 10
N (PD cases/Controls)								
Total Mass/Fat%	**(77/2830)**	**(73/2536)**	**(60/2360)**	**(59/2170)**	**(51/2077)**	**(49/1896)**	**(39/1483)**	**(32/1303)**
Total Lean/Fat Mass	**(80/2917)**	**(77/2635)**	**(64/2470)**	**(61/2288)**	**(54/2195)**	**(52/2005)**	**(42/1602)**	**(35/1437)**
Total Mass (Kg), Mean (SD)								
PD cases	75.9(12.4)	76.1(11.9)	75.7(13)	75.2(13.4)	75.4(13.1)	74.6(13.6)	74.1(14.1)	73.7(15.1)
Non-PD participants	75.5(14.9)	75.3(14.9)	75.1(14.9)	75.0(14.9)	74.5(14.8)	74.4(14.9)	73.6(14.8)	72.5(14.7)
Total Lean Mass (Kg), Mean (SD)								
PD cases	48.4(9.1)	48.3(8.9)	48.4(9.7)	47.9(9.4)	47.9(9.4)	48.4(9.4) *	49.2(9.1) **	48.0(9.8)
Non-PD participants	46.7(10.0)	46.5(9.8)	46.1(9.8)	45.9(9.7)	45.5(9.6)	45.4(9.6)	45.3(9.7)	44.9(9.5)
Total Fat Mass (Kg),Mean (SD)								
PD cases	25.6(7.1)	25.8(7.3)	25.3(6.8)	25.5(7.2)	25.6(7.9)	24.4(7.4) *	23.4(8.1) **	23.3(9.1) *
Non-PD participants	26.8(8.8)	26.6(8.8)	26.9(8.9)	27.0(8.9)	26.9(8.7)	26.8(8.6)	26.3(8.5)	25.6(8.3)
Percentage of fat mass (%), Mean (SD)								
PD cases	33.5(7.0)	33.6(7.5)	33.4(7.0)	33.6(7.0)	33.5(7.6)	32.1(6.9) **	30.8(7.3) **	31.8(8.5) *
Non-PD participants	35.0(7.9)	34.9(7.8)	35.3(7.7)	35.6(7.7)	35.6(7.6)	35.7(7.5)	35.3(7.4)	34.9(7.4)

PD cases and non-PD participants denote those developed PD and had not developed PD during the follow-up, respectively.
P-values were calculated through the Kruskal-Wallis test when comparing means between PD cases and non-PD participants within the same year.
* P-values < 0.05, ** P-values < 0.01.

Table 7 Parameter estimates of the longitudinal linear mixed model for the time-trend of changes in relative body composition measures of PD cases compared with non-PD participants

	Year 2	Year 3	Year 4	Year 5	Year 6	Year 8	Year 10	P†
N (PD cases/Controls)								
Total Mass/Fat%	**(73/2536)**	**(60/2360)**	**(59/2170)**	**(51/2077)**	**(49/1896)**	**(39/1483)**	**(32/1303)**	
Total Lean/Fat Mass	**(77/2635)**	**(64/2470)**	**(61/2288)**	**(54/2195)**	**(52/2005)**	**(42/1602)**	**(35/1437)**	
Total mass (Kg), estimate (SE)								
PD cases	-0.1 (0.3)	-0.3 (0.5)	-0.5 (0.5)	-0.6 (0.6)	-1.0 (0.7)	-1.0 (0.9)	-1.5 (0.8)	0.58
Non-PD participants	Ref	Ref	Ref	Ref	Ref	Ref	Ref	
Total lean mass (Kg), estimate (SE)								
PD cases	-0.02 (0.2)	0.03 (0.2)	-0.09 (0.2)	0.01 (0.3)	0.09 (0.3)	0.3 (0.4)	0.4 (0.4)	0.86
Non-PD participants	Ref	Ref	Ref	Ref	Ref	Ref	Ref	
Total fat mass (Kg), estimate (SE)								
PD cases	0.1 (0.2)	-0.2 (0.3)	-0.4 (0.3)	-0.5 (0.4)	-0.8 (0.5)	-1.1 (0.6)	-1.4 (0.6) *	0.15
Non-PD participants	Ref	Ref	Ref	Ref	Ref	Ref	Ref	
Percentage of fat mass (%), estimate (SE)								
PD cases	0.009 (0.2)	-0.3 (0.3)	-0.4 (0.3)	-0.6 (0.3)	-1.0 (0.4) *	-1.4 (0.5) **	-1.8 (0.5) **	0.03
Non-PD participants	Ref	Ref	Ref	Ref	Ref	Ref	Ref	

† P-values for trend comparison. * p<0.05, ** p<0.01.
The unstructured main model fully adjusted for age, sex, race, height, clinic site, PD status, and the interaction between PD status, age, height and clinic visit year. Variance-covariance matrixes are unstructured and separate for PD cases and non-PD participants.
Estimate is the β coefficient for the effect of the interaction term between PD status and clinic visit year on body composition, which represents the relative body composition change from baseline for PD cases compared to non-PD participants.

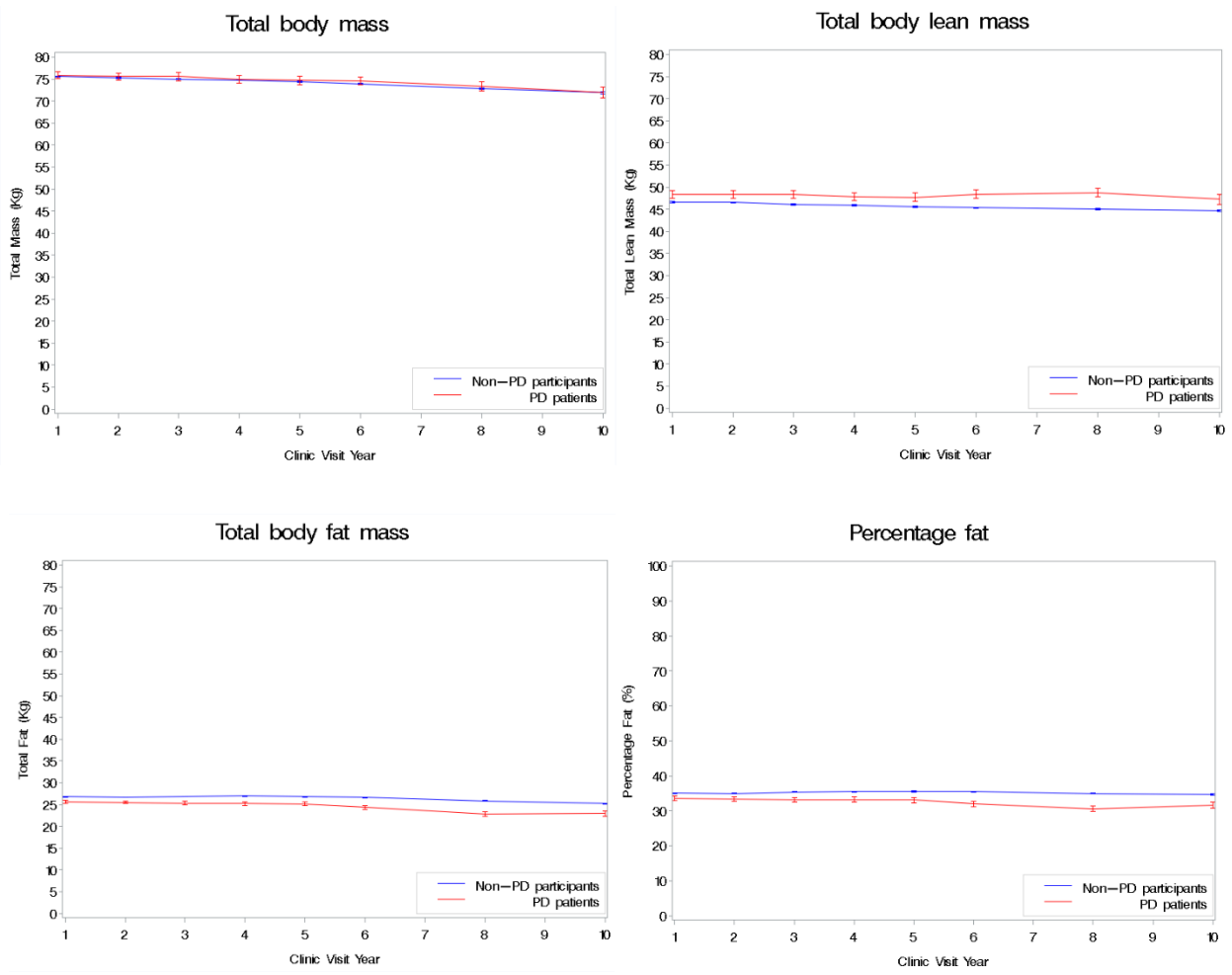


Figure 4 Fitted means (SD) in various body composition measures of PD cases and non-PD participants by clinic visit year

Trend Analysis on the Timescale of Years from PD Diagnosis

Inverse variance-weighted fixed-effect linear model Figure 5 displays changes in body composition measures before and after PD diagnosis by IVW method. The Y-axis represents the difference between the observed body composition values of PD cases and their expected values if they had not developed PD. A significant linear downtrend was observed for total body mass, lean mass, fat mass, and percentage of fat mass several years before the disease diagnosis (Figure 5; Table 8, the p-value for linear trends are 0.003, 0.008, <0.0001, and <0.0001, respectively). Nine years after diagnosis, the average cumulative loss was about 8.0 kg for total mass with about 7.0 kg for fat mass compared to 11 years before diagnosis (Figure 5). In comparison, the loss of total lean mass of PD cases was mild over the entire period (Figure 5, a loss of about 3.0 kg by year 9).

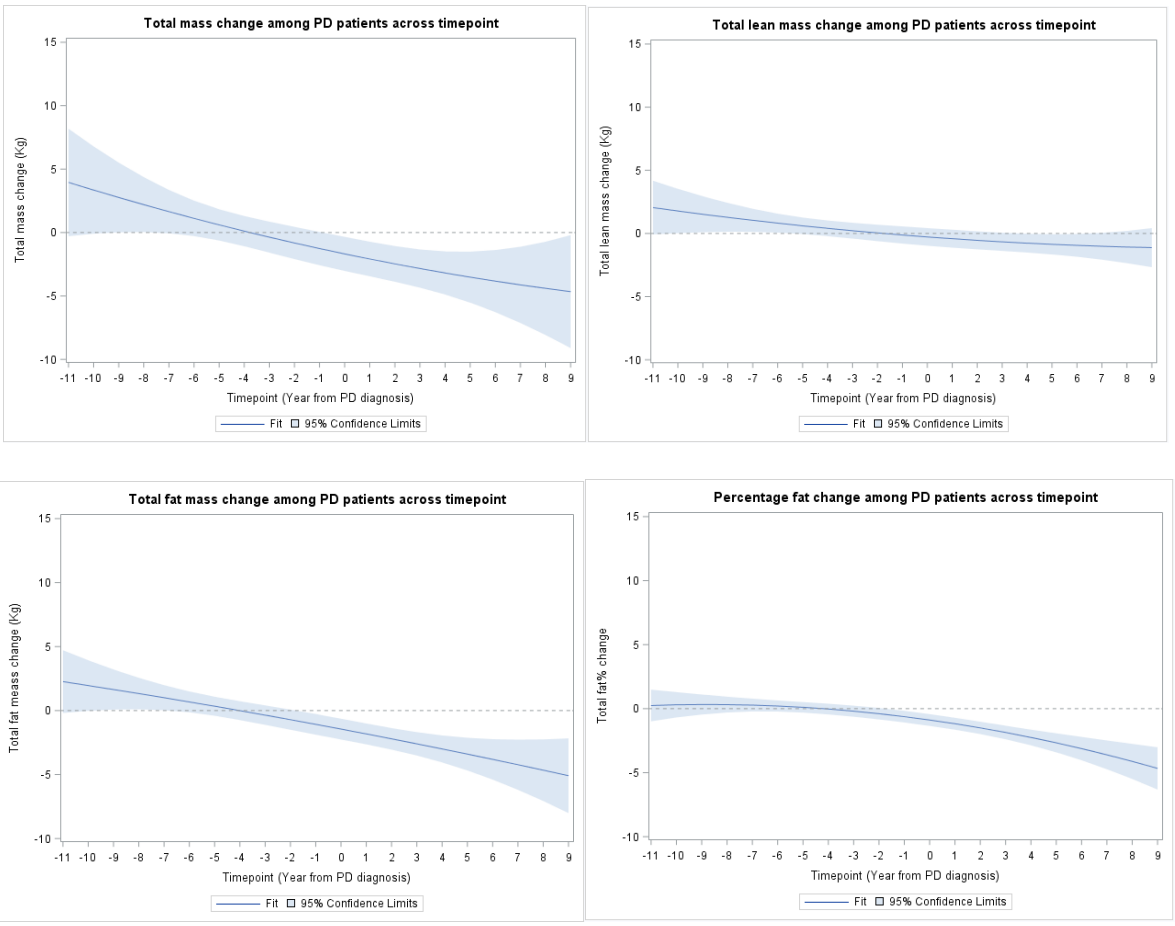


Figure 5 Changes in various body composition measures among PD cases before and after diagnosis by IVW method

All comparisons were made to their expected values at the given timepoint, which estimated based on data from non-PD participants. The horizontal reference line 0 represents standardized expected changes in body composition measures over years for non-PD cases, and for PD cases these refers to the expected changes if they had not developed PD.

Table 8 Parameter estimates of the IVW and CRE model for the time-trend of changes in relative body composition measures of PD cases compared with their expected values

	Intercept			Linear Timepoint			Quadratic Timepoint		
	Estimate (SE)	**95% CI**	**P**	**Estimate (SE)**	**95% CI**	**P**	**Estimate (SE)**	**95% CI**	**P**
Total mass (Kg)									
IVW	-1.7 (0.6)	(-3.0, -0.3)	0.02	-0.4 (0.1)	(-0.7, -0.2)	<0.01	0.009 (0.02)	(-0.03, 0.05)	0.7
CRE	-2.0 (1.3)	(-2.7, 2.3)	0.9	-0.4 (0.1)	(-0.6, -0.2)	<0.01	-0.03 (0.01)	(-0.06, -0.01)	<0.01
Total lean mass (Kg)									
IVW	-0.3 (0.3)	(-1.0, 0.4)	0.4	-0.1 (0.05)	(-0.2, -0.04)	<0.01	0.006 (0.009)	(-0.01, 0.02)	0.5
CRE	0.6 (0.5)	(-0.5, 1.6)	0.3	-0.007 (0.05)	(-0.1, 0.09)	0.9	-0.006 (0.005)	(-0.02, 0.004)	0.3
Total fat mass (Kg)									
IVW	-1.4 (0.4)	(-2.3, -0.6)	<0.01	-0.4 (0.07)	(-0.5, -0.2)	<0.01	-0.003 (0.01)	(-0.03, 0.02)	0.8
CRE	-0.5 (0.8)	(-2.1, 1.1)	0.6	-0.3 (0.07)	(-0.5, -0.2)	<0.01	-0.03 (0.008)	(-0.04, -0.02)	<0.01
Percentage of fat mass (%)									
IVW	-0.9 (0.2)	(-1.3, -0.4)	<0.01	-0.3 (0.04)	(-0.4, -0.2)	<0.01	-0.02 (0.007)	(-0.03, -0.001)	0.04
CRE	-0.7 (0.6)	(-1.8, 0.4)	0.2	-0.3 (0.06)	(-0.4, -0.2)	<0.01	-0.02 (0.007)	(-0.04, -0.01)	<0.01

IVW refers to inverse variance-weighted fixed-effects linear regression model. CRE refers to linear mixed model with crossed random effects.

Timepoint refers to years from PD diagnosis which was included as a linear term and a quadratic term.

Estimate is the fixed effects of timepoint on the difference between actual body composition of PD cases and their expected body composition had they not had PD.

Crossed random effects in CRE included random intercepts for clinic visit year and for participants, and by-participant random slope on timepoint.

CI = confidence interval. P = P-values.

Mixed models with crossed random effects With the inclusion of crossed random effects, the results from the CRE model are believed less biased than the IVW model and are selected to be our final results. Figure 6 shows the results from the analysis using mixed models with crossed random effects. The horizontal reference line 0 represents the standardized expected changes in body composition measures of non-PD participants because at any given point their residuals should be 0. Therefore, this line also represents the expected changes in body composition of PD cases had they did not develop PD. In this analysis, all body composition measures of PD cases were comparable to their expected values prior to PD diagnosis; however, total body mass, fat mass, and percentage of fat mass began to decrease around a few years prior to PD diagnosis which persisted through the all periods after PD diagnosis (Table 8, p-value for trends were 0.003, <0.001, and <0.001, respectively). By year 5, 3, and 2 after diagnosis, total body mass, fat mass, and percentage of fat mass, respectively, became statistically different from their expected values and the decreasing trend persisted. Nine years after diagnosis, the average cumulative loss was 6.0 kg for total mass with 5.6 kg for fat mass compared to 11 years before diagnosis (Table 9). In comparison, the total lean mass of PD cases was stable over the entire period (Table 8, p-value for trends=0.26; Table 9, a gain of 0.09 kg by year 9).

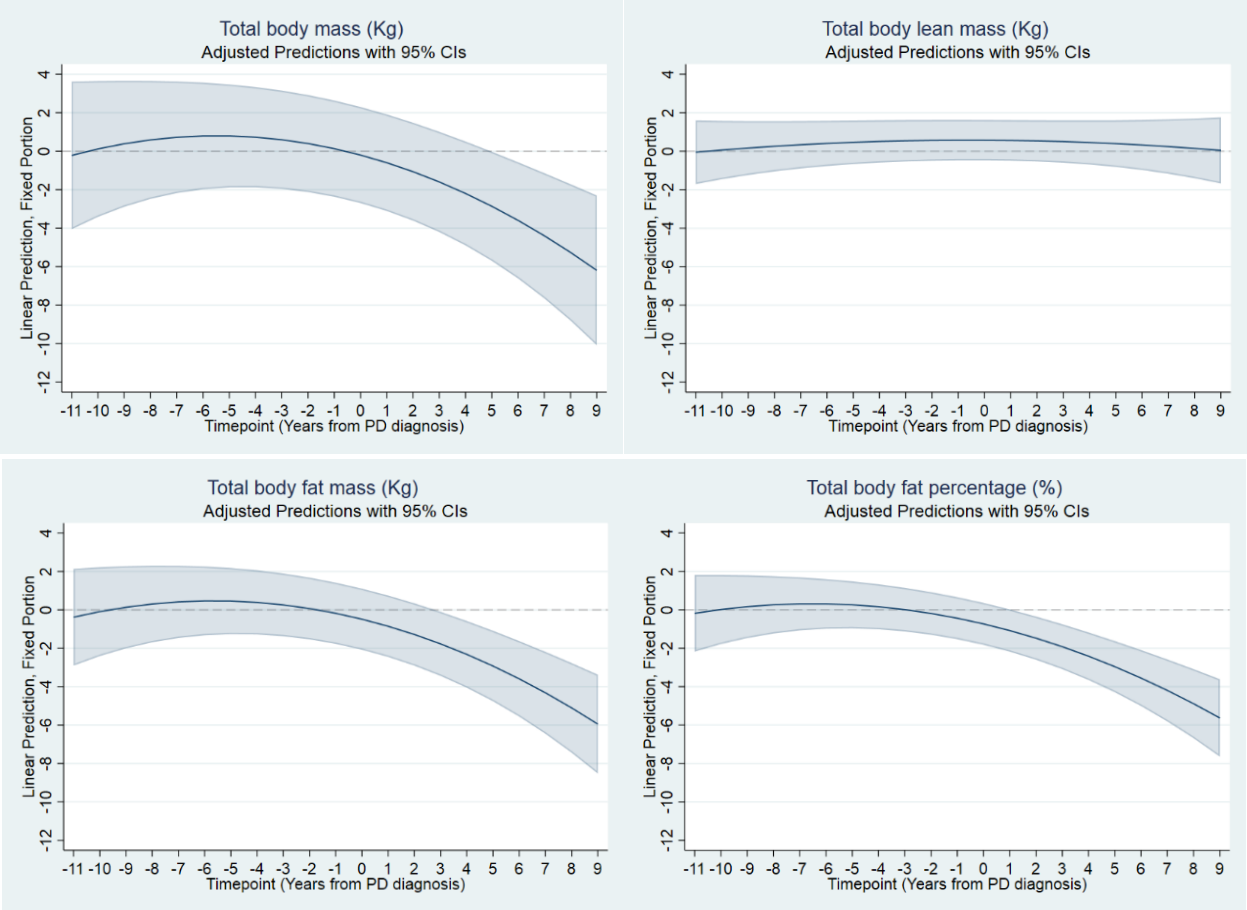


Figure 6 Changes in various body composition measures of PD cases before and after diagnosis by CRE model.

All comparisons were made to their expected values at the given timepoint, which estimated based on data from non-PD participants. The horizontal reference line 0 represents standardized expected changes in body composition measures over years for non-PD cases, and for PD cases these refers to the expected changes if they had not developed PD.

Table 9 CRE model predicted body composition differences at each timepoint in PD cases in reference to had they not had PD

	Total mass (Kg)		Total lean mass (Kg)		Total fat mass (Kg)		Percentage of fat mass (%)	
Time-point	Estimates	95% CI	Estimates	95% CI	Estimates	95% CI	Estimates	95% CI
-11	-0.2	(-4.1, 3.6)	-0.1	(-1.7, 1.6)	-0.4	(-2.9, 2.1)	-0.2	(-2.2, 1.8)
-10	0.1	(-3.4, 3.6)	0.1	(-1.5, 1.6)	-0.1	(-2.4, 2.2)	0.0	(-1.8, 1.8)
-9	0.4	(-2.9, 3.7)	0.2	(-1.2, 1.6)	0.1	(-2, 2.3)	0.2	(-1.5, 1.8)
-8	0.6	(-2.5, 3.7)	0.3	(-1, 1.6)	0.3	(-1.7, 2.3)	0.3	(-1.2, 1.8)
-7	0.7	(-2.2, 3.6)	0.3	(-0.9, 1.6)	0.4	(-1.5, 2.3)	0.3	(-1.1, 1.7)
-6	0.8	(-2, 3.6)	0.4	(-0.8, 1.6)	0.5	(-1.3, 2.3)	0.3	(-1, 1.6)
-5	0.8	(-1.9, 3.5)	0.5	(-0.7, 1.6)	0.5	(-1.3, 2.2)	0.3	(-1, 1.5)
-4	0.7	(-1.9, 3.3)	0.5	(-0.6, 1.6)	0.4	(-1.3, 2.1)	0.2	(-1, 1.3)
-3	0.6	(-2, 3.2)	0.5	(-0.5, 1.6)	0.3	(-1.4, 1.9)	0.0	(-1.1, 1.1)
-2	0.4	(-2.1, 2.9)	0.6	(-0.5, 1.6)	0.1	(-1.5, 1.7)	-0.2	(-1.3, 0.9)
-1	0.1	(-2.4, 2.6)	0.6	(-0.5, 1.6)	-0.2	(-1.8, 1.4)	-0.4	(-1.5, 0.7)
0	-0.2	(-2.7, 2.3)	0.6	(-0.5, 1.6)	-0.5	(-2.1, 1.1)	-0.7	(-1.8, 0.4)
1	-0.6	(-3.1, 1.9)	0.6	(-0.5, 1.6)	-0.9	(-2.5, 0.7)	-1.1	(-2.2, 0)
2	-1.1	(-3.6, 1.5)	0.5	(-0.5, 1.6)	-1.3	(-2.9, 0.3)	-1.5*	(-2.6, -0.3)
3	-1.6	(-4.2, 1)	0.5	(-0.6, 1.6)	-1.8*	(-3.4, -0.1)	-1.9**	(-3.1, -0.7)
4	-2.2	(-4.9, 0.5)	0.5	(-0.7, 1.6)	-2.3**	(-4.1, -0.6)	-2.4**	(-3.6, -1.2)
5	-2.9*	(-5.7, 0)	0.4	(-0.8, 1.6)	-2.9**	(-4.8, -1.1)	-3.0**	(-4.3, -1.6)
6	-3.6*	(-6.6, -0.6)	0.3	(-1, 1.6)	-3.6**	(-5.6, -1.6)	-3.5**	(-5, -2.1)
7	-4.4**	(-7.6, -1.1)	0.2	(-1.2, 1.7)	-4.3**	(-6.4, -2.2)	-4.2**	(-5.8, -2.6)
8	-5.3**	(-8.8, -1.7)	0.1	(-1.4, 1.7)	-5.1**	(-7.4, -2.8)	-4.9**	(-6.7, -3.1)
9	-6.2**	(-10.1, -2.3)	0.0	(-1.7, 1.8)	-5.9**	(-8.5, -3.4)	-5.6**	(-7.6, -3.6)

Timepoint, years from PD diagnosis. Estimate is the mean estimate of differences between actual and expected body composition of PD cases at each timepoint. CI = confidence interval.* P-values < 0.05, ** P-values < 0.01.

DISCUSSION

In this longitudinal study with repeated measures of body composition, we found that the body weight of PD cases began to decrease a few years prior to diagnosis which persisted in the years after diagnosis. By year 9, post PD diagnosis, the cumulative loss was on an average of about 6 kg. We further found that on an average 93% of the loss was due to loss of fat mass while the lean mass in PD cases was well preserved.

Substantial evidence from case-control and cross-sectional studies have shown that PD cases had lower body weight and body mass index than controls[3,5,7,11,35–39]. Some studies further reported that PD cases lose more weight after diagnosis than individuals without PD. The first report was published in 1976 and found a significant loss of body weight in seven levodopa-treated PD cases compared to healthy controls, with an average loss of 6.28 kg over 1-3 years[40]. In 1995, Beyer et al.[3] reported that PD cases were four times more likely to report a significant weight loss of >10 pounds since disease diagnosis. Two later longitudinal studies reported PD cases lost an average of 1.8 kg over 1 year[7] and 4.1 kg over 6.3 years of follow-up[5]. While the clinical and epidemiological data on weight loss in PD is substantial, they are not entirely consistent. Several studies have reported stable or even weight gain in PD cases. For example, Wills et al. found that stable body weight was common in relatively young PD cases who participated in a PD clinical trial[8]. Two other studies with small sample sizes also reported a stable weight among PD cases over 1-3 years of follow-up[41,42]. This inconsistency across studies may in part be explained by differences in study design, case characteristics, sample size, length of

follow-up, treatment strategies used for cases, and lack of repeated objective assessments of body weight.

Several studies further suggest that body weight in PD cases may begin to decline prior to disease diagnosis. In the Health Professionals Follow-up Study (HPFS) and the Nurse's Health Study (NHS), Chen et al. analyzed body weights reported every 2 years by 468 PD cases which were compared to participants who had not developed PD using similar methods as in this study. They found that PD cases began to lose weight 2-4 years before PD diagnosis despite accompanying increased energy intake[4] and decreased physical activities[43]. These trends persisted after diagnosis. To the best of our knowledge, this is the only longitudinal study in which the body weights of PD cases were repeatedly assessed prior to disease diagnosis. The finding is supported by another small clinical study of 49 PD cases with body weight prior to disease diagnosis abstracted from medical records. The authors reported PD cases had a mean weight loss of 1.19% in an average period of 2.4 years prior to the time of PD diagnosis[6]. Therefore, the existing evidence, while limited, suggests that weight loss in PD cases may start in the prodromal stage.

Compared to previous studies, the current study has several notable strengths. Like the HPFS and NHS, the Health ABC study repeatedly measured body weight of PD cases up to about a decade before and after the diagnosis. Beyond HPFS and NHS, body weight in Health ABC was not self-reported but more accurately assessed using DXA assays along with measures of body fat and lean mass, allowing analyses to examine differential changes in body composition compartments. Our study confirmed that a persistent weight loss in PD cases likely starts a few years prior to PD clinical diagnosis. We further found

that the loss was almost entirely due to the loss of fat mass. This observation of a preferential loss of fat mass in PD cases is supported by some but not all of the earlier small clinical studies. For example, in one study, 19 of 26 PD cases lose weight over a year which was mainly due to loss of fat mass as measured by DXA[7]. In contrast, another study of 58 PD cases reported gain in average body weight and fat mass over 3 years measured by bioelectrical impedance analysis[41]. Compared to this existing literature, our study is population-based, had a longer follow-up, repeatedly assessed body composition using DXA, and our finding of a persistent loss of fat mass in PD cases is robust.

Although the exact reasons for weight loss in PD are yet to be identified, it may relate to the complex symptoms and signs that develop as PD pathogenesis progresses. For example, in the prodromal stage of PD, poor olfaction may adversely affect the diet and nutrition which in turn gradually leads to changes in body composition. In support of this, Purdy et al. reported that poor olfaction was associated with faster weight loss in older adults[44]. Experiments in transgenic mice have also shown poor olfaction to trigger a metabolic response that leads to increased catabolic energy utilization and a subsequent loss of body weight[45]. As PD progresses, its motor signs and complications such as tremor[46], muscle rigidity[46,47], and Levodopa-induced dyskinesias[48] may further lead to increased energy expenditure and thus weight loss. Throughout the course of PD, these may further be complicated by other PD symptoms and signs such as depression[49], gastrointestinal disorders[50], cognitive impairment[10,51], dysphagia[52], and adverse events of dopaminergic treatment[51,53], all of which may lead to a progressive and persistent weight loss in PD cases. The preferential loss of fat mass in PD is also intriguing, and may relate to accelerated biological aging[54,55] which is associated with the reduction of subcutaneous

fat and the deposition of fat in non-adipose tissues in late adulthood by reducing adipogenesis through age-related activation of cellular stress response pathways and increased preadipocyte cytokine generation[56,57]. Further investigations are warranted to investigate the causes of weight loss and preferential loss of fat mass in PD cases.

This persistent weight loss in PD cases noted above should not be neglected as it may have important adverse health consequences. In PD cases, weight loss is often associated with higher Hoehn & Yahr stage[9], a higher score on unified Parkinson's disease rating scale[8], the lower density of nigrostriatal dopaminergic neurons[58], lower cognitive function[9–11], declined quality of life[9], a higher number of comorbidities[9], and increased risk of dependency and mortality[5]. Moreover, the reduction of fat mass and the subsequent redistribution of adipocytes into muscle tissues may result in a higher risk of reduced muscle function and the development of frailty in PD cases. In support of this possibility, several studies reported lower fat mass was associated with more severe motor impairment in PD cases[35,59]. Further, one recent study found that PD cases had a higher MRI measured fat content in the bilateral psoas and thigh muscles than their age- and sex-matched healthy controls, which was associated with disease severity and frailty[12]. Finally, a preferential loss of fat tissue in PD cases may also contribute to PD pathogenesis itself. As a storage tissue for the lipid-soluble neurotoxicants such as organochlorine pesticides, the loss of fat mass may release such neurotoxicants to circulation[60,61] which may in turn contribute to the progressive loss of dopaminergic neurons in PD[62,63]. Therefore, changes in weight and body composition in PD cases may inform disease progression and prognosis, offering an opportunity to improve the health and survival of PD cases.

The present study has several limitations. First, although our study identified 81 PD cases, their diagnosis was at different time points of the follow-up, and therefore the actual sample sizes were relatively small and varied at each specific time point in reference to PD clinical diagnosis. Nevertheless, our study revealed persistent weight loss in PD cases which was highly statistically significant. Second, the Health ABC study participants were old at enrollment (range 70-79). As body composition changes is part of the aging process, our findings may not be readily generalizable to younger PD cases. Third, both PD diagnostic adjudication and time of diagnosis were retrospectively adjudicated based on cohort data collection, and thus inadvertent errors are likely. However, our findings of persistent weight loss and when it starts are very consistent with those of HPFS and NHS, which prospectively adjudicated PD diagnosis and further conducted rigorous medical record reviews by gathering diagnostic information from study participants and their treating physicians[4,43]. Fourth, we were unable to account for energy intake and expenditure in our analyses due to limited data availability because such information is often difficult to capture in large populations of older adults.

In summary, in this longitudinal cohort of older adults with objective and repeated measurement of body composition, we found a persistent loss of fat mass in PD cases that started a few years before disease diagnosis. Future studies are needed to understand the underlying mechanisms of this change in the body composition of PD cases and its potential impacts on disease progression and survival.

REFERENCES

REFERENCES

1. Prevalence of Parkinson's disease across North America. Accessed February 18, 2020. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6039505/

2. Kalia LV, Lang AE. Parkinson's disease. *The Lancet*. 2015;386(9996):896-912. doi:10.1016/S0140-6736(14)61393-3

3. Beyer PL, Palarino MY, Michalek D, Busenbark K, Koller WC. Weight Change and Body Composition in Cases With Parkinson's Disease. *J Am Diet Assoc*. 1995;95(9):979-983. doi:10.1016/S0002-8223(95)00269-3

4. Chen H, Zhang SM, Hernán MA, Willett WC, Ascherio A. Weight loss in Parkinson's disease. *Ann Neurol*. 2003;53(5):676-679. doi:10.1002/ana.10577

5. Cumming K, Macleod AD, Myint PK, Counsell CE. Early weight loss in parkinsonism predicts poor outcomes. *Neurology*. 2017;89(22):2254-2261. doi:10.1212/WNL.0000000000004691

6. Uc EY, Struck LK, Rodnitzky RL, Zimmerman B, Dobson J, Evans WJ. Predictors of weight loss in Parkinson's disease. *Mov Disord Off J Mov Disord Soc*. 2006;21(7):930-936. doi:10.1002/mds.20837

7. Lorefält B, Toss G, Granérus A-K. Weight loss, body fat mass, and leptin in Parkinson's disease. *Mov Disord*. 2009;24(6):885-890. doi:10.1002/mds.22466

8. Wills A-MA, Pérez A, Wang J, et al. Association Between Change in Body Mass Index, Unified Parkinson's Disease Rating Scale Scores, and Survival Among Persons With Parkinson Disease. *JAMA Neurol*. 2016;73(3):321-328. doi:10.1001/jamaneurol.2015.4265

9. Akbar U, He Y, Dai Y, et al. Weight Loss and Impact on Quality of Life in Parkinson's Disease. *PLoS ONE*. 2015;10(5). doi:10.1371/journal.pone.0124541

10. Kim HJ, Oh ES, Lee JH, et al. Relationship between changes of body mass index (BMI) and cognitive decline in Parkinson's disease (PD). *Arch Gerontol Geriatr*. 2012;55(1):70-72. doi:10.1016/j.archger.2011.06.022

11. De Lucia N, Peluso S, Esposito M, et al. Frontal defect contribution to decreasing of body mass index in Parkinson's disease cases. *J Clin Neurosci Off J Neurosurg Soc Australas*. 2020;72:229-232. doi:10.1016/j.jocn.2019.11.034

12. Chiang P-L, Chen Y-S, Lin AW-C. Altered Body Composition of Psoas and Thigh Muscles in Relation to Frailty and Severity of Parkinson's Disease. *Int J Environ Res Public Health*. 2019;16(19). doi:10.3390/ijerph16193667

13. Logroscino G, Sesso HD, Paffenbarger RS, Lee I-M. Body Mass Index and Risk of Parkinson's Disease: A Prospective Cohort Study. *Am J Epidemiol*. 2007;166(10):1186-1190. doi:10.1093/aje/kwm211

14. Jeong S-M, Han K, Kim D, Rhee SY, Jang W, Shin DW. Body mass index, diabetes, and the risk of Parkinson's disease. *Mov Disord Off J Mov Disord Soc*. 2020;35(2):236-244. doi:10.1002/mds.27922

15. National Institute on Aging. Introducing the Health ABC Study: The Dynamics of Health, Aging, and Body Composition. Accessed March 3, 2020. https://healthabc.nia.nih.gov/

16. Visser M, Fuerst T, Lang T, et al. Validity of fan-beam dual-energy X-ray absorptiometry for measuring fat-free mass and leg muscle mass. *J Appl Physiol*. 1999;87(4):1513-1520. doi:10.1152/jappl.1999.87.4.1513

17. Salamone LM, Fuerst T, Visser M, et al. Measurement of fat mass using DEXA: a validation study in elderly adults. *J Appl Physiol*. 2000;89(1):345-352. doi:10.1152/jappl.2000.89.1.345

18. Murphy RA, Patel KV, Kritchevsky SB, et al. Weight change, body composition, and risk of mobility disability and mortality in older adults: a population-based cohort study. *J Am Geriatr Soc*. 2014;62(8):1476-1483. doi:10.1111/jgs.12954

19. Schoeller DA, Tylavsky FA, Baer DJ, et al. QDR 4500A dual-energy X-ray absorptiometer underestimates fat mass in comparison with criterion methods in adults. *Am J Clin Nutr*. 2005;81(5):1018-1025. doi:10.1093/ajcn/81.5.1018

20. National Institute on Aging. Year 1 documentation. https://healthabc.nia.nih.gov/sites/default/files/Y1_Documentation.pdf

21. Chen H, Shrestha S, Huang X, et al. Olfaction and incident Parkinson disease in US white and black older adults. *Neurology*. 2017;89(14):1441-1447. doi:10.1212/WNL.0000000000004382

22. Fleiss J. Review papers : The statistical basis of meta-analysis. *Stat Methods Med Res*. 1993;2(2):121-145. doi:10.1177/096228029300200202

23. Lau J, Ioannidis JPA, Schmid CH. Quantitative Synthesis in Systematic Reviews. *Ann Intern Med*. 1997;127(9):820-826. doi:10.7326/0003-4819-127-9-199711010-00008

24. Marín-Martínez F, Sánchez-Meca J. Weighting by Inverse Variance or by Sample Size in Random-Effects Meta-Analysis. *Educ Psychol Meas*. 2010;70(1):56-73. doi:10.1177/0013164409344534

25. Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am J Hum Genet*. 2010;86(1):6-22. doi:10.1016/j.ajhg.2009.11.017

26. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010;1(2):97-111. doi:10.1002/jrsm.12

27. Cochran WG. The Combination of Estimates from Different Experiments. *Biometrics*. 1954;10(1):101. doi:10.2307/3001666

28. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychol Methods*. 1998;3:486-504.

29. Smith VK, Hall TW. A Comparison of Maximum Likelihood Versus Blue Estimators. *Rev Econ Stat*. 1972;54(2):186-190. doi:10.2307/1926280

30. Lee CH, Cook S, Lee JS, Han B. Comparison of Two Meta-Analysis Methods: Inverse-Variance-Weighted Average and Weighted Sum of Z-Scores. *Genomics Inform*. 2016;14(4):173-180. doi:10.5808/GI.2016.14.4.173

31. Quené H, van den Bergh H. Examples of mixed-effects modeling with crossed random effects and with binomial data. *J Mem Lang*. 2008;59(4):413-425. doi:10.1016/j.jml.2008.02.002

32. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang*. 2008;59(4):390-412. doi:10.1016/j.jml.2007.12.005

33. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Corrected edition. Springer; 2001.

34. Rabe-Hesketh S, Skrondal A. *Multilevel and Longitudinal Modeling Using Stata, Third Edition*. Stata Press; 2012.

35. Tan AH, Hew YC, Lim S-Y, et al. Altered body composition, sarcopenia, frailty, and their clinico-biological correlates, in Parkinson's disease. *Parkinsonism Relat Disord*. 2018;56:58-64. doi:10.1016/j.parkreldis.2018.06.020

36. Suzuki K, Okuma Y, Uchiyama T, et al. Determinants of Low Body Mass Index in Cases with Parkinson's Disease: A Multicenter Case-Control Study. *J Park Dis*. 2020;10(1):213-221. doi:10.3233/JPD-191741

37. Davies KN, King D, Davies H. A study of the nutritional status of elderly cases with Parkinson's disease. *Age Ageing*. 1994;23(2):142-.

38. Marck MA van der, Dicke HC, Uc EY, et al. Body mass index in Parkinson's disease: A meta-analysis. *Parkinsonism Relat Disord*. 2012;18(3):263-267. doi:10.1016/j.parkreldis.2011.10.016

39. Barichella M, Cereda E, Cassani E, et al. Dietary habits and neurological features of Parkinson's disease cases: Implications for practice. *Clin Nutr*. 2017;36(4):1054-1061. doi:10.1016/j.clnu.2016.06.020

40. Vardi J, Oberman Z, Rabey I, Streifler M, Ayalon D, Herzberg M. Weight loss in cases treated long-term with levodopa. Metabolic aspects. *J Neurol Sci*. 1976;30(1):33-40. doi:10.1016/0022-510x(76)90253-7

41. Vikdahl M, Carlsson M, Linder J, Forsgren L, Håglin L. Weight gain and increased central obesity in the early phase of Parkinson's disease. *Clin Nutr*. 2014;33(6):1132-1139. doi:10.1016/j.clnu.2013.12.012

42. Lindskov S, Sjöberg K, Hagell P, Westergren A. Weight stability in Parkinson's disease. *Nutr Neurosci*. 2016;19(1):11-20. doi:10.1179/1476830515Y.0000000044

43. Chen H, Zhang SM, Schwarzschild MA, Hernán MA, Ascherio A. Physical activity and the risk of Parkinson disease. *Neurology*. 2005;64(4):664-669. doi:10.1212/01.WNL.0000151960.28687.93

44. Purdy F, Luo Z, Gardiner JC, et al. Olfaction and changes in body composition in a large cohort of older US adults. *J Gerontol A Biol Sci Med Sci*. Published online April 8, 2020. doi:10.1093/gerona/glaa085

45. Riera CE, Tsaousidou E, Halloran J, et al. The Sense of Smell Impacts Metabolic Health and Obesity. *Cell Metab*. 2017;26(1):198-211.e5. doi:10.1016/j.cmet.2017.06.015

46. Levi S, Cox M, Lugon M, Hodkinson M, Tomkins A. Increased energy expenditure in Parkinson's disease. *BMJ*. 1990;301(6763):1256-1257.

47. Markus HS, Cox M, Tomkins AM. Raised resting energy expenditure in Parkinson's disease and its relationship to muscle rigidity. *Clin Sci*. 1992;83(2):199-204. doi:10.1042/cs0830199

48. Bachmann CG, Zapf A, Brunner E, Trenkwalder C. Dopaminergic treatment is associated with decreased body weight in cases with Parkinson's disease and dyskinesias. *Eur J Neurol*. 2009;16(8):895-901. doi:10.1111/j.1468-1331.2009.02617.x

49. Pilhatsch M, Kroemer NB, Schneider C, et al. Reduced body mass index in Parkinson's disease: contribution of comorbid depression. *J Nerv Ment Dis*. 2013;201(1):76-79. doi:10.1097/NMD.0b013e31827ab2cc

50. Fasano A, Visanji NP, Liu LWC, Lang AE, Pfeiffer RF. Gastrointestinal dysfunction in Parkinson's disease. *Lancet Neurol*. 2015;14(6):625-639. doi:10.1016/S1474-4422(15)00007-1

51. Wills A-M, Li R, Pérez A, Ren X, Boyd J, on behalf of the NINDS NET-PD Investigators. Predictors of weight loss in early treated Parkinson's disease from the NET-PD LS-1 cohort. *J Neurol*. 2017;264(8):1746-1753. doi:10.1007/s00415-017-8562-4

52. Nozaki S, Saito T, Matsumura T, Miyai I, Kang J. [Relationship between weight loss and dysphagia in cases with Parkinson's disease]. *Rinsho Shinkeigaku*. 1999;39(10):1010-1014.

53. Pålhagen S, Lorefält B, Carlsson M, et al. Does L-dopa treatment contribute to reduction in body weight in elderly cases with Parkinson's disease? *Acta Neurol Scand*. 2005;111(1):12-20. doi:10.1111/j.1600-0404.2004.00364.x

54. Hou Y, Dan X, Babbar M, et al. Ageing as a risk factor for neurodegenerative disease. *Nat Rev Neurol*. 2019;15(10):565-581. doi:10.1038/s41582-019-0244-7

55. Horvath S, Ritz BR. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease cases. *Aging*. 2015;7(12):1130-1142.

56. Cartwright MJ, Tchkonia T, Kirkland JL. Aging in adipocytes: Potential impact of inherent, depot-specific mechanisms. *Exp Gerontol*. 2007;42(6):463-471. doi:10.1016/j.exger.2007.03.003

57. Kirkland JL, Tchkonia T, Pirtskhalava T, Han J, Karagiannides I. Adipogenesis and aging: does aging make fat go MAD? *Exp Gerontol*. 2002;37(6):757-767. doi:10.1016/S0531-5565(02)00014-1

58. Lee JJ, Oh JS, Ham JH, et al. Association of body mass index and the depletion of nigrostriatal dopamine in Parkinson's disease. *Neurobiol Aging*. 2016;38:197-204. doi:10.1016/j.neurobiolaging.2015.11.009

59. Pisciotta MS, Fusco D, Grande G, et al. Untangling the relationship between fat distribution, nutritional status and Parkinson's disease severity. *Aging Clin Exp Res*. Published online March 15, 2019. doi:10.1007/s40520-019-01166-x

60. Imbeault P, Chevrier J, Dewailly É, et al. Increase in plasma pollutant levels in response to weight loss in humans is related to in vitro subcutaneous adipocyte basal lipolysis. *Int J Obes*. 2001;25(11):1585-1591. doi:10.1038/sj.ijo.0801817

61. La Merrill Michele, Emond Claude, Kim Min Ji, et al. Toxicological Function of Adipose Tissue: Focus on Persistent Organic Pollutants. *Environ Health Perspect*. 2013;121(2):162-169. doi:10.1289/ehp.1205485

62. Baltazar MT, Dinis-Oliveira RJ, de Lourdes Bastos M, Tsatsakis AM, Duarte JA, Carvalho F. Pesticides exposure as etiological factors of Parkinson's disease and other neurodegenerative diseases--a mechanistic approach. *Toxicol Lett*. 2014;230(2):85-103. doi:10.1016/j.toxlet.2014.01.039

63. Van Maele-Fabry G, Hoet P, Vilain F, Lison D. Occupational exposure to pesticides and Parkinson's disease: a systematic review and meta-analysis of cohort studies. *Environ Int*. 2012;46:30-43. doi:10.1016/j.envint.2012.05.004