Towards a phenome-wide view of Parkinson’s disease

Authors
Abby L. Olsen, MD PhD1,2,3, Joseph J. Locascio, PhD1,2,4, Idil Tuncali, MS1,2,3, Nada Laroussi1,2,3, Elena Abatzis1,2,3, Polina Kamenskaya1,2,3, Yuliya Kuras, PhD1,2,3, Tom Yi1,2,3, Aleksandar Vidovic, MD, MSc4, Michael T. Hayes, MD3, Gary P. H. Ho, MD, PhD3, Jordan Paulson, MD3, Vikram Khurana, MD, PhD3, Todd M. Herrington, MD, PhD3, Bradley T. Hyman, MD, PhD4, Dennis J. Selkoe, MD3, John H. Growdon, MD4, Stephen N. Gomperts, MD, PhD4, Trond Riise5, PhD, Michael A. Schwarzschild, MD, PhD4, Albert Y. Hung, MD, PhD3,4, Anne-Marie Wills, MD, MPH4, and Clemens R. Scherzer, MD1,2,3,4,6*

1 APDA Center for Advanced Parkinson Research, Harvard Medical School, Brigham and Women’s Hospital, Boston, MA 02115, USA
2 Precision Neurology Program, Harvard Medical School, Brigham and Women’s Hospital, Boston, MA 02115, USA
3 Department of Neurology, Brigham and Women’s Hospital, Boston, MA 02115, USA
4 Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
5 Department of Global Public Health and Primary Care, University of Bergen, Norway.
6 Aligning Science Across Parkinson’s (ASAP) Collaborative Research Network, Chevy Chase, MD, 20815

Background and Objectives: Many studies have examined the relation between PD and environmental variables serially --- one candidate association at a time. In the real world however, both environmental exposures and patients are much more complex, including correlated environmental exposures, polypharmacy, and complex comorbidities. Here we begin to characterize a holistic view of environmental, health, and pharmacological traits linked to patients with PD.

Methods: The Harvard Biomarkers Study (HBS) is a large case-control study of PD patients and healthy controls that includes an extensive questionnaire covering past medical and social history data and is thus well-suited for such an exploratory study. Sixty-four environmental, pharmacological, and clinical features were evaluated for associations with PD using logistic regression analysis with backward elimination.

Results: Male gender, coronary artery disease, depression, anxiety, restless leg syndrome, head trauma, ibuprofen use, co-enzyme Q10 use, and vitamin D supplementation were significantly positively associated with PD. By contrast, asthma/chronic obstructive pulmonary disease (COPD), naproxen, ezetimibe, and smoking were significantly negatively associated with PD.

Discussion: This study shows that unbiased, data-rich exploration of the Parkinson phenome has the promise to uncover, prioritize, and clarify associations between environment, multi-system health phenotypes, and PD in a patient-centric manner. Associations with coronary artery
disease, mood disorders, and the cholesterol-absorption inhibitor ezetimibe were revealed that have been largely neglected in traditional hypothesis-driven investigations. Interestingly, asthma/COPD was inversely associated with PD, and this was independent of smoking history. Furthermore, well-established associations were confirmed for male gender, smoking, head trauma, and restless legs syndrome.

**Abbreviations**
PD = Parkinson’s disease, HBS = Harvard Biomarkers Study
Introduction

Parkinson’s disease (PD) is the second most common neurodegenerative disorder. Although a minority of cases are familial, the underlying disease driver for most so-called idiopathic PD cases is unknown. PD is likely to arise through a complex interplay of genetic and environmental factors. Many traditional epidemiological studies have examined the relation between PD, health data, and environmental variables serially. In the real world, however, both environmental exposures and patients are much more complex, including phenotypic diversity, complex comorbidities, and polypharmacy. Here we begin to characterize the phenome of PD, that is, a holistic picture of environmental, clinical, and pharmacological traits linked to patients with PD.

We developed the Harvard Biomarkers Study (HBS) as a resource for identifying genes, and biomarkers for PD. HBS includes an extensive questionnaire regarding past medical history, medication and supplement use history, social history, and environmental exposures. This includes data on exposure to some previously reported putative risk or protective factors (e.g. smoking, pesticides) though not all (e.g. dairy intake, exercise).

In this report, we perform logistic regression with backward elimination to determine which of these health variables are positively or negatively associated with PD. With a few notable exceptions, most epidemiologic studies have largely examined one or a few variables at a time. Without adjusting predictor variables for each other in a multivariate model, complex inter-correlations can produce spurious associations, suppression effects, and other obscuring statistical phenomena. Thus, our study represents a rare attempt to more comprehensively characterize the phenome of PD.

Methods

Harvard Biomarkers Study
The Harvard Biomarkers Study (HBS) is a case-control study including 3,000 patients with various neurodegenerative diseases as well as healthy controls (HC). Informed consent was obtained for all participants. The study protocol was approved by the institutional review board of Mass General Brigham. The HBS questionnaire is included as Supplemental Table 1. For more information on the HBS see:

This analysis was limited to a subgroup of the larger HBS, consisting of 1224 total subjects: 933 with PD and 291 healthy controls. Healthy controls consisted chiefly of spouses, friends, and non-blood relatives, who accompanied PD patients to office visits. Diagnosis of PD was made by a board-certified neurologist with fellowship training in movement disorders. Subjects completed a detailed questionnaire including information on past medical history, current and prior medication use, nutritional supplement use, environmental exposures, Parkinson’s disease risk factors, and social history. Each item in the questionnaire is phrased as a binary YES/NO question, which is followed up by quantitative questions in some instances. For example, the question “Do you drink caffeinated coffee?” is followed by questions asking how many cups per day on average and whether the consumption has changed over the past 10 years. For purposes of this initial study, we have limited the analysis to the single binary YES/NO question for each variable. Additionally, we have examined data from a single
timepoint, that of enrollment in the study. At the time of enrollment, average disease duration was 3.8 years (Table 1).

Statistical differences in demographic data (age, sex, and race) between the cases and controls were determined using a Satterwaite t-test, Chi-square test, or Fisher’s exact test, as appropriate (Table 1). Included in the table are the mean Unified Parkinson’s Disease Rating Scale (UPDRS) and the Mini-Mental State Exam (MMSE) scores. The UPDRS is a 4-part clinical rating scale used to measure severity and progression of PD. In the UPDRS motor subscale (part 3), motor signs were assessed by a trained examiner. The MMSE is a 30-question cognitive battery featuring questions on orientation, registration, attention and calculation, recall, language, and copying.

Variable correlation
To get a preliminary sense of which subsets of the predictor variables are statistically similar and distinct in this data set, we subjected the Pearson correlation matrix of the entire set of 64 variables to a hierarchical clustering algorithm (employing the Corrplot package in R, version 1.1.423 and the graphical software GraphPad Prism, version 8.43).

Logistic regression
Multivariate logistic regression with backward elimination was performed in SAS version 9.4, where cases versus controls (PD or healthy controls) served as the binary dependent variable and the initial predictor set included the 64 binary phenome variables, in addition to the covariates of sex and age. Each predictor variable is statistically adjusted for all other variables still in the model at each given step in the backward elimination. The threshold for retaining predictor variables was p < 0.05. After excluding subjects with missing values, environmental data from 481 PD and 139 healthy controls (620 total subjects) were included in the final logistic regression model (Table 2, Figure 2).

For the supplemental analyses, we determined the univariate p-value for the relation of each individual predictor variable to group status (PD versus healthy controls), (Column 2, Supplemental Table 2). We also determined the p-value for each individual variable controlled for sex and age (but not all other environmental variables) (Column 3, Supplemental Table 2) using logistic regression. Finally, we include the p-value for each individual variable, controlling for sex and age as well as multiple testing using the Sidak step-down method or False Discovery Rate method (Columns 4-5, Supplemental Table 2).

Results

Demographics of the HBS cohort are shown in Table 1. The goal of this study was to evaluate the feasibility and promise of comprehensively characterizing the “PD phenome” by examining 64 variables spanning the categories of past medical history, medications and supplements, exposures previously reported to be PD risk factors, and social history factors (Figure 1).

Because some variables may be correlated with others, we first explored potential correlations using a correlation matrix (Figure 2A). Overall correlation between variables was low: the largest positive correlation coefficient was 0.43 between calcium and vitamin D.
supplement use, and the largest negative correlation coefficient was -0.16 between full dose aspirin (325 mg) and baby aspirin (81 mg) use. However, the hierarchical clustering analysis does demonstrate that several sets of variables are correlated with $P$ values below 0.01, forming clusters of pharmacological, disease, and exposure features that tag biologically related phenomena. The largest clusters in our dataset include a linked sleep and mood disorder cluster, a metabolic syndrome cluster, and a vitamin use cluster (Figure 2B).

We performed a logistic regression analysis with backward elimination to determine which variables were associated with PD. In total, the input included 64 binary health variables (Figure 1) as well as age and sex, assessed in 291 healthy controls and 933 subjects with PD. After the backward elimination and excluding subjects with missing values, 481 PD and 139 HC were included in the final model. The analysis identified 14 statistically significant variables (each significant after simultaneously adjusting for the other 13) after backward elimination (Table 2, Figure 3). Male sex, coronary artery disease (CAD), depression, anxiety, restless leg syndrome, head trauma, ibuprofen use, co-enzyme Q 10 use, and vitamin D supplement use were positively associated with PD. Older age was also associated with PD; odds ratios for each 1-year, 5-year, or 10-year increment in age are shown in Figure 3. Asthma/chronic obstructive pulmonary disease (COPD), naproxen use, ezetimibe use, and smoking were inversely associated with PD. The prevalence of each variable in PD versus healthy controls is shown in Table 2. Additional statistical analyses for all variables are shown in Supplemental Table 2, including the univariate $P$-value and the $P$-value after adjustment for sex and age only.

Among the environmental variables linked to PD, small but statistically significant correlations were found between restless leg syndrome and depression, restless leg syndrome and coronary artery disease, anxiety and depression, depression and asthma/COPD, ezetimibe use and coronary artery disease, and asthma/COPD and ibuprofen use (Figure 4).

Many patients with PD may have more than one factor that is associated with PD, such as being diagnosed with both restless leg syndrome and depression. Indeed, 38% of PD subjects (and 15% of healthy control subjects) were positive for two or more of the eight phenome variables that were identified as positively associated with PD in the logistic regression (Figure 5A). We examined the prevalence and odds ratio (OR) for each pairwise combination of variables positively associated with PD, in the PD population (Figure 5B-C). Among the pairwise combinations, the most prevalent variable found in combination with others was depression, and the single most prevalent combination was having both depression and anxiety (OR 3.00). Furthermore, depression became increasingly enriched in the PD subjects as their total number of positive variables grew. That is, the higher the number of positive variables present in a given subject, the more likely one of those variables was to be depression (Figure 5D). In total, depression was twice as prevalent in PD subjects compared to healthy controls (Table 2).

Discussion

This study reveals an initial multi-dimensional and data-rich view of phenome associations with PD. The goal of this study was to evaluate the feasibility and promise of this approach. As this is a case-control study, the results are not intended to imply direction of causality. That is, while some associations may arise because a phenotypic variable affects risk of PD, others likely arise because of having a diagnosis of PD affects risk of the phenotypic variable.
One strength of our analysis lies in replicating some previously reported associations, including the inverse association between smoking and PD and positive association between head trauma and PD. Of all environmental factors that have been inversely associated with PD, the relationship between smoking and PD is perhaps the best established. This was documented as early as the late 1960’s\textsuperscript{13} and has been reproduced by numerous subsequent epidemiologic studies\textsuperscript{14}. The mechanisms underlying this association are unknown, though there is ample evidence for a protective role of nicotine in dopaminergic neurons in animal models of PD\textsuperscript{15–18}, but this has not been successfully confirmed in human clinical trials\textsuperscript{19–21}. Interestingly, in our data, smoking was associated with a larger metabolic syndrome cluster, which also includes atorvastatin and pravastatin use (Fig. 2B). Several epidemiologic studies have examined statin use in PD\textsuperscript{10,22–25}, with mixed results. Thus, these clusters highlight the need for epidemiologic studies to consider related variables in linked groups.

The evidence for a positive association between head trauma and PD has been more mixed. While several studies including a 2013 meta-analysis have demonstrated an association\textsuperscript{26–28}, multiple large population level Scandinavian studies have not\textsuperscript{29–31}. One potential explanation for the varied results are gene-environment interactions, as head trauma has been suggested to influence risk of PD specifically in subjects with higher genetic risk\textsuperscript{32–34}. Another complicating factor is recall bias as well as the timing of the injury. For example, Rugbjerg et al\textsuperscript{35} found that any association between head trauma and PD could be explained by injuries in the months leading up to diagnosis, suggesting that head injury occurs in the setting of prodromal PD. In contrast, Taylor et al\textsuperscript{36} excluded cases of trauma in the 10 years preceding the diagnosis of PD and found that head injury early in life was associated with PD. Thus, further studies are needed.

Although this analysis confirmed relationships between smoking and head trauma and PD, other previously reported associations, such as the inverse association between caffeine and PD, did not reach significance in the current analysis. Of note, in a prior study of the HBS cohort, we reported an inverse relationship between caffeine and PD\textsuperscript{9}. The discrepancy between that study and the current analysis is explained by the fact that we here simplistically reduced all variables to binary yes/no exposures (considering the large number of variables analyzed), whereas in our prior analysis, caffeine intake was quantified. This highlights the need for future studies examining phenome variables holistically and quantitatively rather than categorically.

Beyond addressing these previously reported associations, our analysis demonstrates novel associations that warrant future study. First, we found an inverse relationship between PD and a past medical history of asthma or COPD. To our knowledge, this is the first report of such an association, though there is one study in a Tawainese population demonstrating increased risk of PD in the setting of asthma or COPD\textsuperscript{37,38}.

It is important to consider whether smoking could explain the inverse association between asthma/COPD and PD. Interestingly, in our deeply phenotyped cohort, a diagnosis of asthma/COPD was not associated with smoking ($P$ value = 0.51 with correlation coefficient of 0.019, Figure 2A); rather, smoking was associated with a larger metabolic syndrome cluster (Figure 2B). This inverse association between asthma/COPD and PD is interesting in light of prior work demonstrating that beta2-adrenoreceptor (beta2-AR) agonists, used to treat asthma and COPD, were associated with repression of alpha-synuclein expression and reduced risk of developing PD, whereas the beta 2 antagonist propranolol (when given for cardiovascular indication and excluding use for essential tremor) was associated with increased alpha-synuclein expression and increased risk of developing PD\textsuperscript{39–41}. Consistent with this, a recent meta-analysis of eight studies indicated that beta2-AR agonist use was associated with reduced PD risk (RR =
0.859, 95% confidence interval [CI] 0.741–0.995), while beta2-AR antagonist use was associated with an increased risk of PD (RR = 1.490, 95% CI, 1.195 to 1.857). In HBS, use of many oral medications was carefully recorded, but inhaled medications such as a beta2-adrenoreceptor agonists used for asthma were not recorded. Thus, the current analysis cannot further clarify, whether asthma medications or asthma/COPD diagnosis mediates this observed association.

Another striking finding in our analysis was the strong prevalence of a history of depression in our PD population (Figure 5). Depression was especially enriched in PD subjects who had multiple positive variables associated with PD (as identified by our logistic regression analysis). The interplay between mood disorders and PD is complex, with some evidence suggesting that these represent either pre-motor or very early non-motor co-comorbidities. These data highlight the importance of screening PD patients for mood disorders, which have a large impact on quality of life and are amenable to treatment.

Finally, an unexpected finding in our study pertains to non-steroidal anti-inflammatory drug (NSAID) use. Non-aspirin NSAIDs have been associated with reduced risk of PD in some but not all studies. In our cohort, naproxen use was negatively associated with PD, whereas ibuprofen use was positively associated with PD (Figure 3, Table 2). (We found no significant association between aspirin or celecoxib use and PD). One potential explanation for this apparent discrepancy between aspirin or celecoxib use and PD). One potential explanation for this apparent discrepancy is that ibuprofen use was most highly correlated with sertraline use (coefficient 0.14), which is in turn correlated with depression (0.18) and anxiety (0.21) (Figure 2), which were strongly associated with PD (Table 2). This example highlights the complicated relationship between different environmental exposures as well as their presumed mechanism of action: NSAIDs are not only anti-inflammatory but affect numerous cellular processes and are highly effective analgesics, which is relevant given the high co-morbidity between pain and depression.

In summary, here we have provided an initial comprehensive characterization of the PD phenome using a cohort from the two major Harvard-affiliated hospitals. Our results confirm some previously reported associations as well as highlight other novel associations. Many of the health variables we have examined here are modifiable, meaning that these results may someday have implications for personalized medicine. Future work will require mechanistic studies to identify gene-environment interactions, to determine which factors are truly causative, and to discover whether modifying them has a neuroprotective or symptomatic benefit. As the only patient cohort with this extensive collection of environmental exposure data combined with whole genome sequencing, the Harvard Biomarkers Study represents an essential resource for undertaking these future studies.

Acknowledgements

The Harvard Biomarkers Study (HBS) (https://www.bwhparkinsoncenter.org) is a collaborative initiative of Brigham and Women’s Hospital and Massachusetts General Hospital, co-directed by Dr. Clemens Scherzer and Dr. Bradley T. Hyman. The HBS Study Investigators are:

Co-Directors: Brigham and Women’s Hospital: Clemens R. Scherzer, Massachusetts General Hospital: Bradley T. Hyman; Investigators and Study Coordinators: Brigham and Women’s Hospital: Yuliya Kuras, Nada Laroussi, Elena Abatzias, Polina Kamenskya. Study Investigators: Brigham and Women’s Hospital: Michael T. Hayes,
We thank all study participants and their families for their invaluable contributions. HBS was seeded by generous support from the Harvard NeuroDiscovery Center, with partial contributions from APDA, the Michael J Fox Foundation, NINDS U01NS082157, U01NS100603, and the Massachusetts Alzheimer’s Disease Research Center NIA P50AG005134.

C.R.S.’s work is supported by NIH grants U01NS095736, U01NS100603, R01AG057331, and R01NS115144, and the American Parkinson Disease Association Center for Advanced Parkinson Research.

The study is funded by the joint efforts of The Michael J. Fox Foundation for Parkinson’s Research (MJFF) and the Aligning Science Across Parkinson’s (ASAP) initiative. MJFF administers the grant [ASAP-000301] on behalf of ASAP and itself.

Author Contributions
ALO: manuscript writing, data analysis, interpretation of results. JYL, TR: statistical consultation, manuscript revision, interpretation of results. CRS: HBS design, project design, manuscript revision, interpretation of results. YK, PK, TY, EA, IT, NL: clinical data management and sample collection. AV, MTH, GPH, JP, VK, TMH, BH, DS, JHG, STG, MAS, AYH, A-M W: referring patients to the study, manuscript review.

Conflicts of Interest
CRS is named as co-inventor on a US patent application on sphingolipids biomarkers that is jointly held by Brigham & Women’s Hospital and Sanofi. CRS has consulted for Sanofi Inc. and Calico; has collaborated with Pfizer, Opko, and Proteome Sciences, Genzyme Inc., and Lysosomal Therapies; is on the Scientific Advisory Board of the American Parkinson Disease Association; has served as Advisor to the Michael J. Fox Foundation, NIH, Department of Defense, and Google; has received funding from the NIH, the U.S. Department of Defense, the Michael J. Fox Foundation, and the American Parkinson Disease Association.
References


**Figure Legends**

Figure 1: Schematic demonstrating the variables examined in this study. Variables span domains of past medical history, medication and supplement use, social history, and environmental exposures. They capture functions of many organ systems.

Figure 2: Heat map demonstrating correlations between the individual environmental variables. A. Pearson correlation matrix. B. Several sets of variables are positively correlated, forming clusters of biologically related phenomena. The largest clusters in our dataset include a linked sleep and psychiatric factor cluster, a metabolic syndrome cluster, and a vitamin use cluster. P-values indicating the significance of the correlation are shown.

Figure 3: Results of logistic regression with backward elimination. Male sex, coronary artery disease, depression, anxiety, restless leg syndrome, head trauma, ibuprofen, co-enzyme Q10, and vitamin D were over-represented in PD. Asthma/chronic obstructive pulmonary disease (COPD), naproxen, ezetimibe use, and smoking as under-represented in PD. Odds ratio with 95% confidence limit for each significant variable in the logistic regression is shown. For the continuous variable of age, the odds ratio is listed for incremental increases of 1 year, 5 years, or 10 years.

Figure 4: Among the variables identified in the logistic regression, small but statistically significant correlations exist. P-values <0.01 are shown within the corresponding squares.

Figure 5: Odds ratio for pairwise combinations of variables enriched in PD. A. Distribution of variables positively associated with PD across the HC and PD groups. In the HC group, the majority of subjects are positive for none of the variables associated with PD. In the PD group, the majority of subjects are positive for 1 or more variables associated with PD. B. Distribution of pairwise combinations of variables positively associated with PD among the PD group. C. Odds ratio and confidence limits for pairwise combinations of variables positively associated with PD. Note: For some combinations, OR cannot be calculated because no HC subjects were positive for that pairwise combination. D. Prevalence of depression increases with increasing number of positive variables in PD subjects.
**Figure 1**

- **Medications (26)**
  - Diabetes mellitus
  - Hypothyroidism
  - Multivitamin
- **Past medical history (16)**
  - Restless leg syndrome
  - Periodic limb movements of sleep
  - REM behavior disorder
  - Obstructive sleep apnea
  - Mild cognitive impairment
  - Depression
  - Anxiety
  - Bupropion
  - Alprazolam
  - Sertraline
  - Duloxetine
  - Vitamin E
  - Head trauma
  - Neuroleptic use
  - Metal poisoning
  - Industrial profession
  - Pesticide exposure
  - Encephalitis
  - Alcohol
- **Vitamins and supplements (9)**
  - Zinc
  - Vitamin C
  - Omeprazole
  - Pantoprazole
  - Lansoprazole
  - Famotidine
  - Coffee
  - Tea
  - Soda
  - Vegetarian
- **Environmental exposures (8)**
  - Asthma/COPD
  - Smoking
- **Social history (5)**
  - Hypertension
  - Hyperlipidemia
  - Coronary artery disease
  - Heart failure
  - Aspirin (325 mg)
  - Aspirin (81 mg)
  - Lisinopril
  - Losartan
  - Hydrochlorothiazide
  - Atenolol
  - Ezetimibe
  - Ezetimibe/Simvastatin
  - Atorvastatin
  - Pravastatin
  - Simvastatin
  - Lovastatin
  - Rosuvastatin
  - Fluvastatin

*Total = 64*
Figure 3

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>Lower limit</th>
<th>Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAD</td>
<td>6.15</td>
<td>1.53</td>
<td>49.12</td>
</tr>
<tr>
<td>RLS</td>
<td>3.69</td>
<td>1.46</td>
<td>11.41</td>
</tr>
<tr>
<td>Depression</td>
<td>3.23</td>
<td>1.73</td>
<td>6.37</td>
</tr>
<tr>
<td>Male</td>
<td>3.11</td>
<td>1.97</td>
<td>4.94</td>
</tr>
<tr>
<td>Anxiety</td>
<td>2.99</td>
<td>1.36</td>
<td>7.36</td>
</tr>
<tr>
<td>Co-Q</td>
<td>2.68</td>
<td>1.27</td>
<td>6.31</td>
</tr>
<tr>
<td>Ibuprofen</td>
<td>2.52</td>
<td>1.27</td>
<td>5.36</td>
</tr>
<tr>
<td>Head trauma</td>
<td>2.27</td>
<td>1.18</td>
<td>4.73</td>
</tr>
<tr>
<td>Vit D</td>
<td>1.94</td>
<td>1.86</td>
<td>3.89</td>
</tr>
<tr>
<td>Age (10)</td>
<td>1.53</td>
<td>1.25</td>
<td>1.89</td>
</tr>
<tr>
<td>Age (5)</td>
<td>1.24</td>
<td>1.12</td>
<td>1.38</td>
</tr>
<tr>
<td>Age (1)</td>
<td>1.04</td>
<td>1.02</td>
<td>1.07</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.46</td>
<td>0.29</td>
<td>0.72</td>
</tr>
<tr>
<td>Asthma/COPD</td>
<td>0.36</td>
<td>0.11</td>
<td>0.80</td>
</tr>
<tr>
<td>Naproxen</td>
<td>0.30</td>
<td>0.11</td>
<td>0.69</td>
</tr>
<tr>
<td>Ezetimibe</td>
<td>0.05</td>
<td>0.004</td>
<td>0.47</td>
</tr>
</tbody>
</table>
Figure 4

The figure shows a correlation matrix with various conditions and medications. The values range from -0.35 to 0.45, with significance levels indicated for some comparisons. For example, the correlation between RLS and Anxiety is 0.001, and between Head trauma and Smoking is <0.0001. The significance levels are indicated by color: red for p < 0.0001, blue for p < 0.0004, and white for p > 0.002.
Figure 5

A

HC

PD

Total=291

Total=933

Total=229

B

PD

C

D

No depression

Depression

% with depression

Number of variables

0

1

2

3

4

5

6

Odds Ratio

[Graph showing data and odds ratios for different variables associated with depression.]
Table 1: Demographics of HBS. Statistical differences in demographic data (age, sex, and race) between the cases and controls were determined using a Satterwaithe t-test, Chi-square test and Fisher’s exact test, as appropriate. UPDRS = Unified Parkinson’s Disease Rating Scale. MMSE = Mini-Mental State Exam.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Healthy Controls (HC)</th>
<th>Parkinson’s disease (PD)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Male</td>
<td>125 (43%)</td>
<td>620 (66.5%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>166 (57%)</td>
<td>312 (33.5%)</td>
<td></td>
</tr>
<tr>
<td><strong>Age at enrollment (years)</strong></td>
<td>63.5 ±11.7</td>
<td>66.1 ± 9.9</td>
<td>0.0006</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>2 (0.7%)</td>
<td>9 (0.7%)</td>
<td>1.000</td>
</tr>
<tr>
<td>Black/African American</td>
<td>13 (4.9%)</td>
<td>9 (0.7%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Hawaiian/Pacific</td>
<td>1 (0.3%)</td>
<td>0</td>
<td>0.1920</td>
</tr>
<tr>
<td>More than one race</td>
<td>3 (1%)</td>
<td>3 (0.2%)</td>
<td>0.0888</td>
</tr>
<tr>
<td>Unknown</td>
<td>4 (1.4%)</td>
<td>38 (3.1%)</td>
<td>0.1154</td>
</tr>
<tr>
<td>White</td>
<td>268 (92.1%)</td>
<td>1166 (95.2%)</td>
<td>0.0432</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>8 (2.7%)</td>
<td>17 (1.4%)</td>
<td>0.1211</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>277 (95.2%)</td>
<td>1168 (95.3%)</td>
<td>0.8779</td>
</tr>
<tr>
<td>Unknown</td>
<td>6 (2.1%)</td>
<td>40 (3.3%)</td>
<td>0.3447</td>
</tr>
<tr>
<td><strong>Average education level (years, S.D.)</strong></td>
<td>15.0 ± 1.77</td>
<td>15.2 ± 1.76</td>
<td>0.159</td>
</tr>
<tr>
<td><strong>Age at onset</strong></td>
<td>N/A</td>
<td>62.3 ± 10.4</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Disease duration</strong></td>
<td>N/A</td>
<td>3.8 ± 4.4</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Hoehn and Yahr stage</strong></td>
<td>N/A</td>
<td>2.1 ± 0.64</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>UPDRS motor subscale</strong></td>
<td>N/A</td>
<td>19.8 ± 10.2</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>UPDRS total</strong></td>
<td>N/A</td>
<td>33.2 ± 15.1</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>MMSE</strong></td>
<td>28.99 ± 1.26</td>
<td>28.21 ± 2.30</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Table 2: Results of multivariate logistic regression after backward elimination. P-value, odds ratio, 95% confidence limit, and prevalence in HC versus PD for the 13 retained variables are shown.

<table>
<thead>
<tr>
<th>No.</th>
<th>Feature</th>
<th>P-value (adjusted for all other variables)</th>
<th>Odds ratio estimate</th>
<th>95% C.I.</th>
<th>Healthy Controls</th>
<th>Parkinson’s disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Age (1 year increase)</td>
<td>&lt;0.0001</td>
<td>1.044</td>
<td>1.022 - 1.066</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Age (5 year increase)</td>
<td>1.237</td>
<td>1.117 - 1.375</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Age (10 year increase)</td>
<td>1.531</td>
<td>1.248 - 1.890</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Male gender</td>
<td>&lt;0.0001</td>
<td>3.107</td>
<td>1.972 - 4.944</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Coronary artery disease</td>
<td>0.0313</td>
<td>6.154</td>
<td>1.539 - 49.123</td>
<td>9 (3.21%)</td>
<td>66 (7.42%)</td>
</tr>
<tr>
<td>5</td>
<td>Depression</td>
<td>0.0004</td>
<td>3.225</td>
<td>1.726 - 6.359</td>
<td>34 (12.2%)</td>
<td>229 (25.7%)</td>
</tr>
<tr>
<td>6</td>
<td>Anxiety</td>
<td>0.0101</td>
<td>2.992</td>
<td>1.360 - 7.364</td>
<td>18 (6.5%)</td>
<td>140 (15.8%)</td>
</tr>
<tr>
<td>7</td>
<td>Asthma/COPD</td>
<td>0.0109</td>
<td>0.364</td>
<td>0.168 - 0.801</td>
<td>23 (8.2%)</td>
<td>53 (6.0%)</td>
</tr>
<tr>
<td>8</td>
<td>Restless leg syndrome</td>
<td>0.0112</td>
<td>3.691</td>
<td>1.464 - 11.410</td>
<td>5 (3.0%)</td>
<td>75 (12.6%)</td>
</tr>
<tr>
<td>9</td>
<td>Ibuprofen</td>
<td>0.0113</td>
<td>2.520</td>
<td>1.271 - 5.357</td>
<td>28 (9.9%)</td>
<td>127 (14.2%)</td>
</tr>
<tr>
<td>10</td>
<td>Naproxen</td>
<td>0.0250</td>
<td>0.304</td>
<td>0.109 - 0.888</td>
<td>12 (4.2%)</td>
<td>26 (2.9%)</td>
</tr>
<tr>
<td>11</td>
<td>Ezetimibe</td>
<td>0.0107</td>
<td>0.050</td>
<td>0.004 - 0.470</td>
<td>3 (1.9%)</td>
<td>3 (0.52%)</td>
</tr>
<tr>
<td>12</td>
<td>Head trauma</td>
<td>0.0202</td>
<td>2.269</td>
<td>1.175 - 4.732</td>
<td>25 (9.1%)</td>
<td>176 (19.8%)</td>
</tr>
<tr>
<td>13</td>
<td>Coenzyme Q</td>
<td>0.0149</td>
<td>2.681</td>
<td>1.269 - 6.308</td>
<td>22 (7.7%)</td>
<td>170 (19.0%)</td>
</tr>
<tr>
<td>14</td>
<td>Vitamin D</td>
<td>0.0092</td>
<td>1.937</td>
<td>1.856 - 3.888</td>
<td>34 (23.6%)</td>
<td>202 (40.3%)</td>
</tr>
<tr>
<td>15</td>
<td>Smoking</td>
<td>0.0007</td>
<td>0.459</td>
<td>0.291-0.717</td>
<td>142 (50.1%)</td>
<td>365 (41.1%)</td>
</tr>
</tbody>
</table>