

1 **Population fraction of Parkinson's disease attributable to preventable risk** 2 **factors**

3
4 Haydeh Payami*^{1,2}, Gwendolyn Cohen^{1,2}, Charles F Murchison^{1,2,4}, Timothy R Sampson^{2,3}, David G
5 Standaert^{1,2}, Zachary D Wallen¹

6
7 ¹ Department of Neurology, University of Alabama at Birmingham, Birmingham, AL, 35233, USA

8 ² Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, MD,
9 20815, USA

10 ³ Department of Cell Biology, Emory University School of Medicine, Atlanta GA 30329, USA

11 ⁴ Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, 35233, USA

12

13 *Senior and Corresponding author: haydehpayami@uabmc.edu

14

15 **Abstract**

16 Parkinson's disease is the fastest growing neurologic disease with seemingly no means for prevention. Intrinsic
17 risk factors (age, sex, genetics) are inescapable, but environmental factors are not. We studied population
18 attributable fraction and estimated fraction of PD that could be reduced if modifiable risk factors were eliminated.
19 Assessing several known risk factors simultaneously in one study, we demonstrate that all were operative and
20 independent, underscoring etiological heterogeneity within a single population. We investigated repeated blows to
21 head in sports or combat as a potential new risk factor, and found it was associated with two-fold increased risk of
22 PD. Considering modifiable risk factors, 23% of PD cases in females were attributable to pesticides/herbicides
23 exposure, and 30% of PD cases in males was attributable to pesticides/herbicides, Agent Orange/chemical
24 warfare, and repeated blows to the head. Thus, one-in-three cases of PD in males, and one-in-four cases in
25 females could have potentially been prevented.

26 Disability and death due to Parkinson's disease are rising faster than for any other neurological disease in
27 the world.¹ In 2016, six million people had PD globally, which was double the number of cases reported
28 in 1990.^{1,2} By 2040, the prevalence is expected to triple to 17 million². These prevalence rates may be
29 underestimated due to lack of awareness of PD and limited access to health care in many populations. The
30 economic burden of PD in 2017 was \$51.9 billion in US alone³. In 2022 The World Health Organization
31 declared "A pressing need for effective preventive actions ... to slow the rising incidence of PD before
32 the burden and costs of treatment overwhelm country health services"⁴.

33
34 PD is an etiologically heterogenous syndrome that affects many systems and is progressively
35 debilitating^{5,6}. The earliest signs are often gastrointestinal and olfactory dysfunction, depression, and
36 sleep disorder. Movement disorders, the cardinal signs required for diagnosis, appear after decades of
37 prodromal disease and neurodegeneration. By the time of diagnosis, over 80% of dopaminergic cells are
38 already destroyed. At later stages, many develop psychosis and dementia. There is currently no
39 prevention and no treatment to stop neurodegeneration. All clinical trials for disease modifying treatments
40 have failed partly because PD is heterogenous, subtypes are clinically indistinguishable, and biomarkers
41 to select appropriate patient population for trials are lacking. PD is likely not a single disease, rather a
42 syndrome with a range of causes, all converging to produce the phenotype we call PD^{5,7}. Three to five
43 percent of PD are caused by pathogenic mutations in any one of causal genes, which are rare, and have
44 seemingly different functions, e.g., *SNCA*, *PRKN*, *LRRK2*, *PINK1*⁶. The vast majority of PD remains
45 idiopathic, thought to be caused by environmental triggers interacting with intrinsic risk factors such as
46 genetic susceptibility^{6,8}.

47
48 Currently, the most compelling weapon against the global rise in PD is elimination of risk factors⁹.
49 Intrinsic risk factors, which are not readily amenable to prevention, include increasing age, the male sex,
50 positive family history of PD, and genetic causes and susceptibility variants^{6,10-15}. Among environmental
51 factors, neurotoxicant exposure, especially pesticide and herbicides, has been robustly associated with
52 increased risk of PD^{8,16-18}. Head injury is also common prior to development of PD but its causal
53 relationship is questionable¹⁹⁻²². Cigarette smoking and coffee drinking are associated with reduced risk of
54 PD^{23,24}, and their protective effects are independent and additive such that non-smokers and caffeine users
55 have extremely low risks^{25,26}.

56
57 Here, we investigated several factors simultaneously in a single population. Doing so enabled us to
58 compare their effects side-by-side, test inter-dependence and overlap among them, and assess the burden
59 of disease due to modifiable risk factors. We investigated family history of PD, exposure to

60 pesticides/herbicides, Agent Orange (an herbicide used in the Vietnam war)/chemical warfare, mild or
61 moderate traumatic brain injury (MTBI)/concussion, and repeated blows to the head in sports or military
62 as a potential new risk factor. We addressed the following questions. (1) Are all risk factor operative in
63 one population? If only one or few risk factors predominate in a population, that would be an opening to
64 teasing out etiologic heterogeneity. (2) Do repetitive hits to the head that do not have an immediate
65 clinical manifestation increase risk of PD? There is growing evidence linking repetitive hits to the head in
66 collision sports to neurodegenerative disease later in life^{22,27}. Prior PD studies have defined head injury as
67 hospital-treated MTBI or concussion¹⁹⁻²². It is not known if repetitive hits to the head with no immediate
68 signs of injury increase PD risk. (3) Are the risk factors independent of each other? The few studies that
69 have addressed interdependence among protective and risk factors found them to be independent^{25,26,28},
70 and suggested the risk factors mark etiological subtypes of PD²⁸. (4) What percentage of PD could
71 potentially be reduced if modifiable risk factors were eliminated? Population attributable fraction (PAF)
72 is an epidemiological metric commonly used to estimate the fraction of disease that is attributed to a risk
73 factor^{29,30}. It has been applied widely to many disorders including dementia³¹, but not to PD. One study
74 has calculated population attributable risk (PAR not PAF) for PD reporting that 54% of PD risk is
75 explained by the combined effects of family history of PD, not smoking, and occupational exposure to
76 lead, copper, and insecticides³². Here we calculate PAF, with intervention as the goal, for potentially
77 modifiable risk factors. To our knowledge, this is the first study of PAF for PD.

78
79 The study was conducted in Birmingham, Alabama, in the center of the Deep South of the United States.
80 The Deep South is understudied. Little is known about epidemiology of PD in the Deep South except that
81 in surveys of United States, the Deep South states were found to have high incidence of PD¹⁰, and high
82 rates of mortality from PD³³. We show that all risk factors including repeated blows to the head, but
83 possibly not MTBI/concussion, are operative in this geographically and culturally defined population, the
84 risk factors are independent, and some are actionable candidates for disease prevention, and plausible
85 markers for disease subtyping to reduce heterogeneity in clinical trials.

86 87 **Results**

88 89 *Subject characteristics and PD features*

90
91 Study included 808 persons with PD (PwP) and 415 neurologically healthy controls (NHC). Nearly all
92 study participants were living in the Deep South and peripheral states and 77% of them were born in the
93 Deep South. This relatively homogenous population, compared to metropolitan regions of the world,

94 provided a convenient sample to determine which features and risk factors are operative in a single
95 population. All known features of PD that were tested were captured in this population with statistical
96 significance (**Table 1**). Notably, among PwP, 63% were males ($P < 1E-5$). Constipation was five-fold more
97 prevalent in PwP than in NHC ($P = 4E-25$). Twice as many PwP reported having lost >10 pounds in the
98 prior year than NHC ($P = 4E-7$). Ten percent of PwP vs. none of the NHC reported having had rapid eye
99 movement sleep behavior disorder (RBD), a rare disorder that often leads to PD ($P = 2E-6$). Recapitulating
100 these features establishes robustness of the data.

101

102 *Risk factors*

103

104 We collected data and tested association of the following risk factors with PD simultaneously in a single
105 study population of 808 PwP and 415 NHC: family history of PD, exposure to pesticides/herbicides,
106 exposure to Agent Orange/warfare chemicals, and MTBI/concussion. We also investigated repeated
107 blows to the head in sports or combat as a potential risk factor. The exact questions that were used to
108 collect the data are shown in **Table 2**.

109

110 We detected a strong and robust association signal for every risk factor examined except
111 MTBI/concussion. **Table 3** shows raw unadjusted odds ratios (OR), as if each risk factor was assessed
112 individually, first for all subjects, and then stratified by sex. **Table 4** shows results of conditional analysis
113 where we included all risk factors in one model and estimated adjusted OR and P values for each risk
114 factor conditioned on all other risk factors including age. Sex could not be included in conditional
115 analysis because of collinearity of male sex with repeated blows to the head (all male) and exposure to
116 Agent Orange/warfare chemicals (all male except one female).

117

118 Family history of PD was three- to four-times more common in PwP than NHC and was highly
119 significant and robust to sex-stratification (**Table 3**), adjustments in conditional analysis, and sex-
120 stratified conditional analysis (**Table 4**).

121

122 Pesticide/herbicide exposure was associated with a robust two- to three-fold increased risk of PD, in all
123 subjects, in both sexes (**Table 3**), and when adjusted for other risk factors (**Table 4**). Duration of
124 exposure was also significantly longer, almost double, for PwP than NHC (**Table 3**).

125

126 Unadjusted result for MTBI/concussion was initially significant showing a two-fold increase in PD
127 (**Table 3**). However, the signal weakened in sex-stratified data (**Table 3**) and was further reduced and lost

128 significance in conditional analysis (**Table 4**). Prior studies have questioned the causal relationship of
129 MTBI/concussion with PD because it is possible that balance problems in prodromal PD may be causing
130 falls and head injury (reverse causality). One study showed that association of MTBI/concussion with PD
131 was limited to the events occurring within 10 years of PD diagnosis²⁰. In our study also, excluding
132 MTBI/concussion within 10 years of PD onset, had a notable effect on weakening the signal (**Table 3**).

133

134 The unadjusted results for Agent Orange/chemical warfare and repeated blows to the head are skewed.
135 Note that (a) there is a sex differential in this study, caused initially by inevitable increased prevalence of
136 PD in males, and exaggerated by preferential selection of spouses as controls, and (b) sex could not be
137 included in the conditional models because of strong collinearity of male-sex with repeated blows to head
138 and Agent Orange/chemical warfare. Hence, the sex-stratified analyses are critical for interpreting the
139 results.

140

141 Repeated blows to the head was initially estimated as four-fold more common in PwP than NHC (**Table**
142 **3**). However, repeated blows to the head was reported by males only, and in males, it was associated with
143 two-fold increased risk and was significant both individually (**Table 3**) and after adjusting for other risk
144 factors including MTBI/concussion (**Table 4**).

145

146 Exposure to Agent Orange/chemical warfare yielded an unadjusted eight-fold increased risk (**Table 3**).
147 Only one female, who had PD, reported exposure to Agent Orange/chemical warfare, the rest of the
148 exposed were males. In males, Agent Orange/chemical warfare yielded an unadjusted four-fold increased
149 risk. Interestingly, Agent Orange/chemical warfare association gained strength in conditional analysis,
150 yielding a 10-fold increase in all subjects, and six-fold increase in males (**Table 4**).

151

152 We conclude that family history, exposure to pesticides/herbicides, Agent Orange/chemical warfare (as
153 seen predominantly in males), and repeated blows to the head (as seen only in males) are all operative in
154 this population, and that they have distinct independent effects on PD risk. Demonstration of
155 independence does not preclude possibility of interaction, for example, between a certain genetic locus
156 (captured by family history) and a chemical compound (captured by pesticides/herbicides). It does
157 however establish each risk factor as a significant contributor to risk, independent of other factors.

158

159 *PAF*

160

161 PAF is an epidemiological metric to assess public health impact of a risk factor. PAF incorporates the
162 strength of the risk (OR) with prevalence of risk factor to provide an estimate for the percentage of
163 disease that would be reduced if risk factor were eliminated. We calculated PAF for risk factors that are
164 modifiable, in the appropriate groups i.e., pesticide/herbicide exposure in males and females, Agent
165 Orange/chemical warfare in males, and repeated blows to head in males. In calculating PAF we used
166 adjusted OR (from **Table 4**) that were conditioned on all risk factors including age, family history,
167 MTBI/concussion, pesticide/herbicide exposure, Agent Orange/chemical warfare, and repeated blows to
168 head, as applicable. Hence, all PAF estimates are adjusted PAFs. Adjusted PAF for pesticide/herbicide
169 exposure was 23% [12%-33%] for females and 17% [7%-27%] for males. In addition, among males,
170 adjusted PAF was 10% [0.3%-20%] for repeated blows to the head, and 6% [0.03%-12%] for Agent
171 Orange/chemical warfare exposure.

172
173 In women, only one modifiable risk factor was detected. In men, there were three. To assess fraction of
174 disease in males that was attributed to all three risk factors, we calculated joint PAF for the combination
175 of pesticide/herbicide exposure, Agent Orange/chemical warfare, and repeated blows to head. Joint PAF
176 assumes that risk factors are independent. In conditional analysis (**Table 4**) we established that each risk
177 factor has an independent effect on PD risk. To assess if there is interaction (e.g., synergism), we tested
178 pairwise interaction for the three risk factors in males. There was no evidence for interaction in this
179 dataset. The interaction P values were $P=0.99$ for [pesticide/herbicide exposure* Agent Orange/chemical
180 warfare], $P=0.99$ for [Agent Orange/chemical warfare*repeated blows], and $P=0.56$ for
181 [pesticide/herbicide exposure* repeated blows to head]. Having established independence and lack of
182 interaction, we calculated joint PAF for Agent Orange, pesticide/herbicide exposure, and repeated blows
183 to head in sports or military in males as 30% [7%-49%].

184

185 **Discussion**

186

187 **Summary**

188

189 We investigated multiple risk factors simultaneously in a single population and addressed four questions:
190 (1) Are all major risk factor of PD operative in one population? We had hoped that if only one or few risk
191 factors predominate in a population, that would be an opening to teasing out etiologic heterogeneity.
192 Nearly all features and risk factors tested were associated with PD. Thus, the extensive heterogeneity of
193 PD is encapsulated even in this relatively isolated population in the Deep South. (2) Do repetitive hits to
194 the head that do not have an immediate clinical manifestation increase risk of PD? Repeated blows to the

195 head, which is common in collision sports and do not normally require medical care, was associated with
196 two-fold increases risk of PD. (3) Are the risk factors independent of each other? Family history, toxicant
197 exposure, and repeated blows to the head are independent risk factors with little overlap, which is
198 consistent with a prior report²⁸. (4) What percentage of PD could potentially be reduced if modifiable risk
199 factors were eliminated? One in three cases of PD in males, and one in four cases in females were
200 attributed to modifiable risk factors and could have potentially been prevented. Results have two broad
201 implications. First is prevention. PD is a complex disease with both intrinsic (genetics, aging) and
202 extrinsic (environmental, life-choices) risk factors. While intrinsic factors are unescapable, extrinsic
203 triggers may be preventable, thereby averting the disease. PAF estimates suggest a substantial fraction of
204 PD is potentially preventable. Second is potential markers for disease subtypes. Clinical trials for
205 neuroprotective treatments have all failed, due in part to lack of biomarkers to distinguish subtypes of PD.
206 That the risk factors for PD are independent indicates family history of PD, repeated blows to head, and
207 pesticide/herbicide exposure tag etiologic subtypes of PD.

208

209 *PAF*

210

211 PAF bridges science and decision making. It provides an estimate of excess risk that could potentially be
212 eliminated, a starting point for discussing intervention. For PAF to be meaningful, several requirements
213 must be met. Intervention must be possible. To that point, we calculated PAF only for modifiable risk
214 factors and not for age, family history, or sex. The term “attributable” has a causal interpretation. The risk
215 factors selected for PAF in this study have been causally linked to PD³⁴⁻³⁶. There is probably variation in
216 genetic susceptibility that make people more or less vulnerable to potential damage caused by toxicant
217 exposure or head injury. The cause is technically gene*environment interaction. However, for PAF, the
218 environmental trigger meets the definition/assumption of causality. Another requirement for accurate
219 estimation of PAF is no bias in the study design and data being valid and accurate. We assert the
220 robustness of our data from the strict QC measures that were implemented, subject selection was
221 sequential and from only one movement disorder clinic, methods were standardized, and the resulting
222 data recapitulated every known feature and risk factor of PD with high statistical significance and
223 reasonable effect size. ORs were adjusted for all risk factors in the study, and adjusted PAFs were
224 calculated using the adjusted ORs. Since risk factors are not mutually exclusive (a farmer exposed to
225 toxicants could also have played a collision sport and had repeated blows to the head) the overlaps were
226 accounted for by correct calculation of joint PAF^{30,37}, having first established independence and absence
227 of interaction, and then using adjusted ORs that were conditioned on all other risk factors. Finally, the
228 interpretation of PAF is dependent on how exactly the exposure was defined; to that end, we have

229 provided the exact questions that were asked to solicit the data in **Table 2**. In short, we believe we have
230 met the assumptions of PAF, conducted a rigorous study, and results are robust, suggesting a substantial
231 fraction of PD could have been prevented if these individuals had not been exposed to heavy uses of
232 pesticide/herbicides or Agent Orange/chemical warfare and had avoided activities that caused repeated
233 blows to the head. We note that this is the first attempt at estimating PAF for PD and point out that the
234 PAF estimates will be population specific depending on the distribution and frequency of risk factors.

235

236 *Pesticides/herbicide*

237

238 Pesticides and herbicides are among the most robust risk factors for PD^{8,16-18}. Some pesticide and
239 herbicides are inherently neurotoxic³⁸. For instance, particular insecticides have been designed and
240 developed to target neuronal pathways that are shared between insects and animals. PAF estimates
241 obtained here suggest that exposure to pesticides/herbicide accounted for 23% of PD cases among
242 females and 17% among males. This is an underestimate because our calculations were based on self-
243 reported exposures to heavy uses of pesticides/herbicides and did not include exposure through
244 contaminated soil, water, and consuming contaminated fruits, vegetables, meat, and fish³⁹⁻⁴¹. US allows
245 use of 72 pesticides that are banned in EU, and their usage in US has stayed constant or increased in the
246 past decade⁴². Pesticide levels in surface waters are higher in US than EU and China⁴³. Given such a
247 robust risk and prevalence, and the potential for disease prevention, it is prudent that steps be taken to
248 regulate access to and limit exposure to these toxicants, both acutely and chronically.

249

250 Our study does not identify any one particular chemical that is directly associated with PD. In the US, few
251 states document the use of pesticides and herbicides, with very few making such data publicly accessible.
252 Thus, it is difficult to pinpoint the risk effect of one specific chemical, outside of few locations (such as
253 California⁴⁴). Mandatory and enforced registration of pesticide application would be a first step to better
254 understand exactly which chemicals we are exposed to, at what dose, and for how long. We still depend
255 on certain pesticides and herbicides for modern agriculture and human health. Pyrethroids, for example,
256 are important tools to control mosquito populations for limiting malaria, West Nile and Dengue viruses³⁸.
257 However, for certain chemicals with robust evidence of long-term harm, such as paraquat that has been
258 firmly linked to PD in both human and experimental settings^{5,38}, complete bans, as done in other
259 countries, should seriously be considered.

260

261 *Head injury*

262

263 The association of PD with repeated blows to the head that do not cause concussion or require medical
264 care was a notable finding in this study. Prior PD studies defined head injury as MTBI that caused
265 concussion or needed medical care. MTBI/concussion is often found to be more common among PwP,
266 however, causality has been questionable^{19,20}. Is it MTBI/concussion that increases risk of PD or is it
267 impaired balance in prodromal PD that increases risk of falls and head injuries? MTBI/concussion was
268 more common in PwP in our study as well, but it lost its effect and significance when adjusted for other
269 risk factors. We also saw a weakening signal when we excluded MTB/concussion within 10 years prior to
270 PD onset. MTBI/concussion was the only risk factor that did not retain its signal after adjusting for other
271 risk factors, even when all injuries up to onset of PD were included. Repeated blows to the head,
272 however, had a robust signal of association that withstood adjustment for all other risk factors including
273 MTBI/concussion. The most common source of repeated blows to the head is collision sports. Studies
274 have shown adverse consequences of collision sports (mainly American football and soccer) on brain
275 health later in life^{22,27,45}. It is well-publicized in the popular media that football players go on to suffer
276 high rates of depression, mood and behavioral problems, cognitive loss, dementia, and chronic traumatic
277 encephalopathy. Despite high-profile national coverage by media on the dangers, American football
278 remains highly popular in the US. In our study population, only men reported having had repeated blows
279 to the head as youths. The popularity of collision sports is not waning among boys and is rising among
280 school age girls. It is therefore imperative to implement effective safety measures for football, soccer,
281 and all collision sports.

282

283 ***Caveats, Limitations, Confounders***

284

285 Repeated blows to the head as a risk factor for PD needs to be tested and confirmed in other populations.
286 PAF estimates will vary across populations depending on the prevalence of the risk factors.

287

288 The PAF estimates reported here were based on a population of older adults and their past experiences
289 that may have put them at risk. They are not necessarily predictive of the future. For example, in this
290 cohort of older Southerners, no female reported repeated blows to the head due to sports or military, and
291 only one female (with PD) had been exposed to Agent Orange. In the 1900's when these women were
292 young, women did not partake in collision sports or military, as they do now. The PAFs for future will
293 change, for the better or worse, depending on the actions we take now to clean our environment and
294 improve health and safety standards.

295

296 As noted in sections above, we incorporated guards against potential confounders in our study design and
297 statistical analyses. Here, we address two other commonly cited potential confounders in epidemiological
298 studies of PD: reverse causality, and risk aversion. MTBI/concussion, although it was eliminated
299 statistically, is a good example of confounding by reverse causation. Difficulty with balance may be
300 present in prodromal PD and cause falls resulting in head injury, hence the possibility of PD causing head
301 injuries, not vice versa. Reverse causality is not likely for repeated blows to head because they occur
302 many decades prior to onset of PD, in children and young adults who partake in collision sports or
303 military. Risk aversion, noted to be a premorbid personality trait⁴⁶, complicates interpretation of negative
304 associations such as smoking with PD. For example, while nicotine is known to be neuroprotective, it is
305 still unclear if smoking is protective against PD or individuals who will eventually develop PD are averse
306 to smoking due to its associated risks. This argument does not apply equally to risk factors as it does to
307 protective factors. If anything, risk averse individuals would avoid collision sports, military, and toxicant
308 exposure. Thus, if risk avoidance is a confounder for association of PD with repeated blows to the head,
309 or with toxicant exposure, their true effect sizes could be larger than our estimates.

310

311 *Disease subtypes*

312

313 PD is a heterogenous syndrome with more than one cause^{5,6}. Currently, the subtypes within idiopathic
314 PD are indistinguishable. Disease heterogeneity, and unavailability of biomarkers to distinguish disease
315 subtypes, is considered the main reason that clinical trials for disease modifying PD drugs fail. It has
316 been proposed that the major risk factors for PD, that are independent of each other in their association
317 with PD, represent markers for etiological subtypes²⁸. Our study was set in the Deep South, which is a
318 culturally unique and relatively isolated and stable population in the US. Over 95% of Subjects in this
319 study live in the Deep South and 77% of them were born here. The relative uniformity in this population
320 compared to more cosmopolitan regions of the world provided a desirable setting to test all risk factors
321 simultaneously, in search of population-specific risk factors. We found all risk factors and features of PD
322 robustly associated with PD in our study population. Our results indicate risk heterogeneity is present, to
323 full extent, within even a relatively homogenous population.

324

325 Extensive efforts and cost are being spent in search of molecular biomarkers of disease subtypes in blood,
326 cerebral spinal fluid, and microbiome. In a breakthrough, alpha-synuclein seed amplification assay (SAA)
327 was recently described as a diagnostic biomarker in symptomatic PD and some high-risk populations⁴⁷.
328 SAA's detection power varies by genetics and clinical features, which underscores the etiological
329 heterogeneity of PD, although SAA does not distinguish disease subtypes. Currently, the only markers

330 that can distinguish some PD subtypes are the highly penetrant pathogenic mutation in a causal gene
331 (*SNCA*, *PRKN*, *LRRK2*, etc.), which are rare. For common idiopathic PD, there is currently no clear way
332 to define and divide the subtypes⁴⁸. As efforts for molecular biomarker discovery progress, we propose a
333 simple, crude but potentially powerful alternative for the meantime. Here, we demonstrated that repeated
334 blows to the head, exposure to pesticides/herbicides, and having a family history of PD are each an
335 independent risk factor for PD. They all precede PD. Each is a biologically plausible cause. Their
336 independence suggests they could represent markers for different etiological subtypes of PD. Belvisi et
337 al.²⁸ also found family history of PD and exposure to pesticides (and metals and anesthesia) to be
338 independent risk factors, and surmised that they mark etiological subtypes of PD. We agree. To that end,
339 we propose that taking a brief history – as simple as the questions in **Table 2** - could help gauge and
340 potentially reduce heterogeneity in clinical trials and improve odds of success.

341

342 **Methods**

343

344 *Enrollment of Study participants*

345

346 We have complied with all relevant ethical regulations. The study was approved by the Institutional
347 Review Board (IRB) for Protection of Human subjects at the University of Alabama at Birmingham
348 (UAB) and by the Office of Human Research Oversight (OHRO) of United States Department of Defense
349 (DoD, funding agency). All subjects signed informed consent. No compensation was provided for
350 participating in the study.

351

352 We enrolled 981 persons with PD (PwP) and 485 neurologically healthy controls (NHC). Inclusion
353 criteria for PwP were diagnosis of PD by a movement disorder specialist neurologist at UAB according to
354 UK Brain Bank criteria⁴⁹, and informed consent. PwP were enrolled sequentially as they were seen at
355 clinic. Controls consisted of patients' spouses and community volunteers. Inclusion criteria for controls
356 were informed consent and self-report of not having PD, stroke, ataxia, multiple sclerosis, Alzheimer's
357 disease, dementia, dystonia, autism, bipolar disorder, amyotrophic lateral sclerosis, or epilepsy (hence,
358 neurologically healthy controls, NHC).

359

360 Enrolment occurred in two waves: July 2015-July 2017 (N=316 PwP and 181 NHC) and October 2018-
361 March 2020 (N=665 PwP and 304 NHC). The sample size was determined by the arrival of COVID19
362 pandemic in the Deep South, when we stopped enrolling.

363

364 ***Data collection***

365

366 Subjects were asked to fill out two questionnaires. The Environmental and Family History Questionnaire
367 (EFQ) was first developed 1994 for the NeuroGenetics Research Consortium (NGRC), and has been
368 published for three decades^{12,50}. The Gut Microbiome Questionnaire (GMQ, the source for gastrointestinal
369 data) was developed in 2014 and has also been published⁵⁰. Questionnaires were self-administered with
370 no interference from research staff. In the first enrollment wave, subjects completed the questionnaires in
371 the clinic. In the second wave, subjects took the questionnaires home, and 74% of PwP (N=492) and 77%
372 of NHC (N=234) completed and returned them. The final analytic sample size with completed
373 questionnaires was 808 PwP and 415 NHC (**Table 1**).

374

375 Questions that were asked to collect the data used in this analysis are shown in **Table 2**. Positive family
376 history was defined as having at least one first- or second-degree relative with PD; negative family history
377 was absence of PD in first- and second-degree relatives; and indication of family history without
378 specifying the degree was classified as unknown^{12,50}. MTBI/concussions events (N=19) and exposures to
379 pesticides/herbicides (N=5) which occurred in the same year or after onset of PD were not counted as
380 events. When limiting MTBI/concussions to incidences that occurred more than 10 years prior to PD
381 onset, the incidences within 10 years were treated as no event.

382

383 ***Quality control***

384

385 QC checkpoints were implemented at every step possible during data collection, data entry and data
386 analysis. Briefly, questionnaires were reviewed at arrival for inconsistency across forms, missing data,
387 and potential mix-up at home of case-control couples. Issues were resolved by contacting the subjects or
388 reviewing medical records, otherwise the datapoint was excluded from analysis. All data (100%) were
389 entered in computer twice, once in a Progeny database (**Supplementary Material/Software**), and once in
390 Microsoft Excel, then downloaded and compared, data entry errors identified and corrected. Progeny was
391 used as the main database. When subjects provided a non-standard answer (e.g., “teenager” instead of
392 specifying age), standard rules were created and applied uniformly to all such answers for all subjects (ex.
393 teenager=16). When in doubt, data were recorded as “Unknown” and made distinct from “No”. Statistical
394 analyses were conducted by different analysts independently, one using manual calculation and one using
395 a code and software pipeline, and were cross-checked for accuracy. Here, we present the results obtained
396 via pipeline and the accompanying code.

397

398 *Statistics*

399
400 The number of subjects with data for each item can be extracted from the **Supplementary Material/Data**
401 and are also given for each test in the Tables. For software, the URLs and version used are given in
402 **Supplementary Material/Software**. The code for data analysis is available on Zenodo [link will be
403 provided in peer reviewed paper].

404
405 Differences in frequencies of features (constipation, RBD, weight loss) in PwP vs. NHC were tested using
406 Yates continuity-corrected OR, using the function `'Prop.or'` from the `pairwiseCI` R package specifying
407 `"CImethod='Woolf'`.

408
409 Association between risk factors and disease was tested under the null hypothesis of no difference and the
410 alternative hypothesis that risk factor is associated with increased risk. There was no reason to think or
411 test if they may be associated with reduced risk of PD, because these risk factors were chosen specifically
412 because they have been robustly associated with increased PD risk in other populations. Given that the
413 alternative hypothesis was one-sided (positive direction), tests were conducted one-sided. To test each
414 risk factor individually, OR, lower bound (LB) of 95% confidence intervals (CI), and one-sided P values
415 were calculated with no covariate adjustment from logistic regression models using `'glm'` in R. To assess
416 strength and significance of each risk factor in the presence of, and adjusted for, other risk factors we used
417 conditional analysis, where we included all relevant risk factors in the regression model, as shown in
418 equations (EQ) (1)-(3):

419
420 EQ (1) PD/Control ~ [age, family history, MTBI/concussion, repeated blows to head in sports or
421 military, pesticide/herbicide exposure, Agent Orange/chemical warfare]

422 EQ (2) PD/Control | Male ~ [age, family history, MTBI/concussion, repeated blows to head in sports or
423 military, pesticide/herbicide exposure, Agent Orange/chemical warfare]

424 EQ (3) PD/Control | Female ~ [age, family history, MTBI/concussion, pesticide/herbicide exposure]

425
426 Sex was not included in conditional models due to its collinearity with repeated blows to head, and Agent
427 Orange/chemical warfare. One-sided P value was calculated using `'glm'` specifying `'family="binomial"`
428 in R. Adjusted odds ratio was calculated using equation `'exp(x)'` in R where x is a vector of coefficients
429 extracted from `'glm'` in R. The lower-bound for the one-sided 95% CI were calculated by taking the lower
430 bound of a two-sided 90% Wald CI, using `'exp(confint.default(x, level=0.9))'` in R where x is a `'glm'`

431 object from running the conditional models, and level=0.9 specifies the two-sided 90% interval. Analyses
432 were repeated for males and females separately.

433

434 Population attributable fraction was calculated using Miettinen formula²⁹:

435

436 EQ (4) $PAF = p_c (1 - 1/RR)$

437

438 Where p_c is prevalence of risk factor in cases and RR is relative risk approximated by OR under the rare
439 disease assumption. To adjust PAF for potential confounding, we used adjusted OR values derived from
440 conditional analysis³⁰. For the 95% CI, we used the function `AFglm` from the AF R package on the
441 regression models, specifying `case.control=TRUE`, using a sandwich estimator for the variance of
442 PAF³⁷.

443

444 To assess the fraction of disease attributed to multiple risk factors, we calculated joint PAF (also known
445 as summary PAF)³⁰. Prior to calculating joint PAF, in order to meet the assumption of independence and
446 absence of moderating synergy, we tested for pair-wise interaction between the three risk factors
447 associated with PD in males. We used EQ (2) and added an interaction term, as follows:

448

449 EQ (5) PD/Control | male ~ [age, family history, MTBI/concussion, repeated blows to head in sports or
450 military, pesticide/herbicide exposure, Agent Orange/chemical warfare, Agent Orange/chemical
451 warfare*repeated blows to head in sports or military]

452 EQ (6) PD/Control | male ~ [age, family history, MTBI/concussion, repeated blows to head in sports or
453 military, pesticide/herbicide exposure, Agent Orange/chemical warfare, Agent Orange/chemical
454 warfare*pesticide/herbicide exposure]

455 EQ (7) PD/Control | male ~ [age, family history, MTBI/concussion, repeated blows to head in sports or
456 military, pesticide/herbicide exposure, Agent Orange/chemical warfare, repeated blows to head in
457 sports or military*pesticide/herbicide exposure]

458

459 Having established independence and no interaction, we used EQ (8)³⁰ to calculate joint PAF

460

461 EQ(8) $PAF_{Joint} | Males = 1 - \prod_i (1 - PAF_i)$

462 $PAF_{Joint} | Males = 1 - [(1 - PAF_{Agent\ Orange/chemical\ warfare}) * (1 - PAF_{repeated\ blows\ to\ head}) * (1 - PAF_{pesticide/herbicide})]$

463

464 The upper and lower bounds of 95% CI of joint PAF were calculated using EQ (8) with the set of three
465 upper bounds and three lower bounds of the individual adjusted PAFs.

466

467 **Data availability**

468

469 Data will be publicly available without restriction upon peer reviewed publication. The entire dataset at
470 individual level, de-identified, will be provided in **Supplementary Material/Data**. As per Human
471 Subject Research Privacy considerations, any subject over age 90 years will be denoted as 90 in publicly
472 shared data although true age was used in this analysis. For software used, the URLs and software
473 versions are given in **Supplementary Material/Software**. The code for reproducing the results can be
474 found on Zenodo [link will be provided in peer reviewed paper].

475

476 **Acknowledgements**

477

478 We thank research volunteers for participating in the study. Funding for data collection was provided by a
479 grant from The U.S. Army Medical Research Materiel Command endorsed by the U.S. Army through the
480 Parkinson's Research Program Investigator-Initiated Research Award under Award No.
481 W81XWH1810508 to Haydeh Payami. Time and effort spent on this project were funded by grants from
482 Aligning Science Across Parkinson's [ASAP-020527] through the Michael J. Fox Foundation for
483 Parkinson's Research (MJFF) to Haydeh Payami, NIH/NIEHS 1R01ES032440 to Timothy Sampson, and
484 NIH/NINDS P50 NS108675 to David Standaert. For the purpose of open access, the authors have applied
485 a CC-BY public copyright license to all Author Accepted Manuscripts arising from this submission. The
486 funding agencies had no role in writing of the manuscript or the decision to submit the paper for
487 publication. Opinions, interpretations, conclusions, and recommendations are those of the authors and are
488 not necessarily endorsed by the funding agencies.

489

490 **Competing interests**

491

492 All authors declare no financial or non-financial competing interests.

493

494 **Author contribution**

495

496 HP conceptualized and designed the study, and oversaw data collection, quality control, and statistical
497 analysis. DGS provided clinic access for patient recruitment and took responsibility for clinical aspects of

498 study. HP, GC and ZDW directly accessed and verified the underlying data. HP, GC, CFM, ZDW
499 processed and analyzed the data. TRS provided expertise on toxicants. HP and TRS wrote the first draft
500 of the manuscript, and all authors reviewed, edited, and approved the final version of the manuscript.

501

502 **References**

503

- 504 1 Collaborators, G. B. D. P. s. D. Global, regional, and national burden of Parkinson's disease,
505 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol*
506 **17**, 939-953 (2018). [https://doi.org/10.1016/S1474-4422\(18\)30295-3](https://doi.org/10.1016/S1474-4422(18)30295-3)
- 507 2 Dorsey, E. R., Sherer, T., Okun, M. S. & Bloem, B. R. The Emerging Evidence of the Parkinson
508 Pandemic. *J Parkinsons Dis* **8**, S3-S8 (2018). <https://doi.org/10.3233/JPD-181474>
- 509 3 Yang, W. *et al.* Current and projected future economic burden of Parkinson's disease in the U.S.
510 *NPJ Parkinsons Dis* **6**, 15 (2020). <https://doi.org/10.1038/s41531-020-0117-1>
511 <https://www.who.int/publications/i/item/9789240050983>.
- 512 4 Obeso, J. A. *et al.* Past, present, and future of Parkinson's disease: A special essay on the 200th
513 Anniversary of the Shaking Palsy. *Mov Disord* **32**, 1264-1310 (2017).
514 <https://doi.org/10.1002/mds.27115>
- 515 6 Bloem, B. R., Okun, M. S. & Klein, C. Parkinson's disease. *Lancet* **397**, 2284-2303 (2021).
516 [https://doi.org/10.1016/S0140-6736\(21\)00218-X](https://doi.org/10.1016/S0140-6736(21)00218-X)
- 517 7 Hardy, J. Genetic analysis of pathways to Parkinson disease. *Neuron* **68**, 201-206 (2010).
518 <https://doi.org/10.1016/j.neuron.2010.10.014>
- 519 8 Cannon, J. R. & Greenamyre, J. T. Gene-environment interactions in Parkinson's disease: specific
520 evidence in humans and mammalian models. *Neurobiol Dis* **57**, 38-46 (2013).
521 <https://doi.org/10.1016/j.nbd.2012.06.025>
- 522 9 De Miranda, B. R., Goldman, S. M., Miller, G. W., Greenamyre, J. T. & Dorsey, E. R. Preventing
523 Parkinson's Disease: An Environmental Agenda. *J Parkinsons Dis* **12**, 45-68 (2022).
524 <https://doi.org/10.3233/JPD-212922>
- 525 10 Willis, A. W. *et al.* Incidence of Parkinson disease in North America. *NPJ Parkinsons Dis* **8**, 170
526 (2022). <https://doi.org/10.1038/s41531-022-00410-y>
- 527 11 Hirsch, L., Jette, N., Frolkis, A., Steeves, T. & Pringsheim, T. The Incidence of Parkinson's
528 Disease: A Systematic Review and Meta-Analysis. *Neuroepidemiology* **46**, 292-300 (2016).
529 <https://doi.org/10.1159/000445751>

- 530 12 Payami, H., Larsen, K., Bernard, S. & Nutt, J. Increased risk of Parkinson's disease in parents and
531 siblings of patients. *Ann Neurol* **36**, 659-661 (1994).
- 532 13 Thacker, E. L. & Ascherio, A. Familial aggregation of Parkinson's disease: A meta-analysis. *Mov*
533 *Disord* **23**, 1174-1183 (2008).
- 534 14 Klein, C. & Westenberger, A. Genetics of Parkinson's disease. *Cold Spring Harb Perspect Med* **2**,
535 a008888 (2012). [https://doi.org:10.1101/cshperspect.a008888](https://doi.org/10.1101/cshperspect.a008888)
- 536 15 Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for
537 Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* **18**, 1091-
538 1102 (2019). [https://doi.org:10.1016/S1474-4422\(19\)30320-5](https://doi.org/10.1016/S1474-4422(19)30320-5)
- 539 16 Wang, A. *et al.* Parkinson's disease risk from ambient exposure to pesticides. *Eur J Epidemiol* **26**,
540 547-555 (2011).
- 541 17 Ascherio, A. *et al.* Pesticide exposure and risk for Parkinson's disease. *Ann Neurol* **60**, 197-203
542 (2006).
- 543 18 Perrin, L. *et al.* Pesticides expenditures by farming type and incidence of Parkinson disease in
544 farmers: A French nationwide study. *Environ Res* **197**, 111161 (2021).
545 [https://doi.org:10.1016/j.envres.2021.111161](https://doi.org/10.1016/j.envres.2021.111161)
- 546 19 Marras, C. *et al.* Systematic review of the risk of Parkinson's disease after mild traumatic brain
547 injury: results of the International Collaboration on Mild Traumatic Brain Injury Prognosis. *Arch*
548 *Phys Med Rehabil* **95**, S238-244 (2014). [https://doi.org:10.1016/j.apmr.2013.08.298](https://doi.org/10.1016/j.apmr.2013.08.298)
- 549 20 Rugbjerg, K., Ritz, B., Korbo, L., Martinussen, N. & Olsen, J. H. Risk of Parkinson's disease
550 after hospital contact for head injury: population based case-control study. *BMJ* **337**, a2494
551 (2008). [https://doi.org:10.1136/bmj.a2494](https://doi.org/10.1136/bmj.a2494)
- 552 21 Camacho-Soto, A. *et al.* Traumatic brain injury in the prodromal period of Parkinson's disease: A
553 large epidemiological study using medicare data. *Ann Neurol* **82**, 744-754 (2017).
554 [https://doi.org:10.1002/ana.25074](https://doi.org/10.1002/ana.25074)
- 555 22 Mackay, D. F. *et al.* Neurodegenerative Disease Mortality among Former Professional Soccer
556 Players. *N Engl J Med* **381**, 1801-1808 (2019). [https://doi.org:10.1056/NEJMoa1908483](https://doi.org/10.1056/NEJMoa1908483)
- 557 23 Bellou, V., Belbasis, L., Tzoulaki, I., Evangelou, E. & Ioannidis, J. P. Environmental risk factors
558 and Parkinson's disease: An umbrella review of meta-analyses. *Parkinsonism Relat Disord* **23**, 1-
559 9 (2016). [https://doi.org:10.1016/j.parkreldis.2015.12.008](https://doi.org/10.1016/j.parkreldis.2015.12.008)
- 560 24 Hernan, M. A., Takkouche, B., Caamano-Isorna, F. & Gestal-Otero, J. J. A meta-analysis of
561 coffee drinking, cigarette smoking, and the risk of Parkinson's disease. *Ann Neurol* **52**, 276-284
562 (2002).

- 563 25 Powers, K. *et al.* Combined effects of smoking, coffee and NSAIDs on Parkinson's disease risk.
564 *Mov Disord* **23**, 88-95 (2008).
- 565 26 Kim, I. Y. *et al.* Integration of risk factors for Parkinson disease in 2 large longitudinal cohorts.
566 *Neurology* **90**, e1646-e1653 (2018). <https://doi.org/10.1212/WNL.0000000000005473>
- 567 27 Canseco, J. A. *et al.* Overview of Traumatic Brain Injury in American Football Athletes. *Clin J*
568 *Sport Med* **32**, 236-247 (2022). <https://doi.org/10.1097/JSM.0000000000000918>
- 569 28 Belvisi, D. *et al.* Risk factors of Parkinson disease: Simultaneous assessment, interactions, and
570 etiologic subtypes. *Neurology* **95**, e2500-e2508 (2020).
571 <https://doi.org/10.1212/WNL.0000000000010813>
- 572 29 Miettinen, O. S. Proportion of the disease caused or prevented by a given exposure, trait or
573 prevention. *Am J Epidemiol* **99**, 325-332 (1974).
- 574 30 Khosravi, A. & Mansournia, M. A. Methodological issues estimating population attributable
575 fraction related to 'Health care cost of overweight-related diseases in Bangladesh'. *Public Health*
576 *Nutr* **24**, 4195 (2021). <https://doi.org/10.1017/S1368980020003225>
- 577 31 Mukadam, N., Sommerlad, A., Huntley, J. & Livingston, G. Population attributable fractions for
578 risk factors for dementia in low-income and middle-income countries: an analysis using cross-
579 sectional survey data. *Lancet Glob Health* **7**, e596-e603 (2019). [https://doi.org/10.1016/S2214-109X\(19\)30074-9](https://doi.org/10.1016/S2214-109X(19)30074-9)
- 580
- 581 32 Gorell, J. M., Peterson, E. L., Rybicki, B. A. & Johnson, C. C. Multiple risk factors for
582 Parkinson's disease. *J Neurol Sci* **217**, 169-174 (2004). <https://doi.org/10.1016/j.jns.2003.09.014>
583 <https://www.cdc.gov/nchs/>.
- 584 34 Delic, V., Beck, K. D., Pang, K. C. H. & Citron, B. A. Biological links between traumatic brain
585 injury and Parkinson's disease. *Acta Neuropathol Commun* **8**, 45 (2020).
586 <https://doi.org/10.1186/s40478-020-00924-7>
- 587 35 McKnight, S. & Hack, N. Toxin-Induced Parkinsonism. *Neurol Clin* **38**, 853-865 (2020).
588 <https://doi.org/10.1016/j.ncl.2020.08.003>
- 589 36 Gardner, R. C. *et al.* Mild TBI and risk of Parkinson disease: A Chronic Effects of Neurotrauma
590 Consortium Study. *Neurology* **90**, e1771-e1779 (2018).
591 <https://doi.org/10.1212/WNL.0000000000005522>
- 592 37 Dahlgvist, E., Zetterqvist, J., Pawitan, Y. & Sjolander, A. Model-based estimation of the
593 attributable fraction for cross-sectional, case-control and cohort studies using the R package AF.
594 *Eur J Epidemiol* **31**, 575-582 (2016). <https://doi.org/10.1007/s10654-016-0137-7>
- 595 38 Richardson, J. R., Fitsanakis, V., Westerink, R. H. S. & Kanthasamy, A. G. Neurotoxicity of
596 pesticides. *Acta Neuropathol* **138**, 343-362 (2019). <https://doi.org/10.1007/s00401-019-02033-9>

- 597 39 James, K. A. & Hall, D. A. Groundwater pesticide levels and the association with Parkinson
598 disease. *Int J Toxicol* **34**, 266-273 (2015). [https://doi.org:10.1177/1091581815583561](https://doi.org/10.1177/1091581815583561)
- 599 40 El-Nahhal, I. & El-Nahhal, Y. Pesticide residues in drinking water, their potential risk to human
600 health and removal options. *J Environ Manage* **299**, 113611 (2021).
601 [https://doi.org:10.1016/j.jenvman.2021.113611](https://doi.org/10.1016/j.jenvman.2021.113611)
- 602 41 Craddock, H. A., Huang, D., Turner, P. C., Quiros-Alcala, L. & Payne-Sturges, D. C. Trends in
603 neonicotinoid pesticide residues in food and water in the United States, 1999-2015. *Environ*
604 *Health* **18**, 7 (2019). [https://doi.org:10.1186/s12940-018-0441-7](https://doi.org/10.1186/s12940-018-0441-7)
- 605 42 Donley, N. The USA lags behind other agricultural nations in banning harmful pesticides.
606 *Environ Health* **18**, 44 (2019). [https://doi.org:10.1186/s12940-019-0488-0](https://doi.org/10.1186/s12940-019-0488-0)
- 607 43 Fang, W., Peng, Y., Muir, D., Lin, J. & Zhang, X. A critical review of synthetic chemicals in
608 surface waters of the US, the EU and China. *Environ Int* **131**, 104994 (2019).
609 [https://doi.org:10.1016/j.envint.2019.104994](https://doi.org/10.1016/j.envint.2019.104994)
- 610 44 Ritz, B. R., Paul, K. C. & Bronstein, J. M. Of Pesticides and Men: a California Story of Genes
611 and Environment in Parkinson's Disease. *Curr Environ Health Rep* **3**, 40-52 (2016).
612 [https://doi.org:10.1007/s40572-016-0083-2](https://doi.org/10.1007/s40572-016-0083-2)
- 613 45 Mez, J. *et al.* Clinicopathological Evaluation of Chronic Traumatic Encephalopathy in Players of
614 American Football. *JAMA* **318**, 360-370 (2017). [https://doi.org:10.1001/jama.2017.8334](https://doi.org/10.1001/jama.2017.8334)
- 615 46 Paulson, G. W. & Dadmehr, N. Is there a premorbid personality typical for Parkinson's disease?
616 *Neurology* **42**, 77-80 (1991).
- 617 47 Siderowf, A. *et al.* Assessment of heterogeneity among participants in the Parkinson's
618 Progression Markers Initiative cohort using alpha-synuclein seed amplification: a cross-sectional
619 study. *Lancet Neurol* **22**, 407-417 (2023). [https://doi.org:10.1016/S1474-4422\(23\)00109-6](https://doi.org/10.1016/S1474-4422(23)00109-6)
- 620 48 Fereshtehnejad, S. M., Zeighami, Y., Dagher, A. & Postuma, R. B. Clinical criteria for subtyping
621 Parkinson's disease: biomarkers and longitudinal progression. *Brain* **140**, 1959-1976 (2017).
622 [https://doi.org:10.1093/brain/awx118](https://doi.org/10.1093/brain/awx118)
- 623 49 Gibb, W. R. G. & Lee, A. J. A comparison of clinical and pathological features of young- and
624 old-onset Parkinson disease. *Neurology* **38** (1988).
- 625 50 Wallen, Z. D. *et al.* Metagenomics of Parkinson's disease implicates the gut microbiome in
626 multiple disease mechanisms. *Nat Commun* **13**, 6958 (2022). [https://doi.org:10.1038/s41467-022-](https://doi.org/10.1038/s41467-022-34667-x)
627 [34667-x](https://doi.org/10.1038/s41467-022-34667-x)
- 628
- 629

630 **Table 1. Enrollment, subject characteristics, and features of PD.**

631

Subject Characteristics	PD		NHC		PD features	
	N with data	Summary statistics	N with data	Summary statistics	OR[95%CI]	P
Sample size	808	-	415	-	-	-
Age	808	68.3±8.7	415	66.1±8.6	-	-
PD age at onset	808	59.3±11.0	-	-	-	-
Jewish ancestry	597	16 (3%)	313	9 (3%)	-	-
Black or African American	804	20 (2%)	410	6 (1%)	-	-
Hispanic or Latino	787	13 (2%)	403	3 (0.7%)	-	-
White	804	772 (96%)	410	398 (97%)	-	-
Residents of Deep South	799	764 (96%)	403	376 (93%)	-	-
Residents of DS Periphery	799	32 (4%)	403	26 (6%)	-	-
PD Features						
Sex (N & % male)	808	512 (63%)	415	125 (30%)	-	<1E-5
Constipation	775	345 (45%)	402	57 (14%)	4.8 [3.5-6.7]	4E-25
RBD	479	47 (10%)	229	0 (0%)	50.4 [3.1-821.5]	2E-6
Weight loss	790	212 (27%)	406	56 (14%)	2.3 [1.7-3.1]	4E-7

632

633 A total of 1466 individuals were enrolled, of whom 1223 completed the questionnaires and were used in the present
634 study, including 808 persons with PD and 415 neurologically healthy controls (NHC). Characteristics are shown
635 with the number of individuals for whom specified data were available and analyzed (N with data). Summary
636 statistics is the mean and standard deviation for quantitative traits, and the number of individuals and % of total who
637 are positive for the characteristic for dichotomous traits. Deep South (DS) includes the states of Alabama,
638 Mississippi, Georgia, South Carolina, and Louisiana. DS periphery states are Texas, Tennessee, North Carolina, and
639 Florida. Question on REM sleep behavior disorder (RBD) was added at the start of the second wave of data
640 collection and was not asked of 316 PD and 181 NHC of the first wave. To assess robustness of data, features that
641 are known to be more prevalent in PD were tested. Preponderance of male sex among PD was tested as proportion
642 of persons with PD who were male (63%) vs. expected for null hypothesis of no sex difference (50%), using Z test
643 (one proportion test) and showing the significance (P). Constipation, RBD and wight loss were tested comparing PD
644 to NHC, using odds ratio (OR), 95% confidence interval (CI) of OR, and significance (P) denoting the difference
645 between PD and NHC.

646 **Table 2. Questions used to collect the data**

Question	Response options
Sex	<input type="checkbox"/> M <input type="checkbox"/> F
Today's date (<i>as subject is completing form</i>)	_____
Date of birth	_____
Do you have Parkinson's disease?	<input type="checkbox"/> No <input type="checkbox"/> Yes
How old were you when you first noticed a sign of Parkinson's disease (age at onset)?	_____ years old
Do you have rapid eye movement sleep behavior disorder (RBD)?	<input type="checkbox"/> No <input type="checkbox"/> Yes
Have you had a stroke, ataxia, multiple sclerosis, Alzheimer's disease, dementia, dystonia, autism, bipolar disorder, amyotrophic lateral sclerosis (ALS), or epilepsy?	<input type="checkbox"/> No <input type="checkbox"/> Yes
Are you often constipated (fewer than 3 bowel movements per week occurring frequently)?	<input type="checkbox"/> No <input type="checkbox"/> Yes
Have you lost more than 10 pounds in the last year?	<input type="checkbox"/> No <input type="checkbox"/> Yes
Have you had a head injury that caused loss of consciousness or required medical care?	<input type="checkbox"/> No <input type="checkbox"/> Yes
If yes, how old were you when it first happened	_____ years old
Did you have repeated blows to the head such as in sports or military?	<input type="checkbox"/> No <input type="checkbox"/> Yes
Were you ever exposed to Agent Orange or other chemical warfare?	<input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Don't know
Were you ever exposed to heavy uses of pesticides or herbicides, for example, did you live on or near farms that did crop dusting?	<input type="checkbox"/> No <input type="checkbox"/> Yes
If yes, how old were you (give the range, for example, from birth until age 16)	_____
Are you Hispanic or Latino?	<input type="checkbox"/> No <input type="checkbox"/> Yes
What race do you most identify yourself with?	<input type="checkbox"/> White <input type="checkbox"/> Black or African American <input type="checkbox"/> American Indian/Alaskan Native <input type="checkbox"/> Asian <input type="checkbox"/> Native Hawaiian or other Pacific Islander <input type="checkbox"/> More than one race
Are you of Jewish ancestry?	<input type="checkbox"/> No <input type="checkbox"/> Yes
Do you have any blood relatives who have Parkinson's disease (including your parents, grand-parents, siblings, aunts, and uncles)?	<input type="checkbox"/> No <input type="checkbox"/> Yes
If yes, for each relative who has or had PD, list their relationship to you	_____

647

648 Age was calculated by subtracting date of birth from today's date. Positive family history was defined as having at
649 least one first or second degree relative with PD. Question on having PD was mainly for QC purposes; PD diagnosis
650 was made by one of our movement disorder specialist neurologists at University of Alabama at Birmingham.
651 Question on neurological diseases (stroke, ataxia, etc.) was the exclusion criteria for controls. The questions on
652 repeated blows to the head, exposure to heavy uses of pesticide/herbicide, and Agent Orange/chemical warfare were
653 added midway in the first wave of data collection. The question on RBD, and duration of exposure to
654 pesticides/herbicides were added at the start of second wave. The numbers of subjects for each question or
655 combinations can be extracted from the dataset which is provided in supplement.

656 **Table 3. Assessing association of risk factors with PD in a single population**

657

	N with data		Summary statistic		PD vs. NHC	
	PD	NHC	PD	NHC	OR [95%CI LB]	P
ALL SUBJECTS						
Age	808	415	68.3±8.7	66.1±8.6	1.03 [1.02]	2E-05
Family history of PD	751	403	245 (33%)	59 (15%)	2.82 [2.17]	6E-11
MTBI/concussion	777	410	157 (20%)	45 (11%)	2.05 [1.52]	4E-05
MTBI/concussion >10 years prior to PD onset	777	410	134 (17%)	45 (11%)	1.69 [1.25]	2E-03
Repeated blows to head in sports or military	571	290	76 (13%)	10 (3%)	4.3 [2.44]	1E-05
Pesticide/herbicide exposure	594	314	188 (32%)	45 (14%)	2.77 [2.05]	1E-08
Pesticide/herbicide exposure duration (years)	436	224	5.5±13.6	2.9±10.6	1.02 [1.01]	9E-03
Agent Orange/chemical warfare exposure	597	314	28 (5%)	2 (1%)	7.68 [2.29]	3E-03
Males						
Age	512	125	68.8±8.6	68.2±9.2	1.01 [0.99]	0.22
Family history of PD	480	121	145 (30%)	11 (9%)	4.33 [2.51]	5E-06
MTBI/concussion	493	124	112 (23%)	20 (16%)	1.53 [0.99]	0.06
MTBI/concussion >10 years prior to PD onset	493	124	97 (20%)	20 (16%)	1.27 [0.82]	0.18
Repeated blows to head in sports or military	356	95	76 (21%)	10 (11%)	2.31 [1.28]	0.01
Pesticide/herbicide exposure	375	100	112 (30%)	15 (15%)	2.41 [1.47]	2E-03
Pesticide/herbicide exposure duration (years)	275	65	5.2±13.7	3.1±10.9	1.01 [0.99]	0.13
Agent Orange/chemical warfare exposure	373	99	27 (7%)	2 (2%)	3.78 [1.12]	0.04
Females						
Age	296	290	67.5±8.9	65.3±8.3	1.03 [1.01]	1E-03
Family history of PD	271	282	100 (37%)	48 (17%)	2.85 [2.04]	1E-07
MTBI/concussion	284	286	45 (16%)	25 (9%)	1.97 [1.27]	5E-03
MTBI/concussion >10 years prior to PD onset	284	286	37 (13%)	25 (9%)	1.56 [1.00]	0.05
Repeated blows to head in sports or military	215	195	0 (0%)	0 (0%)	-	-
Pesticide/herbicide exposure	219	214	76 (35%)	30 (14%)	3.26 [2.19]	6E-07
Pesticide/herbicide exposure duration (years)	161	159	5.9±13.3	2.8±10.5	1.02 [1.01]	0.02
Agent Orange/chemical warfare exposure	224	215	1 (0.4%)	0 (0%)	-	-

658

659 Each risk factor was tested individually as if they were the only factor under investigation (OR are not adjusted for
660 potential effect of other risk factors. See **Table 4** for adjusted OR). Total number of subjects (N with data) and the
661 number and percentage of those who were positive for risk factor (summary statistics) are shown for PD and
662 neurologically healthy control (NHC). PD vs. NHC were one-sided tests with null hypothesis of no difference and
663 alternative hypothesis that risk factor is positively associated (increased) in PD. Odds ratio (OR) and the lower

664 bound (LB) of 95% confidence interval (CI) are shown. Upper bound of CI is infinity. OR is indicative of strength
665 of association approximating fold-difference in prevalence of the risk factors in PD vs. NHC, and one-sided P is the
666 statistical significance of association. Family history of PD: 1st or 2nd degree relative with PD. MTBI/concussion:
667 mild traumatic brain injury that caused loss of consciousness or required medical care, analyzed first excluding all
668 incidences occurring around or after the time of onset, and once excluding incidences that occurred within 10 years
669 prior to PD onset. Agent orange: herbicide used in the Vietnam war.

670 **Table 4. PD Risk factors are independent of each other.**

671

Conditional analysis model	N PD	N NHC	Risk factors in model	Adjusted OR [95%CI LB]	P
All subjects					
PD/Control ~ [age, family history, MTBI/concussion, repeated blows to head in sports or military, pesticide/herbicide exposure, Agent Orange/chemical warfare]	471	272	age	1.03 [1.02]	3E-04
			family history	2.87 [2.05]	1E-07
			MTBI/concussion	1.30 [0.89]	0.13
			repeated blows to head in sports or military	3.68 [2.00]	2E-04
			pesticide/herbicide exposure	2.20 [1.56]	8E-05
			Agent Orange/chemical warfare	10.81 [1.97]	0.01
Males					
PD/Control ~ [age, family history, MTBI/concussion, repeated blows to head in sports or military, pesticide/herbicide exposure, Agent Orange/chemical warfare]	297	88	age	1.02 [0.99]	0.10
			family history	4.84 [2.51]	4E-05
			MTBI/concussion	1.20 [0.68]	0.30
			repeated blows to head in sports or military	2.00 [1.04]	0.04
			pesticide/herbicide exposure	2.52 [1.37]	6E-03
			Agent Orange/chemical warfare	6.03 [1.08]	0.04
Females					
PD/Control ~ [age, family history, MTBI/concussion, pesticide/herbicide exposure]	193	206	age	1.03 [1.01]	8E-03
			family history	2.62 [1.75]	4E-05
			MTBI/concussion	1.35 [0.79]	0.18
			pesticide/herbicide exposure	2.85 [1.87]	2E-05

672

673 Five risk factors (from **Table 3**) and age were included in one conditional analysis model and tested simultaneously
674 for association with PD. N PD and N NHC are the numbers of persons with PD and neurologically healthy controls
675 for whom we had data on all risk factors in the specified model. Sex was not included in the model because of
676 collinearity (all repeated blows to head, and all except one exposure to Agent Orange/chemical warfare were males),
677 instead, analyses were repeated for males and females separately. The tests were one-sided with null hypothesis of
678 no difference and alternative hypothesis of risk factor is increased in PD. Results are shown for each risk factor as
679 adjusted odds ratio (OR), lower bound (LB) of 95% confidence interval (CI), upper bound is infinity, and
680 significance (one-sided P) of association with PD conditioned on the other risk factors in the model. Age, family
681 history (1st or 2nd degree relative with PD), repeated blows to the head, pesticides/herbicide exposure, and Agent

682 Orange/chemical warfare retained their effect sizes and significance; they have distinct, independent signals. Mild
683 traumatic brain injury (MTBI)/concussion lost significance.

684 **Table 5. Population attributable fraction (PAF) of preventable risk factors**

685

Risk factor	Group	PAF [95%CI]	<i>Determinants of PAF</i>	
			<i>adjusted OR</i>	<i>Prevalence in PwP</i>
PAF for Individual risk factors				
Pesticide/herbicide exposure	All	16% [9%-24%]	2.20	30%
Pesticide/herbicide exposure	Females	23% [12%-34%]	2.85	35%
Pesticide/herbicide exposure	Males	17% [7%-27%]	2.52	28%
Agent Orange/chemical warfare	Males	6% [0.03%-12%]	6.03	7%
Repeated blows to head in sports or military	Males	10% [0.3%-20%]	2.00	20%
Joint PAF for combination of risk factors				
Repeated blows to head in sports or military, pesticide/herbicide & Agent Orange/chemical warfare	Males	30% [7%-49%]	-	43%

686

687 PAF is the fraction of disease that is attributed to one or combination of risk factors and is interpreted as the fraction
 688 of disease that would be reduced if risk factor was eliminated. It is calculated using prevalence of risk factor in
 689 patient population and relative risk (or odds ratio (OR) if disease is rare as is PD). PAF estimate for each risk factor
 690 was adjusted for other risk factors, by using adjusted OR from **Table 4**.