

1 **Disease progression strikingly differs in research and real-world Parkinson’s populations**

2
3 Brett K. Beaulieu-Jones^{1,2,3,4,5}, Francesca Frau⁶, Sylvie Bozzi⁷, Karen J. Chandross⁸, M Judith
4 Peterschmitt⁹, Caroline Cohen¹⁰, Catherine Coulovrat¹¹, Dinesh Kumar¹¹, Mark J. Kruger⁹, Scott
5 L. Lipnick¹, Lane Fitzsimmons¹, Isaac S. Kohane¹, Clemens R. Scherzer^{2,3,5}

6
7 1 Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA.

8 2 APDA Center for Advanced Parkinson Research of Harvard Medical School and Brigham and
9 Women’s Hospital, Boston, MA 02115, USA.

10 3 Precision Neurology Program of Brigham & Women's Hospital, Harvard Medical School,
11 Boston, MA 02115, USA.

12 4 Department of Medicine, University of Chicago, Chicago, IL 60615 USA.

13 5 Aligning Science Across Parkinson’s (ASAP) Collaborative Research Network, Chevy Chase
14 MD, 20815

15 6 Sanofi R&D, Data and Data Science, Frankfurt, Germany.

16 7 Sanofi Health Economics and Value Assessment, Sanofi, Paris, France.

17 8 Sanofi R&D, Bridgewater, NJ, United States.

18 9 Sanofi Genzyme, Clinical Development Neurology, Cambridge, MA, United States.

19 10 Sanofi R&D, Paris, France

20 11 Sanofi Translational Sciences, Framingham, 01701 USA

21

22 **Abstract**

23

24 Characterization of Parkinson's disease (PD) progression using real-world evidence could guide
25 clinical trial design and identify subpopulations. Efforts to curate research populations, the
26 increasing availability of real-world data and recent advances in natural language processing,
27 particularly large language models, allow for a more granular comparison of populations and the
28 methods of data collection describing these populations than previously possible.

29 This study includes two research populations and two real-world data derived (RWD)
30 populations. The research populations are the Harvard Biomarkers Study (HBS, N = 935), a
31 longitudinal biomarkers cohort study with in-person structured study visits; and Fox Insights (N
32 = 36,660), an online self-survey-based research study of the Michael J. Fox Foundation. Real-
33 world cohorts are the Optum Integrated Claims-electronic health records (N = 157,475),
34 representing wide-scale linked medical and claims data and de-identified data from Mass
35 General Brigham (MGB, N = 22,949), an academic hospital system. Structured, de-identified
36 electronic health records data at MGB are supplemented using natural language processing with
37 a large language model to extract measurements of PD progression. This extraction process is
38 manually validated for accuracy.

39 Motor and cognitive progression scores change more rapidly in MGB than HBS (median
40 survival until H&Y 3: 5.6 years vs. >10, $p < 0.001$; mini-mental state exam median decline 0.28
41 vs. 0.11, $p < 0.001$; and clinically recognized cognitive decline, $p = 0.001$). In the real-world
42 populations, patients are diagnosed more than eleven years later (RWD mean of 72.2 vs. research
43 mean of 60.4, $p < 0.001$). After diagnosis, in real-world cohorts, treatment with PD medications is
44 initiated 2.3 years later on average (95% CI: [2.1-2.4]; $p < 0.001$).

45 This study provides a detailed characterization of Parkinson's progression in diverse
46 populations. It delineates systemic divergences in the patient populations enrolled in research
47 settings vs. patients in the real world. These divergences are likely due to a combination of
48 selection bias and real population differences, but exact attribution of the causes is challenging
49 using existing data. This study emphasizes a need to utilize multiple data sources and to
50 diligently consider potential biases when planning, choosing data sources, and performing
51 downstream tasks and analyses.

52 Introduction

53
54 Despite substantial investment and increased biological understanding, disease-modifying
55 therapies for Parkinson’s disease (PD) have not yet proven successful.¹ Therapies recently
56 considered to be promising in preclinical and early clinical trials have failed at later stages.²
57 There are numerous challenges to developing disease-modifying therapies for PD. Perhaps most
58 importantly, we lack a complete understanding of the molecular processes driving disease
59 pathogenesis.^{3–5} A wide range of PD onset ages and differing patterns of disease progression
60 suggest heterogeneity.^{6–8} Recently, genome-wide survival studies have identified six loci ---
61 *GBA*, *APOE*, *RIMS2*, *TMEM108*, *WVOX*, and the mitochondrial super macro haplogroup J,T,U -
62 -- associated with genetic heterogeneity in the progression from PD to Lewy body dementia.^{9–13}
63 This is important because homogeneous patient populations and an accurate measurement of
64 disease progression would greatly facilitate clinical trial design. More homogenous
65 subpopulations may make biomarker associations more powerful, clarify the genetic basis for
66 disease, and lead to more successful therapeutic development.^{12,14}

67 Characterization of Parkinson’s disease (PD) progression using real-world evidence
68 could guide clinical trial design and identify subpopulations. Most prior work examining PD
69 progression focused on research data from major academic centers,¹⁵ but this may not accurately
70 reflect the population at large.^{8,16} It is important to determine whether analyses performed in one
71 particular data modality will lead to generalizable results. Large scale adoption of electronic
72 health records and recent advances in natural language processing (NLP) could be used to
73 address this knowledge gap. Individual health systems (e.g., Mass General Brigham, MGB)¹⁷
74 and large national managed care providers (e.g., UnitedHealth Group, Optum)¹⁸ have developed
75 robust databases in the form of electronic health records (EHRs) through routine care. Even after
76 filtering to enrich for newly diagnosed PD patients—those who had at least two years of data
77 prior to diagnosis—Optum includes 157,475 patients with at least two PD diagnoses and MGB
78 has a further 13,257 patients. It is important to note that while large, inclusion in these datasets is
79 driven by access to care (e.g., insurance, socioeconomic factors) at MGB or within UnitedHealth
80 Group. In addition to these new sources of real-world data (RWD), natural language processing
81 and especially large language models have recently achieved improved performance on
82 benchmark evaluations by substantial margins. In this work, we leverage these improved models
83 to supplement structured EHR data at MGB by extracting phenotypic information, particularly
84 clinical scores from clinical notes.^{19–23} To ensure the model’s extraction process is accurate, it is
85 straightforward to perform manual chart review on a sample set of notes.

86 In this study we provide a comparison between two research studies and two RWD
87 sources. While the research cohorts were actively recruited, the RWD are “passively” collected
88 because they are derived from typical clinical care and not explicitly for research purposes. We
89 characterized PD progression to better understand “typical” progression across the four research
90 and real-world data sources. We evaluated the strengths and weaknesses of these data sources to
91 characterize PD progression and their potential to enable more precise endophenotyping.

92 In performing this evaluation, we acknowledge and emphasize substantial potential
93 sources of bias related to the way the different datasets were collected. For example, in the Fox
94 Insight study, patients are recruited through online advertising, and they provide their data
95 through an online portal. Within HBS, patients must have a Mini-Mental State Examination
96 (MMSE) score consistent with the ability to provide informed consent or their next of kin must
97 be present to provide consent. While ethically critical, criteria like this may systematically

98 exclude certain participants, those that are older may be less likely to have and use the internet
 99 on their own, may have lower MMSE scores and be less likely to be in close contact or
 100 proximity with their next of kin. On the other hand, RWD are available only in the context of an
 101 individual’s interactions with healthcare and can be biased by access to healthcare and other
 102 socioeconomic factors. Additionally, analyzing RWD requires specifying rules for inclusion,
 103 clinical event, and outcome definitions. Reasonable changes to these rules have the potential to
 104 significantly change results.²⁴

105 This study examines the differences of PD progression between these data modalities
 106 with the intent of providing guidance for future studies. By better understanding potential biases
 107 in data, future study of potential disease-modifying therapies can be designed to minimize these
 108 biases. This can be done by choosing the appropriate data for the question at hand and including
 109 multiple data modalities to help determine whether effects are artifacts of the data source or
 110 accurate measures of progression. This has the potential to aid future studies by providing
 111 guidance of when changes in progression are significant, the determination of appropriate
 112 clinical endpoints, and baselines for the identification of patient subpopulations of interest.
 113

114 **Results**

115 The study took place across four different datasets, two research populations: the 1)
 116 Harvard Biomarkers Study (HBS), and the 2) Fox Insight longitudinal health study (Fox Insight);
 117 and two real-world populations: 3) Mass General Brigham electronic medical records (MGB),
 118 and 4) Optum Claims-EHR dataset (Optum). This study was approved by the Mass General
 119 Brigham IRB. HBS, Fox Insight and Optum are de-identified and MGB electronic medical
 120 records were accessed through the Research Patient Data Registry. All work was retrospective
 121 review without additional contact or intervention in accordance with the IRB. No linking
 122 between datasets was performed.
 123

124 **Table 1.** Cohort Characteristics for HBS, MGB, Optum, and Fox Insights cohorts

	HBS	Fox Insight	MGB	Optum Claims-EHR
Total Population with PD	935	36,660	22,949	157,475#
Population with data since initial PD diagnosis	935	36,660	13,257	157,475#
Percentage of population with data from initial PD diagnosis who are Male	67.0%	55.8%	57.8%	56.7%
Mean Age at diagnosis (years) [25% and 75%]	62.5 [56.0 – 70.0]	60.3 [53.6-61.3]	72.9 [66.5 – 81.2]	72.1 [66.0-80.0]
Mean Years from Diagnosis to Study Enrollment [25%, 75%]	3.7 [0.0 – 5.0]	3.5 [1.2-7.5]	NA	NA
Mean Initial H&Y Stage if recorded within 1 year of diagnosis (MGB) or	2.1 [2, 2, 2.5] n=853	NA	2.3 [2, 2, 3] n=1,071	NA

at enrollment (HBS) [25%, 50%, 75%]				
Mean Length of Follow-up from disease diagnosis (years) [25%, 75% percentile]	6.2 [2.0-9.0]	1 [0-1.0]*	3.8 [0.7 – 5.6]	3.1 [0.6-4.9] ^

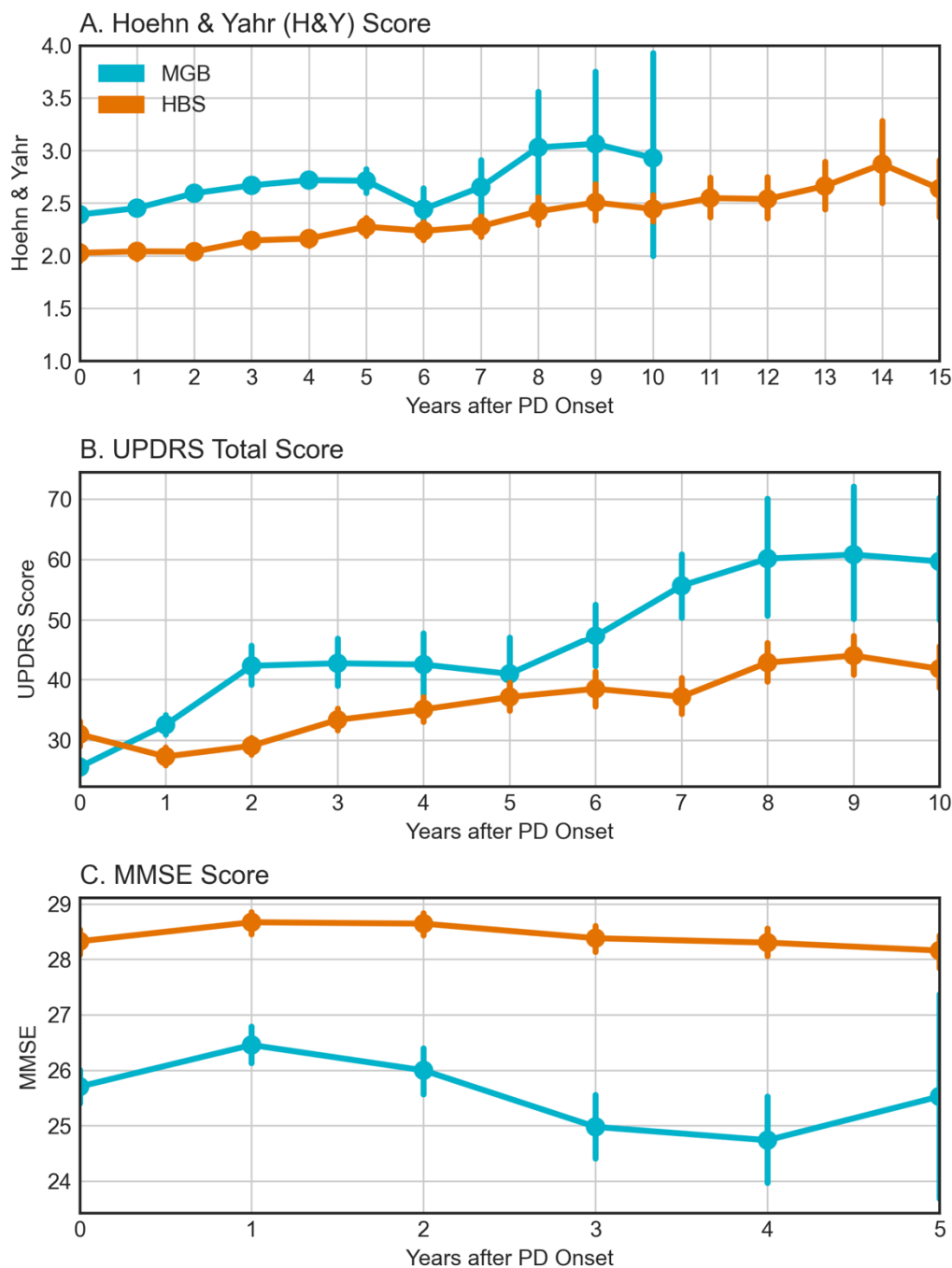
125 * Years enrolled within Fox Insight. Due to survey nature of data source, historical data are available for certain
126 information.

127 # Subjects at MGB and Optum were only included if they had at least two years of data coverage prior to initial PD
128 diagnosis. This quiescence period was designed to filter for de novo diagnoses.

129 The research populations (HBS and Fox Insight) had lower age of PD diagnosis than the
130 real-world populations (mean 60.4 vs. 72.2, $p < 0.001$, Table 1). Fox Insight had a shorter length
131 of follow-up after enrollment, limited by the fact that it is a new resource. HBS demonstrated the
132 potential for research data sources to offer significant longitudinal data that is thus far difficult to
133 find in RWD (mean follow-up time of 6.1 years vs. 3.2 years, $p < 0.001$). Finally, because the
134 research cohorts were actively enrolled, there is the potential to collect additional history, for
135 example the timing of an initial PD diagnosis is available for all participants of HBS and Fox
136 Insight, but it must be extracted with rules from the RWD sources.

137 Within HBS, clinical rating scales are actively collected, recorded as structured data and
138 are available in most patients (Table 1, 853/935: 91.2%). In MGB they are passively collected
139 and were extracted from the clinical notes for a small proportion of patients (1,071/13,257: 8.1%,
140 extraction process described in methods). Clinical scores are not available in Optum. To evaluate
141 the effects of potential interventions, it is important to understand the typical progression of PD.
142 Clinical rating scales, including H&Y, MMSE, UPDRS and MoCA represent common clinical
143 study endpoints for Parkinson's progression. We compared progression of these measures
144 between HBS, an actively assessed research cohort and MGB, a real-world cohort where data are
145 available only if they are collected as a part of routine clinical care (i.e., passive data collection).
146 We compared the cohorts by stratifying patients based on their time from initial diagnosis and
147 compared clinical scores (Figure 1). We observe several key results: first, scores within HBS
148 show slower movement than MGB in general than HBS. Outside of the UPDRS total score for
149 MGB, the scores do not change substantially during the length of time of typical clinical trials
150 (1-5 years). MMSE was significantly higher throughout the first five years within HBS ($p < 0.001$,
151 Figure 1C). Finally, the biggest change in progression is observed in the UPDRS total score, with
152 the median score at MGB doubling by year 6.

153



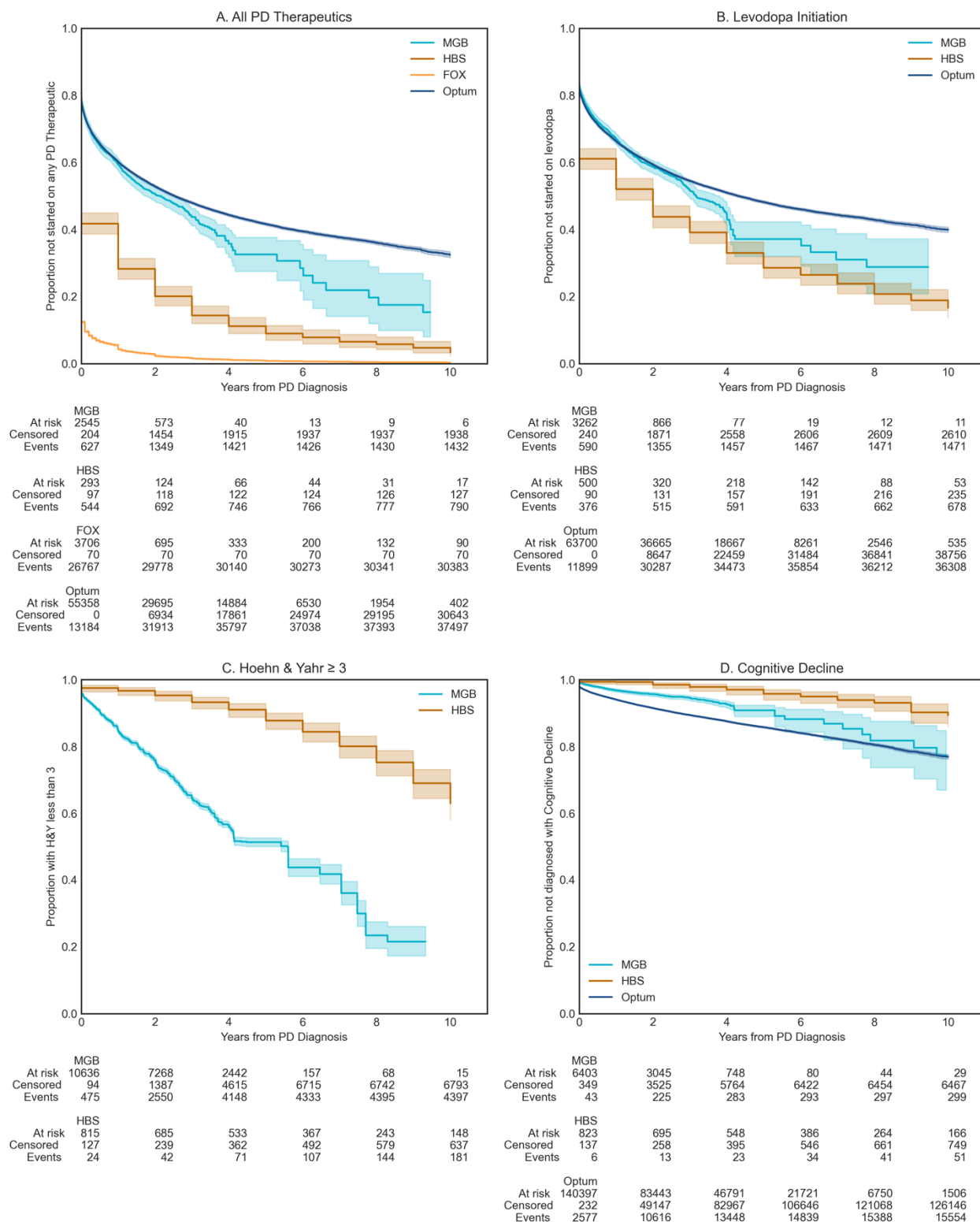
154
155 **Figure 1.** Parkinson's Progression as measured by clinical rating scales (Teal = MGB, Orange =
156 HBS). MGB were only available for up to ten years due to patient size drop-off. **A.)** Hoehn &
157 Yahr Stage (MGB & HBS), **B.)** UPDRS Total Score (MGB & HBS), **C.)** MMSE Score (MGB &
158 HBS),
159

160 Survival against key clinical events including levodopa initiation, reaching H&Y Stage 3,
161 and discharge to long term care, first falls, or fractures after PD diagnosis, diagnoses of
162 depression, and diagnoses of dyskinesias were evaluated in each of the data sources. We
163 observed significantly earlier start of PD medications in the research populations compare to the

164 real-world populations ($p < 0.001$). This was particularly pronounced in the Fox Insights data
165 (Figure 2A). A potential explanation for this outlier result in Fox Insights could be that it entirely
166 relies on patient self-reports. Interestingly, patients in the EMR data (MGB) were overall started
167 on PD medications later than the research cohorts, however, initiation of levodopa therapy
168 specifically overlapped with the research cohorts (Figure 2B). This may indicate that newer non-
169 levodopa PD medications may be more aggressively prescribed in the patients participating in
170 the research cohorts. Patients in RWD reached a H&Y score of 3 or more (Figure 2C, $p < 0.001$)
171 or matched the definition for cognitive decline (Figure 2D, Optum vs. HBS, $p < 0.001$, MGB vs.
172 HBS, $p=0.001$) sooner after diagnosis than those in the research settings.

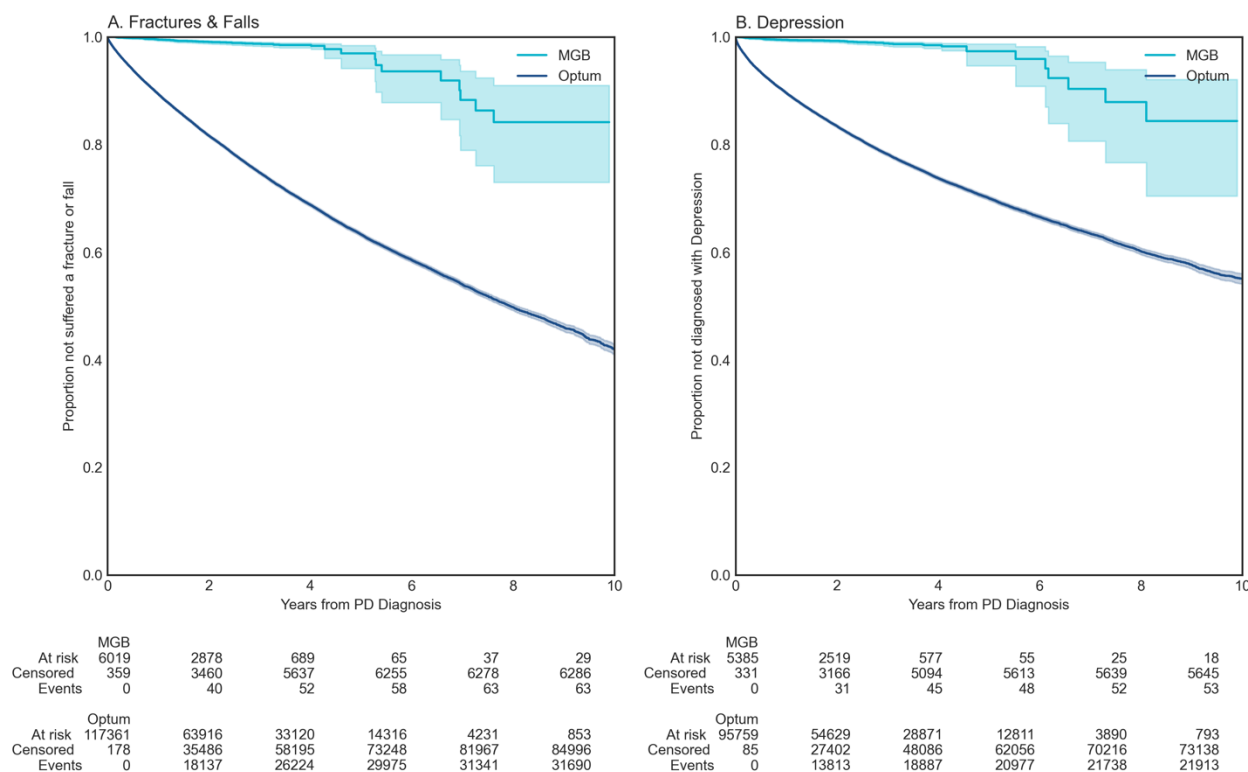
173 Additionally, we examined the association between disease duration (e.g., years since PD
174 diagnosis) and the clinical rating scale score conditioning on the age of diagnosis, sex, and
175 number of years of education (years of education is available in HBS only). For UPDRS, after
176 adjusting for covariates, the intercept was 18.93 in MGB and 28.25 in HBS. At the intercept, the
177 MGB slope was more than double that of HBS (3.87 vs. 1.54 within HBS, $p < 0.001$). Prior
178 studies have estimated a 3 to 14 total UPDRS change per year, making the observed values in
179 HBS lower than previously reported.²⁵ Slopes for H&Y were not observed to be significantly
180 different between MGB and HBS (Supplemental Table 4). While H&Y ranges from 1-5, few
181 patients were observed on either extreme, indicating that this measure may not be sensitive
182 enough to observe differences in the follow-up period. Finally, after adjusting for covariates,
183 significant differences were observed between MMSE scores both at the time of diagnosis (HBS:
184 28.7, MGB: 25.8) and in terms of progression (HBS: 0.11 annual decline, MGB: 0.28 annual
185 decline). Thus, both survival models and the linear regression model (Supplemental Table 4) are
186 consistent with more rapid motor and cognitive decline in real-world vs. research population
187 data. The change in H&Y is in line with what is previously reported in the literature
188 (Supplemental Table 5).

189 Additionally, we broke down the percentage of patients who started therapy with
190 levodopa by their H&Y stage (Supplemental Table 6). By H&Y stage 2.5 >80% of patients had
191 initiated levodopa and we did not see any further increase. This is consistent with clinical
192 guidelines indicating levodopa should be started in patients over 65 upon mild symptoms and for
193 all symptoms as moderate or severe symptoms present.²⁶ This plateau of ~80-85% levodopa
194 initiation during the study follow-up period indicates an upper limit on the utility of the surrogate
195 measure of levodopa initiation to track PD progression. Levodopa provides a benchmark for the
196 treatment of many PD symptoms which can be used as a tool to compare experimental disease
197 modifying therapies or to plan future clinical trials.



198
 199 **Figure 2.** Survival against key clinical events. **A.)** Proportion of patients not started on
 200 Levodopa, **B.)** Proportion of patients not started on PD medications, **C.)** Proportion of patients
 201 with HY ≤ 3 (smaller or equal) (MGB, HBS), **D.)** Proportion of patients who have not
 202 experienced Cognitive Decline Proxy (MGB, HBS, Optum)

203 We then examined the incidence of complications with major effects on patients' quality
 204 of life, that is, fractures and falls; and depression.^{27,28} We observed a substantially higher and
 205 earlier incidence of fractures and falls as well as depression diagnoses in Optum compared to
 206 MGB data ($p < 0.001$, Figure 3A, 3B). The interpretation of this divergence is complex. The
 207 difference could be due to distinct care utilization; primary vs. tertiary point of care; general
 208 neurologist vs. movement disorders specialty care; or other systemic differences in the patient
 209 populations such as socio-economical divergence. This result indicates the need to examine
 210 results from RWD in the context of healthcare delivery.
 211



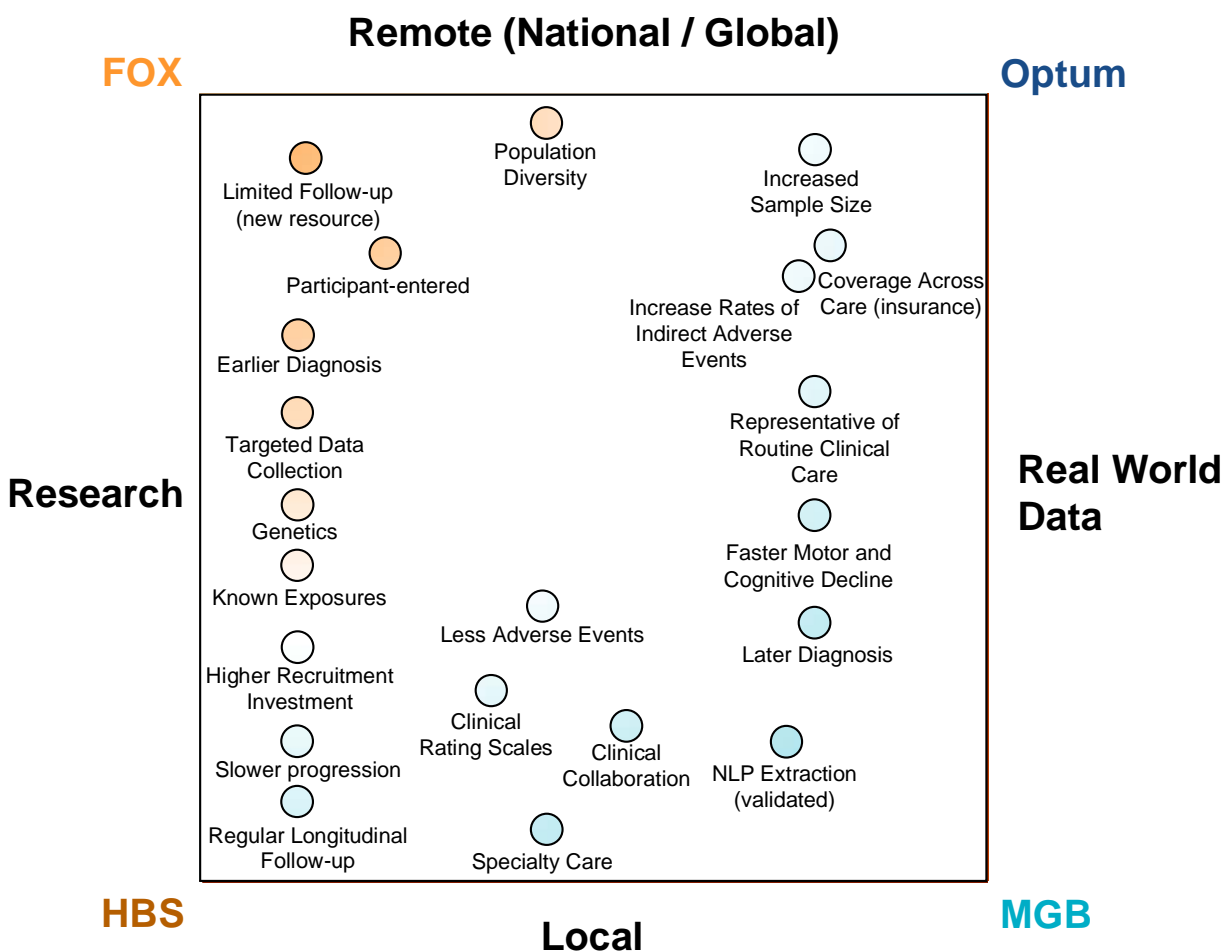
212
 213 **Figure 3. Survival against indirect adverse events in RWD. A.) Fractures & Falls (MGB,**
 214 **Optum), B.) Depression (MGB, Optum).**
 215

216 Discussion

217 This work examines PD progression across four different datasets, two research
 218 populations and two real-world data populations. It demonstrates the ability to utilize large
 219 language models to extract clinical rating scores from clinical notes as part of routine care and to
 220 compare these scores to a research-level, longitudinal clinical observational study. This work
 221 shows progression patterns in clinical rating scales and fit models to estimate “typical”
 222 progression across the irregularly spaced events common in healthcare. A key purpose of this
 223 study is to illustrate that while we observe these differences, it is difficult to attribute these
 224 differences to specific sources. They are likely to be driven at least in part by biases specific to
 225 each data source and collection mechanism, the populations within each data set as well as
 226 clinical care provided within each data source. By performing these analyses across four
 227 different data modalities, this work aims to provide a diverse set of measures accessible to
 228 researchers with access to different data sources. We hope that it can help researchers choose the

229 appropriate data modality, or ideally multiple modalities, for their research question and to
 230 approach their analyses paying careful attention to sources of bias.

231 The study shows systematic differences and potential directional biases between research
 232 and real-world data sets (Figure 4). Patients in research populations are diagnosed much earlier,
 233 are started earlier on levodopa and other PD medications, and have slower changes in clinical
 234 scales of motor and cognitive progression. RWD-based populations are diagnosed at an older
 235 age, start medications later than research cohorts and have faster changes in clinical scales and
 236 events geared at measuring motor and cognitive decline. RWD-based populations offer
 237 substantially larger sample sizes, do not require active recruitment, and may be available
 238 immediately for past populations. Most importantly, RWD-based populations may be more
 239 representative of routine clinical care and thus may be more reflective of the general population.
 240



241 **Figure 4** Comparison of PD phenotypic resources, research populations vs. real-world data and
 242 local vs. remote demonstrating the pros and cons of each data modality for PD research.
 243

244 Which variables may underlie the divergent progression features in the real vs. the
 245 research world? We hypothesize that several factors may play a role:

246 1.) The way cohorts were recruited (e.g., actively recruited research datasets or passively
 247 generated real world data extracted from routine clinical care). This may lead to recruitment bias
 248 for younger and more educated patients into research studies.^{16,29} We previously observed
 249

250 “proxy” diagnoses where a non-specialist might diagnose a patient with a less specific gait,
251 tremor, or cognitive disorder as justification for a referral to a movement disorders specialist.³⁰
252 One potential explanation for this pattern could be a hesitation to make a diagnosis as severe as
253 PD without being completely confident. Populations enriched for Mendelian forms of PD have
254 been observed to have an earlier disease onset and could be over-represented among the research
255 populations. Their earlier diagnosis of PD may be the result of greater access to specialists
256 resulting in a faster transition from a “proxy” diagnosis to PD. Due to critical ethical
257 requirements, HBS requires cognitive competency or a next of kin to provide informed consent.
258 This requirement may exclude clinical subtypes of PD more likely to have substantial cognitive
259 decline before they are able to be recruited and enrolled.³¹ In contrast, patients with later
260 diagnosis and patients with the postural instability-gait difficult (PIGD) form of PD have both
261 been observed to have faster UPDRS part II³² as well as general cognitive³³. There is a
262 relationship between PIGD and disease duration^{34,35}, so accurate assessment of disease onset is
263 critical for future studies of subtypes. This study emphasizes the challenges in differentiating
264 potential phenotypic subtypes from confounding factors such as disease duration.

265 2.) The specialization of care, particularly in terms of what is deemed relevant for
266 treating (MGB) or studying PD (HBS) may dictate which data are collected and how they are
267 collected in comparison to networks that should have comprehensive coverage of all care
268 (Optum). This discrepancy may be a large component of the differences in survival against
269 depression as well as falls and fractures.

270 3.) The source of the data (e.g., provider vs. patient provided) may have led to
271 exaggerated differences between Fox Insight and the other data sources. This was especially
272 relevant regarding PD-therapeutic initiation. In contrast to data from each of the other sources,
273 most patients within Fox Insight believed they immediately started a PD-therapeutic. The
274 different data sources also dictated which analyses were possible, this was especially seen in
275 longitudinal analyses reliant on follow-up periods. Fox Insight represents a new data source and
276 therefore the length of follow up was limited. However, HBS demonstrates the potential for
277 research data sources to offer significant longitudinal data that is difficult to find in RWD (mean
278 follow-up time of 6.1 years vs. 3.2 years, $p < 0.001$). This comparison between data modalities
279 may be related and may help to quantify potential “hawthorne effects”, where greater interaction
280 between participant and physician can lead to differing outcomes.^{36,37}

281 As RWD becomes increasingly available, it is critical to understand the strengths and
282 weaknesses of different sources of data and analytical approaches. This study identified
283 important limitations for each of the datasets. HBS has strengths in terms of the quality of
284 targeted data (e.g., environmental exposures, clinical rating scales, genetics) but given its nature
285 as longitudinal biomarkers study with structured study visits, it is comparatively small in
286 participant numbers. As it is based at a single site in Boston, it also may not be representative of
287 the diversity of PD across the nation. MGB shares this issue of representation but has
288 approximately 25 times as many participants as HBS (22,949 vs. 935) and includes data
289 extracted from typical clinical care records for PD (i.e., real-world data). Patients at MGB and
290 HBS are predominantly of European descent, while ethnicity data were not available in the
291 licensed Optum dataset. This meant, that despite being large data sources, we were not able to
292 subset people with PD according to race or ethnicity in search of different progression patterns.
293 We do however, suspect that there are substantial biases related to access to care within both
294 MGB and Optum. MGB is a highly ranked academic health system and those with means such as
295 insurance, ability to travel, and the support of caregivers who can afford time away to help may

296 be more likely to seek out care at MGB as opposed to whichever provider is closest. Similarly,
297 while geographically broad, Optum is still highly enriched for private insurance and the Midwest
298 region. These factors are not easy for single organizations, even organizations with as many
299 resources as these to mitigate. Less biased data may require system-wide intervention or efforts
300 such as government supported registries universal access to care. Studying neurodegenerative
301 conditions that frequently emerge around the age of retirement is especially challenging as
302 people frequently switch insurers or even move to new locations for retirement. In both cases,
303 neither the claims data nor provider data alone can be relied upon. The population that has
304 continuous data is likely to be biased by those that retire later and among other potential biases
305 may be healthier or have increased socioeconomic needs to continue working.

306 Efforts such as Fox Insight have the potential to alleviate some of these challenges by
307 recruiting for a large, diverse, patient population but this will require intentional effort.
308 Importantly, Fox Insight also collects genetic data allowing for more accurate determination of
309 ethnicity. Otherwise, limitations of EMR data sources like MGB include the fact that genetics are
310 only available on a small subset of patients. Consequently, even where ethnicity data are
311 available, analyses are often dependent on patient reported ethnicity rather than genetic ancestry.
312 However, even this effort is subject to substantial bias since patients must interact with the study
313 online, meaning the patient or someone close to the patient needs to have internet access. Adding
314 to the access question, Fox Insight leverages online advertising and the populations who spend
315 more time online may be more likely to discover Fox Insight.

316 Additionally, clinical scores are not typically captured in structured EHR data, there is no
317 guarantee of completeness of care, there are censoring challenges, and these scores are based on
318 subjective assessments. Future assessments using biomarkers (e.g., RT-QuIC³⁸) may enable the
319 measurement of progression in settings without experts trained in different clinical scales and
320 may provide faster and more objective measurement of progression even when experts are
321 available. Optum, as a combination of EMR and administrative claims data, improves upon the
322 challenges of completeness of care data but has less per-patient granularity than identified EMR
323 data. For example, in some cases it may be possible to extract clinical rating scores, but because
324 end users do not have access to Optum's proprietary natural language processing extraction code
325 or the raw notes the values are extracted from, it's not possible for an end user to validate the
326 extraction. Within MGB it was possible to compare the extracted values to manual chart review-
327 extracted values to validate that the NLP extraction process worked accurately. We are actively
328 working to generalize this approach to share it as open-source software after evaluating in
329 additional settings and cohorts. Without being able to validate the extraction process in Optum,
330 it's difficult to assess whether the extracted values can be used for downstream analyses with
331 sufficient confidence. For this reason, we excluded the clinical rating scales available within
332 Optum from our analyses. Despite these challenges, Optum includes nearly eight times the
333 number of patients as a single site EMR like MGB and is likely more representative of national
334 demographics of PD. We believe that future studies would benefit greatly from comparing both
335 between RWD-sources as well as between the cohort of interest and background control
336 populations. A limitation of this study was the lack of access to control populations in resources
337 like Optum due to data licensing.

338 Finally, the Fox Insight study, is relatively new and therefore has limited follow-up time.
339 Fox insight's web-based recruitment and information collection model have demonstrated
340 promise, rapidly recruiting tens of thousands of PD patients. Because it is survey-based, there is
341 likely to be heterogeneity in response quality (e.g., therapeutic initiation), and unlike the other

342 data sources, it is patient-provided. Data are not entered by physicians and clinical research
343 affiliates. Additionally, Fox Insight, currently collects a history of the presence of clinical events
344 but does not currently collect the date of occurrence. This may be difficult for subjects to
345 remember but would be useful in characterizing progression (i.e., survival analyses require dates
346 of clinical events). We hope that this work can provide guidance to future efforts of this
347 emerging resource as well as aid in the design of future studies.

348 Due to the heterogeneity of PD, a better understanding of the “ground truth” of
349 progression is critical for the development of disease-modifying therapeutics. This understanding
350 may enable the identification of subpopulations for target discovery and improve the design of
351 clinical trials (e.g., selection of more appropriate endpoints). Finally, without a better
352 understanding of current progression, it is difficult to value therapeutics which may slow
353 progression according to patient quality of life and healthcare cost savings.

354
355

356 **Methods**

357 ***Data Sources and Study Population***

358 *Harvard Biomarkers Study*

359 The Harvard Biomarkers Study (HBS) is a longitudinal case-control registry which aims
360 to accelerate the development of molecular diagnostics which track or predict progression of
361 early-stage PD.^{10,39–41} The study includes participants with PD, mild cognitive impairment,
362 Alzheimer’s disease in addition to healthy controls. Participants are recruited from throughout
363 Mass General Brigham and enroll for at least three years through the Memory and Movement
364 Disorders Units at Brigham & Women’s Hospital and Massachusetts General Hospital. Full study
365 information is available in study publications^{9,40,42} and online
366 (<https://www.bwhparkinsoncenter.org/>).

367 PD cases are defined according to the UK PD Society Brain Bank Criteria¹ or a
368 movement disorders specialists’ assessment. Cases were required to have Mini-Mental State
369 Examination (MMSE) scores >21 or the next of kin present to provide informed consent. PD
370 cases were excluded for presence of blood disorders, hematocrit levels <30%, and active ulcers
371 or colitis. Clinical disease severity is measured in the “operational on state” according to
372 clinician expert judgement, but patients were not necessarily at the peak response (“true on
373 state”).

374 At the time of this analysis, the study included a total of 1,914 participants including 935
375 with PD and 979 healthy controls. The biobank includes genotyping (e.g., 42 with *GBA*
376 mutations), medical record data, as well as repeat clinical score and survey
377 measurements. Within HBS, year of diagnosis (but not day and month) were recorded. Thus, all
378 analyses are done at the year of diagnosis level.

379 *Fox Insight*

381 Fox Insight is a web questionnaire-based, longitudinal study including both people
382 having and not having Parkinson’s disease (see publication for full description)⁴³. At the time of
383 study, the Fox Insight database included 36,660 people with Parkinson’s disease. The
384 questionnaire covers the patient’s medical and family histories, potential environmental
385 exposures, as well as ongoing functional status. Additionally, genetic testing has been performed
386 on a subset of the population.

387 *Mass General Brigham (MGB) Electronic Medical Record*

388 The MGB EMR (previously known as Partner’s Health Care) includes > 6.8 million total
389 active records from 1990 to 2020. Available data include encounters, diagnoses, medications,
390 procedures, observations, and labs. The MGB EMR also includes a biobank with >36,000
391 patients who have been genotyped with Illumina Multi-Ethnic Global Array (MEGA) microarray
392 chip.

394 The full EMR includes 22,949 patients with at least two diagnoses of PD as in Yuan et
395 al.³⁰ (ICD9: 332, 332.0, ICD10: G20). To identify new onset PD, we required a two-year
396 quiescence period (730 days) where a patient had encounters with the health system prior to
397 having an initial diagnosis of PD and consider the time from initial (Supplemental Figure 1).
398 There were 13,257 with encounters at least two years prior to their initial PD diagnosis. Full
399 criteria for initial PD diagnosis are in the Study Design and Statistical Analysis section under
400 *Initial Parkinson’s Disease Diagnosis*.

401

402 *Optum® Integrated Claims-EHR Data*

403 Optum-provided description (quoted directly)⁴⁴: “Optum’s Integrated Claims-Clinical
404 dataset combines adjudicated claims data with Optum’s Electronic Health Record data. Optum’s
405 longitudinal clinical and claims repository from Optum Analytics is derived from more than 50
406 healthcare provider organizations in the United States, that include more than 700 Hospitals and
407 7000 Clinics: treating more than 102 million patients receiving care in the United States.
408 Optum’s Integrated Claims-EHR dataset is statistically de-identified under the Expert
409 Determination method consistent with HIPAA and managed according to Optum® and customer
410 data use agreements. The Integrated dataset, which leverages a proprietary Optum algorithm that
411 uses both salting and cryptographic hashing, links both claims and clinical data for
412 approximately 24M matched individuals. The Integrated dataset includes historical, linked
413 administrative claim data from pharmacy claims, physician claims, and facility claims, with EHR
414 information, inclusive of medications prescribed and administered, lab results, vital signs, body
415 measurements, diagnoses, procedures, and information derived from clinical notes using Natural
416 Language Processing.”

417 We identified a total of 157,475 patients with at least two distinct PD diagnoses after at
418 least 730 days between coverage beginning and an initial PD diagnosis. Available data for this
419 cohort include encounters, diagnoses, medications procedures, observations, and labs. Full
420 criteria for initial PD diagnosis are in the Study Design and Statistical Analysis section under
421 *Initial Parkinson’s Disease Diagnosis*.

422

423 *Study Design and Statistical Analysis*

424 *Natural Language Processing Extraction of Clinical Rating Scales*

425 We evaluated PD progression in accordance with common clinical scales: Hoehn & Yahr
426 (H&Y)⁴⁵, Mini-Mental State Examination (MMSE)⁴⁶, Montreal Cognitive Assessment
427 (MoCA)⁴⁷, and the Unified Parkinson’s Disease Rating Scale (UPDRS).⁴⁸ Scores were available
428 as structured data within HBS but needed to be extracted at MGB using natural language
429 processing. While each scale includes subjective assessments, they have undergone validation
430 studies^{10,48-58} and are commonly used in the assessment of PD and/or general cognitive
431 progression.

432 To extract scores, we used the Flan-PaLM instruction-finetuned model²² (flan-t5-xxl) as
433 the base for a few shot (n=10 with clinical scores, 10 without clinical scores) approach with
434 exemplars. The prompt provided exemplars of a positive example and a negative example with
435 the form (italics replaced with actual note content):

436 “Q: Answer the following question with a value between 1 and 5 or NA. What is the Hoehn &
437 Yahr score for the patient in the following note? Note: *Note Content A*: 3 Q: Answer the
438 following question with a value between 1 and 5 or NA. What is the Hoehn & Yahr score for the
439 patient in the following note? Note: *Note Content A*:”

440

441 We provided 10 prompts with exemplars that had clinical rating values present and 10
442 prompts with exemplars that did not have clinical rating scales for each extraction. We then
443 compared the outputs between the inferences and flagged any notes with less than 9 out of 10
444 matching outputs. To verify that the model was appropriately finding scores when they were
445 present, we manually reviewed the extracted scores for:

446

447 Hoehn & Yahr: 200 notes where the reviewer identified a score, 200 notes where the
448 model identified a score, 200 notes where no score was found by the model and 200 notes
449 where we manually identified that no score was found. For the 800 notes, extraction
450 accuracy was 99.75%; 2 notes were incorrect. Both cases were appropriately flagged, as
451 multiple Hoehn & Yahr values were recorded, and across the 10 runs only 4/10 and 5/10
452 scores matched.

453
454 UPDRS: 100 notes where the reviewer identified a score, 100 notes where the model
455 identified a score, 100 notes where no score was found by the model and 100 notes where
456 we manually identified that no score was found. For the 400 notes, model accuracy was
457 98%.

458
459 MMSE: 100 notes where the reviewer identified a score, 100 notes where the model
460 identified a score, 100 notes where no score was found by the model and 100 notes where
461 we manually identified that no score was found. For the 400 notes model accuracy was
462 99.50%

463
464 Across all 3 scores, no extraction with 9 or more matching outputs was incorrect in the
465 randomly sampled evaluation sets. We therefore focused on reviewing all notes not included in
466 our manual review pool that had inconsistent outputs across exemplar runs. This included 133
467 H&Y, 41 UPDRS, and 37 MMSE notes. The majority prediction (but less than 9) was correct for
468 94/133 H&Y notes, 36/41 UPDRS notes and 29/37 MMSE notes. We manually corrected the
469 notes that did not meet our threshold with the human annotated value.

470
471 ***Definitions of Health Care Utilization and Clinical Events***
472 Other definitions used to describe healthcare utilization and clinical events are provided in full in
473 Supplemental Tables 1-3 and described here.

474
475 ***Initial Parkinson's Disease Diagnosis***
476 The initial Parkinson's disease diagnosis date is explicitly stated in both of the research
477 data cohorts (HBS, Fox Insight). Within the RWD, we required a 730-day quiescence period to
478 enrich for newly diagnosed PD (Supplemental Figure 1). Within Optum, this quiescence period
479 was defined by taking the first observed PD diagnosis (Defined in Supplemental Table 2) and
480 checking to see whether the subject had continuous coverage for the 730 days prior. To be
481 included within MGB, an individual needed to have at least one encounter 730 or more days
482 prior. In both RWD sources, subjects were then excluded if they did not have an additional PD
483 diagnosis at least 30 days after the initial PD diagnosis in an attempt to filter out codes used for
484 referral or diagnostic purposes.

485 Within the RWD, subjects diagnosed with conditions that commonly result in clinical
486 parkinsonism symptoms during the two-year quiescence period were excluded.³⁰ This included
487 those with encephalitis, Alzheimer's disease, or other cognitive conditions mimicking idiopathic
488 PD. Additionally those with conditions such as schizophrenia, Parkinson's-related disorders,
489 metabolic neurogenic anomalies (e.g., Wilson's Disease), or other degenerative ailments
490 resulting in clinical parkinsonism symptoms (e.g., Multiple system atrophy, Progressive
491 supranuclear palsy) were also excluded. The codes corresponding to these conditions are detailed
492 in Supplementary Table 2 and derived from Yuan et al.³⁰

493 This exclusion was not performed within HBS as subjects were screened by study
494 personnel and it was not possible to perform within Fox Insights.

495
496 *Cognitive Decline*

497 Cognitive decline was defined as any diagnosis of mild cognitive impairment (ICD9:
498 331.83, ICD10: G31.84) a prescription for a medication typically used to treat cognitive
499 impairment during an encounter with a PD diagnosis (Aricept, Exelon, Namenda, Galantamine
500 using national drug codes derived with RxNorm^{59,60}), or a referral for a neurology specialist for
501 cognitive decline. For HBS, an MMSE score with the cut-off of ≤ 25 was defined as an indicator
502 of significant global cognitive impairment as recommended by the International Parkinson and
503 Movement Disorders Society (MDS) Task Force. The RWD definition was developed in
504 conjunction with clinicians (led by CRS) and trial statisticians (JP, CC) to correlate as closely as
505 possible to the definition using MMSE with data available in RWD. To validate this metric, we
506 employed the Phi coefficient (ϕ), to quantify the degree of association between two binary
507 variables (matching cognitive decline definition, and MMSE threshold). We compared this
508 coefficient for 200 patients across 912 visits where both the RWD-cognitive decline definition
509 was available and a MMSE score is available. We obtained a ϕ value of 0.848, suggesting a
510 strong correlation between the two definitions.

511
512 *PD Therapeutics & Levodopa*

513 Within Fox Insights and HBS the timing of PD therapeutic is explicitly asked of the
514 participant or entered by study staff respectively. Within HBS Levodopa initiation is explicitly
515 asked, but within Fox Insights it is not. Therefore, Fox Insights was not included in the analyses
516 for levodopa initiation.

517
518 For the RWD (Optum, MGB), PD therapeutic initiation (Supplemental Table 1) and levodopa
519 initiation was considered to be first date of a sequence prescription fills meeting all of the
520 following criteria:

- 521 a. Initial fill of a therapeutic listed in linked to an encounter with a PD diagnosis.
522 b. > 90 total days of pills.
523 c. 2+ fills occurring over a period of > 90 days but within 180 days from the first fill of
524 the sequence.

525 These criteria aim to a.) ensure that the prescription is for PD, b.) the patient is not
526 receiving the medication as part of an evaluation or receiving a single 90-day fill, and c.) identify
527 the true starting point of treatment as opposed to evaluation.

528
529 *Clinical Events - Depression & Falls and Fractures*

530 Adverse clinical events were represented by operational criteria for health care utilization
531 and are defined in Supplemental Table 3. These criteria were developed by the Health Value &
532 Economics team within Sanofi.⁶¹ Patients were required to have at least one code for the
533 specified event as either the primary or admitting code for a visit which included an inpatient
534 stay or treatment (procedures or medications). Patients were excluded from this analysis if they
535 experienced diagnoses for the clinical event in question during the quiescence period.

536
537 *Statistical Analyses*

538 *Statistical Tests and Survival Analyses*

539 Quantitative population attributes (e.g., age of diagnosis) were first plotted using
540 Quantile-Quantile Plots followed by a Shapiro-Wilk test for normality. After determining
541 normality, Levene's test was used to assess whether variances were equal. In the case of age and
542 follow-up time, the variances were not equal and therefore a Welch's t-test was performed to
543 determine whether there were significant differences between the populations.

544 For Clinical rating scores (MMSE), after determining that the values were not normally
545 distributed, a Mann-Whitney U test was employed to compare scores between HBS and the
546 extracted scores from MGB.

547 Survival analyses against clinical events were conducted using the Kaplan-Meier method
548 to estimate the time-to-event distributions. These analyses were performed using the 'lifelines'
549 python package.⁶² For each patient, the time to clinical event (or last observed encounter) was
550 recorded along with an binary indicator denoting whether the event was observed or censored.
551 The resulting Kaplan-Meier survival curves visualize the time-to-event distributions, and we
552 used a log-rank test to compare these distributions between different groups.

553 Finally, to ascertain an average progression of UPDRS, MMSE, and H&Y we fit a linear
554 regression model accounting for the irregularity of sampling intervals. To do this we included the
555 age at diagnosis, time since diagnosis, sex, and number of years of education (within HBS only)
556 as covariates and fit a model to the appropriate clinical rating scale. Residual plots were
557 examined to verify the homoscedasticity and linearity assumptions. The normality of residuals
558 was assessed using quantile-quantile (QQ) plots and the Shapiro-Wilk test. Multicollinearity
559 among predictors was evaluated using the variance inflation factor (VIF), ensuring all VIF values
560 were below 10.

561 562 *Missing Data*

563 Within the two research databases, there was minimal missing data due to the overlap of
564 their study designs and the analyses in this work. Within HBS, data are entered by study
565 personnel who are instructed to complete all of fields used in this study. All participants had an
566 age at study visit, age at diagnosis, gender, race, and ethnicity at each study visit. Across all
567 study visits, H&Y was recorded in all visits and MMSE was recorded in all but 8 visits (99.8%).
568 Visits without a MMSE value were excluded from analyses. Within Fox Insights, participants
569 answer questions in panels in order as they sign up. Participants were excluded if they did not
570 complete the initial set of questions required for this study. In the RWD, data are only available
571 through normal clinical care when patients have interactions with the healthcare system. Study
572 inclusion criteria were selected to maximize the likelihood that patient encounters would be
573 captured. To this end, patients were required to either have coverage (Optum) or have prior visits
574 (MGB) at least two years prior to their initial PD diagnosis. This was put in place to both filter
575 for new diagnoses of PD as well as to limit patients who were referred after an initial diagnosis.

576 Additionally, many visits do not include a full PD-specific work up and/or do not record
577 PD clinical rating scores. Less than 10% of patients had an initial H&Y score included in their
578 note (1,071 of 13,257 patients, 8.08%). Similar to prior work⁶³, we trained simple machine
579 learning classifiers to predict whether a score would be present based on a patient's age, sex, age
580 at initial diagnosis, binary presence of comorbidities (rolled up to PheWAS codes),⁶⁴ distance
581 from the clinic, which clinic, and physician specialty. Outside of physician specialty, no feature
582 was determined to be significant and even with physician specialty, performance was only
583 marginally better than random (AUC: 0.56). Due to the relative rarity, lack of a clear discernable
584 pattern of presence, and the fact that it could not be determined that missingness was dependent

585 on outcome, we elected not to perform imputation and instead to include clinical rating scores
586 only where they were available in the notes. This approach comes with an important caveat, the
587 patients with scores may not be representative of the entire cohort. The reasons for missing
588 scores may not be observed, whether related to disease severity, accessibility to care, or other
589 clinical factors, could systematically differ between those with and without scores. This study
590 aims to mitigate these challenges by using multiple datasets and presents a comparison between
591 these datasets to better understand these biases.

592
593

594 ***Standard Protocol Approvals, Registrations, and Patient Consents***

595 This study was approved by the Mass General Brigham Institutional Review Board
596 (#2017P002452). This study received only de-identified, retrospective data from research
597 datasets (HBS and Fox). Research dataset participants are consented for study and publication as
598 part of the recruitment process. Real world datasets (MGB and Optum) are retrospective de-
599 identified data and therefore individual consent is not required for publication.

600

601 ***Data Availability***

602 Authors had primary access to all data sources used in this study (BKB). All data used in this
603 study are secondary use and therefore the authors must direct those interested in using any data
604 involved to their source:

605

- 606 1. Harvard Biomarkers set is accessible through Accelerating Medicines Partnership
607 program for Parkinson's disease <https://amp-pd.org/unified-cohorts/hbs> and through
608 <https://www.bwhparkinsoncenter.org/biobank/>
- 609 2. Mass General Brigham EMR; [https://rc.partners.org/about/who-we-are-risc/research-](https://rc.partners.org/about/who-we-are-risc/research-patient-data-registry)
610 [patient-data-registry](https://rc.partners.org/about/who-we-are-risc/research-patient-data-registry)
- 611 3. Optum® Integrated Claims-EHR Data⁴⁴
612 ([https://www.optum.com/content/dam/optum3/optum/en/resources/fact-](https://www.optum.com/content/dam/optum3/optum/en/resources/factsheets/Integrated-Data-product-sheet.pdf)
613 [sheets/Integrated-Data-product-sheet.pdf](https://www.optum.com/content/dam/optum3/optum/en/resources/factsheets/Integrated-Data-product-sheet.pdf)),
- 614 4. Fox Insight⁴³ <https://foxinsight.michaeljfox.org/>

615

616

617 **Author Contributions**

	Conception or design of the work	Acquisition of data	Analysis of Data	Interpretation of Data	Manuscript draft	Revision
Brett K. Beaulieu-Jones ¹	x	x (pre-processing)	x	x	x	x
Francesca Frau	x		x	x		x
Sylvie Bozzi	x		x	x	x	x
Karen J. Chandross	x			x	x	x
M Judith Peterschmitt			x	x	x	x
Caroline Cohen			x	x	x	x
Catherine Coulovrat			x	x	x	x
Dinesh Kumar				x		x
Mark Kruger		x		x	x	x
Scott L. Lipnick	x	x			x	x
Lane Fitzsimmons					x	x
Isaac S. Kohane	x			x	x	x
Clemens R. Scherzer	x	x		x	x	x

618

619 **Acknowledgements**

620 This work was funded in part by the Sanofi iDEA Awards Initiative (BKB) as well the National
621 Institute of Neurological Disorders and Stroke grant number (K99NS114850, R00NS114850).
622 CRS's work is supported in part by NIH grants NINDS/NIA R01NS115144, U01NS095736,
623 U01NS100603, the U.S. Department of Defense, and the American Parkinson Disease
624 Association Center for Advanced Parkinson Research. The Fox Insight Study (FI) is funded by
625 The Michael J. Fox Foundation for Parkinson's Research. This research was funded in part by
626 Aligning Science Across Parkinson's [ASAP-000301] through the Michael J. Fox Foundation for
627 Parkinson's Research (MJFF). We would like to thank the Parkinson's community for
628 participating in this study to make this research possible. We would like to acknowledge Cliona
629 Molony and Sanofi Digital, for facilitating data access through the Sanofi Darwin platform to
630 support this project. We would also like to thank all of the participants and organizers of the
631 Harvard Biomarker Study as well as the organizers of the Research Patient Data Registry at
632 Mass General Brigham. For the purpose of open access, the author has applied a CC BY public
633 copyright license to all Author Accepted Manuscripts arising from this submission.

634

635 **Competing Interest**

636 FF, SB, KJC, MJP, CC, DK, CC are employees of Sanofi and may hold shares and/or stock
637 options in the company. All authors declare no other competing financial or non-financial
638 interests.

639

640 **References**

- 641 1. Kalia, L. V., Kalia, S. K. & Lang, A. E. Disease-modifying strategies for Parkinson's
642 disease. *Mov. Disord.* **30**, 1442–1450 (2015).
- 643 2. Dawson, V. L. & Dawson, T. M. Promising disease-modifying therapies for Parkinson's
644 disease. *Sci. Transl. Med.* **11**, (2019).
- 645 3. Zheng, B. *et al.* PGC-1 α , a potential therapeutic target for early intervention in Parkinson's
646 disease. *Sci. Transl. Med.* **2**, 52ra73 (2010).
- 647 4. Iwaki, H. *et al.* Genomewide association study of Parkinson's disease clinical biomarkers in
648 12 longitudinal patients' cohorts. *Mov. Disord.* **34**, 1839–1850 (2019).
- 649 5. Blauwendraat, C. *et al.* Parkinson's disease age at onset genome-wide association study:
650 Defining heritability, genetic loci, and α -synuclein mechanisms. *Mov. Disord.* **34**, 866–875
651 (2019).
- 652 6. van Rooden, S. M. *et al.* The identification of Parkinson's disease subtypes using cluster
653 analysis: A systematic review. *Mov. Disord.* **25**, 969–978 (2010).
- 654 7. Thenganatt, M. A. & Jankovic, J. Parkinson Disease Subtypes. *JAMA Neurol.* **71**, 499
655 (2014).
- 656 8. Jankovic, J. *et al.* Variable expression of Parkinson's disease: A base-line analysis of the
657 DAT ATOP cohort. *Neurology* **40**, 1529–1529 (1990).
- 658 9. Liu, G. *et al.* Specifically neuropathic Gaucher's mutations accelerate cognitive decline in
659 Parkinson's. *Ann. Neurol.* **80**, 674–685 (2016).
- 660 10. Liu, G. *et al.* Prediction of cognition in Parkinson's disease with a clinical–genetic score: a
661 longitudinal analysis of nine cohorts. *Lancet Neurol.* **16**, 620–629 (2017).
- 662 11. Liu, G. *et al.* Genome-wide survival study identifies a novel synaptic locus and polygenic
663 score for cognitive progression in Parkinson's disease. *Nat. Genet.* **53**, 787–793 (2021).

- 664 12. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications.
665 *Nat. Genet.* **47**, 856–860 (2015).
- 666 13. Liu, G. *et al.* Mitochondrial haplogroups and cognitive progression in Parkinson’s disease.
667 *Brain* **146**, 42–49 (2023).
- 668 14. Mittal, S. *et al.* β 2-Adrenoreceptor is a regulator of the α -synuclein gene driving risk of
669 Parkinson’s disease. *Science* **357**, 891–898 (2017).
- 670 15. Parkinson Progression Marker Initiative. The Parkinson Progression Marker Initiative
671 (PPMI). *Prog. Neurobiol.* **95**, 629–635 (2011).
- 672 16. Beaulieu-Jones, B. K. *et al.* Examining the Use of Real-World Evidence in the
673 Regulatory Process. *Clin. Pharmacol. Ther.* **318**, 703 (2019).
- 674 17. Karlson, E. W., Boutin, N. T., Hoffnagle, A. G. & Allen, N. L. Building the Partners
675 HealthCare Biobank at Partners Personalized Medicine: Informed Consent, Return of
676 Research Results, Recruitment Lessons and Operational Considerations. *J Pers Med* **6**,
677 (2016).
- 678 18. Wallace, P. J., Shah, N. D., Dennen, T., Bleicher, P. A. & Crown, W. H. Optum Labs:
679 building a novel node in the learning health care system. *Health Aff.* **33**, 1187–1194 (2014).
- 680 19. Agrawal, M., Hegselmann, S., Lang, H., Kim, Y. & Sontag, D. Large language models are
681 few-shot clinical information extractors. in *Proceedings of the 2022 Conference on*
682 *Empirical Methods in Natural Language Processing 1998–2022* (Association for
683 Computational Linguistics, 2022).
- 684 20. Singhal, K. *et al.* Large Language Models Encode Clinical Knowledge. *arXiv [cs.CL]*
685 (2022).
- 686 21. Chowdhery, A. *et al.* PaLM: Scaling Language Modeling with Pathways. *arXiv [cs.CL]*
687 (2022).
- 688 22. Chung, H. W. *et al.* Scaling Instruction-Finetuned Language Models. *arXiv [cs.LG]* (2022).
- 689 23. Brown, T. B. *et al.* Language Models are Few-Shot Learners. *arXiv [cs.CL]* (2020).
- 690 24. Huang, Y., Yuan, W., Kohane, I. S. & Beaulieu-Jones, B. K. Illustrating potential effects of
691 alternate control populations on real-world evidence-based statistical analyses. *JAMIA Open*
692 **4**, ooab045 (2021).
- 693 25. Holden, S. K., Finseth, T., Sillau, S. H. & Berman, B. D. Progression of MDS-UPDRS
694 Scores Over Five Years in De Novo Parkinson Disease from the Parkinson’s Progression
695 Markers Initiative Cohort. *Mov Disord Clin Pract* **5**, 47–53 (2018).
- 696 26. Spindler, M. A. & Tarsy, D. Initial pharmacologic treatment of Parkinson disease. in
697 *UpToDate* (2019).
- 698 27. Talman, L. Quality of Life for Patients With Parkinson Disease. *Neurology* **98**, e2293–
699 e2295 (2022).
- 700 28. Bock, M. A., Brown, E. G., Zhang, L. & Tanner, C. Association of Motor and Nonmotor
701 Symptoms With Health-Related Quality of Life in a Large Online Cohort of People With
702 Parkinson Disease. *Neurology* **98**, e2194–e2203 (2022).
- 703 29. Ludmir, E. B. *et al.* Factors Associated With Age Disparities Among Cancer Clinical Trial
704 Participants. *JAMA Oncol* **5**, 1769–1773 (2019).
- 705 30. Yuan, W. *et al.* Accelerating diagnosis of Parkinson’s disease through risk prediction. *BMC*
706 *Neurol.* **21**, 201 (2021).
- 707 31. Fereshtehnejad, S.-M. *et al.* New Clinical Subtypes of Parkinson Disease and Their
708 Longitudinal Progression: A Prospective Cohort Comparison With Other Phenotypes.
709 *JAMA Neurol.* **72**, 863–873 (2015).

- 710 32. Jankovic, J. & Kapadia, A. S. Functional decline in Parkinson disease. *Arch. Neurol.* **58**,
711 1611–1615 (2001).
- 712 33. Burn, D. J. *et al.* Motor subtype and cognitive decline in Parkinson’s disease, Parkinson’s
713 disease with dementia, and dementia with Lewy bodies. *J. Neurol. Neurosurg. Psychiatry*
714 **77**, 585–589 (2006).
- 715 34. Alves, G., Larsen, J. P., Emre, M., Wentzel-Larsen, T. & Aarsland, D. Changes in motor
716 subtype and risk for incident dementia in Parkinson’s disease. *Mov. Disord.* **21**, 1123–1130
717 (2006).
- 718 35. Kotagal, V. Is PIGD a legitimate motor subtype in Parkinson disease? *Ann Clin Transl*
719 *Neurol* **3**, 473–477 (2016).
- 720 36. Braunholtz, D. A., Edwards, S. J. L. & Lilford, R. J. Are randomized clinical trials good for
721 us (in the short term)? Evidence for a “trial effect.” *J. Clin. Epidemiol.* **54**, 217–224 (2001).
- 722 37. McCarney, R. *et al.* The Hawthorne Effect: a randomised, controlled trial. *BMC Med. Res.*
723 *Methodol.* **7**, 30 (2007).
- 724 38. Vivacqua, G. *et al.* Salivary α -Synuclein RT-QuIC Correlates with Disease Severity in de
725 novo Parkinson’s Disease. *Mov. Disord.* **38**, 153–155 (2023).
- 726 39. Ding, H. *et al.* Unrecognized vitamin D3 deficiency is common in Parkinson disease:
727 Harvard Biomarker Study. *Neurology* **81**, 1531–1537 (2013).
- 728 40. Huh, Y. E. *et al.* β -Glucocerebrosidase activity in GBA-linked Parkinson disease: The type
729 of mutation matters. *Neurology* **95**, e685–e696 (2020).
- 730 41. Locascio, J. J. *et al.* Association between α -synuclein blood transcripts and early,
731 neuroimaging-supported Parkinson’s disease. *Brain* **138**, 2659–2671 (2015).
- 732 42. *al., A. O. et.* The Harvard Biomarker Study: Biomarkers for Parkinson’s. *In prep.*
- 733 43. Smolensky, L. *et al.* Fox Insight collects online, longitudinal patient-reported outcomes and
734 genetic data on Parkinson’s disease. *Sci Data* **7**, 67 (2020).
- 735 44. Optum. Optum’s de-identified Integrated Claims-Clinical dataset (2007-2019). (2020).
- 736 45. Hoehn, M. M. & Yahr, M. D. Parkinsonism: onset, progression, and mortality. 1967.
737 *Neurology* **50**, 318 and 16 pages following (1998).
- 738 46. Folstein, M. F., Robins, L. N. & Helzer, J. E. The Mini-Mental State Examination. *Arch.*
739 *Gen. Psychiatry* **40**, 812 (1983).
- 740 47. Nasreddine, Z. S. *et al.* The Montreal Cognitive Assessment, MoCA: a brief screening tool
741 for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**, 695–699 (2005).
- 742 48. Movement Disorder Society Task Force on Rating Scales for Parkinson’s Disease. The
743 Unified Parkinson’s Disease Rating Scale (UPDRS): status and recommendations. *Mov.*
744 *Disord.* **18**, 738–750 (2003).
- 745 49. Goetz, C. G. *et al.* Movement Disorder Society Task Force report on the Hoehn and Yahr
746 staging scale: status and recommendations the Movement Disorder Society Task Force on
747 rating scales for Parkinson’s disease. *Mov. Disord.* **19**, 1020–1028 (2004).
- 748 50. Gallagher, D. A., Goetz, C. G., Stebbins, G., Lees, A. J. & Schrag, A. Validation of the
749 MDS-UPDRS Part I for nonmotor symptoms in Parkinson’s disease. *Mov. Disord.* **27**, 79–
750 83 (2012).
- 751 51. Müller, J. *et al.* Progression of Hoehn and Yahr stages in Parkinsonian disorders: a
752 clinicopathologic study. *Neurology* **55**, 888–891 (2000).
- 753 52. Siciliano, M. *et al.* Cognitive impairment is associated with Hoehn and Yahr stages in early,
754 de novo Parkinson disease patients. *Parkinsonism Relat. Disord.* **41**, 86–91 (2017).

- 755 53. Martinez-Martin, P. *et al.* Expanded and independent validation of the Movement Disorder
756 Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS). *J. Neurol.* **260**, 228–236
757 (2013).
- 758 54. Goetz, C. G., Stebbins, G. T. & Tilley, B. C. Calibration of unified Parkinson's disease
759 rating scale scores to Movement Disorder Society-unified Parkinson's disease rating scale
760 scores. *Mov. Disord.* **27**, 1239–1242 (2012).
- 761 55. Skorvanek, M. *et al.* Differences in MDS-UPDRS scores based on Hoehn and Yahr stage
762 and disease duration. *Mov. Disord. Clin. Pract.* **4**, 536–544 (2017).
- 763 56. Hoops, S. *et al.* Validity of the MoCA and MMSE in the detection of MCI and dementia in
764 Parkinson disease. *Neurology* **73**, 1738–1745 (2009).
- 765 57. Fountoulakis, K. N., Tsolaki, M., Chantzi, H. & Kazis, A. Mini Mental State Examination
766 (MMSE): A validation study in Greece. *Am. J. Alzheimers. Dis. Other Demen.* **15**, 342–345
767 (2000).
- 768 58. Beaman, S. R. de *et al.* Validation of a modified version of the Mini-Mental State
769 Examination (MMSE) in Spanish. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.*
770 **11**, 1–11 (2004).
- 771 59. Liu, S., Ma, W., Moore, R., Ganesan, V. & Nelson, S. RxNorm: prescription for electronic
772 drug information exchange. *IT Prof.* **7**, 17–23 (2005).
- 773 60. Nelson, S. J., Zeng, K., Kilbourne, J., Powell, T. & Moore, R. Normalized names for
774 clinical drugs: RxNorm at 6 years. *J. Am. Med. Inform. Assoc.* **18**, 441–448 (2011).
- 775 61. Giladi, N. *et al.* Safety and efficacy of venglustat in GBA1-associated Parkinson's disease:
776 an international, multicentre, double-blind, randomised, placebo-controlled, phase 2 trial.
777 *Lancet Neurol.* **22**, 661–671 (2023).
- 778 62. Davidson-Pilon, C. *et al.* *Lifelines*. (2018). doi:10.5281/ZENODO.1210821.
- 779 63. Beaulieu-Jones, B. K. *et al.* Characterizing and Managing Missing Structured Data in
780 Electronic Health Records: Data Analysis. *JMIR Med Inform* **6**, e11 (2018).
- 781 64. Denny, J. C. *et al.* PheWAS: Demonstrating the feasibility of a phenome-wide scan to
782 discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).