

1. Dataset

Extraction

The data contained in Enroll-HD PDS5 was extracted from the Enroll-HD electronic data capture (EDC) database on **October 31, 2020**, at 23:00 UTC.

Data sources

The PDS5 dataset encompasses data exclusively from Enroll-HD participants, collected from several sources. These sources are the Enroll-HD study, the REGISTRY study, and clinical data collected in adhoc visits outside of the aforementioned studies.

Enroll-HD is an observational cohort study and global clinical research platform designed to facilitate Huntington's disease (HD) clinical research. It includes participants from North America, Europe, Australasia, and Latin America. The study started in 2012 and is still active and actively recruiting.

REGISTRY is an observational cohort study of HD conducted in Europe. The study started in 2004 and concluded in 2015. As Enroll-HD began, REGISTRY sites and participants began to transition into Enroll-HD. Enroll-HD dataset releases include individuals who initially participated in REGISTRY then consequently enrolled in Enroll-HD and consented to the migration of their REGISTRY data into the Enroll-HD dataset. Registry data are available for a subset of Enroll-HD participants.

Clinical data from additional sources (**Ad Hoc** data) are available for a subset of Enroll-HD participants. These data were collected at routine clinical visits outside of the Enroll-HD and REGISTRY studies, and comprise HD assessment data (e.g., UHDRS Motor). The date of collection of these data typically pre-date a participant's enrolment into REGISTRY or Enroll-HD.

Study specific **protocols**, **annotated eCRFs**, and **data collection guidelines** are housed [here](#).

Participant inclusion

To be included in the PDS5 release, participant's data had to meet several requirements. **Figure 1** illustrates the number of participants whose data met each of the predefined inclusion requirement and illustrates how the final sample size of PDS5 was determined.

Figure 1. Participant flow chart for inclusion in PDS5.



Note that as a result of these requirements, not all participants enrolled in Enroll-HD at the time of PDS5 data cut are included in PDS5. Similarly, not all participants included in PDS4 are included in PDS5. Data from 150 participants who were included in PDS4 were not included in the current PDS release. Enroll-HD is an active, longitudinal study. A participant eligible for inclusion for one release may be ineligible the next (e.g., participant data quarantined). Data for PDS4 participants not included in PDS5 may be available through specified dataset (SPS) request.

2. Sample Size

PDS5 contains data on 21,116 Enroll-HD participants. Sample size by PDS release is presented in **Figure 2**.

Figure 2. Enroll-HD sample size by PDS release.



3. Visits

PDS5 contains data from 71,682 visits (baseline and follow-up visits only; all sources). Of these, 55,975 were Enroll-HD visits. The remainder are from Registry (N = 14,737) and 'Ad Hoc' sources (N = 970). A breakdown of visits by data source is provided in **Table 1**. Number of Enroll-HD visits only by PDS release are illustrated in **Figure 3**.

Table 1. Number of visits in PDS5 by constituent data source. Participant number indicates the number of Enroll-HD participants with visit data available for the indicated data source (maximum N = 21,116).

Data source	Participants	Visits
Enroll-HD	21,116	55,975
Registry 3	4,168	9,708
Registry 2	2,079	5,029
Ad Hoc	302	970
TOTAL		71,682

Figure 3. Enroll-HD visits only (baseline and follow-up only) by PDS release.



Considering baseline and follow-up visits from *Enroll-HD only*, total number of visits per participant in PDS5 ranges from 1 to 9. In **Figure 4**, we illustrate participant counts by maximum number of Enroll-HD visits. Each participant is represented once, included in the bar indicative of their maximum number of visits. In **Figure 5**, we illustrate maximum participant counts for a specific number of visits. This plot is cumulative, the goal being to illustrate largest available sample size for a specific number of visits. For example, the participant with 9 visits is represented in visit bars 1 through 9, the participants with 8 visits are represented in visit bars 1 through 8, and so on.



Figure 4. Participant counts by maximum number of Enroll-HD visits (baseline and follow-up visits only; unscheduled visits and phone contacts excluded). Full sample represented (N = 21,116; Missing N = 0).

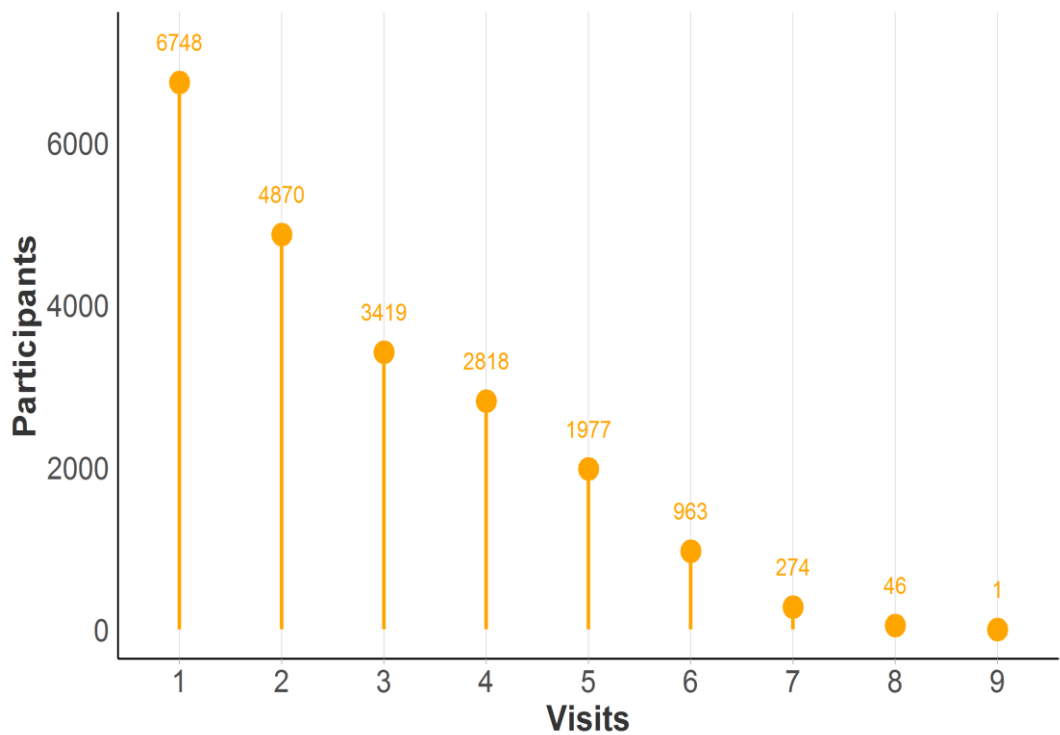
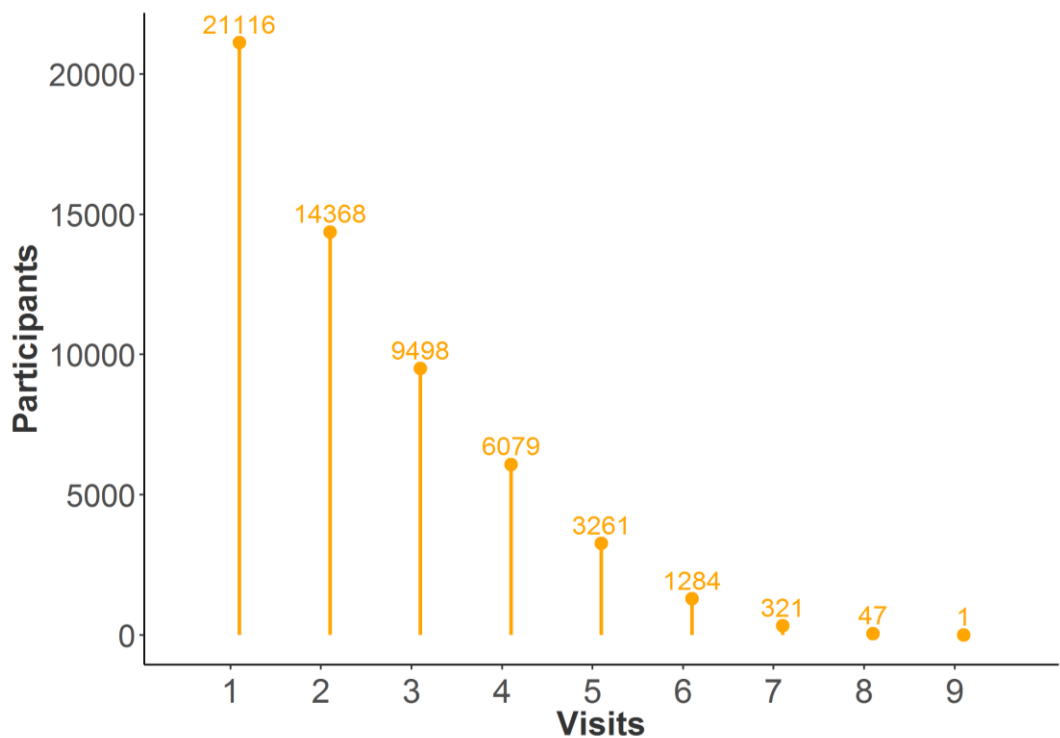
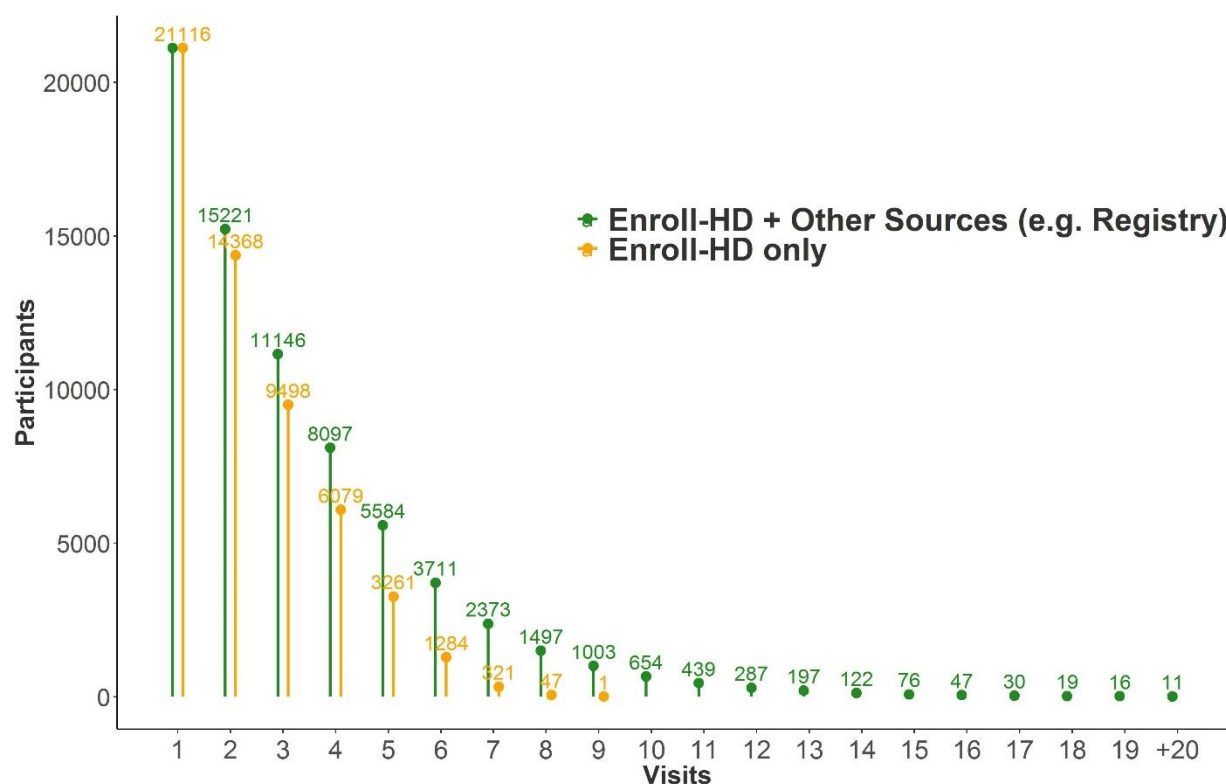


Figure 5. Maximum participant counts for a specific number of Enroll-HD visits (baseline and follow-up visits only; unscheduled visits and phone contacts excluded). Cumulative plot. Full sample represented (N = 21,116; Missing N = 0).



Considering baseline and follow-up visits from *all data sources* (Enroll-HD, REGISTRY, ad hoc visits), total number of visits per participant ranges from 1 to >20. Maximum participant counts by visit number are presented in **Figure 6**.

Figure 6. Maximum participant counts for a specific number of visits; Enroll-HD, REGISTRY, Adhoc (baseline and follow-up visits only; unscheduled visits and phone contacts excluded). Cumulative plot. Full sample represented (N = 21,116; Missing N = 0).



4. Sample Characteristics

The PDS5 sample is characterized below with respect to participant category, sociodemographic variables, and clinical characteristics (Figures 7 to 18).

Figure 7. Participant category (hdcat) at baseline Enroll-HD visit. Full sample represented (N = 21,116; Missing N = 0)

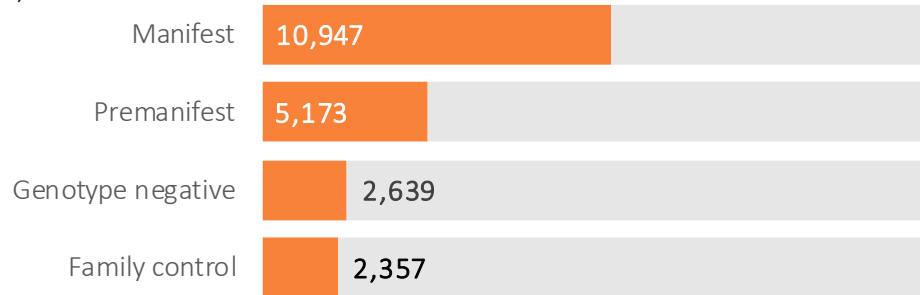


Figure 8. Participant category (hdcat) at latest Enroll-HD visit. Full sample represented (N = 21,116; Missing N = 0)

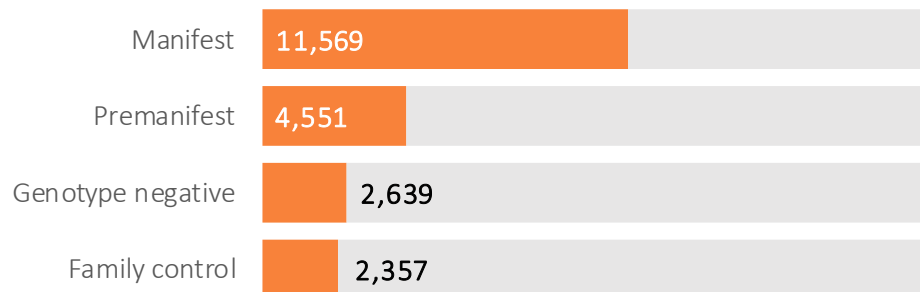


Figure 9. Geographical region (region). Full sample represented (N = 21,116; Missing N = 0)

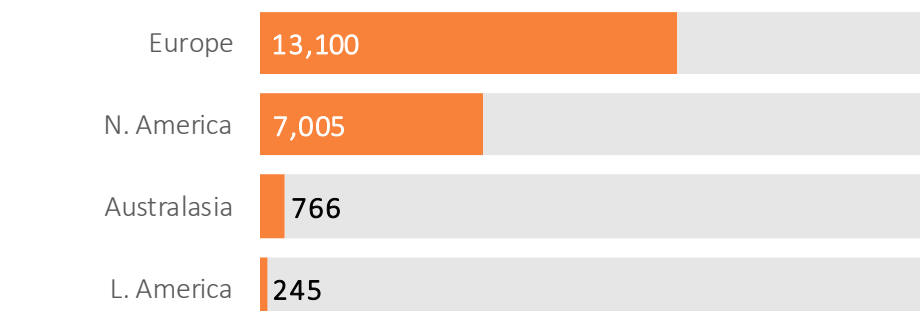


Figure 10. Sex (sex). Full sample represented (N = 21,116; Missing N = 0)





Figure 11. *ISCED (isced) at baseline Enroll-HD visit.* Full sample represented (N = 21,017; Missing N = 99)

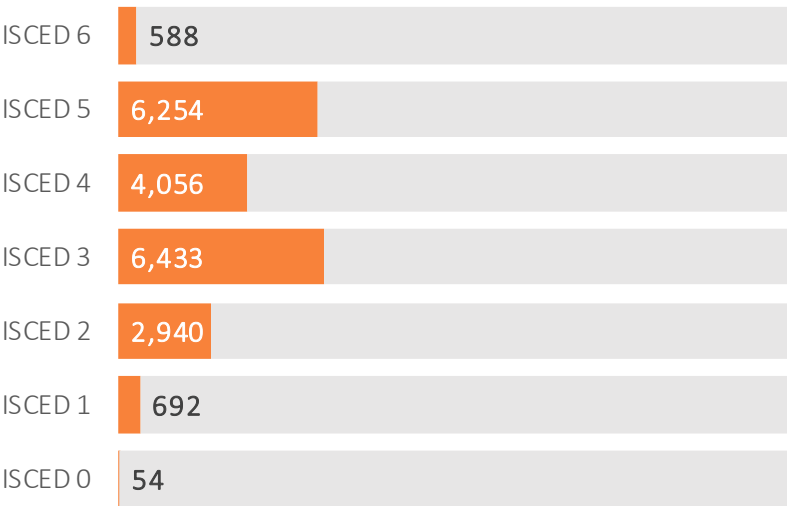


Figure 12. Age at baseline Enroll-HD visit. Full sample represented (N = 21,116; Missing N = 0). Note that in PDS5, age for individuals under the age of 18 years is represented by the value '<18'. These values have been transformed to 17 for inclusion in this histogram.

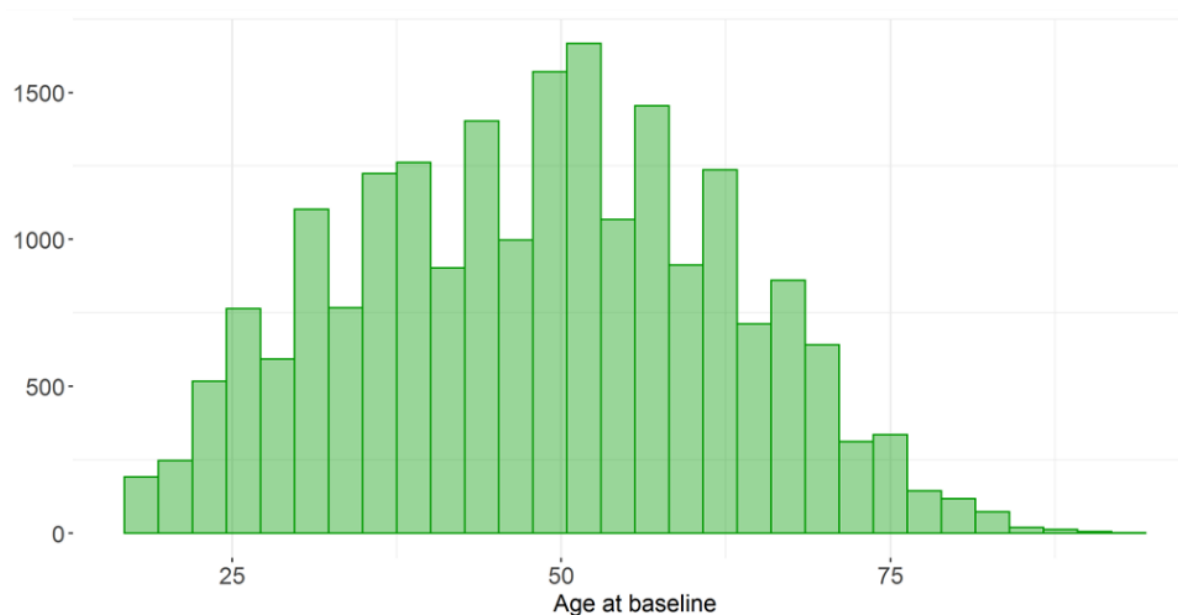


Figure 13. CAG length. HDGECs represented (N = 16,120; Missing N = 0). Note that in PDS5, CAG length for individuals with a CAG length greater than 70 is represented by the value '>70'. These values have been transformed to 71 for inclusion in this histogram.

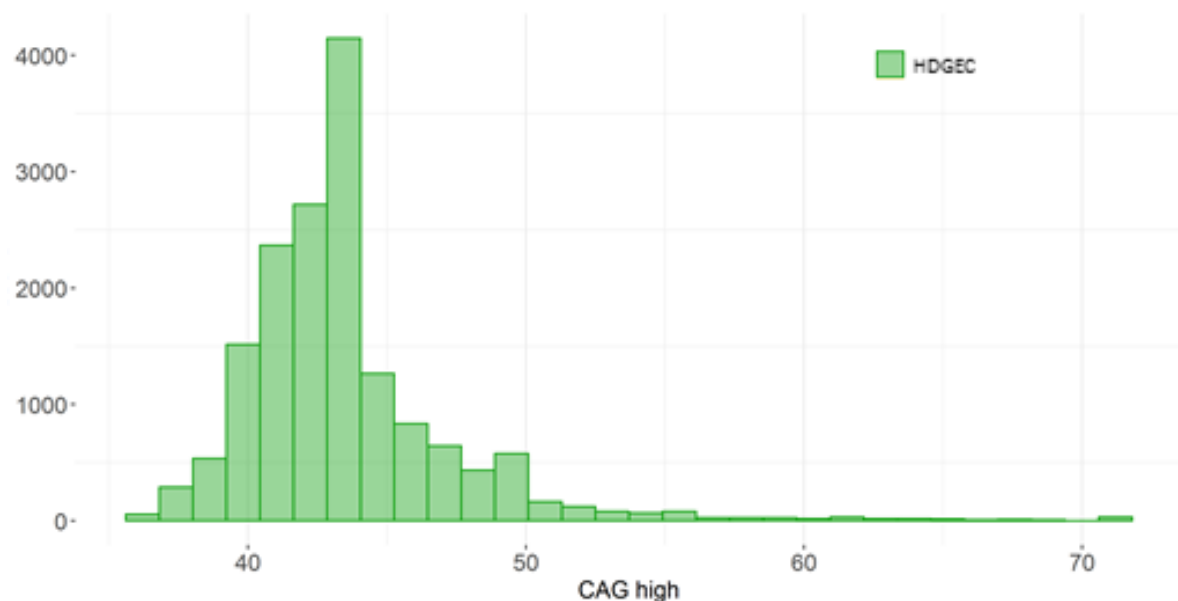


Figure 14. CAP score at baseline Enroll-HD visit. HDGECs represented (N = 16,120; Missing N = 0). Calculated using the Warner et al.¹ CAP score formula.

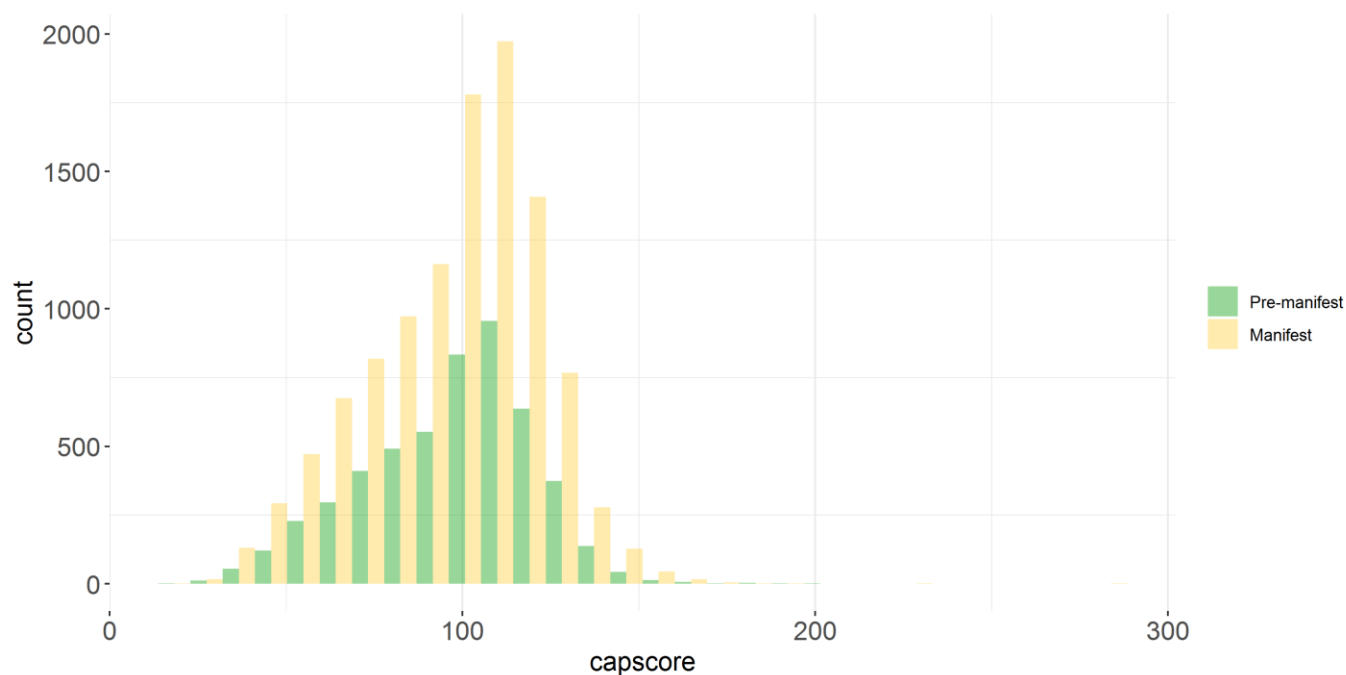
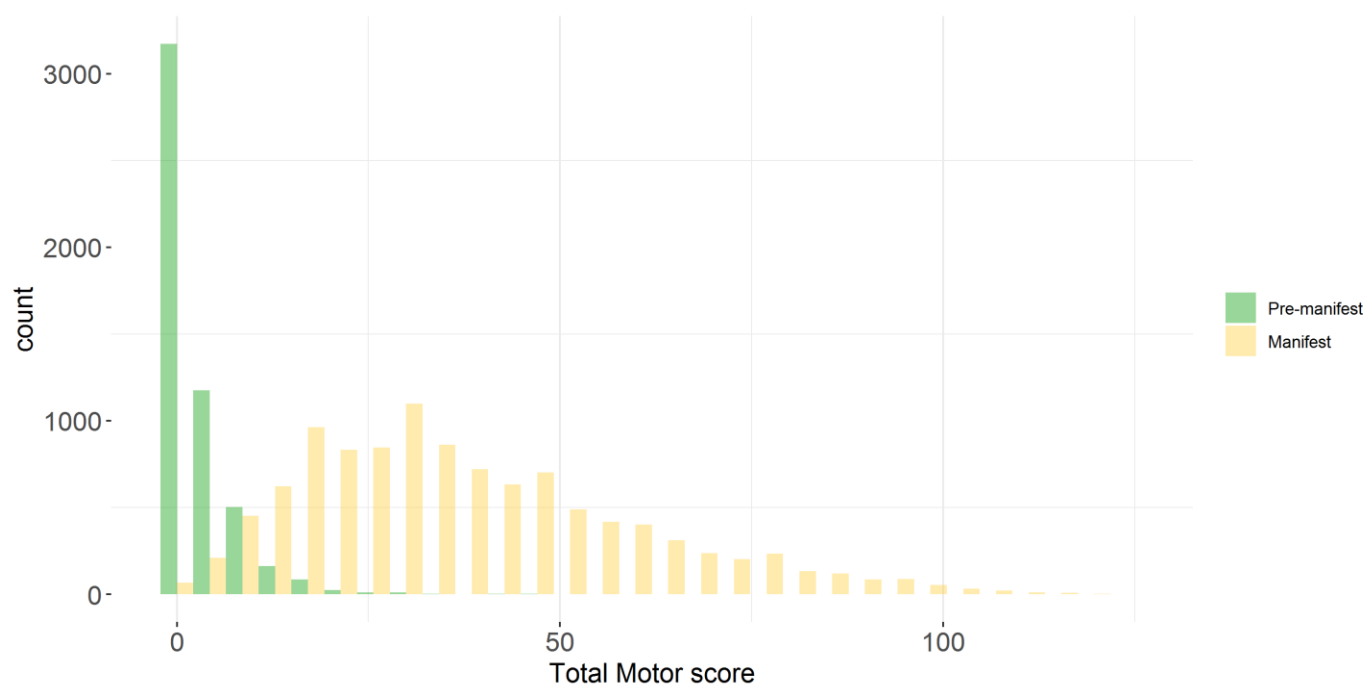


Figure 15. UHDRS total motor score at baseline Enroll-HD visit. HDGECs represented (N = 16,000; Missing N = 120).



¹ Warner, J. H., J. D. Long, J. A. Mills, D. R. Langbehn, J. Ware, A. Mohan, and C. Sampaio. 2020. "Standardizing the CAP Score in Huntington's Disease I: Predicting Age-at-Onset."

Figure 16. UHDRS total functional capacity score at baseline Enroll-HD visit. HDGECs represented (N = 16,071; Missing N = 49).

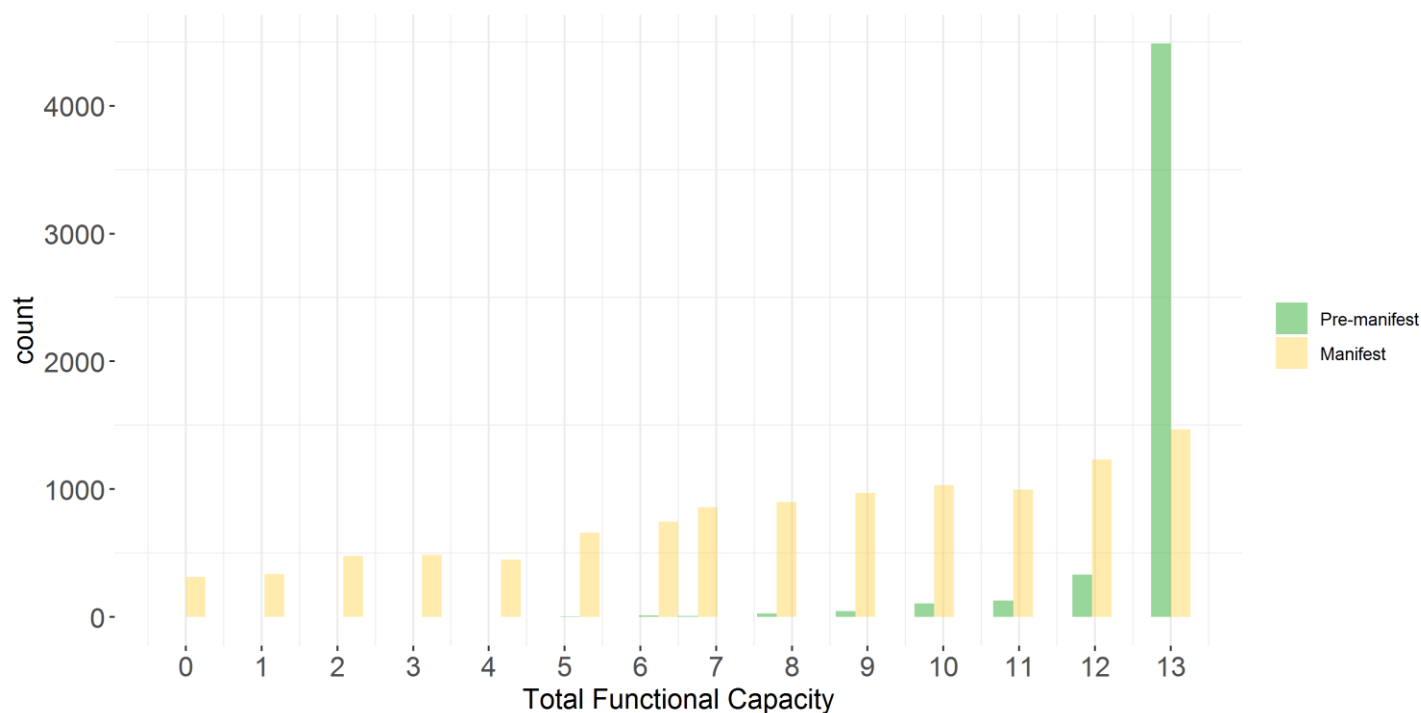


Figure 17. UHDRS functional assessment score at baseline Enroll-HD visit. HDGECs represented (N = 15,744; Missing N = 376).

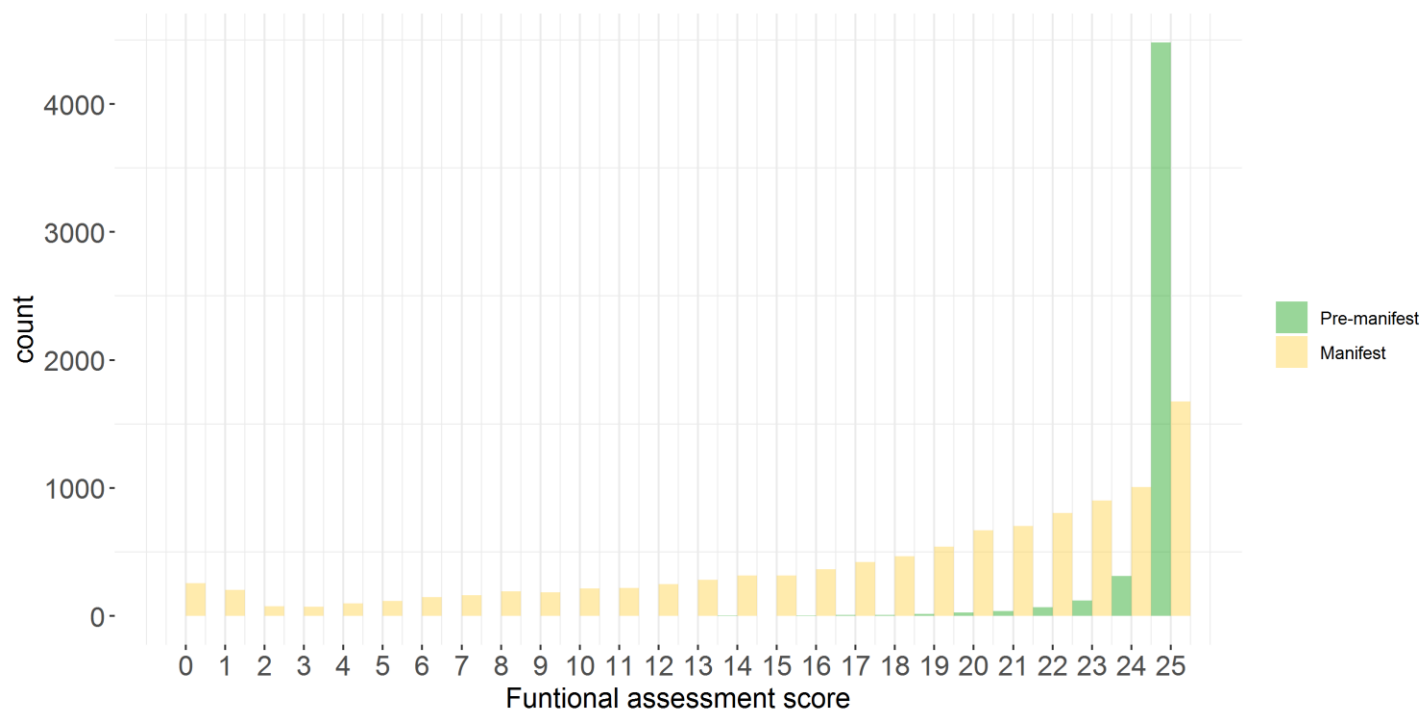
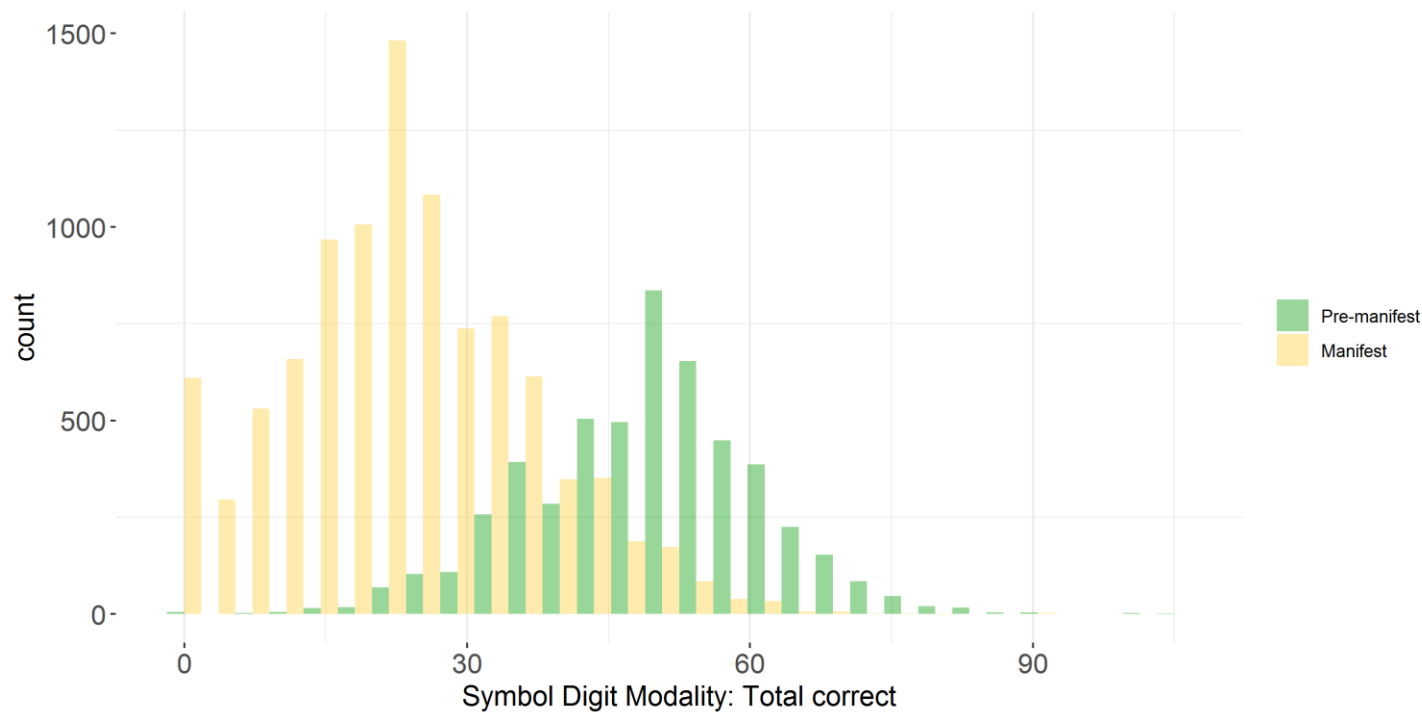




Figure 18. Symbol digit modality test score (total correct) at baseline Enroll-HD visit. HDGECs represented (N = 15,127; Missing N = 993).



5. Data availability and completeness

Figure 19. Completeness of core data elements in Enroll-HD as a function of percentage of total participants (N = 21,116).

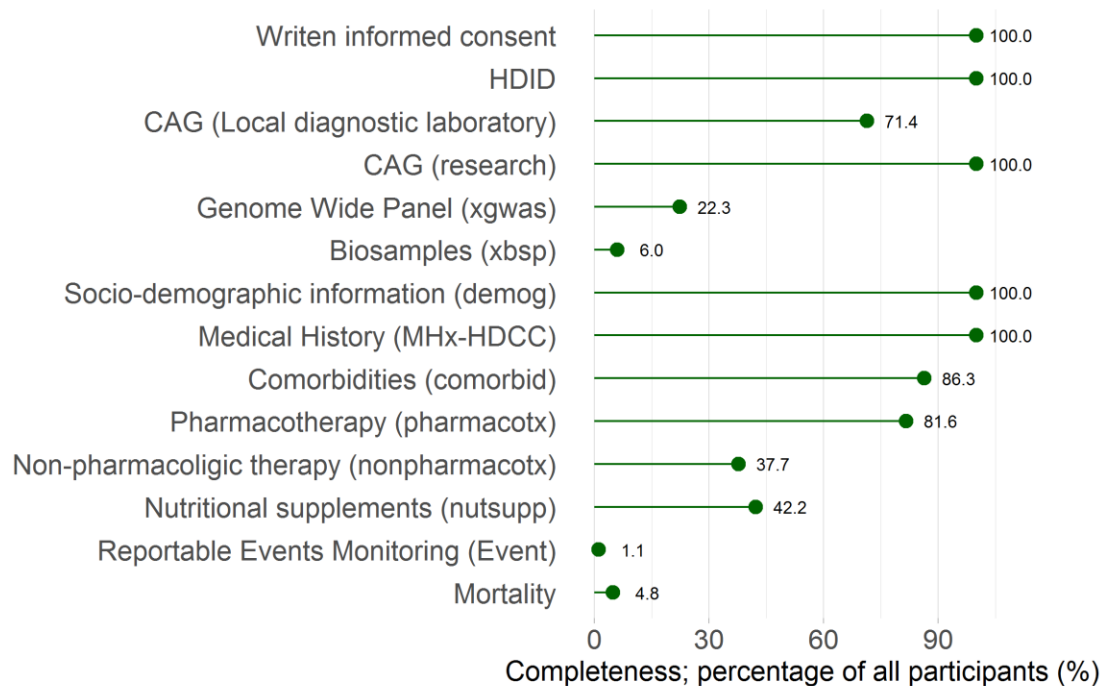
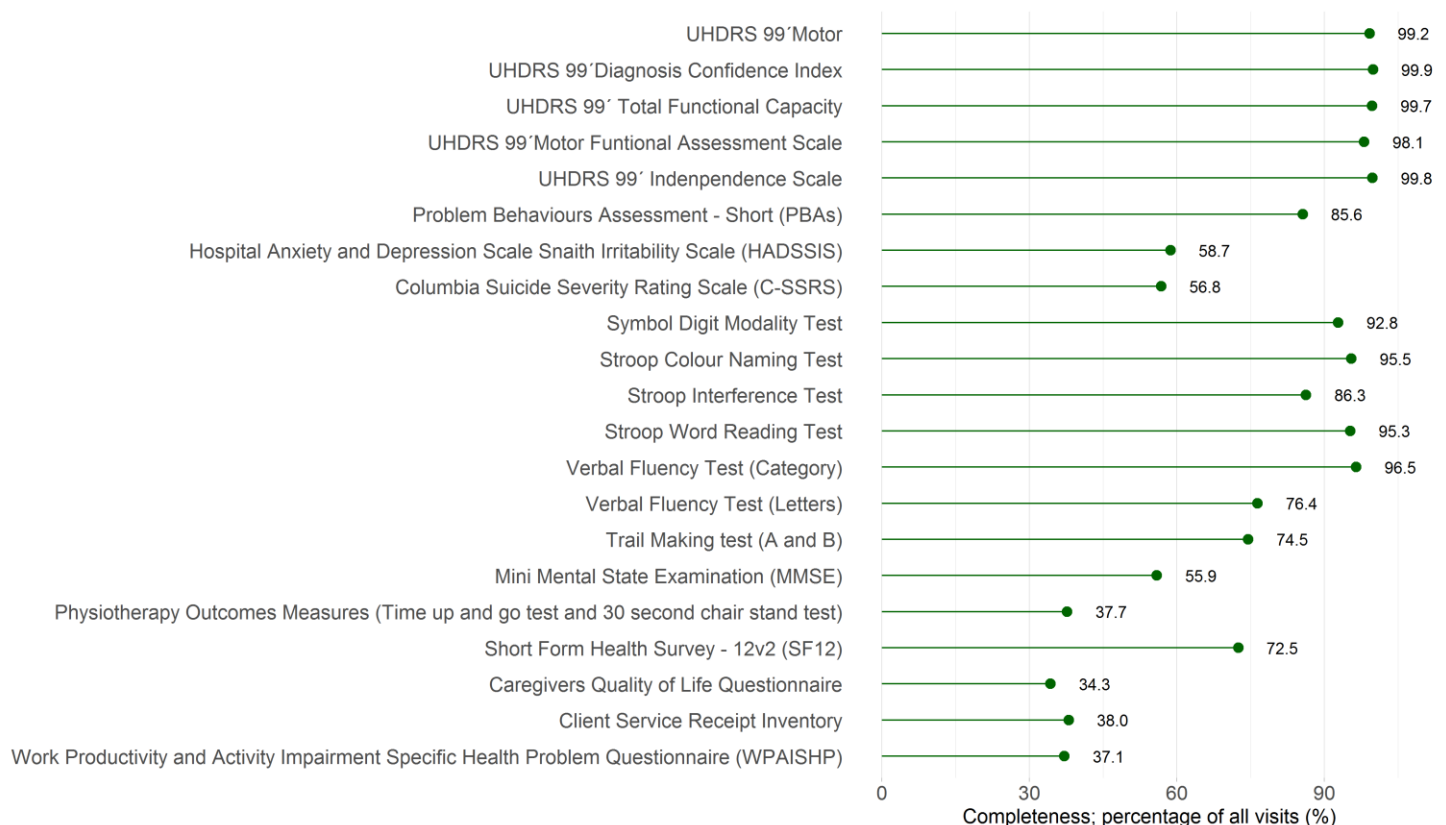


Figure 20. Completeness of assessments (core and extended) as a function of percentage of total Enroll-HD visits (visit N = 55,975). Completeness defined as completion of at least one item of the indicated assessment.



Coverage: data availability by region, participant category, visit count

Coverage charts

Availability of participant data by number of visits, region, and HD participant category, is provided in the coverage charts below (Tables 2 to 4).

Table 2: PDS5 coverage chart (cumulative; latest). Maximum number of participants available for X visits by region and participant category at latest visit. Visit counts consider only Enroll-HD visits (baseline and follow-up only, unscheduled visits and phone contacts excluded). Note that participants with more than 1 visit are represented in multiple columns.

Region	HD Category	Baseline Visit	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9
Australasia	Family control	77	53	31	23	19	10	0	0	0
	Genotype neg	82	65	50	21	16	7	0	0	0
	Manifest	333	276	208	155	112	61	10	0	0
	Premanifest	274	205	146	102	70	37	3	0	0
Europe	Family control	1145	774	493	267	89	19	0	0	0
	Genotype neg	1332	817	522	289	116	26	1	0	0
	Manifest	7912	5636	3696	2361	1191	353	17	0	0
	Premanifest	2711	1726	1062	637	282	60	5	0	0
Latin America	Family control	18	5	1	0	0	0	0	0	0
	Genotype neg	54	17	11	3	0	0	0	0	0
	Manifest	141	74	33	17	4	0	0	0	0
	Premanifest	32	12	5	1	0	0	0	0	0
North America	Family control	1117	803	575	382	249	149	72	13	1
	Genotype neg	1171	741	531	377	245	130	56	8	0
	Manifest	3183	2183	1509	1028	635	312	113	20	0
	Premanifest	1534	981	625	416	233	120	44	6	0
Total		21116	14368	9498	6079	3261	1284	321	47	1

Table 3. PDS5 coverage chart (absolute; latest). Absolute number of participants available for X visits by region and participant category at latest visit. Visit counts consider only Enroll-HD visits (baseline and follow-up only, unscheduled visits and phone contacts excluded). Note that in contrast to Table 2, each participant is represented in a single column only, indicative of their maximum visit count.

Region	HD Category	Baseline Visit	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Total
Australasia	Premanifest	69	59	44	32	33	34	3	0	0	274
	Manifest	57	68	53	43	51	51	10	0	0	333
	Genotype neg	17	15	29	5	9	7	0	0	0	82
	Family control	24	22	8	4	9	10	0	0	0	77
Europe	Premanifest	985	664	425	355	222	55	5	0	0	2711
	Manifest	2276	1940	1335	1170	838	336	17	0	0	7912
	Genotype neg	515	295	233	173	90	25	1	0	0	1332
	Family control	371	281	226	178	70	19	0	0	0	1145
Latin America	Premanifest	20	7	4	1	0	0	0	0	0	32
	Manifest	67	41	16	13	4	0	0	0	0	141
	Genotype neg	37	6	8	3	0	0	0	0	0	54
	Family control	13	4	1	0	0	0	0	0	0	18
North America	Premanifest	553	356	209	183	113	76	38	6	0	1534
	Manifest	1000	674	481	393	323	199	93	20	0	3183
	Genotype neg	430	210	154	132	115	74	48	8	0	1171
	Family control	314	228	193	133	100	77	59	12	1	1117
Total		6748	4870	3419	2818	1977	963	274	46	1	21116

Table 4. PDS5 coverage chart (absolute; baseline). Absolute number of participants available for X visits by region and participant category at baseline visit. Visit counts consider only Enroll-HD visits (baseline and follow-up only, unscheduled visits and phone contacts excluded). Note that in contrast to Table 2, each participant is represented in a single column only, indicative of their maximum visit count.

Region	HD Category	Baseline Visit	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Total
Australasia	Premanifest	69	62	52	38	35	50	3	0	0	309
	Manifest	57	65	45	37	49	35	10	0	0	298
	Genotype neg	17	15	29	5	9	7	0	0	0	82
	Family control	24	22	8	4	9	10	0	0	0	77

Europe	Premanifest	986	731	494	464	298	85	7	0	0	3065
	Manifest	2275	1873	1266	1061	762	306	15	0	0	7558
	Genotype neg	515	295	233	173	90	25	1	0	0	1332
	Family control	371	281	226	178	70	19	0	0	0	1145
Latin America	Premanifest	20	7	4	1	0	0	0	0	0	32
	Manifest	67	41	16	13	4	0	0	0	0	141
	Genotype neg	37	6	8	3	0	0	0	0	0	54
	Family control	13	4	1	0	0	0	0	0	0	18
North America	Premanifest	553	389	259	244	153	103	58	8	0	1767
	Manifest	1000	641	431	332	283	172	73	18	0	2950
	Genotype neg	430	210	154	132	115	74	48	8	0	1171
	Family control	314	228	193	133	100	77	59	12	1	1117
Total		6748	4870	3419	2818	1977	963	274	46	1	21116

Geographical coverage

Enroll-HD PDS5 data were collected from 171 clinical sites located across 20 countries (**Figure 21**).

Figure 21. Enroll-HD PDS5 map. Enroll-HD data in PDS5 were collected from clinical sites in 20 countries.

