



PDS6 Overview

Contents

- 1. Dataset..... 2
- 2. Sample Size..... 4
- 3. Visits 5
- 4. Sample Characteristics..... 8
- 5. Data availability and completeness 14
- 6. Coverage: data availability by region, participant category, visit count..... 16
- 7. Revision history..... 20

1. Dataset

Extraction

The data contained in Enroll-HD PDS6 was extracted from the Enroll-HD electronic data capture (EDC) database on **November 4, 2022**, at 10:00 UTC.

Data sources

The PDS6 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.

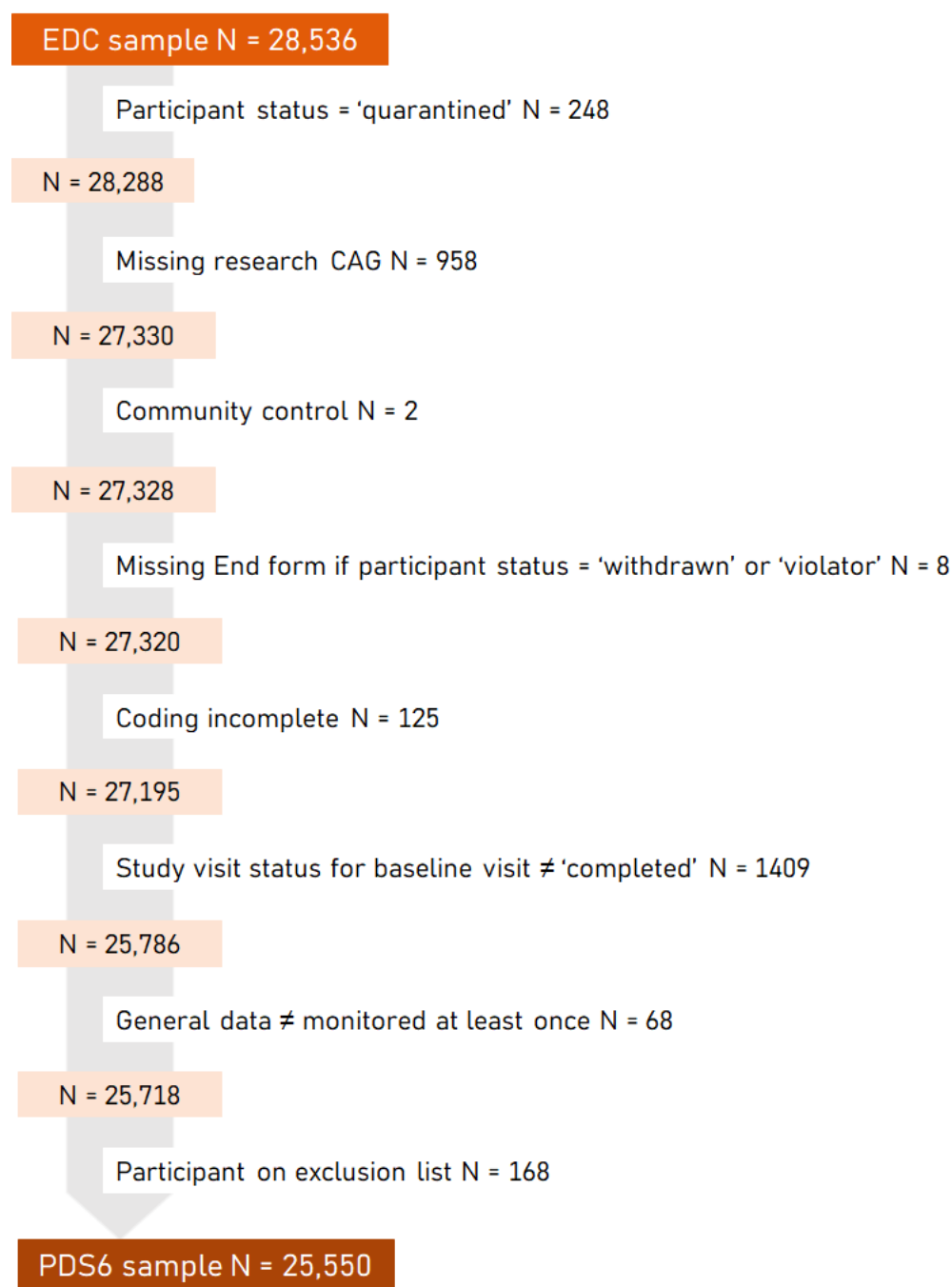
Clinical data from additional sources (**Ad Hoc** data) are available for a subset of Enroll-HD participants. These data were collected at routine clinical visits outside of the Enroll-HD and REGISTRY studies, and comprise HD assessment data (e.g., UHDRS Motor). The date of collection of these data typically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

Study specific **protocols**, **annotated eCRFs**, and **data collection guidelines** are housed on the [Documentation](#) page of the Enroll-HD website.

Participant inclusion

To be included in the PDS6 release, participant's data had to meet several requirements. **Figure 1** illustrates the number of participants whose data met each of the predefined inclusion requirements and illustrates how the final sample size of PDS6 was determined.

Figure 1. Participant flowchart for inclusion in PDS6.



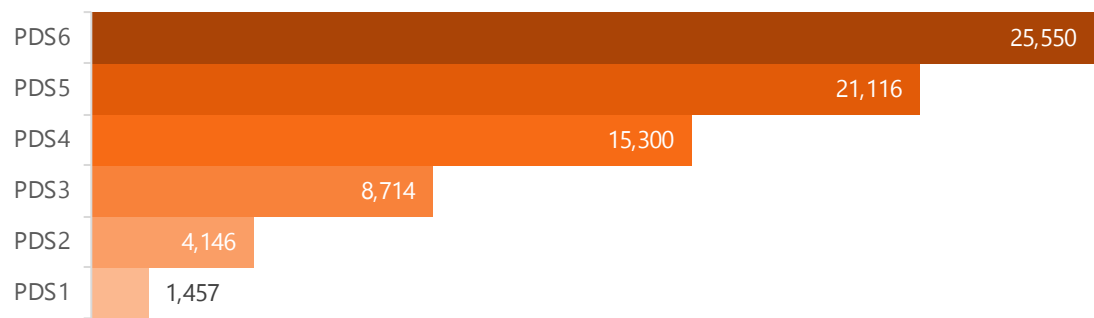
Due to data exclusion requirements, not all participants enrolled in Enroll-HD at the time of PDS6 data cut are included in PDS6. Similarly, not all participants included in PDS5 are included in PDS6. Data from 200 participants who were included in PDS5 were not included in the current PDS release. Enroll-HD 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 PDS5 participants not included in PDS6 may be available through specified dataset (SPS) request.



2. Sample Size

PDS6 contains data on 25,550 Enroll-HD participants. Sample size by PDS release is presented in **Figure 2**.

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





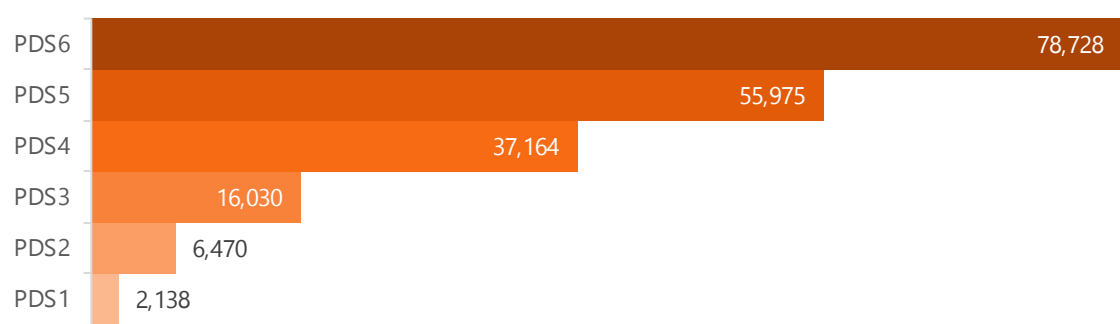
3. Visits

PDS6 contains data from 95,038 visits (baseline and follow-up visits only; all sources). Of these, 78,728 were Enroll-HD visits. The remainder are from Registry (N = 15,292) and ‘Ad Hoc’ sources (N = 1,018). 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 PDS6 by constituent data source. Participant number indicates the number of Enroll-HD participants with visit data available for the indicated data source (maximum N = 25,550).

Data source	Participants	Visits
Enroll-HD	25,550	78,728
Registry 3	4,337	10,114
Registry 2	2,153	5,178
Ad Hoc	316	1,018
TOTAL		95,038

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 PDS6 ranges from 1 to 10. 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 participants with 10 visits are represented in visit bars 1 through 10, the participants with 9 visits are represented in visit bars 1 through 9, 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 = 25,550; Missing N = 0).

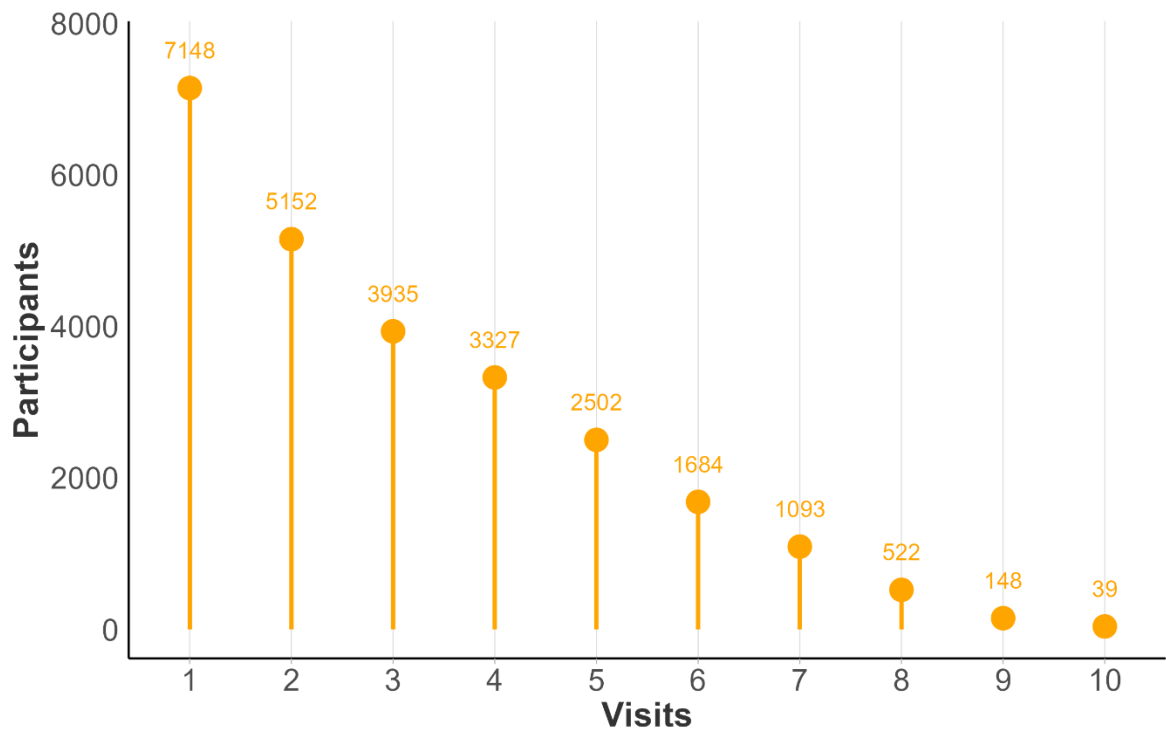
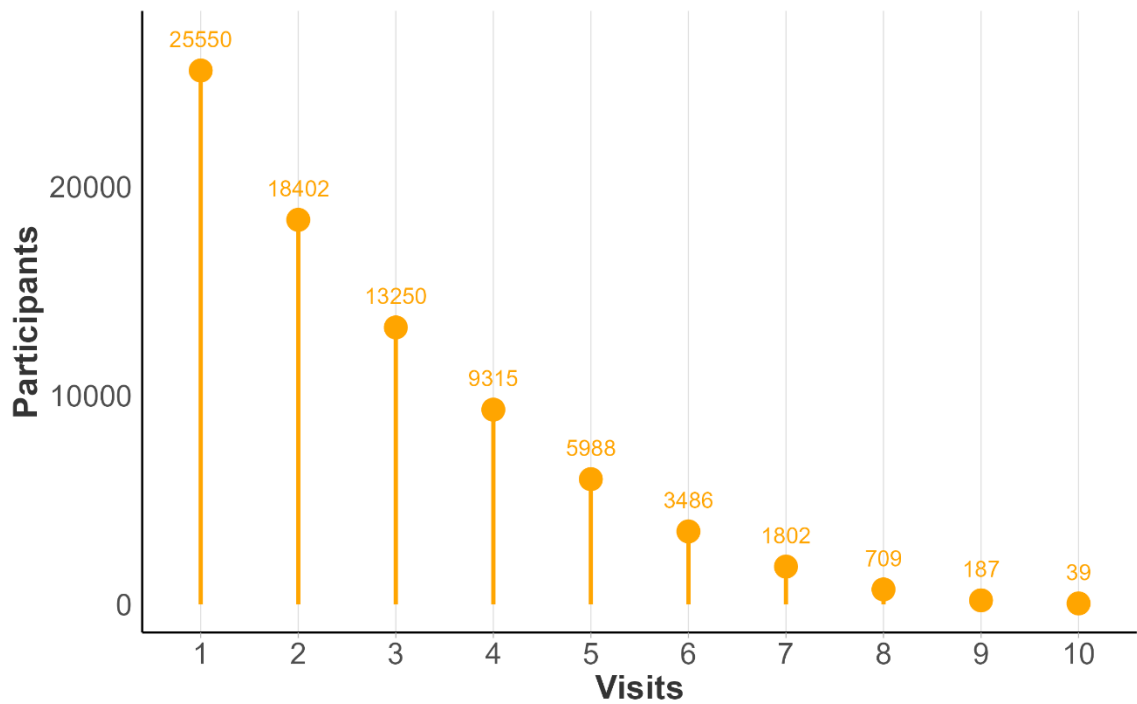
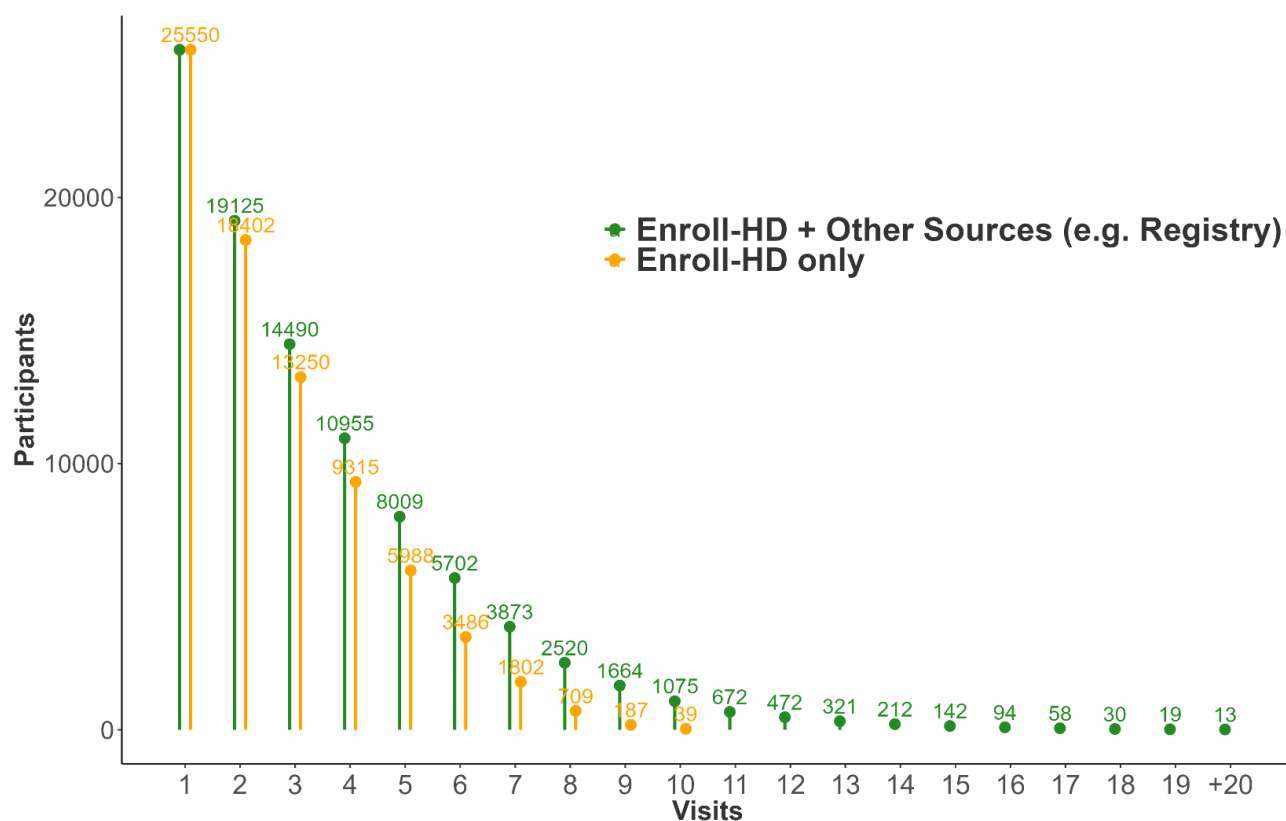


Figure 5. Maximum participant counts for a specific number of Enroll-HD visits (baseline and follow-up visits only; unscheduled visits and phone contacts excluded). Cumulative plot. Full sample represented (N = 25,550; Missing N = 0).



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 = 25,550; Missing N = 0).





4. Sample Characteristics

The PDS6 sample is characterized below with respect to participant category, sociodemographic variables, and clinical characteristics (Figures 7 to 19).

Figure 7. Participant category at baseline Enroll-HD visit (hdcatt_0). Full sample represented (N = 25,550; Missing N = 0)

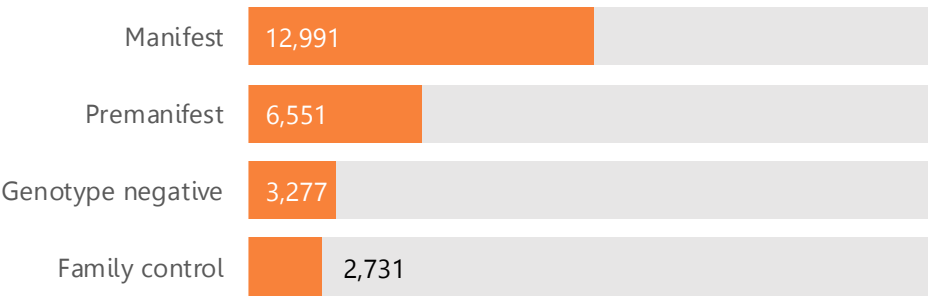


Figure 8. Participant category at latest Enroll-HD visit (hdcatt_1). Full sample represented (N = 25,550; Missing N = 0)

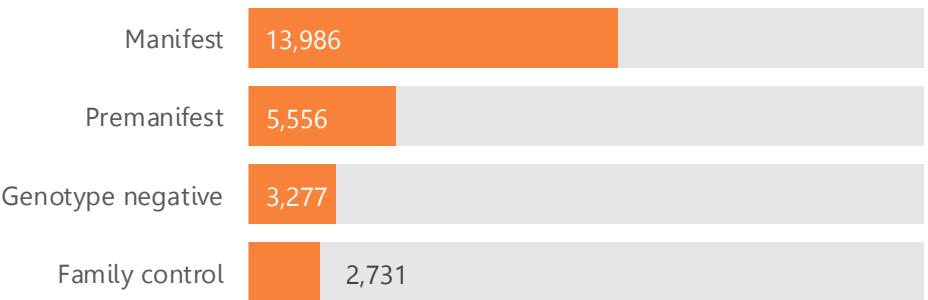


Figure 9. Geographical region (region). Full sample represented (N = 25,550; Missing N = 0)

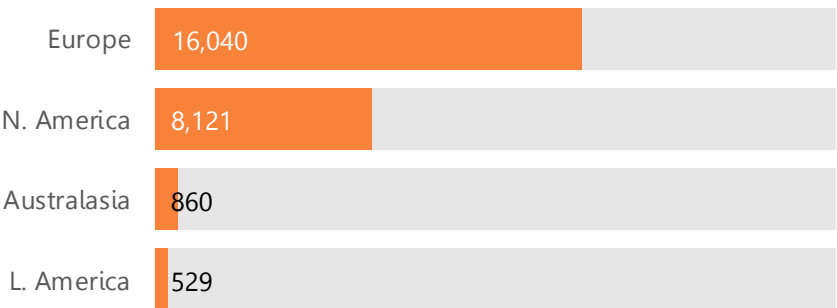


Figure 10. Sex (*sex*). Full sample represented (N = 25,550; Missing N = 0)



Figure 11. ISCED (*isced*) at baseline Enroll-HD visit. Full sample represented (N = 25,435; Missing N = 115)

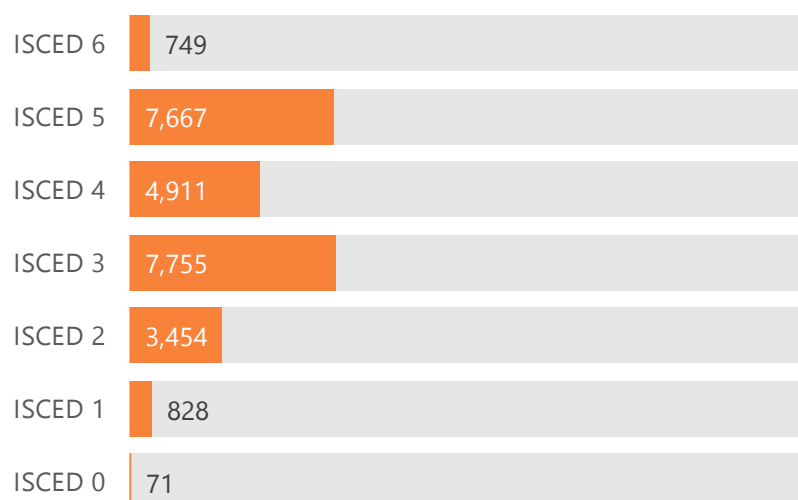


Figure 12. HD integrated staging system (HD-ISS) imputed stage (*hdiss_stage_imp*) at baseline Enroll-HD visit. Individuals with research CAG length of ≥ 40 represented only (N = 17,594).

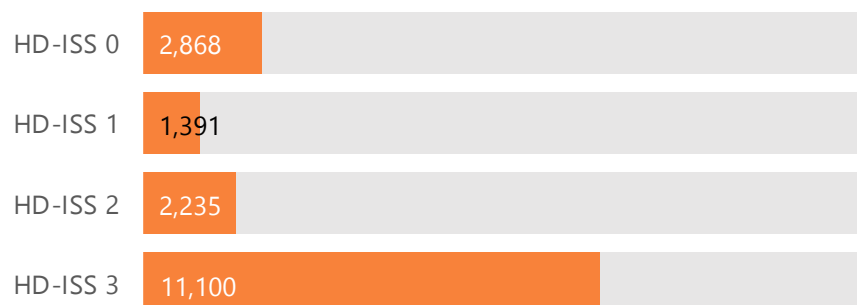




Figure 13. Age at baseline Enroll-HD visit (age_0). Full sample represented (N = 25,550; Missing N = 0). Note that in PDS6, age for individuals under the age of 18 years is represented by the value ‘<18’. These values have been transformed to 17 for inclusion in this histogram.

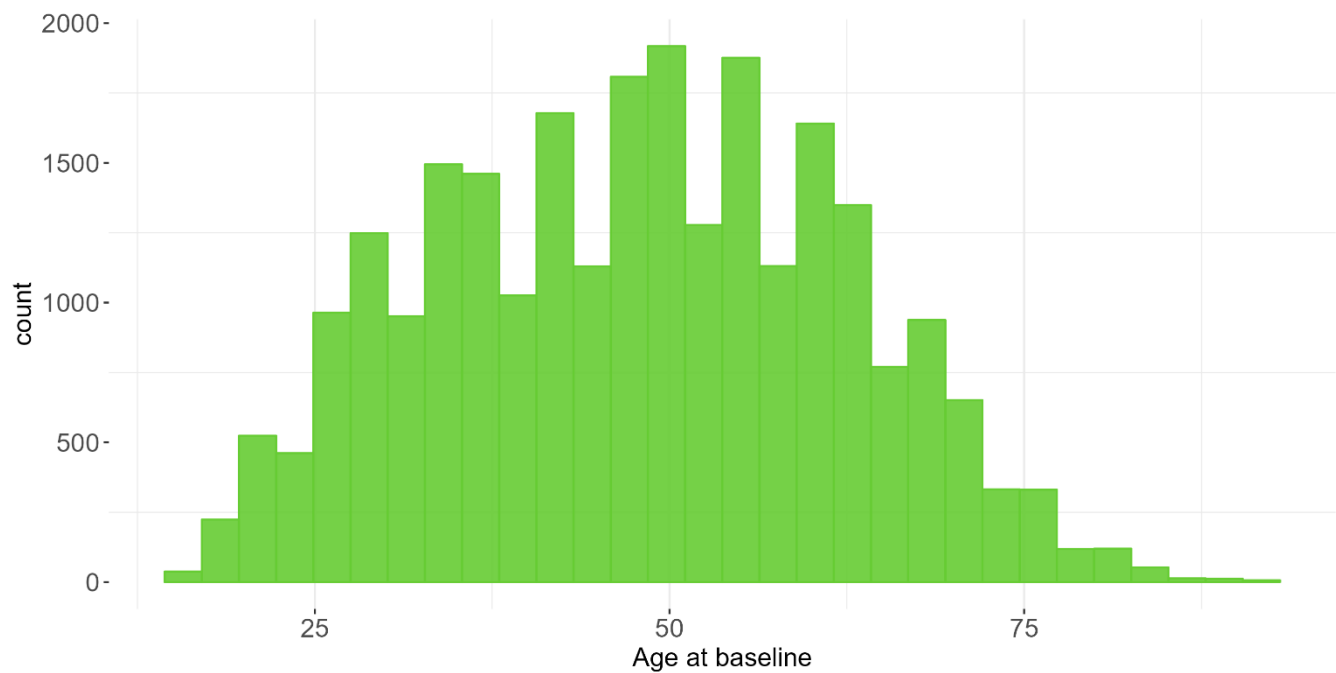


Figure 14. Research CAG length (caghigh). Full sample represented (N = 25,550; Missing N = 0). Note that in PDS6, CAG length for individuals with a CAG length greater than 70 is represented by the value ‘>70’; these values have been transformed to 71 for inclusion in this histogram. HDGEC = individual with CAG \geq 36; Non-HDGEC = individual with CAG < 36.

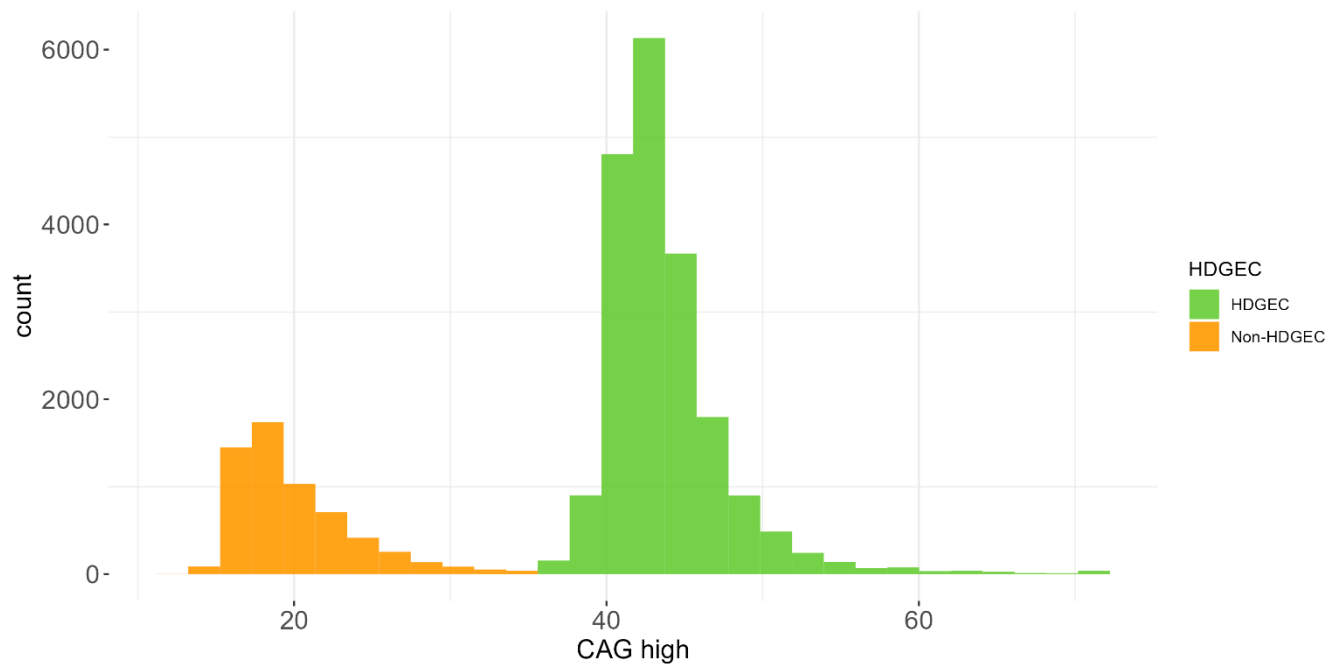


Figure 15. CAP score (capscore) at baseline Enroll-HD visit. Calculated using the CAP score formula in Warner et al.¹ Only individuals with a research CAG of ≥ 36 are represented. *Note:* a subset of these individuals with an aggregated value for age and/or research CAG length (N = 51) are *not* represented; these individuals have blank 'entries' for CAP score in the dataset.

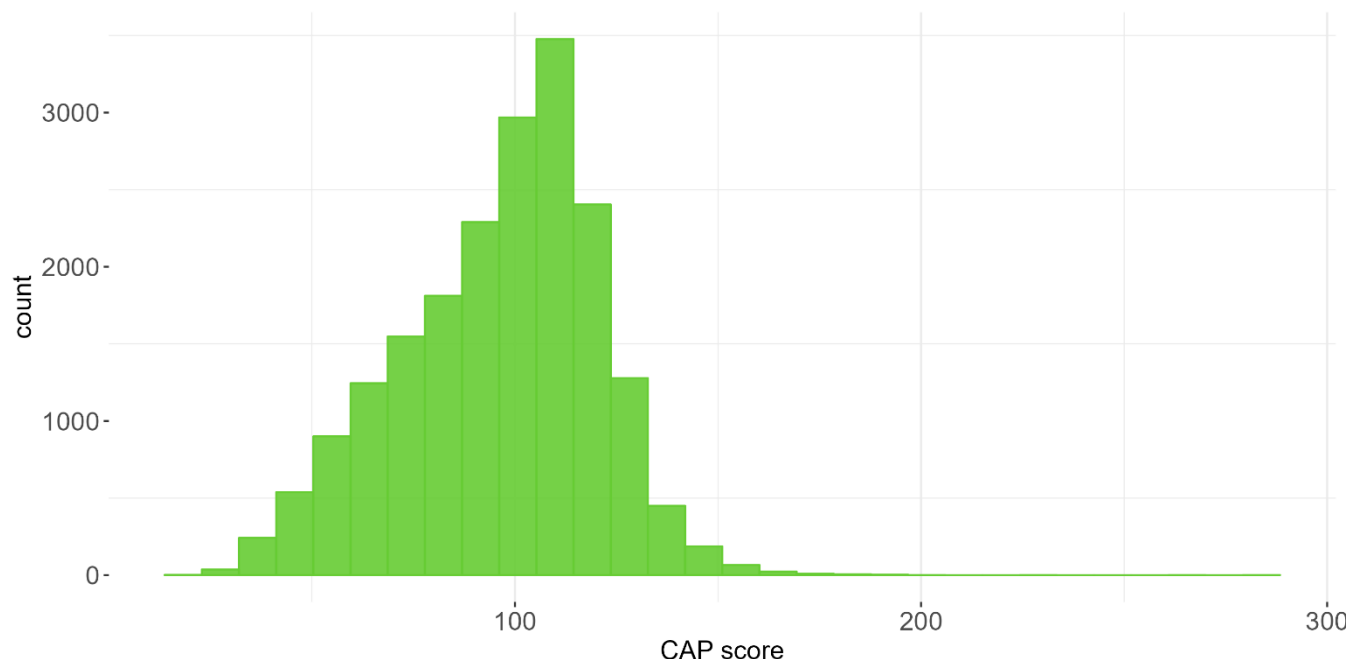
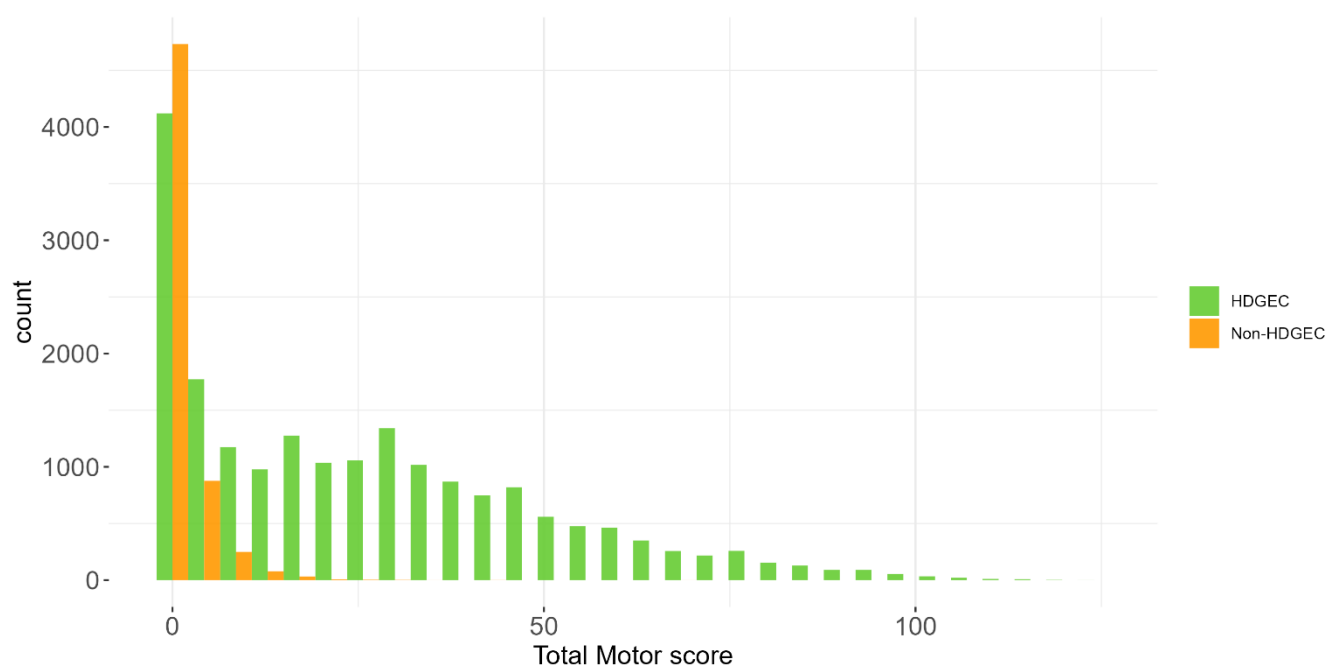


Figure 16. UHDRS total motor score (motscore) at baseline Enroll-HD visit. Full sample represented (N = 25,366; Missing N = 184). HDGEC = individual with CAG ≥ 36 ; Non-HDGEC = individual with CAG < 36.



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 17. UHDRS total functional capacity score (tfcscore) at baseline Enroll-HD visit. Full sample represented (N = 25,463; Missing N = 87). HDGEC = individual with CAG ≥ 36 ; Non-HDGEC = individual with CAG < 36 .

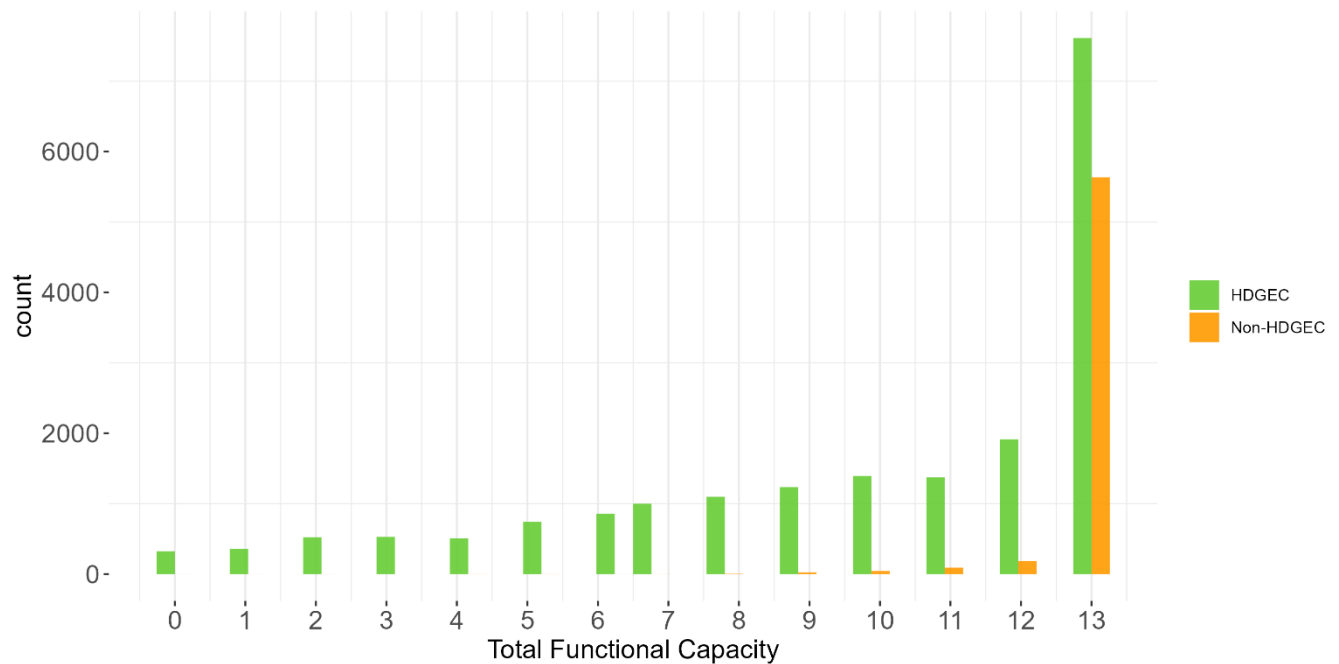


Figure 18. UHDRS functional assessment score (fascore) at baseline Enroll-HD visit. Full sample represented (N = 25,006; Missing N = 544). HDGEC = individual with CAG ≥ 36 ; Non-HDGEC = individual with CAG < 36 .

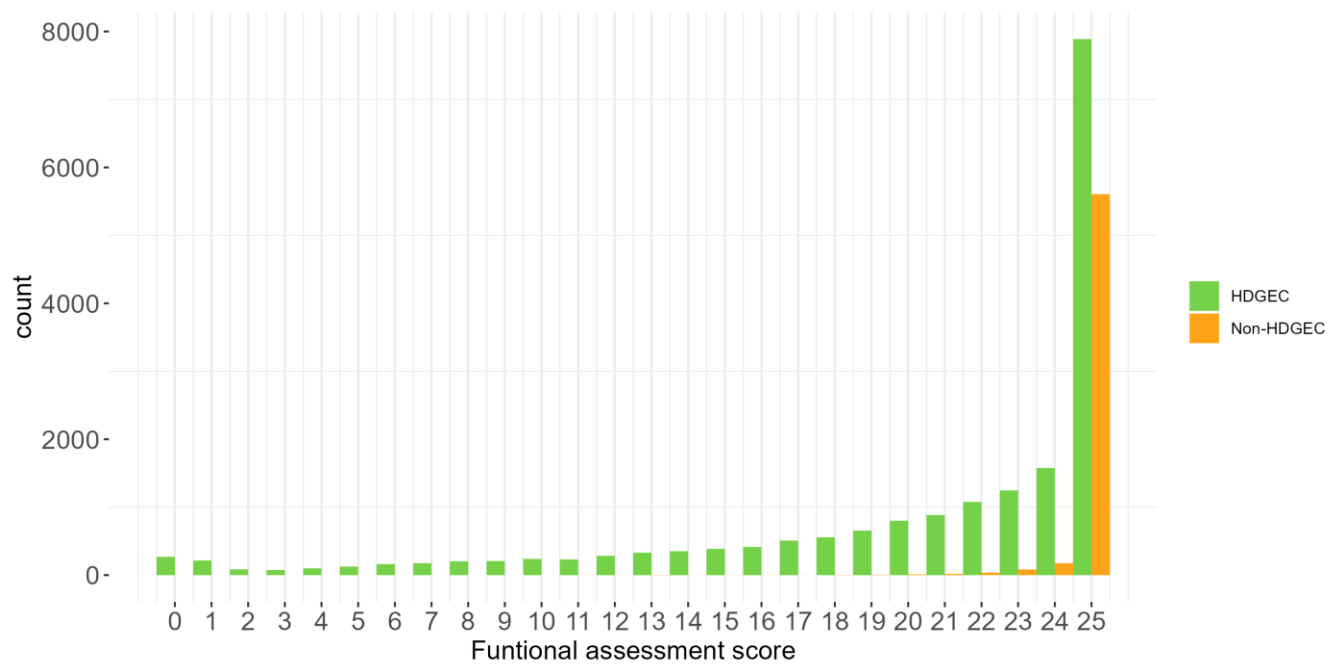
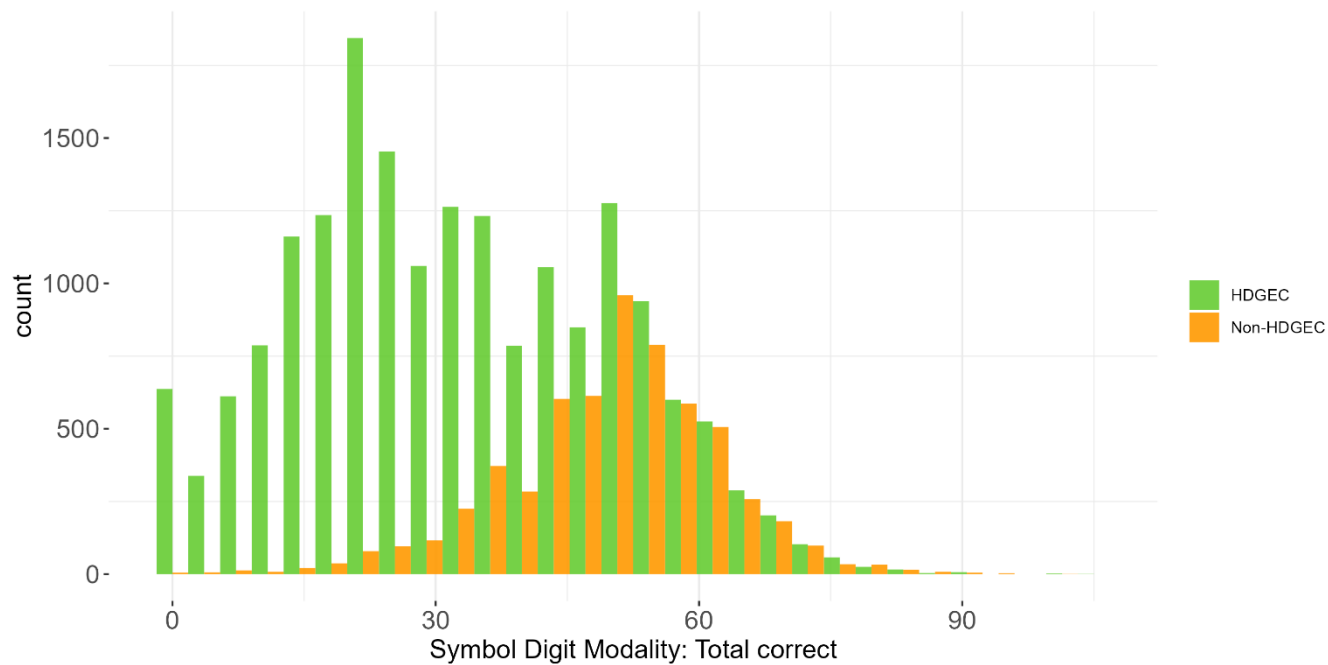




Figure 19. Symbol digit modality test score (total correct) (sdmt1) at baseline Enroll-HD visit. Full sample represented (N = 24,316; Missing N = 1,234). HDGEC = individual with CAG ≥ 36 ; Non-HDGEC = individual with CAG < 36.



5. Data availability and completeness

Figure 20. Completeness of data elements in Enroll-HD as a function of percentage of total *participants* (N = 25,550). Note that the completeness metric for ‘CAG (local)’ is 87% when limited to individuals with a CAG (research) of ≥ 36 .

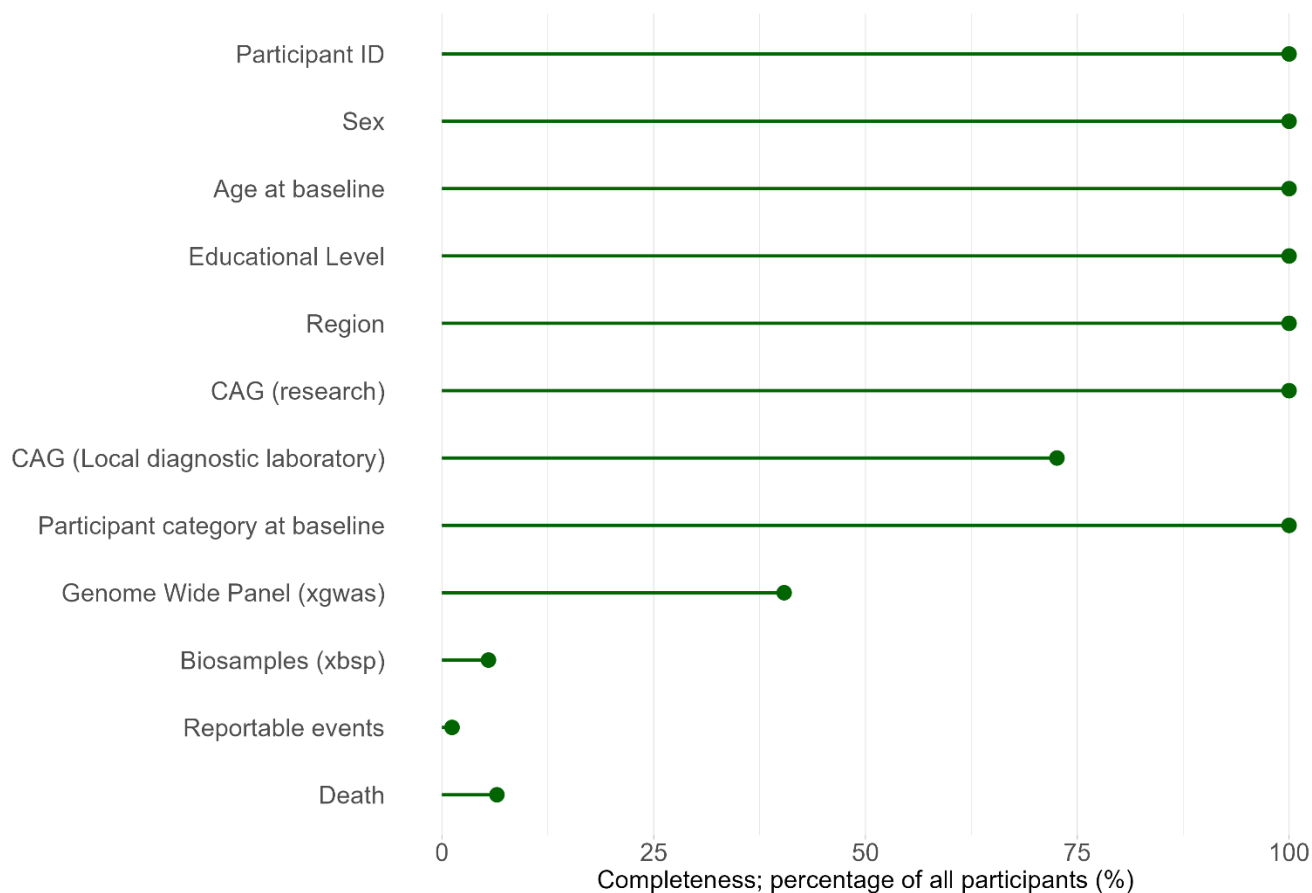
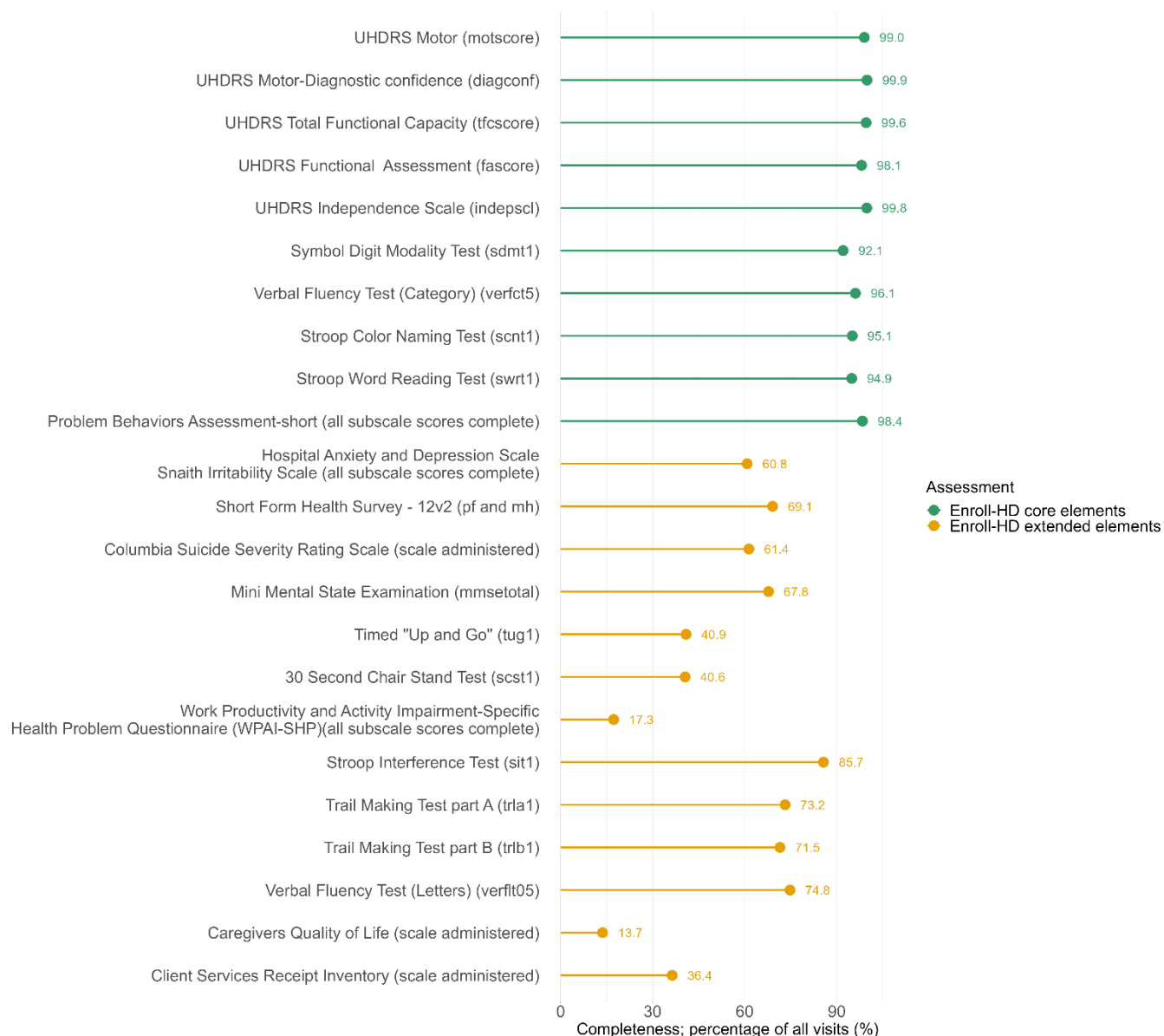


Figure 21. Completeness of assessments (core and extended) as a function of percentage of total Enroll-HD visits (visit N = 78,730). For assessments and scales with a key outcome variable(s), the completeness metric was operationalized as sufficiently completed such that the key outcome variable(s) is available for that visit. Key outcome variables are indicated in parentheses alongside each scale. For scales with no key outcome variable(s), i.e., CSSRS, Caregivers QoL, and CSRI, the completeness metric was ‘scale administered’ (operationalized as at least one variable field completed at that visit).





6. 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: PDS6 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. M = manifest; PM = pre-manifest; FC = family control; GN = genotype negative.

Region	HD category	Baseline visit	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10
Aus.	M	377	332	256	192	148	100	47	27	9	0
	PM	292	233	170	127	90	57	27	11	4	0
	FC	95	71	51	36	27	20	9	4	0	0
	GN	96	83	57	41	29	16	9	4	0	0
Europe	M	9633	7280	5381	3758	2349	1289	679	250	21	0
	PM	3345	2339	1622	1117	678	368	170	46	5	0
	FC	1407	978	705	492	285	153	59	9	4	0
	GN	1655	1086	740	520	315	184	83	18	0	0
Latin Am.	M	237	147	90	45	23	11	6	2	0	0
	PM	101	57	16	10	5	0	0	0	0	0
	FC	26	13	8	4	1	0	0	0	0	0
	GN	165	95	35	12	8	2	1	0	0	0
North Am.	M	3739	2705	1973	1410	947	591	311	144	51	14
	PM	1818	1208	813	563	368	221	123	56	23	6
	FC	1203	904	701	520	365	241	141	69	40	14
	GN	1361	871	632	468	350	233	137	69	30	5
Total		25550	18402	13250	9315	5988	3486	1802	709	187	39



Table 3. PDS6 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. M = manifest; PM = pre-manifest; FC = family control; GN = genotype negative.

Region	HD category	Baseline visit	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10
Aus.	M	45	76	64	44	48	53	20	18	9	0
	PM	59	63	43	37	33	30	16	7	4	0
	FC	24	20	15	9	7	11	5	4	0	0
	GN	13	26	16	12	13	7	5	4	0	0
Europe	M	2353	1899	1623	1409	1060	610	429	229	21	0
	PM	1006	717	505	439	310	198	124	41	5	0
	FC	429	273	213	207	132	94	50	5	4	0
	GN	569	346	220	205	131	101	65	18	0	0
Latin Am.	M	90	57	45	22	12	5	4	2	0	0
	PM	44	41	6	5	5	0	0	0	0	0
	FC	13	5	4	3	1	0	0	0	0	0
	GN	70	60	23	4	6	1	1	0	0	0
North Am.	M	1034	732	563	463	356	280	167	93	37	14
	PM	610	395	250	195	147	98	67	33	17	6
	FC	299	203	181	155	124	100	72	29	26	14
	GN	490	239	164	118	117	96	68	39	25	5
Total		7148	5152	3935	3327	2502	1684	1093	522	148	39



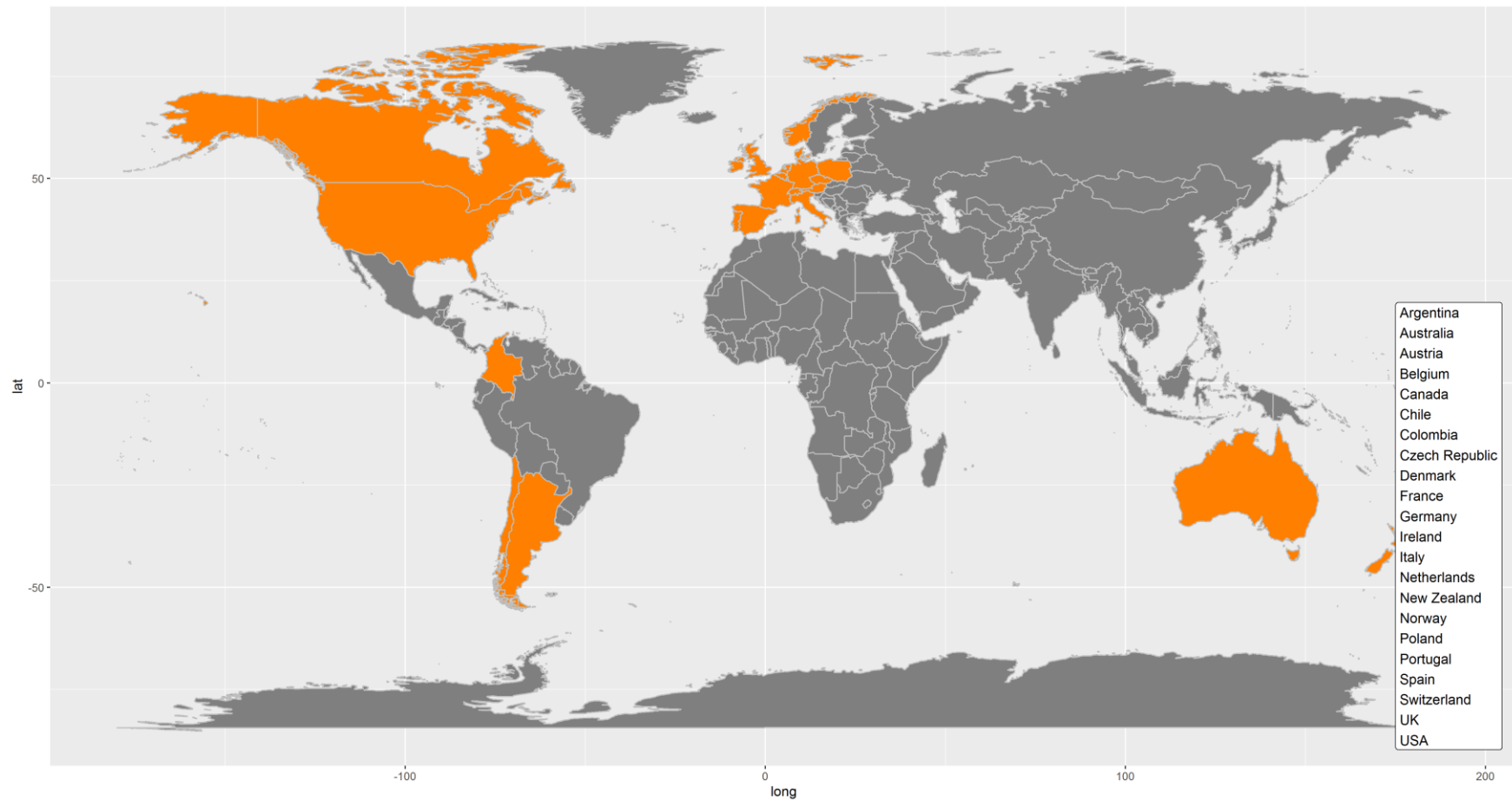
Table 4. PDS6 coverage chart (absolute; baseline). Absolute number of participants available for X visits by region and participant category at baseline visit. Visit counts consider only Enroll-HD visits (baseline and follow-up only, unscheduled visits and phone contacts excluded). Note that in contrast to Table 2, each participant is represented in a single column only, indicative of their maximum visit count. M = manifest; PM = pre-manifest; FC = family control; GN = genotype negative.

Region	HD category	Baseline visit	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10
Aus.	M	45	73	55	39	43	38	18	14	9	0
	PM	59	66	52	42	38	45	18	11	4	0
	FC	24	20	15	9	7	11	5	4	0	0
	GN	13	26	16	12	13	7	5	4	0	0
Europe	M	2352	1835	1518	1300	933	524	362	189	19	0
	PM	1007	781	610	548	437	284	191	81	7	0
	FC	429	273	213	207	132	94	50	5	4	0
	GN	569	346	220	205	131	101	65	18	0	0
Latin Am.	M	90	58	43	21	11	5	4	2	0	0
	PM	44	40	8	6	6	0	0	0	0	0
	FC	13	5	4	3	1	0	0	0	0	0
	GN	70	60	23	4	6	1	1	0	0	0
North Am.	M	1033	686	525	382	306	234	137	57	22	9
	PM	611	441	288	276	197	144	97	69	32	11
	FC	299	203	181	155	124	100	72	29	26	14
	GN	490	239	164	118	117	96	68	39	25	5

Geographical coverage

Enroll-HD PDS6 data were collected from 179 clinical sites located across 22 countries (Figure 21).

Figure 22. *Enroll-HD PDS6 map.* Enroll-HD data in PDS6 were collected from clinical sites in 22 countries.



7. Revision history

PDS6 overview R1 > PDS6 overview R2:

Text, tables, and plots updated to reflect changes implemented for PDS6 R2. Note that visits were removed for two family control participants. For further information on PDS changes, please refer to the ***Change Log***.