PDS4 | V2018-10-R1

Periodic Dataset

Data Handling Manual

ENROLL-HD

A worldwide observational study for Huntington's disease families

A CHDI Foundation Project

CONTENTS

1. PURPOSE OF DOCUMENT

This document provides guidance on use of the Enroll-HD Periodic Dataset (PDS), and describes how data were compiled for the current release. This document has been updated to accompany the Enroll-HD **PDS4** release.

2. DATASET CONSIDERATIONS

2.1 Data Policy

Enroll-HD periodic datasets are made available to verified researchers who have applied for and received an Enroll-HD Clinical Data and Biosamples Access Account on the www.Enroll-HD.org website.

The use of the Enroll-HD periodic dataset is subject to the terms and conditions set forth in the Data Use Agreement.

2.2 PDS4 Overview

The data contained in the current Enroll-HD PDS, **PDS4,** was extracted from the 'live' Enroll-HD database (EDC) on **October 31, 2018**.

All individuals in Enroll-HD PDS4 are Enroll-HD participants. Like PDS3, PDS4 includes data gathered not only from the **Enroll-HD** study, but also integrates data from **REGISTRY** (REGISTRY 2 / R2, and REGISTRY 3 / R3), as well as **Ad Hoc** data. Ad Hoc data, drawn from a variety of different sources, principally comprises UHDRS data, typically gathered prior to a participant's enrollment into Enroll-HD.

To be included in the PDS4 release, **participant data had to meet a number of requirements**. As such, not all participants enrolled in Enroll-HD at the time of PDS4 data cut are included in PDS4. The figure below (**Figure 1**) illustrates the number of participants whose data met each of the

predefined inclusion requirements, and illustrates how the final sample size of PDS4 was determined.

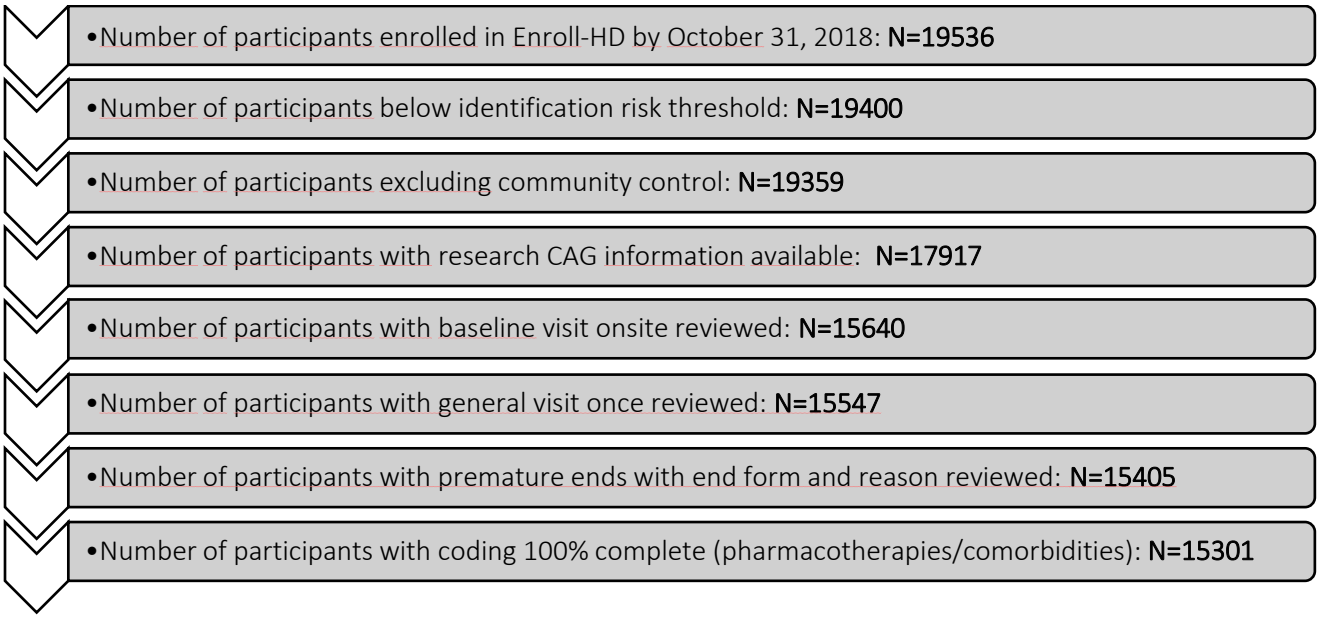

Figure 1 – Participant inclusion filters for PDS4

As mentioned, Enroll-HD PDS4 includes visit data from Enroll-HD as well as other studies. The figure below (**Figure 2**) is a frequency plot illustrating number of participants as a function of number of visits by data source. Column color coding indicates source(s) of visit data. **Green** columns indicate number of participants with data from at least X visits collected under the **Enroll-HD study only**. **Orange** columns indicate number of participants with data from at least X visits considering **all available data sources** (i.e., Enroll-HD, REGISTRY 3, REGISTRY 2 and Ad Hoc). P**articipants with data from more than 1 visit will be represented in multiple columns**. For example, a participant with data from 3 Enroll-HD visits would be represented in both green and orange columns displayed for 1, 2 and 3 visits.

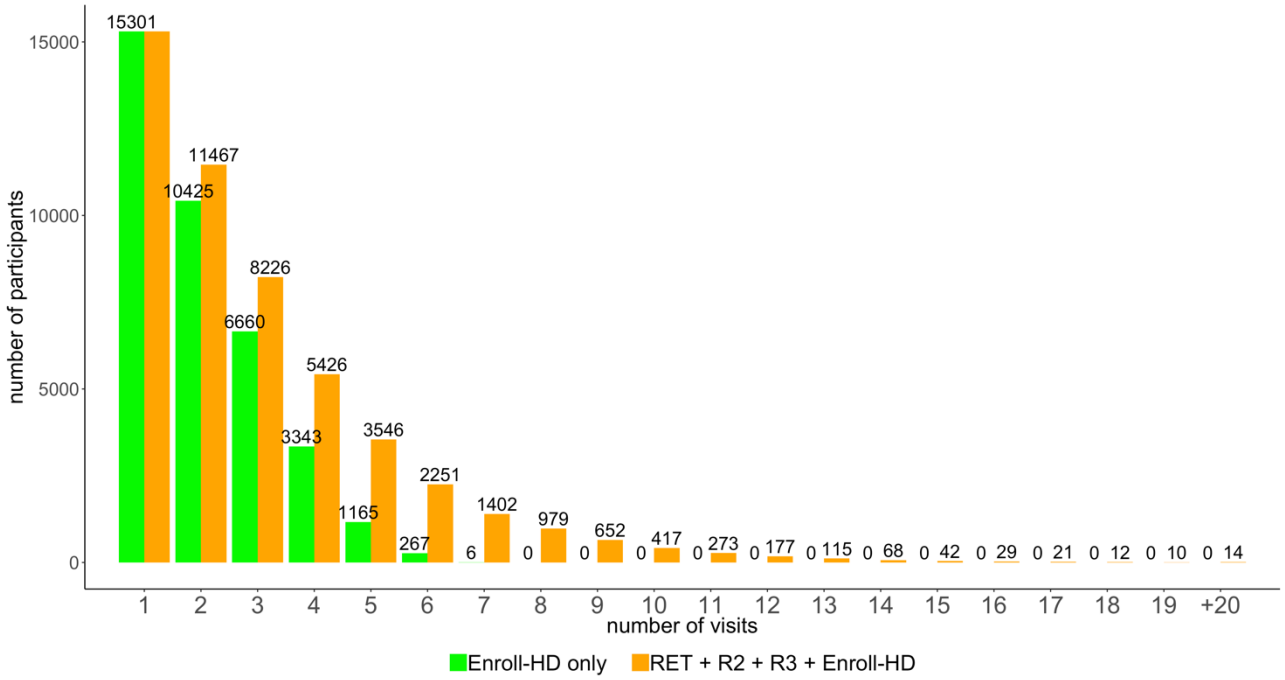

Figure 2 – Frequency plot of number of participants by number of study visits

There are a total of 50,452 visits included in the PDS4 from four data sources from a total of 15,301 independent participants (Table 1).

Table 1 - Number of participants and visits (baselines and follow-ups) for each study included in PDS4

Study	Enrollments	Visits
Enroll-HD	15301	37167
Ad Hoc	258	809
REGISTRY3	3528	7933
REGISTRY2	1827	4543
Total		50452

Data from 244 participants who were included in PDS3 were *not* included in the current PDSrelease. Enroll-HD is an active, longitudinal study. A participant eligible for inclusion one year may be ineligible the next for reasons described above (e.g., data not coded, end visit not reviewed). Data for these 244 participants, as provided in *PDS3*, are available through special request (see Section 3).

2.3 Data Quality Control and De-Identification

Prior to publishing, the Enroll-HD PDS goes through **Quality Control** (QC) and **de-identification** procedures. These are described below.

Quality Control

Data quality control checks are implemented at multiple levels, from **point of data entry**, through to **onsite** and **remote** data monitoring. Prior to a PDS release, data are also subject to an enriched, unique set of remote data review checks. All of these checks aim to maximize data integrity.

Under **remote monitoring** procedures, participant's data (core assessment and selected extended assessments) are subject to monthly cross-sectional QC checks, which include checks for consistency, completeness and plausibility. Participant data are also subject to longitudinal (i.e., within subject) QC checks for a subset of variables (e.g., height, TFC score). See Table 3 for an example.

Table 3 - Example of outlier detected through longitudinal participant QC check

subjid	*studyid*	*visit*	*Age*	*height*
R073515909	R2	Baseline	53	132
R073515909	R2	Follow Up	54	178
R073515909	R2	Follow Up	55	177
R073515909	R2	Follow Up	56	179
R073515909	R2	Follow Up	57	179
R073515909	ENR	Baseline	60	178
R073515909	ENR	Follow Up	61	178

Prior to a PDS release, an enriched set of remote data QC checks are also performed. These include comprehensive checks for **outliers** and bespoke checks for **unusual values**.

Outlying and implausible values are reviewed by the monitoring and/or medical monitoring teams, and queried directly with sites where relevant. In certain instances these values cannot be queried (e.g., observation recorded under Registry protocol – see Table 3 example), or are

queried and confirmed as correct by site staff. In instances such as these, values are provided 'as is', and it is left to the analyst to determine how best to evaluate the data.

Outlying values and unusual entries are detailed in the Unusual Findings document.

While substantial efforts are made to maximize data quality, **researchers are encouraged to visualize the data and perform their own QC checks prior to commencing analyses**.

De-Identification

During the de-identification process, the dataset is assessed to determine the risk of participant identification for each participant based on a pre-specified set of variables. A participant identification risk threshold is used to assess whether a participants' data should be included in the dataset. For genotype unknown participants, the risk threshold is 1%. For all other participants, the threshold is 3%.

For certain variables, an **aggregation approach** has been used to reduce risk of identification and maximize inclusion of participant data (see section 2.4). If an individual falls outside of the acceptable identification threshold after the completing the de-identification process (including data aggregation), that participant is removed from the dataset. These participants may be included in future PDS releases should identification risk fall to an acceptable threshold.

2.4 Aggregated Values

As part of the de-identification process, **data aggregation techniques** were applied to specific variables. The variables, and criteria/thresholds for aggregation, are described in the table below.

Table 4 - Aggregated values in the dataset

Data file	Variable	Variable Label	Criteria for aggregation
Participation	'*age*'	Age at enrollment	All participants with values <18

Enroll	'*age*'	Age at visit	All participants with values <18
Profile	'*caghigh*'	Research larger CAG allele determined from DNA	All participants with values > 70
Profile	'*caglow*'	Research smaller CAG allele determined from DNA	All participants with values > 28
Profile	'*race*'	Ethnicity	Number of cases per ethnicity*

* Seven categories for ethnicity are represented in the dataset: Caucasian (1); American – Black (2); Hispanic or Latino (3); Other (6); American Indian/ Native American/ Amerindian (8); West Asian and East Asian (13 and 14); Mixed (15).

The following are aggregated into "Other (6)": Native Hawaiian or Other Pacific Islander (4), Alaska Native/Inuit (5), South African (11), North African (12).

The number of participants impacted by the data aggregation thresholds described above are listed below:

Table 5 – Number of participants impacted by data aggregation threshold for age and CAG.

Variable	Label	Number of Participants
'age'	<18	31
'caghigh'	>70	28
'caglow'	>28	208

Table 6 - Number of participants impacted by data aggregation threshold for ethnicity.

Ethnicity	Label	Number of Participants
Other*	6	232

* Includes individuals from the following categories: Native Hawaiian or Other Pacific Islander (4), Alaska Native/Inuit (5), South African (11), North African (12)

Please note that **numerical variables (e.g. age, CAG) with aggregated data have been converted to text variables.** In order to convert these variables to a numeric form, **cells that contain >/< values should be replaced with a numeric value.** Mean, median, mode, maximum or minimum values could be used as a replacement value.

Descriptive statistics for the aggregated variables, determined *prior* to aggregation, are available by specified dataset (SPS) request.

Note that aggregation thresholds differ between Enroll-HD PDS releases. Changes in number/type/given values for participants make aggregation adjustments necessary.

Due to reasons relating to de-identification risk, data collected during visits *prior* to Enroll-HD, (i.e., REGISTRY and/or Ad Hoc visits) have been omitted for participants aged <18 years. This information can be requested through a specified dataset (SPS) request.

2.5 Unexpected zero values for dosage / heaviness of use variables

Data on drug, pharmacotherapy, and nutritional supplement use and/or periodic dosage are included in the Enroll-HD PDS. These variables are often **derived** from raw measures of dose and frequency of use. For example, the variable '*packy',* indicative of an individual's cumulative smoking history in terms of pack/years, is derived from daily intake and history of use variables. If one of these raw input values is missing, or is an extremely low value, the derived value may be zero (in the latter case due to rounding). This can be misleading. **We recognize this as an issue, which will be addressed in future PDS releases.** Additionally, if dose is unknown, a zero value may be entered. This is frequently the case for combined drugs/nutritional supplements. A zero value may also be entered for dose if the drug is sporadically used. Finally, dose units are sometimes are set to "WRONG" when the participant has not mentioned the units of the supplement he is taking. In these cases, total dose is not provided and a code for missing data is used instead (see Missing Data section).

The raw data values used to calculate dose / heaviness of use values can be requested through SPS request. To request a specified dataset, send an email to AccountSetup@Enroll-HD.org.

2.6 Dataset File Formats

The Enroll-HD PDS4 is provided in **R** and **tab-delimited CSV** file format.

The CSV file format is readable by Excel and most statistical software packages including R, Stata, and SAS. In many cases, **while importing the data into the analysis software, it is essential to specify that the variables are separated by tabs** as other separation conventions are also commonly used. It is also important that these files are not edited in word processing or other programs that may potentially modify the tab characters as this may damage the integrity of the original files. If needed, CSV files can be saved in other formats that are compatible with other statistical software packages.

The procedure for importing PDS CSV files into Excel is outlined in the Reference Guide "Importing Enroll-HD PDS Files".

2.7 Datasets

Enroll-HD PDS4 contains 11 data files. The structure of the data files is provided in Table 7 and Figure 3 below.

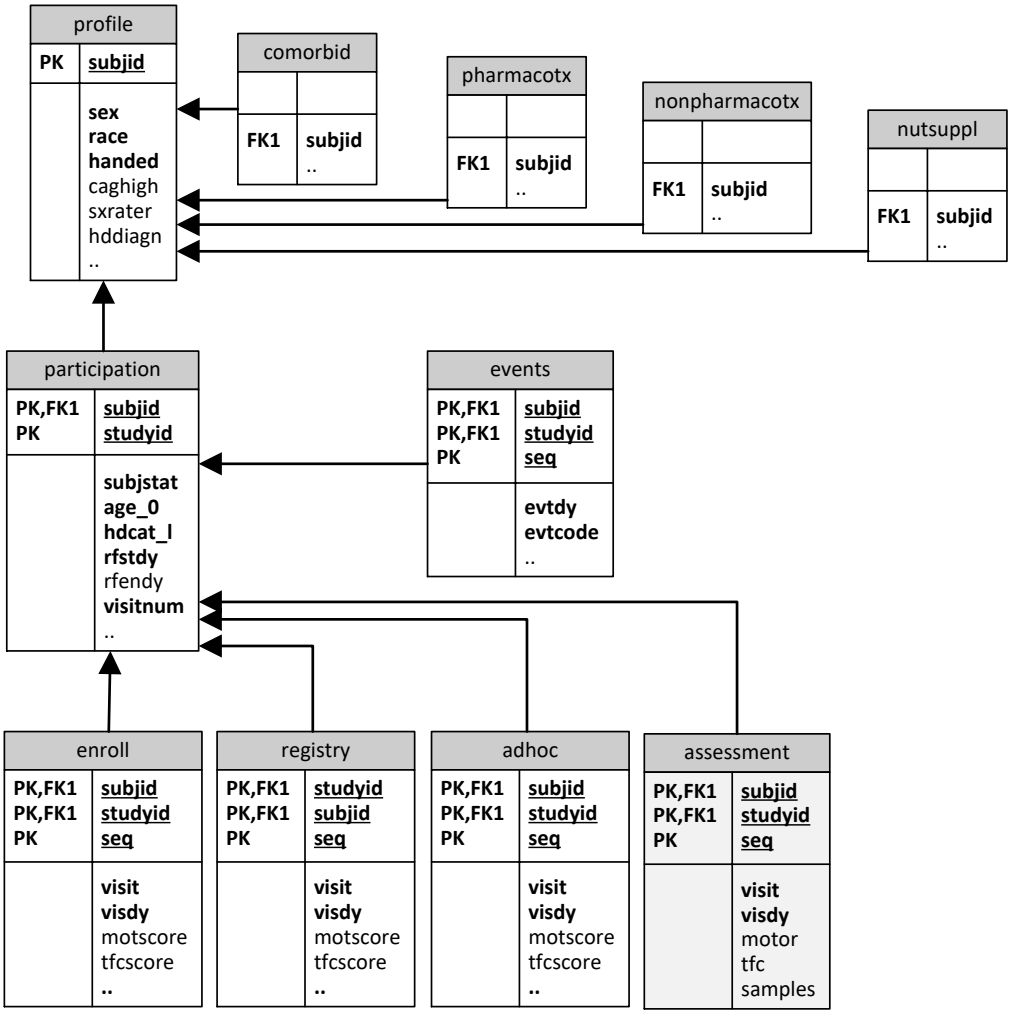

Figure 3 – Entity Relationship Diagram

Variable labels are not included in the PDS files – please consult the Data Dictionary for this information.

The Enroll-HD PDS files are either subject-based, meaning that the data contained in the file describes participant information that is not specific to a visit (e.g. demographic information), or visit-based, meaning that the file contains information regarding baseline or follow-up visits. Study-based data files such as "participation", "assessment" and "events" contain study specific information about a participant within a study. The data file "events" is a special data file for Enroll-HD containing all the reportable events of a participant.

Table 7 - Data files included in PDS4

Files	Type	Studies	Description
Profile	Participant	ENR R3 R2 RET	General and annually updated information. Includes the following forms: Demog, HDCC, CAG, Mortality.
PharmacoTx	Participant	ENR R3 R2 RET	Form: Pharmacotherapy information.
NutSuppl	Participant	ENR R3 R2 RET	Form: Nutritional Supplements information.
NonPharmacoTx	Participant	ENR R3 R2 RET	Form: Non-Pharmacologic therapies information.
Comorbid	Participant	ENR R3 R2 RET	Form: Comorbid conditions information.
Participation	Study	ENR R3 R2 RET	Study specific information about each participant Start, end, reasons, visits, etc. are stored in a general format.
Assessment	Study	ENR R3 R2 RET	Study specific information gathered during onsite and remote visits Provides an overview about number, type and sequence of each assessment perfomed for each study.
Event	Study	ENR	Reportable Event Monitoring that happened during Enroll-HD.

Enroll-HD	Visit	ENR	Enroll-HD file contains information from the Enroll-HD study. This file contains information on Baseline, Follow-up, unscheduled visits and phone contacts.
REGISTRY	Visit	R3 R2	REGISTRY file contains information from REGISTRY 2 and REGISTRY 3 studies. This file contains information on Baseline, Follow-up, and unscheduled visits.
Ad Hoc	Visit	RET	Ad Hoc file contains information colleted during a clincial visit. All this information wascolleted pior to enrollment into any of the studies. All visits available for this study are included, visits can be ordered by using the 'seq' variable which represents the sequence of the visits.

'ENR'= Enroll-HD; 'R2'= REGISTRY 2; 'R3' = REGISTRY 3; 'RET' = Ad Hoc visits.

See annotated CRF for description of each form.

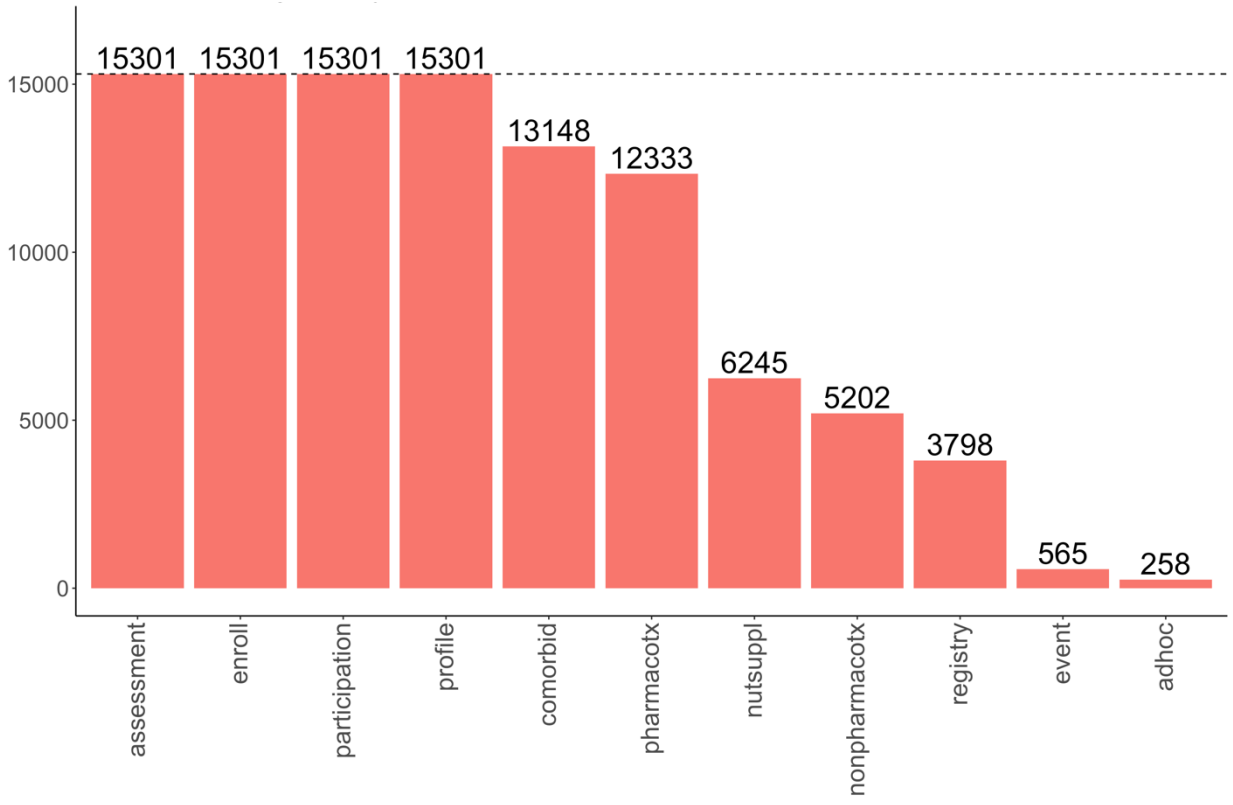

Figure 4 - Number of participants included in each file

2.8 Merging and Aligning Participant Data Across Files

The Enroll-HD PDS4 dataset contains one **key variable**, ***subjid***, and it is included in every data file. This allows the user to merge two or more data files, linking information for each participant across data files. The key variable *subjid*, labeled as *HDID (recoded)*, is obtained by recoding the HDID. Note that the HDID is a unique participant identifier used across multiple HD studies. HDIDs are not included in any PDS release.

To merge longitudinal data available in visit-based data files, it is important to take in consideration the variable ***seq***, as this variable provides information on the visit sequence. Visit day (*visdy*) is also available for sequencing visit data temporally.

WARNING: Merging data files in Excel can cause misalignment. Before analyzing the data, check that the resulting merged data file correctly lines up across appropriate fields. To avoid issues with merging data files, it is highly recommended that you use a reputable statistical software package.

Below we provide guidance on selecting entries/lines using Excel or R, respectively. The exemplar task comprises merging age of HD diagnosis from the "profile" file to "age" at the last visit of each participant in the "enroll" file.

EXCEL: Select the *seq* variable in the "enroll" file using that variable will allow you to select the latest visit for the participant. Then once you select the age for the latest visit of each participant you can easily, using a [vlookup] function, merge the variable age of HD diagnosis in the profile file.

R: Select the highest value in the *seq* variable, then use the [merge] function to merge different files.

2.9 Missing Data

As typical of many studies, data are missing in Enroll-HD. Data may be missing for a number of different reasons.

An entry must be made for all user defined/applicable fields in the Enroll-HD EDC. Where a value is not known or a value known to be wrong is entered, a tool may be used to mark the field or field entry as unknown, not applicable or wrong. These so-called 'exceptional values' are defined as user-defined missing values.

1. Unknown

For some yes/no questions where 'unknown' was a very common response, a third response option, 'Unknown', has been added to facilitate data entry (e.g., Mother affected? yes/no/unknown). In the dataset these values are presented as 9999.

2. Missing value

Refers to mandatory values for which data collection was not performed. This can occur for a number of different reasons, such as the value is not known, the participant refuses to share the data, or the collection of data (as opposed to entry of data) has been omitted accidentally. In the dataset, these missing data are marked as 9998.

3. Not applicable

Refers to empty values present for fields/questions that could not be answered by the participant because they do not apply due to certain circumstances or characteristics (e.g., age of onset of the affected parent if the parent is still pre-symptomatic). In the dataset, the not applicable data is marked as 9997.

4. Wrong

Refers to values which are entered into the EDC even though they are known to be wrong or highly questionable by the data entry person. This may be because they were collected under unreliable conditions (e.g. cognitive tests, started but aborted), by the wrong person (e.g. assessment done by an untrained site member), or the values are questionable due to faulty instrumentation (e.g. the weight of a participant measured on a scale that is later found to be inaccurate). It means that a value is entered in the live EDC but should not be used for analysis.

Note, **wrong values are never exported into the PDS dataset**. In the data set the wrong values are marked as **9996**.

As described above, within PDS4, these user-defined missing values are represented by unique values depending on category of 'missingness' and variable type (i.e., numeric, text/string, date) and illustrated below:

Table 8 – User-defined missing value table

User missing value	Unknown	Missing	Not applicable	Wrong
Numeric values	9999	9998	9997	9996
Text values	UNKNOWN	MISSING	NOTAPPL	WRONG
Date values	----	9998-09-09	9997-09-09	9996-09-09

Transformation of missing data entries should be considered depending on the software package you are using. Note that inclusion of text values for a variable will result in such variables being be read as text variables in most statistical software packages.

2.10 Transformation of Date Variables

As part of the de-identification process, all variables that represent dates in the Enroll-HD dataset were changed from date format to numeric. The numeric variable represents the number of days after the **baseline visit date of Enroll-HD study**.

As dates have been transformed to numeric values relative to the date of the baseline visit, date values can be **negative numbers**. This is typical for start dates of medications and comorbid conditions.

Since many date variables do not require an exact date to be entered (e.g. "YYYY-MM-DD"), **rules were required to establish a numeric value for incomplete date entries**. The following rules were applied to incomplete date fields:

- Value entered as "YYYY-MM," change to "YYYY-MM-15";
- Value entered as "YYYY," change to "YYYY-07-01".

Examples for an enrollment date of 11/01/2014:

Table 9 - Example of date transformation

Entered Date	Value Completed Date	Representation Dataset	Precision in Dataset
11/01/14	11/01/2014	0	Day
11/12/14	11/12/2014	11	Day
11/14	11/15/2014	15	Month
2014	07/01/2014	-123	Year

Note that because of this rule, events with clear temporal definition sometimes appear out of order, for example, end dates appearing prior to start dates for entries on '*PharmacoTx*' and '*Comorbid*' forms.

2.11 Ordering of Visit Data:

The Enroll-HD PDS4 data files 'enroll', 'registry' and 'adhoc' contain data for all baseline and follow-up visits, for each participant, for each study. There is not a separate file for each follow-up visit.

The files 'enroll', 'registry' and 'adhoc' include a variable called *seq*. This variable refers to the **sequence of the visits** and will enable the data analyst to order visits temporally. The *seq* value is in accordance with number of days after the baseline visit (*visdy*), where *seq*=1 refers to the baseline visit, *seq* =2 refers to the 1st follow up visit, *seq*=3 to the 2nd follow up visit, and so on, including unscheduled and phone contact visits.

Table 10 - Visit sequencing example

subjid	*studyid*	*visit*	*seq*	*visdy*
R001084542	ENR	Baseline	1	0

R001084542	R3	Baseline	1	-728
R001084542	R3	Follow up	2	-363

Phone contact visits only occur in the 'enroll' file. These visits contain missed visit information, reason for missed follow-up visit and participant's availability to continue the study. If these data are not required for your analyses, these visits can be filtered out.

Unscheduled visits occur in the 'enroll' and 'registry' file. These visits take contains all the information as a regular follow-up visit and are executed when the participant presents to the clinic before the define time window. If these data are not required for your analyses, these visits can be filtered out.

2.12 HD Classification and Disease Severity Variables

HD classification

The Enroll-HD dataset contains a variable *hdcat* which reports the HD category of each participant at each visit as classified by the site staff. Participants of the group "genotype unknown" are reclassified for the purpose of this data set release.

HDcat categories are:

- **Premanifest/premotor-manifest HD ('hdcat'=2)**:
- **Manifest HD patient/motor-manifest HD ('hdcat'=3)**:
- **Genotype negative ('hdcat'=4):**

The baseline and most recent categorization of the participant, based on the last follow-up evaluation, are available in the data file '*participation*'.

The variable *hdcat* refers to the **baseline evaluation** and the variable *hdcat_l*, which is included in the '*participation*' data file, refers to the **most recent** HD category information for that participant.

Please note that the *hdcat* variable is available for Enroll-HD and REGISTRY 3, but it is not available for REGISTRY 2 and Ad Hoc since these two studies did not use an HD classification system.

Note that some participants classified as HD-manifest in Enroll-HD (i.e., *hdcat*=3) have high values for Total Motor Score (> 10). In these instances, the manifest categorization may be due to psychiatric or cognitive symptom onset as opposed to motor.

Disease severity:

CAP score, derived from CAG and age, can be used to determine disease severity. CAP score can be calculated at each time point.

CAP score is not included in PDS4 but can be calculated as follows;

CAP score = Age * (CAG – L)/K, where L=30 and K=6.49 [1] [Variable '*cap_score*']

2.13 Distinguish between Enroll only participants from those coming from REGISTRY

In order to distinguish Enroll-HD participants who migrated from the REGISTRY study from those who did not, thus Enroll only participants, there is several ways to select these participants, herein we aim to propose a simple way to achieve this.

Using Excel the function [vlookup] to identify participants that are included in the 'registry' file too, after identify these participants you can easily remove them from the analysis.

3. SPECIFIED DATASET REQUESTS

Specified dataset (SPS) requests can be made to obtain data that is not provided in the current PDS. **SPS requests are subject to approval by the Enroll-HD Scientific Publication Review Committee (SPRC).**

[1] John Warner (2017)

To use data collected from your own site, researchers must submit an SPS request. However researchers are encouraged to use PDS alternatively.

To request a specified dataset, send an email to AccountSetup@Enroll-HD.org.

3.1 HD Category SPS Request

Enroll-HD classifies participants into six categories, *hdcat*: Genotype Negative, Pre-manifest, Manifest, Genotype Unknown, Family Controls, and Community Controls. The Enroll-HD PDS includes data from Family Controls but does *not* include data from Community Controls.

Approximately 10 percent of the participants included in the Enroll-HD EDC are categorized as Genotype Unknown, meaning neither the participant nor their physician knows the genetic status of the participant. However, in the PDS all genotype unknown are re-categorized into the appropriate category as follows:

- **Genotype Negative**: research genotype larger CAG allele < 36;
- **Pre-manifest**: research genotype larger CAG allele ≥ 36 and classification (in the enrollment form) not set to manifest. Alternatively, the classification can be based on motor signs ('certainty') not set to 4 (Diagnostic Confidence Level from the UHDRS);
- **Manifest**: research genotype larger CAG allele ≥ 36 and classification (in the enrollment form) set to manifest. Alternatively, the classification can be based on motor signs ('certainty') set to 4 (Diagnostic Confidence Level from the UHDRS).

Data from participants in the Genotype Unknown group may be obtained by special request, subject to approval.

To request a specified dataset, send an email to AccountSetup@Enroll-HD.org

3.2 Aggregated Data Specified Request

Enroll-HD PDS4 contains aggregated data (see Section 2.4) that can be de-aggregated and obtained by special request, subject to approval.

To request a specified dataset, send an email to AccountSetup@Enroll-HD.org

3.3 Precision Information Specified Request

Dates: PDS4 does not contain dates. Dates have been transformed to decrease risk of participant identification (see Section 2.10). Since these transformations used an automatic rule for missing day and month, an additional variable containing precision information (d, m, and y) can be requested. The precision variable identifies the level of date completeness:

d – for a complete date (precision "days")

m – if day information is missing (precision "months")

y – if day and month information is missing (precision "years")

To request a specified dataset, send an email to AccountSetup@Enroll-HD.org

4. ADDITIONAL CONSIDERATIONS

4.1 Therapies and Comorbidities Coding

In Enroll-HD PDS4, the data files Pharmacotherapy ('PharmacoTx'), Non-Pharmacologic Therapies ('NonPharmacoTx'), Nutritional Supplements ('NutSuppl'), and Comorbid Conditions ('Comorbid') contain a variable labeled 'Ongoing' (coded into 1 'YES' and 0 'NO'). **This value is set at 1 'YES, ongoing' for all conditions and therapies that do not have a stop date.**

The 'pharmacotx' file contains a variable for the drug name that is coded with an internal Enroll-HD drug code (Rx000000001).

More information on therapy and comorbidity coding systems included in the 'PharmacoTx', 'NonPharmacoTx', 'NutSuppl' and 'Comorbid' files can be found in the Enroll-HD PDS Reference Guide "**Coding Systems: Medications, Comorbidities, Occupation**".

Comorbidities are separated into conditions, coded using ICD-10, and procedures coded with an internal Enroll-HD code (e.g., Cx000000001).

The data files 'pharmacotx', 'onpharmacotx', 'nutsuppl' and 'comorbid' may include some duplicate entries. These entries are recorded intentionally and are likely the result of two or more exact medications or comorbidities that share the same start and end date.

4.2 Unified Huntington's Disease Rating Scale (UHDRS) score calculation

Enroll-HD PDS4 contains calculated composite UHDRS scores. Please refer to the following reference for further information:

Huntington Study Group. Unified Huntington's Disease Rating Scale: Reliability and Consistency. Neuropsychiatry Movement Disorders 1996, Vol. II, No. 2, 136-142.

Table 11 - UHDRS sub-score calculation

UHDRS Section	Variable	Sub-Score Calculation
UHDRS Motor	*motscore*	Sum the value of scores
UHDRS Total Functional Capacity	*tfcscore*	Sum the value of scores
UHDRS Functional Assessment	*fascore*	Sum the value of scores

4.3 Problem Behaviors Assessment – Short (PBA-s) score calculation

Enroll-HD PDS4 contains calculated composite PBA-s scores. Please refer to the following reference for further information:

Craufurd D, Thompson JC, Snowden JS. Behavioral changes in Huntington Disease. Neuropsychiatry Neuropsychol Behav Neurol. 2001 Oct-Dec;14(4):219-26.

Table 12 - PBA-s sub-score calculation

PBA-s Section	Variable	Sub-Score Calculation
Depression	*depscore*	depressed mood + suicidal ideation + anxiety
Irritability/Aggression	*irascore*	irritability + angry or aggressive behaviour
Psychosis	*psyscore*	delusions / paranoid thinking + hallucinations
Apathy	*aptscore*	apathy
Executive Function	*exfscore*	perseverative thinking or behaviour + obsessive-compulsive behaviours

These composite scores are calculated by multiplying *severity* by *frequency* for each symptom, which are then summed to create a composite score. For example: Depression = (severity of depressed mood*frequency of depressed mood) + (severity of suicidal ideation*frequency of suicidal ideation) + (severity of anxiety*frequency of anxiety).

4.4 Mini Mental State Examination (MMSE) score calculation

Enroll-HD PDS4 contains calculated MMSE scores. Please refer to the following reference for further information:

Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. Psychiatr Res. 1975 Nov;12(3):189-98

Table 13 - MMSE total score calculation

MMSE Section	Variable	Total Score Calculation
MMSE Score	*mmsetotal*	Sum the value of all scores

4.5 Hospital Anxiety Depression Scale / Snaith Irritability Scale (HADS-SIS) score calculation

The HADS-SIS assessment used in Enroll-HD is formed from **two separate scales**, the **HADS** (Zigmond & Snaith, 1983) and the **SIS** (Snaith, 1978). It is important to recognize that the HADS-SIS is comprised of these two separate scales so that analyses can incorporate the respective subscales and items appropriately.

The following reference provides further information on score calculation:

Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983 Jun;67(6):361-70.

Snaith, RP. A clinical scale for the self-assessment of irritability. Brit J Psychiat. 1978; 132: 164-171.

4.6 Short Form Health Survey - 12v2 (SF-12) score calculation

Enroll-HD PDS4 contains calculated scores for SF-12 scales. These are presented in the 'enroll' data file. The following reference provides further information on score calculation:

Ware JE, Kosinski M, and Keller SD. A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. Medical Care, 1996; 34(3):220-233.

4.7 Short Form Health Survey – 36 v1/v2 (SF-36) score calculation

Enroll-HD PDS4 contains calculated scores for the SF-36 scale (version 1 and version 2) available in 'registry' datafile. The following reference provides further information on global score calculation:

Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992 Jun; 30(6):473-83.

5. Revision History

Document Name	Summary of Changes
Version 2015-01-R1	Initial version for Enroll-HD PDS1
Version 2015-10-R2	Revised version for Enroll-HD PDS2
Version 2016-10-R1	Revised version for Enroll-HD PDS3
Version 2018-10-R1	Revised version for Enroll-HD PDS4