Enroll-HD Plus Data Handling Manual

Version 2016-10-R1

Enroll-HD

A worldwide observational study for Huntington's disease families

A CHDI Foundation Project

CONTENTS

1. PURPOSE OF THIS DOCUMENT 3

2. DATASET CONSIDERATIONS 3

2.1 Data Policy 3

2.2 Data Specifications 3

2.3 Data Quality Control and De-Identification 5

2.4 Dataset File Formats 6

2.5 Datasets 6

2.6 Merging Data Files 8

2.7 Missing Data 8

2.8 Null values 9

2.9 Aggregated Values 9

2.10 Transformation of Date Variables 11

2.11 Follow-Up (FU) Information 12

2.12 HD Classification Variables 13

2.13 Information on Outliers and Unusual Findings 14

3. SPECIFIED DATASET REQUESTS 15

3.1 HD Category Specified Request 15

3.2 Aggregated Data Specified Request 16

3.3 De-identified Data Specified Request 16

3.4 Precision Information Specified Request 16

4. ADDITIONAL CONSIDERATIONS 17

4.1 Therapies and Comorbidities Coding 17

4.2 Outstanding values 17

4.3 Subject Status..... 18

4.4 Unified Huntington's Disease Rating Scale (UHDRS) scores calculation 18

4.5 Problem Behaviors Assessment – Short (PBA-s) score calculation 19

4.6 Mini Mental State Examination (MMSE) score calculation 19

4.7 Hospital Anxiety Depression Scale / Snaith Irritability Scale (HADS-SIS) score calculation 20

4.8 Short Form Health Survey - 12v2 (SF-12) score calculation 20

4.9 Short Form Health Survey – 36 v1/v2 (SF-36) score calculation 21

1. PURPOSE OF THIS DOCUMENT

This document contains guidance on how to use the Enroll-HD Plus periodic dataset and describes how data were compiled for this release.

2. DATASET CONSIDERATIONS

2.1 Data Policy

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

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

2.2 Data Specifications

The third Enroll-HD Plus periodic dataset was extracted on October 31, 2016 from a larger dataset based on predefined requirements.

The Enroll-HD Plus dataset includes data from Ad Hoc visits (RET), REGISTRY (R2 and R3) and Enroll-HD studies (ENR). The Enroll-HD Plus dataset only contains data about Enroll-HD participants who meet the following criteria (see the Enroll-HD Plus User Guide for a full list of the criteria):

- Participants with monitored baseline visit information (remote and onsite monitored);
- All participants included in REGISTRY or Ad Hoc data must be included in the Enroll-HD study.

The figure below indicates the number of participants enrolled in the study as of October 31 2016 and the number of participants whose data met each of the predefined inclusion requirements.

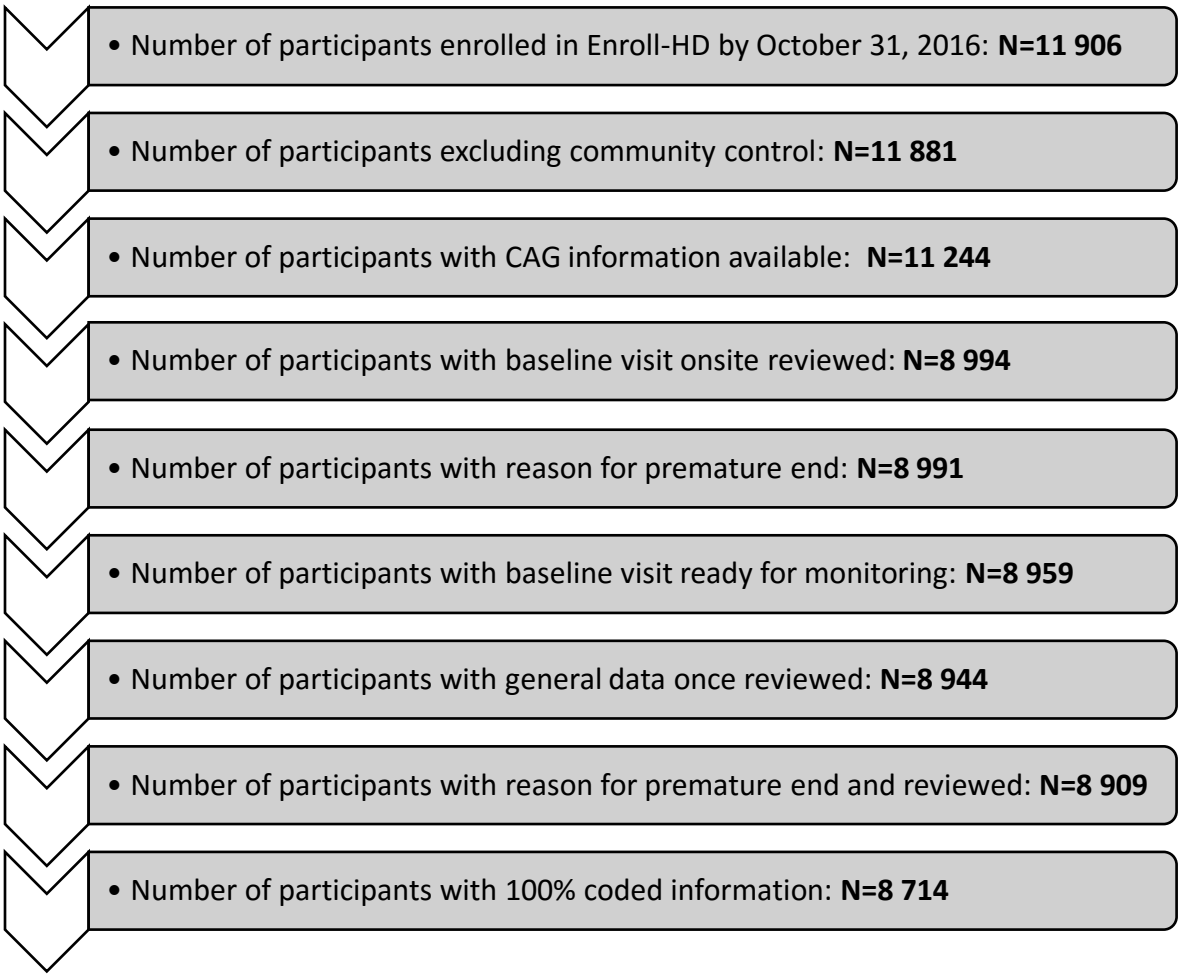

Figure 1 – Flow chart showing requirements for participant exclusion from the dataset

As mentioned, the Enroll-HD Plus dataset includes data from Ad Hoc visits (RET), REGISTRY (R2 and R3) and Enroll-HD studies (ENR). The following chart describes the number of Enroll-HD PDS3 participants by number of study visits. Column color coding indicates source(s) of visit data. Orange columns indicate number of participants with data from at least X visits collected within the Enroll-HD study only. Blue 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, Ad Hoc). This plot is cumulative; participants 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 blue and orange columns displayed for 1, 2 and 3 visits.

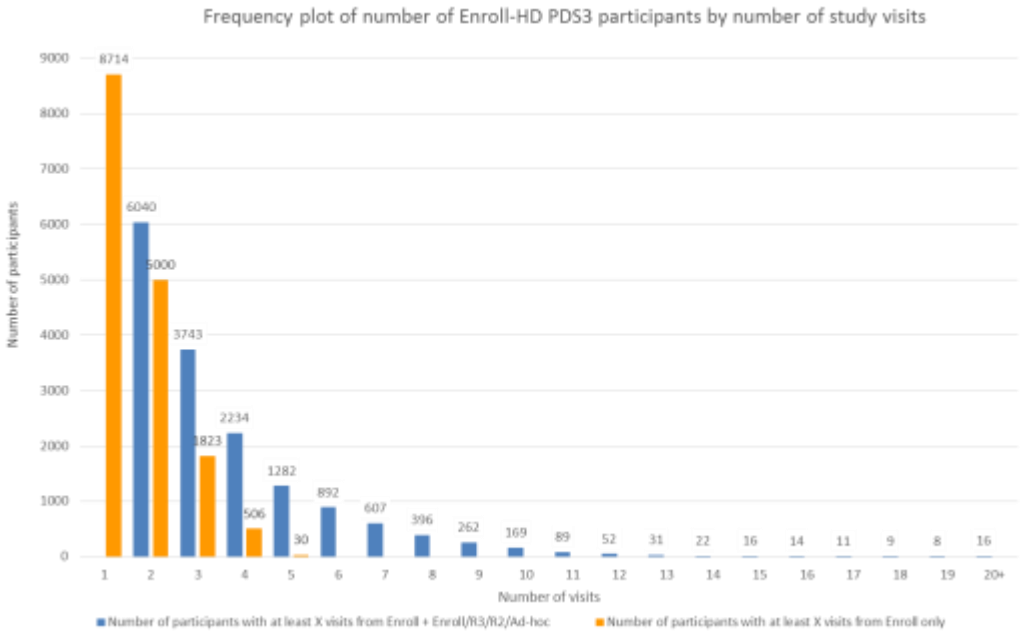

Figure 2 – Frequency plot of number of Enroll-HD PDS3 participants by number of study visits

2.3 Data Quality Control and De-Identification

Prior to publishing the data, the Enroll-HD Plus dataset goes through both a quality control (QC) and de-identification process.

During the QC process, the Enroll-HD Plus dataset is checked for consistency, integrity (data were recorded exactly as intended), correctness (data plausibility) and completeness (amount of missing information).

During the de-identification process, the dataset is assessed to determine the risk of identification for each participant based on their information in the dataset. This process uses predefined cutoff values to determine whether a participants' data should be included in the dataset. For some variables, an aggregation process can be used that reduces the risk of

identification but allows that participants' data to remain in the dataset. If, after the completing the de-identification process, an individual's data still does not meet specific cutoff values for identification risk, then the participant is removed from the dataset.

2.4 Dataset File Formats

The Enroll-HD Plus dataset is provided in two formats: comma-separated values (.csv extension, delivered as tab-separated text files) and in R format for R users.

The .csv files are separated by a tab character, which is normally invisible when inspecting the data in a generic text editor. The .csv file format is readable by most statistical software packages including R, Stata, and SAS. In many cases, it is essential to specify in the program that the variables are separated by tabs, as other separation conventions are also commonly used. It is also important that these files not be edited in word processing or other programs that may potentially modify the tab characters, as this will destroy the ability of the data to be imported correctly. If needed, .csv files can be saved in other formats that are compatible with other statistical software packages. The procedure for importing .csv files into Excel is outlined in the following document: Enroll-HD Periodic Dataset quick reference guide "How to use the csv files".

The dataset is also available in R format and can be directly opened in R.

2.5 Datasets

The Enroll-HD Plus dataset contains 11 data files. The structure of the data files is provided in Table 1 and further explained in the Enroll-HD Plus Periodic Dataset Structure Quick Start-up Guide and in the Data Dictionary of Enroll-HD Plus Periodic Dataset.

The Enroll-HD Plus 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 collected at the baseline or annual visits. Furthermore, the data in the files are arranged in three ways: 1) participant-based – contain

general study-independent information about the participant, 2) study-based – contain general overview study information and 3) visit-based – contain information about each visit each participant attended within a given study.

Table 1: Data files contained in the dataset provided

Entity	Type	Studies	Description
Profile	participant-based	ENR R3 R2 RET	General and annually updated information including the forms: Demog, HDCC, CAG, Mortality.
PharmacoTx	participant-based	ENR R3 R2 RET	Form: Pharmacotherapy.
NutSuppl	participant-based	ENR R3 R2 RET	Form: Nutritional Supplements.
NonPharmacoTx	participant-based	ENR R3 R2 RET	Form: Non-Pharmacologic Therapies.
Comorbid	participant-based	ENR R3 R2 RET	Form: Comorbid Conditions.
Participation	study-based	ENR R3 R2 RET	Study specific information about participant taking part in the study. Start, end, reasons, visits, etc. are stored in a general format.
Assessment	study-based	ENR R3 R2 RET	Study-specific information about existing regular visits performed in any study. Provides an overview about number, type and sequence of visits.
Event	study-based	ENR	Reportable Event Monitoring that happened during Enroll-HD.
Enroll-HD	visit-based	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-based	R3 R2	Registry file contains information from Registry 2 and Registry 3 studies. This file contains information on Baseline, Follow-up, and unscheduled visits.

Entity	Type	Studies	Description
Adhoc	visit-based	RET	Adhoc file contains information from Variable, Motor, Function, TFC, MMSE, Cognitive assessments. All visits available for this study are included, visits can be ordered by using the 'seq' variable which represents the sequence of the visits.

2.6 Merging Data Files

The Enroll-HD Plus dataset contains one key variable that allows the user to merge two or more data files. The key variable is 'subjid' and it is included in every data file. 'subjid' can be used to uniquely link information for each participant across data files. The key variable 'subjid' is labeled as 'HDID (recoded)'.

Variable labels are created only when the dataset is read from the text files into R using the R scripts provided in the R software package. No labels exist in the text data otherwise. For statistical software packages other than R, labels should be created in accordance with the provided Data Dictionary for the Enroll-HD Plus periodic dataset.

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.

2.7 Missing Data

The Enroll-HD Plus periodic dataset contains missing data, not applicable data, or implausible values data. 'NA' and blank cells are used to identify missing data in the csv files. All missing data are coded as 'NA' in the R files.

Values which are set to missing, not applicable or wrong are set to NA. Calculated variables are set to NA if the source variables are set to one of the exceptional values mentioned above. If the calculation was not possible the values are empty. Normal values are empty if they are not entered due to dependencies in the eCRF, e.g. packyears for non-smokers.

Depending on the statistical software package used in the analysis and the type of analysis being performed, the fields designated as 'NA' may need to be modified. Transformations should be considered depending on the statistical software package used. The variables that contain 'NA' values will be read as text variables in most statistical software packages. Some statistical software packages include the option to translate 'NA' values into other non-text values. If this is not an option, Excel can be used to perform the data translation. Excel's global search and replace function makes it easy to replace 'NA' values with other values that are easier for standard statistical packages to interpret.

2.8 Null values

For the variables 'packy', 'hxpacky', 'alcunits' a zero value is occasionally observed. , This is due to rounding of extremely low values. These zero values were confirmed and validated.

The medication total dose is a calculated variable ('cmdostot') that depends on frequency and daily intake. In cases where a zero value is observed, this may be due to either a very low frequency (e.g. medication taken as needed) or a very low daily intake.

2.9 Aggregated Values

As part of the de-identification process, various aggregation techniques were applied to specific variables. The variables, the criteria for aggregation, and the new values are described in the table below.

Table 2: Aggregated values in the dataset

Data file	Variable	Variable Label	Criteria for aggregation	New Value
participation	'age'	Age at enrollment	All participants with age under	<18
enroll	'age'	Age at visit	All participants with age under	<18
profile	'caghigh'	Research larger CAG allele determined from DNA	All participants with their larger CAG allele above	> 70
profile	'caglow'	Research smaller CAG allele determined from DNA	All participants with their smaller CAG allele above	> 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 total numbers of participants in every aggregated group for ethnicity, age and cag can be find in following tables:

Table 3: Number of participants per category

variable	*label*	*number of participants*
'age'	<18	14
'caghigh'	>70	12
'caglow'	>28	129

Table 4: Number of participants per ethnicity

race	*label*	*number of participants*
Caucasian	1	8089
American – Black	2	89
Hispanic or Latino Origin	3	158
Other:	***6***	***131***
Native Hawaiian or Other Pacific Islander	4	1
Alaska Native/Inuit	5	3
Other	6	116
African - South	11	5
African - North	12	6
American Indian/ Native American/ Amerindian	8	89
Mixed	15	93
Asian - West and East:	***16***	***62***
Asian - West	13	37
Asian - East	14	25

Because of de-identification reasons, the information collected at visits prior to Enroll-HD, namely Registry and Adhoc visits, for participants aged <18 years was not included in this release. This information can be requested through a specified dataset (SPS) request.

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. The descriptive statistics for the aggregated variables, prior to aggregation, are available by specified dataset (SPS) request.

The level of aggregation differs between Enroll-HD periodic dataset releases. The changes in number/type/given values for participants make aggregation adjustments necessary.

2.10 Transformation of Date Variables

As part of de-identification process, all variables that represent dates in the Enroll-HD Plus dataset were changed from a date field to a numeric field. The number in the field represents the

number of days with respect to the **baseline date of Enroll-HD study**; for example, if a participant attended a baseline visit on January 1, 2015 and completed the UHDRS Motor/Diagnostic Confidence on January 31, 2015, the date information will appear as 30. This represents the number of days that the evaluation of UHDRS Motor/Diagnostic Confidence was performed after the baseline visit.

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 rule was applied to incomplete date fields: value entered as "YYYY-MM," change to "YYYY-MM-15"; values entered as "YYYY," change to "YYYY-07-01". In other words, "15" was used as day if day was missing, and "1" was used as day and "7" as month if both day and month were missing.

Due to this rule, for the forms 'PharmacoTx' and 'Comorbid', some end days are set prior to start days.

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

Table 5: 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

Because dates have been changed to numeric values relative to the BL visit, dates can be negative numbers. This is typical for start dates of medications and comorbid conditions.

2.11 Follow-Up (FU) Information

The Enroll-HD Plus 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 Enroll-HD Plus files: 'enroll', 'registry' and 'adhoc' include a new variable called 'seq'. This variable refers to the sequence of the visits and will enable the data analyst to align 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 and seq=3 to the 2nd follow up visit and so on, including the unscheduled and phone contact visits.

Table 6: Example of follow up visit sequencing

subjid	studyid	visit	seq	visdy
R001084542	ENR	Baseline	1	0
R001084542	ENR	Follow Up	2	363
R001084542	ENR	Follow Up	3	728

The phone contact visits only occur in Enroll file. These visits contain missed visits information, reason for missed follow-up visit and if participant disposition. If these data are not needed for a given analysis, they can be filtered out.

2.12 HD Classification Variables

Variables used to classify participants:

The Enroll-HD Plus dataset contains a variable, 'hdcat', which reports the HD category of each participant. 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 latest HD category information for that participant.

In a few cases ´hdcat_l´ does not match to the value of the participant in the latest follow up visit file. This is not an error but indicates that there was an additional follow up visit for this participant, but this follow up visit was excluded from the most recent dataset. There are several reasons a given visit would be excluded, but the most likely reason is that the follow up visit data has not yet been monitored. For example, if a participant had an 'hdcat' classification of 2 (pre-manifest) at baseline, but then returned for a follow up visit a year later and their 'hdcat'

classification at follow up was 3 (manifest) – then the hdcat_l variable is revised to 3. In Enroll-HD Plus the 'hdcat_l' will be listed as 3, the 'hdcat' at baseline will be 2, but the follow up visit may or may not be included in the dataset, depending on if the data for follow up visit 1 had completed monitoring.

The 'hdcat' variable is available for Registry 3 but it is not available for Registry 2 and Adhoc since these two studies did not use an HD classification.

HD classification is not an exact science. In general, the following criteria for classification can be used (adapted from 2008 European HD Network):

- Premanifest/premotor-manifest HD ('hdcat'=2): CAG repeat length on larger allele >= 36 and no HD diagnosis and motor score <= 10.
- Manifest HD patient/motor-manifest HD ('hdcat'=3): CAG repeat length on larger allele >= 36 and HD diagnosis > 3 and motor score > 10.
- Genotype negative ('hdcat'=4): CAG repeat length on larger allele <= 35, with family history of HD.
- Family control (without family history of HD) ('hdcat'=5): No signs or symptoms of HD, no family history of HD.

Participant´s classification:

Some participants classified as HD-manifest present with low scores for Total Motor Scores (< 10). This may be due to the symptom onset being of psychiatric nature instead of motor.

2.13 Information on Outliers and Unusual Findings

The Enroll-HD Plus periodic dataset underwent an outlier analysis and clinical statistical monitoring process. Outliers and unusual findings identified during this process were validated by the monitoring and medical monitoring teams. These outliers and unusual entries are detailed in the Unusual Findings document.

3. SPECIFIED DATASET REQUESTS

Specified dataset requests can be made to obtain data not provided in the current periodic dataset.

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

3.1 HD Category Specified Request

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

Approximately 10 percent of the participants included in the periodic dataset were categorized as Genotype Unknown, meaning neither the participant nor their physician knows the genetic status of the participant. However, because each participant is genotyped as part of the study, the Genotype Unknown participants were categorized into an 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 by the Scientific Publication Review Committee (SPRC).

3.2 Aggregated Data Specified Request

The Enroll-HD Plus dataset has aggregated data that can be deaggregated and obtained by special request, subject to approval by the Enroll-HD Scientific Publication Review Committee (SPRC).

The deaggregated data that can be requested is presented in Table 2 (page 10, Aggregated values in the dataset).

3.3 De-identified Data Specified Request

In the Enroll-HD Plus dataset, some medications and comorbidities have been excluded due to risk of identification. These data may be obtained by special request, subject to approval by the Scientific Publication Review Committee (SPRC).

Researchers who are also affiliated to an Enroll-HD study site and who request such data will not be provided with data collected from their own Enroll-HD study site.

3.4 Precision Information Specified Request

The Enroll-HD Plus dataset does not contain dates as these have been transformed to decrease risk of participant identification; see Section 2.9 for a description of how dates were transformed. 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”)

In the Enroll-HD Plus dataset, doses for some medications can appear as zero (0). The raw data values (frequency*daily intake) used to calculate the doses can be requested through special dataset request.

4. ADDITIONAL CONSIDERATIONS

4.1 Therapies and Comorbidities Coding

In the Enroll-HD Plus dataset, 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' for all conditions and therapies that do not have a stop date and are therefore assumed to be ongoing at the time of the visit.

The 'pharmacotx' file contains a variable for the drug name that is coded with the WHO-Drug Dictionary. Special attention should be paid when opening this file when using csv since it is possible that excel misreads this code as a number and this could corrupt the variable (for more information on handling this file please check the Enroll-HD Periodic Dataset quick reference guide "How to use the csv files"). More information on the coding included in the 'PharmacoTx', 'NonPharmacoTx', 'NutSuppl' and 'Comorbid' files can be found on the Enroll-HD Periodic Dataset quick reference guide "How to use the coded data" document.

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

4.2 Outstanding values

Some variables have values outside of the normal range, for example height might vary considerably. Table 5 below shows height for the same individual taken over several years with values ranging from 132 cm to 179 cm. The first value, 132 cm, is from the Registry study and can

no longer be queried. For data such as these, all data are provided and it is left to the analyst to determine how best to evaluate the data.

Table 7: Example for outstanding values for the same individual

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

4.3 Subject Status

Typically, for each participant that has been categorized as 'withdrawn' or 'violator', a reason is provided to explain why the participant has discontinued the study. For some cases these reasons have been removed due to risk of potential participant identification.

4.4 Unified Huntington's Disease Rating Scale (UHDRS) scores calculation

The Enroll-HD Plus dataset contains calculated UHDRS scores. The following reference will provide further information on score calculations:

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

Table 8: UHDRS sub-score calculation

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

UHDRS Independence Scale	indepscl	Sum the value of scores

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

The Enroll-HD Plus dataset contains calculated PBA scores. The following reference will provide further information on score calculations:

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

Table 9: 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*frequency for each symptom and summed up 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).

4.6 Mini Mental State Examination (MMSE) score calculation

The Enroll-HD Plus dataset contains calculated MMSE scores. The following reference provides further information on score calculation:

- 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 10: MMSE total score calculation

MMSE Section	Variable	Total Score Calculation
Mini Mental State Examination Score	mmsetotal	Sum the value of all scores

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

The HADS-SIS assessment used for Enroll-HD Plus is formed from two separate scales, the HADS (Zigmond & Snaith, 1983) and the SIS (Snaith, 1978). It is important to realize the HADS-SIS is made up 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.

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

The Enroll-HD Plus dataset contains calculated scores for the eight scaled scores of SF-12 scale present in the 'enroll' data file. The following reference provides further information on scores 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.

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

The Enroll-HD Plus dataset contains calculated global score 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.

Revision History

Document Name	Summary of Changes
Version 2015-01-R1	Initial version for first Enroll-HD periodic dataset
Version 2015-10-R2	Revised version for second Enroll-HD periodic dataset
Version 2016-10-R1	Revised version for third Enroll-HD Plus periodic dataset