

Enrollees

SAS Dataset Name: Enrollees

General description:

The Enrollees dataset contains demographic and grouping information, much of which does not change from visit to visit, but is needed in analyses with data from every visit. Most of these variables were moved from the SubjectChar00 dataset and put into this new dataset as of release 1.1.1, although they remained in SubjectChar00 until it was re-released as version 0.2.1. The variables can now be found only in the Enrollees dataset, which is re-released with every new public data release.

The image assessment project membership variables have been removed from the Enrollees dataset as of version 21. Since image assessment project membership variables are also included in the Measurement Inventory for Study Planning dataset (formerly named Imaging Inventory), it was decided to keep the image assessment project membership variables in the Measurement Inventory dataset only.

This dataset contains variables for **race (P02RACE), sex (P02SEX), cohort assignment (V00COHORT), and site (V00SITE)**, which may need to be controlled for in many analyses. The variable P02HISP, which indicates whether a participant is Hispanic/Latino or not, was included as of release 0.2.1. It also contains a variable (V01HADINTV) that indicates whether a participant had an 18- or 30-month interim visit. This variable will be helpful for planning analyses using these interim-visit data. As of release version 7 of the Enrollees dataset, the 30-month interim visit was complete, so some participants who previously had V01HADINTV=0 have now become V01HADINTV=2 to indicate that they have now had 30-month visits.

Image group membership variables:

There are now separate image-group indicator variables for each image group that has been released. The variable V00IMGSMPL, originally found in SubjectChar00, was renamed to V00IMAGESA to make it more consistent with later image group indicator variables. Once an image group has been created, there is an indicator variable for each visit with each of the appropriate prefixes. For the two interim visits (18- and 30-month), only the indicator variable pertinent to that visit (V02IMAGESD and V04IMAGESG, respectively) has been created. In addition, starting with release 0.2.2 and 1.2.1, these indicators have been expanded from 0/1 variables, to categorical variables that take the value of 0 if the participant is not in the group, 1 if the participant is in the group but had only MR imaging (no radiographs) at that visit, 2 if they had only radiographs (no MRI), and 3 if they had both. If a participant's images for a particular visit exist in more than

one image group, then each applicable indicator variable takes a value ≥ 1 . For example, although unintentional, there are four participants who are in both Image Group A and Image Group B, so for these four participants $V00IMAGESA \geq 1$, $V01IMAGESA \geq 1$, $V00IMAGESB \geq 1$, and $V01IMAGESB \geq 1$. For more information about the $VxxIMAGESx$ variables, please see the DICOM Image Release Notes (DIRN) document.

For all follow-up visits, the indicator variables include SAS special missing values to indicate when a participant is in the image group but all imaging is missing for that visit, due to the participant's having missed the visit altogether (.V) or having refused all imaging for that visit (.R). A participant who is not part of that image group (e.g., a participant in the second half of the cohort in the case of Image Group C) can only have a value of 0, regardless of whether they missed the visit or refused all imaging at that visit. If a participant is part of the image group (e.g., a participant in the first half of the cohort and Image Group C), the value of the variable may be 1 (part of the group and has MR imaging only at that visit), 2 (part of the group and has X-ray imaging only at that visit), 3 (part of the group and has both X-ray and MR imaging at that visit), .V (part of the group and missed the visit), or .R (part of the group and refused all imaging at that visit).

Data Structure/Contents:

This dataset contains one record per participant. The variable uniquely identifying a record is ID, and the dataset is sorted by ID, which can be used to merge/join to data in other datasets.

A full list and description of all the variables contained in this dataset can be found in the contents.pdf.

Condition of data:

- **Known data errors:** problems/cautions for use are listed by variable in the “Release Comments.”
- **Dataset strengths/weaknesses:** None at this time.

General strategies for use:

When using with other datasets, merge/join by ID.

Analysts are encouraged to always output and view SAS variable labels in their entirety to ensure important information about the variables is not lost. The maximum SAS label length is 160 characters.