

**WHI Memory Study (WHIMS) Limited Access Data Release**  
**Data Preparation Guide**  
**July 15, 2007**

## **1. Introduction**

The WHIMS data sets available on the website include baseline and follow-up data obtained during the course of the clinical trials, through July 7, 2002 for the E+P cohort and February 29, 2004 for the E-alone cohort.

Form 39 was administered annually for WHIMS participants. If a participant scored at or below the cut point (80 for women with 8 or fewer years of formal education and 88 for those with 9 or more years of formal education), she progressed to further cognitive testing and clinical assessment (Phases 2 and 3). Note that before July 1, 1998 the cut points were 72 and 76. A friend or family member, whose name the participant furnished at the beginning of the trial, was also interviewed regarding her cognitive and behavioral status.

Based on clinician judgment, some participants then progressed to Phase 4 requiring additional medical tests and procedures to more fully evaluate and classify cognitive decline. Note that Phases 2/3/4 could be conducted after the end date(s) above, as long as the Form 39 which triggered the progression occurred during the course of the trial(s).

### **Data sets contained in the 7/15/2007 release**

- **WHIMS\_pub (contains Form 39 and adjudication data)**
- **P2form\_pub (Form A, Phase 2: Administration and Scoring Manual)**
- **P2family\_pub (Form C, Phase 2: Friend/Family Member Interview)**
- **P3form\_pub (Form D, Phase 3: Clinical Evaluation)**
- **P4form\_pub (Form D, Phase 4: Laboratory and Imaging Studies)**

## **2. Data File Setup**

Each data set is provided as a version 9.1.3 SAS data set. The data sets can be found in the *data\sas\whims\sas* directory. Each data set is zipped up into a .ZIP file that includes the SAS data set and a SAS formats documentation file.

To use the data you will first need to unzip the .ZIP file for the data set you wish to use.

The data sets do not contain the same participants since not all women progressed to Phases 2/3/4. The first variable in each file is ID. Computed variables have also been included and are described in detail under the variable listings. The form questions used in the computation of the computed variables have been noted in the variable descriptions.

### 3. Data Conventions

#### Dates

No actual dates are included in the data files. All dates have been converted to the number of days since WHI randomization. A negative number of days indicates the date occurred before randomization. Likewise, a positive number indicates occurrence after randomization.

#### Data Edits

At data entry, the built-in features of the study database application prevented entry of most invalid or impossible data values for all categorical variables. Broad range checks applied to continuous variables have set out-of-range responses to missing. There still may be values that appear extreme; **it is up to the user to examine all data before proceeding with data analysis.**

For Form 39 data, the scanned data provided by WHI usually appear in the data sets. In some cases, when scanned data were not available, data sent to the WHIMS Coordinating Center on the Tally Form appear in the data sets. Note that the Tally Form was a WHIMS-specific form used for supplementary Form 39 data collection to provide early notice of a possible progression.

Consistency checks between data items from the scanned data and the Tally Form were done routinely over the course of the study. In some cases, the clinic staff was not able to make corrections to the scanned data so hard-coded revisions were made at the WHIMS Coordinating Center to replace the erroneous scanned values. Therefore, there may be occasional discrepancies in Form 39 scores between WHI data and WHIMS data.

In addition, extensive cleaning was performed for the data collected as part of Phases 2/3/4. For example, selected missing data were checked to verify their missingness or correct data entry errors. Important skip patterns were also examined.

#### Missing Data

Missing data can result from a participant not progressing to Phases 2/3/4, a required form not being completed, a particular question on a form not being answered or not required because it was part of a skip pattern, or an item not appearing on an earlier form version. Missing values in the data files are represented by a single period (“.”).

## 4. Specific Data Set Information

### Form 39 Data

Form 39 data consist of multiple observations over time for each participant. The variables “YEAR” and “NUMBER” distinguish the observations as follows:

YEAR = 0 represents assessments obtained +/- 6 months around the date of randomization;

YEAR = 1 represents assessments obtained +/- 6 months around the first year anniversary of the date of randomization;

YEAR = 2 represents assessments obtained +/- 6 months around the second year anniversary of the date of randomization; and so forth.

The variable “NUMBER” identifies the sequence of observations obtained within each “YEAR”. It assumes integer values 1 and higher. In most cases, NUMBER = 1 but there are instances where multiple assessments were completed within the same “YEAR.”

Included within the Form 39 data are individual item scores, the total computed score and calculated subscales. The subscales are documented in the data dictionary for the WHIMS dataset (Teng EL, Chui HC: The Modified Mini-Mental State (3MS) Examination. Journal of Clinical Psychiatry 48: 314-318, August 1987).

### Progression Data

Progression data consist only of those participants who scored at or below the cut points for Form 39. Women can progress repeatedly, year after year. Once a woman is classified as probable dementia (PD), she then goes to Phase 2 in subsequent years but no longer has to undergo Phases 3/4. From that time onward, she is classified as PD. However, minor cognitive impairment (MCI) can be transient so women once classified as MCI can later be reclassified.

The following forms comprise progression data:

Phase 2: Administration and Scoring (Form A)

Phase 2: Friend/Family Member Interview (Form C) Note: May be more than one form associated with its corresponding Form 39 if multiple family members and/or friends are interviewed.

Phase 3: Clinical Evaluation (Form D)

Phase 4: Laboratory and Imaging Studies, and Case Classification (Form D)

Adjudication Tracking Data: Data tracking the adjudication process and final diagnoses outcomes.

## **5. Merging Data Files**

The progression data (P2form, P3form, P4form) and the friend/family member interview can be merged with the main WHIMS data using two variables: ID and F39DAYS.

If you wish to expand your data analyses to include WHI Clinical Trial data, you can use the ID variable in the WHIMS data set and the ID variable in the WHI Clinical Trial data set to merge data sets. The WHIMS and WHI Clinical Trial data releases use the same participant ID.