WHI Memory Study (WHIMS) Limited Access Data Release
Data Preparation Guide

Introduction

The WHIMS data sets available include data through August 14, 2009. A list of the data sets and a description of each component follows.

The WHIMS data sets include:

1) On trial baseline and follow-up data obtained during the course of the WHI clinical trials, through July 7, 2002 for the E+P cohort and February 29, 2004 for the E-alone cohort.
2) Post trial/extension data obtained after the end of the clinical trials, after July 8, 2002 for the E+P cohort and after February 29, 2004 for the E-alone cohort
3) Cognitive outcome data based on the Supplemental Case Ascertainment Protocol (SCAP) for deceased participants or those for whom only the proxy could be contacted
4) Participant clinical data including consent, screening, medical history, MRI procedure, summary and QC data and MRI scan data on a subset of participants who underwent a brain scan
5) In-depth cognitive data on a subset of women who participated in the Women's Health Initiative Study of Cognitive Aging (WHISCA)

The timing of the various study components is depicted in the table below.

Timing of Data Collection for Study Components

Phase	Beginning and end dates
On trial	1995 through July 8, 2002 for CEE+MPA trial 1995 through February 29, 2004 for CEE-alone trial
Post trial/extension	End-date of trial through September 30, 2007
SCAP	2005 through present
MRI-1	2004-2006
WHISCA	1999 through 2007

On Trial

During the WHI clinical trials, 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).

Post Trial Extension

During the post trial/extension period, the Modified Mini-Mental State Examination (3MSE, Form 39) continued to be administered annually for WHIMS participants through 2007. 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 in-person cognitive testing and clinical assessment (Phases 2 and 3). 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.

Supplemental case Ascertainment Protocol (SCAP)

SCAP was instituted in 2005 to identify cases of Probable Dementia (PD) and Mild Cognitive Impairment (MCI), in the deceased and proxy-dependent participants. SCAP includes the Dementia Questionnaire (DQ) [Silverman, 1986], a standardized, validated instrument used to reliably classify dementia and specifically, Alzheimer's disease in deceased persons. It has demonstrated sensitivity and specificity. The SCAP survey consists of 48 items assessing memory and other cognitive functions, language, daily functioning, insight, and other medical and psychiatric difficulties. The DQ is a semi-structured interview that was administered by telephone to informants previously selected by the participant, who are knowledgeable about the participant's medical history and ante-mortem functional status. SCAP is still in progress.

Magnetic Resonance Imaging (MRI-1)

The WHIMS-MRI study entitled Effects of Hormone Therapy on Subclinical Neurological Pathology was a cross-sectional magnetic resonance imaging substudy in women from the WHIMS Post-Trial Extension. Women who consented underwent a single MRI scan. Total brain volume and ischemic lesion volume were obtained for different brain regions.

WHI Study of Cognitive Aging (WHISCA)

WHISCA is an ancillary study to assess the efficacy of postmenopausal hormone therapy on age related changes in specific cognitive functions in non-demented women at 14 of the 39 WHIMS sites. WHISCA provides more comprehensive annual assessments of cognitive function and mood. Annual follow-up of the WHISCA women continued after the termination of the WHI trials until 2007. The content of the cognitive battery is described below.

Classification of Events

It is important to note that SCAP does not include face-to-face cognitive testing, clinical assessment, or medical testing for adjudicated classification of cognitive decline.

Data Sets Description

Data sets for each component are described below. For further information, please contact Patricia Hogan at phogan@wfubmc.edu.

On Trial

- 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)

Post Trial

- WHIMS_post (contains Form 39 and adjudication data)
- P2form_postc (Form A, Phase 2: Administration and Scoring Manual)
- P2family_postc (Form C, Phase 2: Friend/Family Member Interview)
- P3form_postc (Form D, Phase 3: Clinical Evaluation)
- P4form_postc (Form D, Phase 4: Laboratory and Imaging Studies)

SCAP

- SCAP (contains DQ and adjudication data)

MRI-1

- MRI1_totalvol (Total volume by region)
- MRI1_abnormalvol (Abnormal volume by region)
- MRI1_consent (contains MRI1 consent date)
- MRI1_initial (MRI 1 Initial Contact/Screening Questionnaire)
- MRI1_medical (MRI 1 Medical Questionnaire)
- MRI1_procedure (MRI 1 Procedure Questionnaire)
- MRI1_qc (MRI 1 Quality Confirmation)
- MRI1_summary (MRI 1 Summary Report/Clinic Neuroradiologist Form)

WHISCA

- WHISCA (contains final scores and medical update)

2. Data File Setup

Each data file is a SAS data set. The data sets can be found in the *data\as\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 and subsets of women participated in the other study components. The identifying variable in each file is Participant ID, '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 indicate 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 a 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).

For the Finscor (3MS Score) variable, if a response is unknown, not attempted/disabled or missing, the algorithm adjusts the scoring of the question to prevent the participant from being penalized as if she had given an incorrect response. The maximum points awarded for the question are subtracted from a counter set initially at 100. Once all questions have been tallied, this counter becomes the denominator in determining the score – the percentage of correct answers versus the number of total possible points for that participant. For example, if question 6.3 (What season of the year is it?) is answered not Attempted/disabled or missing, 1 point is subtracted from the Total possible score. If this is the only such question on the Form 39, the score would be computed as a percentage of total Points earned out of 99 possible points, instead of 100. If the total points Discounted as missing, unknown or not attempted/disabled exceeds 20 (the denominator is less than 80), the test is not considered valid.

Progression Data Based on Form 39

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.

SCAP Data

SCAP relies on telephone interview data from a proxy to classify cognitive decline, July 1 was selected as the month and day for all year variables because the Dementia Questionnaire (DQ) asks what year particular events began happening, but not the exact date. The dataset contains time from randomization for all variables. For example, if a proxy noted that a participant began having memory problems in 2001, that date was converted to 7/1/2001 and then calculated days since randomization to 7/1/2001.

Included are the results from the DQ as well as classifications (probable dementia (PD), minor cognitive impairment (MCI), or no dementia (ND)) of those cases that were adjudicated by a central panel of cognitive experts. The variable d_scap has different definitions based on the final classification and the status of the participant (deceased or proxy-dependent) as summarized in the table below. If the participant has been classified as PD or MCI in SCAP and is proxy-dependent, d_scap is the midpoint of the date of the participant's last 3MSE and the date of the DQ interview. If the participant is deceased and classified as PD or MCI, d_scap is the midpoint of the date of the participants last 3MSE and the date of death. On the other hand, if the participant is classified in SCAP as ND and she is proxy-dependent, d_scap is the maximum of the date of the last 3MSE and the date of the DQ interview. If the participant is deceased and classified in SCAP as ND, d_scap is the date of death. In the event that the DQ was not administered or classification is still in process, d_scap is defined as the date of the last 3MSE.

Definition of Date of SCAP (d_scap)

	SCAP Classification		
Participant status	PD or MCI	ND	Missing data
Proxy-dependent	Midpoint of date of last 3MSE and date of DQ	Maximum of date of last 3MSE and date of DQ	Date of last 3MSE
Deceased	Midpoint of date of last 3MSE and date of death	Date of death	Date of last 3MSE

MRI Data

The data consist of total brain volume data and ischemic lesion volume data presented by brain region. In addition, process data is also included from screening, recruitment, and the conduct of the MRI scan.

WHISCA Data

The contents of the battery and variable names are listed below. Note that negative values are possible for Card Rotations Test and PMA Vocabulary.

Contents of Battery

Test/Scale	Variable name	Variable description
Primary Mental Abilities (PMA) Vocabulary	pmavoc	Total correct minus one third number incorrect
Letter fluency (F, A, S)	flulet	Total correct
Category fluency (vegetables, fruits)	flucat	Total correct
Benton Visual Retention Test (BVRT)	bvreatot	Total figures with errors
California Verbal Learning Test (CVLT)	cvltca cvltcb cvlfrs cvlfrl cvlcrs cvlcrl	Total of 3 list A trials Total for List B Total for short-delay free recall Total for long-delay free recall Total for short-delay cued recall Total for long-delay cued recall
Digit Span Test	digit_forwardscore digit_backwardscore	Total score for digits forward Total score for digits backward
Card Rotations Test	crdrot	Total correct minus total incorrect
Finger Tapping Dominant Hand	tapdomin	Total score
Finger Tapping Nondominant	tapnondo	Total score
Positive and Negative Affect Schedule (PANAS)	panaspos panasneg	Mean score for positive Mean score for negative
Geriatric Depression Scale (GDS)	gdstot	Total score

5. Merging Data Files

The progression data (P2form, P3form, P4form) and the friend/family member interview can be merged with the Form 39 and adjudication data using two variables: ID and F39DAYS. The MRI data files can be merged using the ID variable.

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.

6. References

Silverman JM, Breitner JC, Mohs RC, et al. Reliability of the family history method in genetic studies of Alzheimer's disease and related dementias. Am J Psychiatry. 1986;143:1279-1282.