SECTION 18

DATA MANAGEMENT

INTRODUCTION

Data enter the WHI database (WHILMA) via one of three means: key-entry from a study form, direct entry onto a study screen (without a form), or electronically via scanning of a mark-sense form. Quality of data is assured by a variety of methods including review of completed forms before entry, training of data collection and management staff, participant file audits, and edit checks built into the data entry system. The Clinical Coordinating Center (CCC) also performs routine edits of the central database. Clinical Centers (CCs) must store hard copies of data collection forms in a readily accessible and secure manner, and respond to data queries from the CCC. Carefully recorded, entered, and maintained data are critical to the success of the study.

18.1   Form Flow

Overview

At each participant visit, the following general activities take place:

- CC staff place blank forms that will be used at the particular visit in the participant's file. Sheets of barcode labels with the participant's study ID are generated for use on the forms as needed.

- When the participant arrives, any completed forms that she has brought with her are added to the file. Interviewers and clinicians complete the visit-specific forms inside the file during the visit. They may complete additional forms and add them to the file during the visit. Specimen tracking forms, such as *Form 92 – Pap Smear*, may be removed from the file and temporarily filed elsewhere while results are pending. The CC staff transport the file about the CC; the participant file should not be in the possession of the participant.

- At the end of the visit, the file is sent to the area of the CC set aside for review and data entry activities. Avoid leaving files unsecured or in undesignated areas at the end of the day.

- The data entry staff reviews the file for completeness and accuracy, resolves any forms problems with other CC staff, then scans or enters the data. Most self-administered forms such as *Form 35 – Personal Habits Update* and *Form 38 – Daily Life* are reviewed for completeness only. (See *Section 18.2.3 – Review of Forms* for more detailed information.)

- The file is filed in the file room when data entry is complete.

18.1.1   Specimen Tracking

The following forms are used by the CCs to track specimens that are sent to outside labs for analysis, or to record results of tests that are conducted elsewhere:

- *Form 81 - Pelvic Exam/Pap Smear*

- *Form 82 - Endometrial Aspiration*

- *Form 83 - Transvaginal Uterine Ultrasound*

- *Form 85 - Mammogram*

- *Form 92 – Pap Smear*

- *Form 100 - Blood Collection and Processing*

The *Members with Missing Lab Results (WHIP0782)* (see *Vol. 5 - Data System, Appendix D - WHILMA Reports*) can be used to track specimens or pending tests if the encounter information from the top part of the form is key-entered shortly after the specimen is collected or the test is ordered. The *Members with Missing Blood Results* report can be used to track pending blood results. If desired, CCs can maintain a log of pending specimens or tests that includes the name of the facility doing the analysis or test, the date the specimen is sent or the test result requested, and the date the results are received by the CC.

Incomplete specimen forms can be filed in a central area accessible to CC staff members who will be responsible for completing the forms when results are received. As results are received, clerical staff can match them to the corresponding form and file them according to the staff member(s) responsible for completing the particular form. Clinical Centers can key-enter the forms in a manner that best fits the flow of the CC. Options include:

- Key-enter the entire form after the results have arrived and have been recorded on the form.

- Key-enter the encounter information from the form after the specimen is collected or results requested, then wait for the arrival of the results and key-enter the remainder of the form. This option allows you to use WHILMA to assist with specimen tracking.

After key-entry, both the form and results are filed in the participant's file.

18.1.2   Follow-Up Visits

Before Contact

For each follow-up phone contact or CC visit, CC staff pull the participant's file and add any blank forms that will be needed for the contact. Tasks to be performed at each follow-up contact are shown in the *Tasks Required at Visit (WHIP0144)* report that can be printed from WHILMA for any randomized participant. All forms needed for the follow-up visit are labeled with the participant's barcode and the entire file is given to the CC staff person overseeing the contact.

Day of Contact

On the day of the contact, CC staff members complete the applicable forms and review them before filing them in the participant's file and sending the entire file to Data Entry. Specimen tracking forms and test report forms that are initiated at follow-up visits are handled as described in *Section 18.1.1 - Specimen Tracking*.

18.2   Data Collection (Required)

Study data are collected and entered into WHILMA by certified staff at the CCs, using WHI data collection forms or direct entry screens developed at the CCC. CCs should not develop and use their own study forms without permission from the CCC. Exceptions are visit checklists and forms that are not administered to the participant or used to collect data from the participant.

This section primarily covers the collection, review, editing, and entry of data on study forms. For related topics, refer to *Section 2.11 – Interviewer Procedures* and *Vol. 5 – Data System*.

18.2.1   Data Collection Forms

The U.S. Government Printing Office (GPO) prints most of the WHI forms. To order these forms CCs complete *Form 170 – Forms Order* at scheduled times and submit the orders to the CCC. The CCC consolidates the orders from the CCs and sends a printing request to the GPO. Each CC is responsible for ordering a sufficient supply of forms to last until the next specified printing date. Some forms that are used infrequently are not printed. Instead, the CC will be provided with an original to photocopy as needed. The Table of Contents for *Vol. 3 – Forms* indicates which forms CCs receive from the GPO and which to copy at the CC.

Specific guidelines on completing each individual WHI form are contained in *Vol. 3 - Forms*.

Keep original, completed forms in the participant's file.

Note for ancillary study participants: Forms used only for an ancillary study (not WHI) may be filed in a separate ancillary study chart. For example*, Form 39 – Cognitive Assessment* collected at non-WHI specified times, such as AV4 and AV5, may be filed in a separate chart. *Form 39s* collected at AV3, 6, and 9 should be filed in the participant's WHI chart.

18.2.2   Recording Data on Forms

-   Use a pre-printed participant ID label (with barcode) on the front page of each form.

-   Complete key-entry forms in black or blue ink. Complete mark-sense forms using a #2 pencil.

-   Right-justify numerical data, using preceding zeros if necessary, for example:

$$\boxed{0\,|\,6}$$

-   Use standard rounding rules to round values to the appropriate number of decimal points unless otherwise instructed by the form or *Vol. 3 – Forms*. If a value ends in a fraction less than .5, round down to the next lowest whole number. If a value ends in a fraction equal to or greater than .5, round up to the next highest whole number. Record numbers as fractions **only** when there is a decimal point pre-printed on the form.

    If a participant gives a range of numbers in response to a question, and the interviewer is unable to pinpoint a value using interviewing techniques, take the midpoint of the range and use the standard rounding rules to round off as needed.

-   Record times as hours 1-12, with a.m. or p.m. indicated in the box provided. Do not use the 24-hour clock. Use leading zeros as necessary. For example, record 2:15 p.m. as:

    Time: $\boxed{0\,|\,2}$: $\boxed{1\,|\,5}$ ☐AM

                                                                                               ☒PM

If exact time of an event is unknown, use the best approximation. If minutes cannot be determined or estimated, record "00" in the space provided for minutes.

- Record dates in MM-DD-YY or MM-DD-YYYY format, where two digits are recorded for month, two for day, and two for year. In some cases, only month and year are requested (MM-YY or MM-YYYY). Add leading zeros as necessary to fill in the boxes:

$$\boxed{0 \quad 6} \text{-} \boxed{2 \quad 7} \text{-} \boxed{8 \quad 7}$$

(See *Section 13 - ECG Procedures* for specific instructions for entering dates into the MACPC.)

- When boxes are provided for recording an answer, clearly mark your choice with an "X." If the instructions indicate to "mark all that apply," mark as many of the responses following as are appropriate.

- Make sure that each item on a form is answered (except those items appropriately skipped in a skip pattern, e.g., "If yes, go to question 4"). If the information is unknown or unavailable, write "unk" next to the question and initial.

18.2.2.1   Recording "Unknown" Responses on Forms

Occasionally a woman will be unable to remember a date or other event that is requested on a form. Responses to questions involving non-date events that affect eligibility (for example, number of meals prepared outside the home) must be provided by the woman in order to determine eligibility. Dates affecting eligibility can be estimated as described below.

Interviewers and other CC staff should use probing techniques (see *Section 2.11.4.1 – Interview Techniques*) to obtain the woman's "best guess" response to questions. When all attempts fail to obtain an eligibility-related date, the estimation procedure described below is used. This method produces a conservative estimate for dates that affect eligibility; therefore it is best if the CC can encourage a woman to provide a reasonable date rather than estimating one.

- For dates that affect a woman's eligibility status:

  If the woman can't remember the year, record the current year.

  If the woman can remember the year, but not the month:

  > Record the current month if the recorded year is the current year.

  > Record "12" (December) if the recorded year is other than the current year.

  In either case, make a note next to the question indicating that the response is an estimate. Initial and date the note.

- For other "unknown" responses and dates that do not affect a woman's eligibility status:

  Use the above method for estimating the date or leave the response blank. Make a note next to the question on the form to indicate that the answer is an estimate or that the woman was unable to answer the question. Initial and date the note.

18.2.2.2   Recording Data on Mark-Sense Forms

The following guidelines should be used when recording data on mark-sense forms:

- Use a #2 pencil.

- Fill in the bubbles neatly and completely.

- Place the participant's ID label exactly where indicated on the front of the form. The mark-sense scanner cannot read the barcode if the label is crooked.

- Do not staple or punch holes in the *FFQ*. Other mark-sense forms have pre-punched holes. If additional holes are necessary, they should be punched only in the margins of the form.

18.2.3   Review of Forms

Review and edit all completed forms on a daily basis. Interviewers should review and edit forms before giving them to the data entry staff. Review all self-administered forms while the participant is still present in the CC, so that any missing pages can be given to the participant to complete. It is best to review interviewer-administered forms before the participant leaves the CC as well, so that clarifications can be made without waiting until the next visit or trying to reach the participant by phone.

Review clinic-administered forms (*7, 10, 17, 39, 44, 45, 80, 81, 82, 83, 84, 85,* and *100)* for:

- Responses omitted unintentionally.

- Answers that do not follow the specified skip pattern.

- Codes/answers incorrectly marked or circled.

- Inconsistencies between questions.

Review self-administered *Forms 35, 38,* and OS Follow-up *Forms 145, 146, 167,* etc. for problems that would interfere with scanning, such as incomplete erasures and multiple marks. You do not need to edit these forms for incorrect skip patterns or inconsistent answers and you do not need to contact the participant to obtain missing response.

In addition, review mark-sense forms for:

- Stray pencil marks.

- Incorrectly filled or unfilled bubbles.

- Torn or soiled edges that may cause the form to be rejected by the scanner.

- Incomplete erasures.

See *Section 10.2.3.2 – Pre-Scan Edit* for detailed instructions for reviewing *FFQs*.

Review *Forms 10A, 17A, 33,* and *33D for:*

- Responses omitted unintentionally.

- Answers that do not follow the specified skip pattern.

- Codes/answers incorrectly marked or circled.

- Inconsistencies between questions.

Return incomplete forms to a staff person so that they can contact the participant and fill in incomplete information.

18.2.4   Editing Forms

18.2.4.1   Editing Key-Entry Forms

Do not erase data on key-entry forms. Instead, draw a line neatly through the wrong answer, mark or enter the correct one, and place your initials and the date next to the correction. Indicate whether the changed answer is an update of a previously entered correct response or a correction of an error. For coded answers, draw a line through the incorrect code and print the correct code above the wrong code (or mark the correct box if a checkbox question). Use colored ink instead of pencil for making edits.

After reviewing self-administered forms with participants, interviewers record their staff ID codes on the forms. Data coordinators and other data entry staff cannot make changes to these forms. Any problems should be returned to the interviewer for resolution.

The data entry system is set up so that certain routine checks of the data are made at the time of entry at the CC. A warning message is displayed on the screen if a value is outside of the expected range. Warning messages may be overridden by pressing the enter key. A "fatal" error occurs when a value that is not an "allowable response" is entered in WHILMA (for example, a text response is entered when a numeric value is expected). You must delete the incorrect response and enter a correct response before proceeding with key-entry of the form.

18.2.4.2   Editing Mark-Sense Forms

When editing *FFQs*:

- Completely erase any incorrect answers and bubble in the correct answer.
- Remove stray pencil marks.
- Do not write any comments in the margin of the *FFQ*.
- Use white correction fluid to cover stray or incorrect marks that cannot be erased.

When editing other mark-sense forms:

- Before scanning: follow the same editing guidelines as for the *FFQ*.
- After the form has been scanned once, don't change the bubbled items if you are only updating a few responses. Instead:
  1. Write the correct response in the margin along with your initials and the date.
  2. Indicate whether the change is an update or correction.
  3. Give the form to the data entry person who will key-enter the changes.

18.3   Data Entry

As mentioned previously, data enter WHILMA in one of three ways: key-entry from forms, scanning of mark-sense forms and direct data entry (without a form).

Following participant visits to the CC, forms are kept in the participant's file during data entry (except for pending test/specimen forms which may be stored in a central area of the CC to await results).

18.3.1   Data Entry Workspace

Each staff person involved in data entry should have adequate work space to keep participant files and data forms separated according to where they are in the data entry process. It is recommended that each data entry staff person maintain participant files-in-process in these categories:

- Incoming participant files from the CC or "problem" participant files returned by interviewers after resolution.

- Active forms ready to be scanned or key-entered or in process of being key-entered.

In addition, file space is needed for "loose sheets," that is, data forms that are processed outside of the participant's file, for example forms used for specimen tracking (see *Section 18.1.1 - Specimen Tracking*). Loose sheets should also be sorted by the above categories: ready to be reviewed, ready to be (or in the process of being) entered or scanned.

Depending upon how a CC is set up, the following items may be useful in the Data Entry area:

- An "urgent" box, containing participant files for women who have a visit scheduled in the very near future. This would contain files of participants that require priority data entry attention.

- A mailroom-like set of "pigeonholes" or bins labeled with the names of various CC staff members. Files requiring review and attention by the various staff members would be stored here.

18.3.2   Direct Data Entry

Information that will be recorded using direct data entry (without a form) includes:

- Current Medications and Current Supplements

- Medication dispensing (HRT and CaD)

- Medication adherence (HRT and CaD)

- Medication inventory

Paper forms will be provided for back-up in case the computer is not functional, except for selecting HRT and CaD medication bottles for randomized participants and Medication Inventory which can only be done by the computer. For additional information on direct data entry, consult *Vol. 5 - Data System, Section 7 - Data Entry*.

18.3.3   Scanning of Mark-Sense Forms

Scannable forms include *Form 35 – Personal Habits Update, Form 38 – Daily Life,* and *Form 60 - FFQ*. (Others may be developed in the future.) If mark-sense forms do not need to be scanned immediately, they will be placed in the participant's file before sending the file to Data Entry.

18.3.4   Key-Entry

Incoming participant files from the CC can be divided between data entry staff members if there is more than one data entry staff person. Each data entry staff person is responsible for reviewing the files they will key-enter or scan. The following process is recommended:

1. Review all participant forms for completeness and/or accuracy as appropriate (see *Section 18.2.3 - Review of Forms*).

2. Return participant files with problems to the interviewer or clinical staff responsible. The entire participant file should be returned to the interviewer or clinical staff, even if only one or two forms have problems. Problem areas should be clearly flagged (self-adhesive notes are useful). It is recommended that problem participant files be stored in a central area in the CC where they are available to clinicians and interviewers, but where they won't be mistaken for completed participant files and filed by medical records staff. To aid in return of files to staff, it may help to set up bins labeled with names of individual staff members. Files needing attention by various staff members can be placed in these bins. Effort should be made to keep participant files in this central area rather than on individual staff members' desks, except when the files are being reviewed.

3. Problems with participant files are resolved by the appropriate staff persons and returned with flags intact to the data entry staff person who originally reviewed the file. The data entry person then only needs to review the flagged pages since the others were reviewed previously. These returned files will be added to the top of files to be entered, so that "oldest" data is always entered first.

4. After participant files are reviewed and problems resolved, the forms are entered or scanned by a data entry staff person. Enter only those tasks that have been done. Do not enter encounter data for missed tasks. This includes forms and procedures that are required but have been "refused by the participant. Exceptions are:

    •   If a *Form 44 – Current Medications* or *Form 45 – Current Supplements* is required, but at a visit, the participant is not taking any medications and/or supplements, it is appropriate to enter the encounter information with no medication (or supplement) data.

    •   When a participant is having a procedure done outside of the CC (such as a mammogram) or is awaiting results of a Pap smear, it is acceptable to complete the encounter portion of the form and enter it into WHILMA before the results are received. This allows you to use the empty encounter as a tracking device in conjunction with the *Missing Labs Report (WHIP0782)*. These empty encounters should be deleted if the procedure wasn't done or the results will not be available.

5. After entry, the data entry staff person records their initials next to the "K" (to indicate key-entry of a keyed form) or "S" (to indicate that a scannable form has been scanned) on the front page of the form. Once a form is entered, CC staff should not make changes to it without notifying data entry staff. Storage areas for forms and files in the key-entry process must be clearly labeled to avoid accidental disturbance by other staff members. Forms should be entered as soon as possible after they are completed, to reduce the risk of the form getting lost or being missed by data entry staff.

6. After data entry is complete and all problem areas resolved, active CT and OS participant files are sent to the file room for filing. Forms and files for participants who become ineligible during the screening process can be stored separately. If data from ineligible participant's screening forms has been entered into WHILMA, the forms can be discarded. If you do not data enter the forms, they can be archived. Individual participant's files can be condensed and placed in storage boxes in the interest of saving space.

Signed informed consents for ineligible participants should not be discarded before the end of the study. No forms for randomized or enrolled participants should be discarded. Off-site storage for archived files must be readily accessible.

18.3.4.1 Key-entry of Specimen/Test Forms

To facilitate the reconciliation of WHILMA data with subcontractor data, enter all of the following forms into WHILMA for specimens that you collect and procedures that you perform.

- *Form 86 – ECG*

- *Form 87 – Bone Density Scan* (Bone density sites only)

- *Form 100 – Blood Collection and Processing*

- *Form 101 – Urine Collection and Processing* (Bone density sites only)

Note: Form 100 must be entered into WHILMA even if local lab results are not received. You do not need to enter local lab results into WHILMA for participants who dropped out of the study before randomization or enrollment.

18.3.5 Data Security

Staff using WHILMA should log completely out of WHILMA when vacating their workstations and when leaving for the day. Minimizing WHILMA on the task bar does not provide adequate security.

Information contained in participants' files is accessible to all CC staff, but should be maintained in a manner that allows adequate control and tracking of the location of the participant file and prevents unnecessary access to confidential participant information.

Participant files must be accessible for CCC chart audits. If a CC has had a chart audit since mid-1998, the CCC will usually only audit forms from follow-up visits and will not need access to screening forms. However, corrections to older forms may need to be made for CCC central analysis.

Only WHI staff may have access to participant files and WHILMA data. Each staff member must sign a confidentiality statement, agreeing not to disclose any WHI participant information to persons outside of WHI without a signed authorization from the participant.

Non-WHI staff, including ancillary study coordinators may not have access to participant data unless there is a specific release form signed by the participant granting access to the data.

18.4   Quality Assurance

Data management procedures and processes are monitored on-site for quality by CCC Quality Assurance personnel. See *Section 19 - Quality Assurance* for a detailed listing of data management operations that are reviewed during site visits. WHI data quality is monitored centrally at the CCC.

In addition, CCs can run reports from WHILMA that help monitor the quality of data management. Please see *Vol. 7, Table 5.2 – CC Schedule for Data Monitoring (Required)* for a list of WHILMA reports to use in monitoring data quality.

Another helpful report is the *Timeliness of Key Entry (WHIP0774)* report. Clinical Centers can print this report and determine whether they meet the "acceptable" standard of 80% of all self-administered forms entered within two weeks of collection. The report is broken down by form, so that CCs can identify problems with specific forms.

CCs are encouraged to perform routine audits and data verification on a randomly selected sample of charts. The *Questionnaire Responses Report (WHIP0104)* can be printed from WHILMA and compared with the original WHI forms in the chart to identify data entry mistakes. Other things to watch for when auditing participant charts are listed in the document "Chart Audit Codes" which can be found in the "QA Resources" Public Folder of the Microsoft Outlook e-mail system.

This page left blank intentionally

**Section 18
Data Management**

Table of Contents