National Human Exposure Assessment Survey (NHEXAS)

Maryland Study

Quality Systems and Implementation Plan for Human Exposure Assessment

Emory University
Atlanta, GA 30322

Cooperative Agreement CR 822038

Standard Operating Procedure **NHX/SOP-D01**

Title: Data Flow Procedures

Source: Harvard University/Johns Hopkins University

U.S. Environmental Protection Agency
Office of Research and Development
Human Exposure & Atmospheric Sciences Division
Human Exposure Research Branch

Notice: *The U.S. Environmental Protection Agency (EPA), through its Office of Research and Development (ORD), partially funded and collaborated in the research described here. This protocol is part of the Quality Systems Implementation Plan (QSIP) that was reviewed by the EPA and approved for use in this demonstration/scoping study. Mention of trade names or commercial products does not constitute endorsement or recommendation by EPA for use.*

1 Title of Standard Operating Procedure

Harvard University/Johns Hopkins University Standard Operating Procedures:
D01 Data Flow Procedures, Rev. 1.0

2 Overview and Purpose

This SOP describes data flow and data tracking procedures to be used in the Harvard University/ Johns Hopkins University NHEXAS investigation.

3 Discussion

In addition to standard Chain-of-Custody data tracking, we will implement a computerized tracking program that will afford tracking of all samples at all locations. This procedure will be accomplished through data entry, computer networking, and other means so that up-to-date records of sample and sample data locations will be available to all interested parties.

The data file track.xls will be used to monitor the progress of data and each individuals participation on the computer. The chain of custody forms will be the source of information about this progress. See SOP G04 for details on the chain of custody forms and the physical aspects of tracking data. The data file particip.db although not directly involved with data tracking will be updated during the course of data tracking and so will be discussed here. The file particip.db will be used to cross-check the progress of the data. See Quality assurance procedures.

The workbook track.xls will contain two worksheets: the participation summary worksheet or dataset (PSDS) and the custody tracking worksheet or dataset (CTDS). The PSDS will keep tally on each individual's participation at each cycle and the CTDS will track the data collected for those participating. A description of the two worksheets follows:

Participation Summary Data Set: Every household contacted will be asked to respond to the questions in the descriptive questionnaire. This worksheet will contain information on every individual contacted who has at least filled out the Descriptive Questionnaire. Each record of the worksheet will contain 10 columns: Household ID (col 1.), and Yes/No response to whether they agreed to participate in the study (col 2), and if so whether they participated in cycles 1-8 (cols 3-10).

Custody Tracking Data Set: The information in this worksheet is derived from the Chain-of-Custody Forms. The purpose of this worksheet is to consolidate and put on-line the tracking information contained in the chain of custody forms. Each cycle will be divided into 4 two week segments called tracking units. Each tracking unit is assigned to a different worksheet. There will be 32 worksheets by the end of the study. As the samples reach the next stage of data processing it will be noted in the CTDS.

The first four columns of the worksheet identify the type of data being tracked in the record and the rest of the columns identify where the data are at the moment. The field definitions are:

1. Participant ID

2. Cycle Number.
3. Media/Questionnaire - Identifies which type of data that is being tracked in this record. For example, Personal Air Sample or Questionnaire, *etc*.
4. Compound/Questionnaire - The media sample will be measured for many different compounds. This variable identifies the specific compound or questionnaire being tracked in this record. For example, lead or the Baseline questionnaire.
5. Field - Date sampling started for the participant-cycle being monitored in this record.
6. FCC - Date the data is received at the Field Coordination Center.
7. Lab/Data Entry - Date the data is sent to the lab or data entry location. Depending on whether the information is a physical sample or questionnaire the data needs to pass on to the Lab or to data entry personnel. In some cases subdivisions may be needed to mark different stages lab samples take between labs. For example, soil needs to be sieved, extracted and then analyzed and these stages might occur at different labs. The movement through the different labs will be shown through a coded entry in this field.
8. HSPH-In - Date the initial data files have been returned from either the lab or data entry personnel.
9. HSPH-Tr - Date the data has entered the official database.
10. EPA - Date the data has been sent to EPA validation. This is not applicable to lab results from Harvard and SwRI and questionnaire response. This will automatically be noted in database.
11. HSPH-F - Date the data has been validated by the EPA. Once this field has been checked it implies that the data have been completely approved.

If there has been a failure in the data at some point along the line, such as blood samples not taken or sample containers broken in transit, the break in the process should be noted in the CTDS at the point where failure is discovered. For example, if blood samples were not taken, it would be noted in the FCC field.

4.Personnel Responsibilities

The Field Coordinator and Field Staff are responsible for recordkeeping and custody of samples and data in the field and Field Coordination Center (FCC), including shipping.

Data Entry Staff are responsible for receiving, coding, entering, and transmitting data.

Laboratory Staff are responsible for receiving samples and associated data, preparing extracts, analyzing samples, and transmitting data including analytical results.

The Quality Assurance (QA) Officer is responsible for ensuring the quality of samples and data.

The Project Data Coordinator is responsible for receiving and tracking data.

5.Required Equipment and Reagents

Paperwork: chain-of-custody forms, ID labels, logsheets, laboratory record books, questionnaires

Computers and software
- computers
- bar code readers and software
- *Paradox* Relational Database System Software v5.0

6.Procedures

6.1 Records Management System

6.1.1 General Guidelines

The goal is to generate quality data at all steps of the project.

6.1.2 Record Types

Types of records used in the project include:

- questionnaires
- chain-of-custody forms
- logsheets
- lab results
- computer database

6.1.3 Shipping and Receiving Data

The FCC staff will be responsible for collection, maintenance, and shipping of samples and questionnaires. For collection, storage and shipping of samples and logsheets see SOP F02-F12 and SOP G04. For collection, storage and shipping of questionnaires see SOP D02 and SOP G04. Once samples have been collected they will be shipped according to this schedule:

- Physical samples to appropriate lab
- Questionnaires except confidential questions to the Data Entry Company
- Logsheet copy 1 and confidential questions to Project Data Coordinator at Emory
- Logsheet copy 3 to Principle Investigator at Emory

All data records generated from analysis of samples and keypunched at the Data Entry Company will be sent to the Project Data Coordinator at Emory. Data may arrive via Federal Express or U.S. Mail, or by electronic mail. The Project Data Coordinator or his or her designate will check these sources daily for incoming data.

The Project Data Coordinator or his or her designate will log in data received on the log sheet and will be assigned an identification number (see SOP G03, "Identification Numbers for Samples and Forms").

The Project Data Coordinator or his or her designate will initiate a new chain-of-custody form for those data (see SOP G04, "Chain-of-Custody and Sample Tracking" for instructions).

The Project Data Coordinator or his or her designate will copy data received in or copied to magnetic format (computer disk), make a hard copy, and store the original disk with the project data files. He/she will download data received via electronic mail to disk and hardcopy.

6.1.4 Transcribing Data

Logsheet information and confidential questions will be transcribed at Emory by the Data Entry Supervisor and Data Entry Assistant. Details on the methods and quality assurance procedures are given in SOP D04.

Data entry for the questionnaires except the confidential questions will be sent to the Data Entry Company directly from the FCC and the results will be sent to Emory. The Data Entry Supervisor and Data Entry Assistant will take the data files sent from the Data Entry Company and incorporate the results into the CDS. See SOP D02 for details.

Lab results will arrive at Emory after samples have been sent out from the FCC. See SOP G05 for shipping of samples information. The results will be on computer files and the Data Entry Supervisor and Data Entry Assistant will take the data files sent from the lab and incorporate the results into the CDS. See SOP D03 for details.

6.1.5 Storage of Data

Questionnaires: Three hard copies of the Questionnaires minus the confidential questions and two hard copies of the confidential questions are created during the data creation process. The following describes the final location of each copy.

Original copy of confidential questions - Will be sent to Emory from FCC. Upon arrival the Data Entry Staff will store the originals in a secure location designated as "Confidential questions to be entered". Upon entry of questions into the CDS the Data Entry Staff will store the originals in the designated secure location for "Archived confidential questions". These will be saved for a minimum of 5 years. They may be destroyed after that time at the discretion of the EPA.

Photocopy of confidential questions - Will be created by FCC Clerk when data is to be sent to Emory. It will be stored at FCC by FCC Clerk in the designated location for confidential questions. This will be a secure area with access limited to certain FCC staff. These will be destroyed at the end of the study.

Original copy of Questionnaire - Will be sent to Emory from FCC. Upon arrival the Data Entry Staff will store the originals in the designated location for questionnaires. These will be saved for a minimum of 5 years. They may be destroyed after that time at the discretion of the EPA.

Photocopy 1 of Questionnaire - Will be created by FCC Clerk when data is to be sent to Emory. Will be stored at FCC by FCC Clerk in designated location for questionnaires. These will be destroyed at the end of the study.
Photocopy 2 of Questionnaire - Will be created by FCC Clerk when data is to be sent to Emory. Will be sent to Data Entry Company and sent on to Emory with computerized data files. Upon arrival at Emory the photocopies will be stored by the Data Entry Staff in the designated location for photocopies of questionnaires. These may be destroyed at the end of the study at the discrection of the Principle Investigator.

Lab Results: When results have returned from the lab, the Data Entry Staff will copy the computer files immediately to the directory for initial data. The disk and accompanying hard copy, if any, will be stored in a location designated "archived lab results" where they should not need to be removed except for quality assurance purposes.

Logsheets: Three copies of logsheets are created during the sampling process. Before sending out the orignal copy and copy 1 of the logsheets they will temporarily be stored at the FCC. They will be stored in a location designated for logsheets. The following describe the location for each copy.

Original Copy - Is sent by the FCC Clerk to the Principle Investigator at Emory. Upon arrival it will be stored in the location designated for original logsheets. These will be saved for a minimum of 5 years. They may be destroyed after that time at the discretion of the EPA.
Copy 1 - Is sent by the FCC Clerk to the Project Data Entry Coordinator at Emory. Upon arrival the Data Entry Staff will store Copy 1 in a location designated as "Logsheets to be entered". Upon entry of logsheets into the CDS the Data Entry Staff will store Copy 1 in the designated location for "Archived logsheets". These may be destroyed at the end of the study at the discretion of the Principle Investigator.
Copy 2 - Is retained at the FCC and will remain in the location designated for logsheets. These will be destroyed at the end of the study.

Chain of Custody Forms: After The chain of custody forms will be stored in location designated for chain of custody forms. The forms will be stored in this location immediately upon arrival. Each participant will be allocated their own "folder" and the chain of custody forms will then be placed in the appropriate folder. Within a folder, the chain of custody forms for the same data will be combined using paper clips.

Updating Official Database: As updated versions of the official database data files are created, a backup will be copied onto floppy disk and stored in the location designated "official database". Remaining disk space will be tabulated on the disk label and space permitting the disk may be used to add more files of the official database. The list of files on the disk will be entered on the disk label. Once a disk is full there will be no reason to remove the disk except to reinstall data files which may have been corrupted. See D02, D03 for details on official database.

Computerized versions of all data, except confidential questions will be found in three locations: The archive directory and associated back-up disks, the CDS directory and associated back-up disks, and the ADS directory and associated back-up disks. Confidential questions will not be stored on a computer to help guarantee security and instead will have two sets of disks. One mimicing the hard drive version and a back-up disk.

A single file will be maintained providing a link between the identifying number and the respondent. This file will be maintained, in encrypted form, for the duration of the data collection process. This is necessary to maintain records for contact of respondents. Upon completion of the data collection phase, the linking file will be

purged from available computer systems and stored in a locked filing cabinet (along with all hard copies of records with name identifiers) for a minimum of five years. After that time, at the discretion of the Environmental Protection Agency, the records may be destroyed.

Sample storage is covered in HSPH SOP G05 "Sample Storage and Shipping." Prior to disposing of any sample, the data will first be examined and analyzed, e.g. for outliers or any apparent discrepancies. In addition, it will be ensured that the sample analysis was carried out with all QA/QC steps followed (e.g., with regard to the calibration curve R2 and the concentration of the standard reference material analyzed).

6.2 Record Identification

A unique ID number will be assigned for each sample and questionnaire (see SOP G03 "Identification Numbers for Samples and Forms"). The digits of the ID number identify the household, target individual, geographic stratum, Cycle, and sample type.

Printed labels will show the ID number in bar-code and human-readable format. Identical labels will be affixed to a questionnaire or sample container and to all logsheets, Chain-of-Custody forms, and other paperwork corresponding to that questionnaire or sample. All data from questionnaires, activity diaries, etc, will be keyboard-entered using this ID number as an identifier.

Data analysis will proceed with respondents identified only by ID number. No data will be released with names of respondents intact. All publication or dispersal of data will be accomplished through summary information in a manner designed to ensure that no individual records will be identifiable with a respondent. Other procedures may also be implemented in consultation with the EPA or other government entity as may be required.

6.3 Data Tracking

The flow chart gives an overview of the flow of data.

6.3.1 Field Coordination Center (FCC)

The Telephone Interviewer will open a data record and enter household information on a Descriptive Questionnaire for each household telephoned or otherwise for which a contact has been attempted. For households which are contacted the Telephone Interviewer will ask the household to respond to the Descriptive Questionnaire. This procedure also generates a computer record of the household keyed by Household Identification number (HI).

The FCC Clerk will call participants to set an appointment for the start of sample collection for a Cycle. On agreement of a time the FCC Clerk will put together a Field Packet of household paperwork for field use, including labels, logsheets, Chain-of-Custody forms, and other necessary paperwork. See SOP F01 “Field Sampling -- General Information.”

Samples collected in the field are identified and tracked as outlined in the SOPs G03 "Identification Numbers for Samples and Forms" and G04 "Chain-of-Custody and Sample Tracking."

Upon return of samples to the Field Coordination Center, a data entry program will be invoked to allow the Field Technician, Field Interviewer, and Phlebotomist to log in samples. A computer record indicating this will be generated and stored.

The top sheet of each 3-part logsheet will be sent to the data entry location. Of the other two sheets, page 2 will be sent to the Principal Investigator and page 3 will be filed at the FCC.

6.3.2 FCC to Emory and Laboratories

A computer record will be generated indicating shipment of samples to appropriate analytical laboratories.

These records will be transferred via electronic means on a daily basis back to Emory. The Project Data Coordinator or designate will inspect these records on a daily basis. Database management systems described in the discussion will be used to produce summary statistics on rates of sample collection, to track samples, and to otherwise maintain close watch on Chain-of-Custody. Description of data entry into the PSDS and CTDS are given below.

Entering a new Participant-cycle
When the technicians enter a participant's home to set up for the particular cycle new records this information will be sent to Emory. Upon arrival the Data Entry Supervisor will add the information to the particip.db, the PSDS worksheet and the relevant worksheet of the CTDS. The information entered into particip.db and the PSDS worksheet is the participant ID and Cycle number. A template of the first 5 columns of the CTDS can be found in the worksheet *template*. This will be edited to contain the correct information for the new participant-cycle and the block of data copied to the CTDS.

Updating records
As information on the progress of sampling is returned to Emory, the Data Entry Supervisor will enter the progress of the data into the CTDS. This involves opening the track.xls workbook and recording the progress in the appropriate cell of the worksheet.

For cooperating laboratories other than Harvard, the Chain-of-Custody forms will be the main tracking system while the samples and data are in the hands of the laboratory. All data will be shipped to Emory in both hardcopy format and magnetic format (3.5" HD disks). If possible, transfer will also be effected through electronic mail systems (e.g., Internet).

6.3.3 Laboratories

Data to be processed at Harvard laboratories will be logged in on computer systems with direct access to Harvard's Network Systems. Data tracking will be effected

through direct access to the SPARCsystem in the Principal Investigator's Office.

All consumables such as reagents used in the Harvard laboratory will be logged and tracked so that the reagents used in the analysis of any sample can be identified. See SOP L01 "Purchase of Consumables."

6.3.4 Laboratories to Emory and EPA

Data from laboratory analyses will be sent to the Emory and EPA data analysis locations. Data from analysis done at Harvard and SwRI will not be sent to EPA. Data will be sent in hardcopy format, on disks, and by electronic mail.

Emory will do preliminary analyses of the unvalidated data.

EPA will validate the data from all labs except Harvard and SwRI and send validated data to Emory in the same formats as above.

Emory will analyze the validated data.

7.Quality Assurance Procedures

These procedures are, themselves, intrinsic quality assurance operations.

All primary recording documents (whether field logsheets, laboratory notebooks, or computerized records) will be initialed and/or signed as well as dated. This authentication of the primary records will enable the tracing back of collected data to the individual collecting them.

Weekly summaries of data throughput and an analysis of the location of all samples will be done.

8.References

Harvard University/Johns Hopkins University Standard Operating Procedures:
G02 Ensuring Confidentiality of Respondents' Records
G03 Identification Numbers for Samples and Forms
G04 Chain-of-Custody and Sample Tracking
G05 Storage and Shipping of Samples
F01 Field Sampling -- General Information
L01 Purchase of Consumables
D02 Questionnaire Data Entry and Preparation
D03 Lab Result Data Entry and Preparation
D04 Logsheet and Confidential Question Data Entry and Preparation
D05 Data Analysis