

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-D04

Title: Logsheet and Confidential Questionnaire Data Entry and
Preparation

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 Procedure:
D04 Logsheet and Confidential Questionnaire Data Entry and Preparation, Rev 1.1
2. Overview and Purpose
The purpose of this SOP is to describe how data is transcribed from logsheets and confidential questionnaire forms into the official database also known as the Complete Dataset (CDS). This SOP also describes the structure and purpose of the CDS and the Analysis-ready Dataset (ADS).
3. Discussion

Data will be entered into the CDS and ADS incrementally. To help ensure that all data is accounted for during the data entry process each participant's logsheets and confidential questions will be assigned to a *unit* of data. A unit of data will be handled together and processed as a unit. All logsheets and confidential questions collected from individuals who are participating during a specified interval of time will be defined as a unit of questionnaire data. See SOP D02 for more details.

Structure of the CDS.

The CDS is the official database. The CDS will be used to create any database used for analysis, e.g. used to construct the ADS, and will be the database sent when data is requested from other investigators. The confidential questionnaires, while part of the CDS, will be available to a limited number of individuals.

Initial data sets will be *Paradox* tables and each unit of data entered will be entered into a new file which will be appended later to the CDS. The initial data set files will have names nearly identical to the the corresponding CDS files except the file names will start with the letter *i*.

Information from logsheets will be contained in two files. The file logt.db will contain the data entered onto the logsheets required to translate the lab results into typical units of measurement used to analysis the data. The file logc.db will perform a similar, but expanded role as qxcom.db, where comments which flag unusual samples and correction information will be stored.

The confidential questions will be stored in five tables, see Appendix A for a complete list of table names and their contents. These tables contain names, address, telephone numbers of household members, and health and income conditions of the participant.

The CDS will reflect exactly what was written on the confidential questionnaires. Information recorded from logsheets in the CDS will include all data necessary to transform the data to the appropriate units of measurements and comments, if any, on collection problems.

Structure of the ADS

Confidential questions will not be part of the ADS.

Information from the logsheets will be indirectly part of the ADS, see SOP D03 for details.

File naming conventions

File name conventions are the same as for the questionnaires. See SOP D02 for details. Names and contents of the templates for the CDS tables are given in Appendix A of this SOP.

4. Personnel Responsibilities

- 4.1 Project Data Coordinator is responsible for
 - assuring accuracy and consistency of database
 - modifying data entry protocol if necessary, in writing
 - tracking confidential questionnaires and logsheets from collection to entry into database to storage.
 - notifying the Data Entry Supervisor of any modifications that affect working databases.
- 4.2 Data Entry Supervisor is responsible for
 - training and supervising Data Entry Assistants
 - tracking confidential questionnaires and logsheets through data entry and review
 - resolving problems and ambiguities that Assistants are unable to handle.
- 4.3 Data Entry Assistants are responsible for
 - logging in arriving data, checking for ID, storing confidential questionnaires and logsheets
 - data entry
 - reviewing data entry.

The same Assistant may do all of these jobs, but no one verify a stack s/he has entered.

5. Required Equipment and Reagents

- set of confidential questionnaires and logsheets and pen for marking
- computer with Borland Paradox V5.0 for Windows and database directories
- data in hardcopy format
- initial data set disks and back-ups
- CDS backup disks.
- data entry instruction sheet appropriate to the type of data being coded
- data entry log (forms with initials and dates for data entry; kept in Supervisor's logbook)

6. Procedures

Marking the Questionnaires

The interviewer will use a black or blue ink pen to mark the questionnaire. The response will be circled by the interviewer. If a change needs to be made in the marked response, the incorrect mark will be crossed out and initialed. The correct response will then be marked. The interviewer will give these instructions to the participants for questionnaires they mark.

Filling in the Logsheets

Discussion of logsheets and procedures for filling in the logsheets are given in the relevant Field SOP, SOP F01-SOP F12.

Processing Questionnaires and Logsheets at the FCC

The confidential questions will be separated from the questionnaires by the FCC Clerk on return from the field. They will be stored until a unit of data has been collected. See the discussion

section above for the definition of a unit of questionnaires and SOP D01 for details on storage procedures. When a unit of questionnaires has been collected the FCC Clerk will photocopy the questionnaires minus confidential questions twice and the confidential questions once.

The FCC Clerk will store the logsheets in the designated location (see SOP D01) until a unit of data has been collected.

The FCC Clerk will then send the originals of the questionnaires, confidential questions and the original copy of the logsheet to the Project Data Coordinator and copy 1 of the logsheet to the Principle Investigator. The photocopies of the confidential questions will be returned to the secure location described in SOP D01. The FCC Clerk will notify the Project Data Coordinator via e-mail that the confidential questionnaires and logsheets have been shipped.

Data Preparation of Logsheets at Emory

On receipt of the logsheets the Data Entry Assistant will store the logsheets in a labeled folder in the location designated for logsheets which need to be entered into the CDS. The Data Entry Assistant will note the arrival of the logsheets in the logbook and the CTDS.

The procedure will be the same for both tables, itlog.db and iclog.db; the table itlob.db will be used as an example. The steps in preparing the initial data sets for logsheets are:

- 1) The Data Entry Supervisor will assign the folder to an Assistant for entry into the database.
- 2) The Data Entry Assistant who will enter the data will:
 - i) take the folder with the logsheets and the initial data set and backup disks
 - ii) check whether a file with the appropriate filename exists; if not, rename a copy of the template of the itlog.db table to identify it as the latest version of the entered itlog.db table, e.g. itlog1E.db.
 - iii) enter the data from the folder into the appropriate fields in the file, according to the instruction sheet
 - iv) if there are any problems or ambiguities, mark each one with a post-it with an arrow and note pointing out the location of the problem; and continue to the next field until all data has been entered
- 3) the Data Entry Assistant will resolve any ambiguities found in previous step and return to the skipped fields and enter result. Clarification of resolution will be noted on logsheets and explanation of resolution in comments section of logsheet. Ambiguities will be resolved with the assistance of the Data Entry Supervisor and Project Data Coordinator.
- 4) The Data Entry Assistant will then record completion of data entry in the data entry logbook
- 5) The Data Entry Supervisor will assign the folder to a different Assistant to verify the data entry.
- 6) The Data Input Assistant who will verify the data entry will:
 - i) take the folder, data entry instruction sheet, and data and backup disks
 - ii) verify the data entry according to the instructions
 - iii) the Data Entry Assistant will resolve any ambiguities found in previous step and return to the skipped fields and enter result. Clarification of resolution will be noted on logsheets and explanation of resolution in comments section

- of logsheet. Ambiguities will be resolved with the assistance of the Data Entry Supervisor and Project Data Coordinator.
- iv) include in logbook a record of completion
 - v) Rename the file by dropping the *E* from the file name.
 - vi) The Data Entry Assistant will make a backup copy and submit it to the Data Entry Supervisor when data entry has been completed and any ambiguities have been resolved.
 - vii) The Data Entry Assistant will then record completion of verification in the data entry logbook

Data Preparation of Confidential Questionnaires at Emory

It is anticipated that very little data will need to be entered for each cycle for the confidential questionnaire data. This is because we expect only a small percent of participants to have health problems, be on medication, be pregnant or have large changes in income (so that they change income categories). If this presumption turns out to be false, the data entry process will be changed to the same method as described for the logsheets.

Two members of the Data Entry Staff will participate in entry simultaneously. They will follow these steps in data entry.

1. The Data Entry Supervisor will assign the confidential questions to be entered into the tables to two Data Entry Staff (the Data Entry Supervisor may be one of the staff).
2. A staff member will separate out those questionnaires which do not have a response to any of the confidential questions (or no change as a response).
3. The second staff member will check to make sure that the separation was performed correctly.
4. The table will be renamed to reflect that the new table is the next version.
5. One staff member will then enter the data into the appropriate table, while the second member checks the data as it is entered.
6. The new version will be copied to appropriate back-up disks.
7. A record of completion will be entered into the logbook for both staff members.
8. The confidential questions will be returned to the Data Entry Supervisor who will return the confidential question forms and disks to the designated storage location.

Entering new data units to the CDS

New records are added to copies of the latest version of the tables by using the Add/Append functions to put the results from the initial data files into the CDS. An introduction to these methods can be found on p. 180 and p. 187 of the *Paradox* User's Guide.

The procedure will be the same for all tables. The Data Entry Assistant will take the following steps in preparing the CDS tables for confidential questions and logsheets.

1. Gather all new units of initial data files to be entered. Then for each new unit of data:
2. Rename a copy of the last version of the CDS table to identify it as the latest version of the CDS table.
3. Take the initial data files and use the Tools|Utilities|Add command to add the initial table to the latest version of the relevant CDS table created in Step 2.
4. The new CDS will be copied to the CDS back-up disk. See SOP D01 for more details.
5. The old version of the CDS will be removed from the CDS directory so only the current version of the CDS is available.

6. A record of completion will be entered into the data entry logbook.

Correcting the CDS

See Quality Assurance Procedures for methods used to detect errors in the CDS. If a possible error is observed by the Data Entry Assistant, they will notify the Data Entry Supervisor who will notify the Project Data Coordinator. The Project Data Coordinator will decide whether the correction will be made and will notify the Principle Investigator about the decision. If the CDS needs to be corrected a renamed copy of the CDS will be made as described in the subsection *File Naming Conventions* in the discussion above. An example: if the CDS name is techw4.db then the renamed copy will be techw4a.db. The Data Entry Assistant will make the correction directly to the renamed CDS table and the nature of the correction and the reason for the error if known will be noted in the file conqx.db or logc.db. The updated file and the comment files will be copied and stored in the designated official database location. The old version of the CDS will then be archived and removed from the CDS directory. See SOP D01 for storage details.

Creating the ADS

Confidential questions will not be part of the ADS.

Information from the logsheets will indirectly be part of the ADS, see SOP D03 for details.

7. Training of Data Entry Assistants

7.1 General

Training and testing materials will be prepared and approved by the Project Data Coordinator and the Data Entry Supervisor. The Data Entry Supervisor will ensure that all Assistants have been trained and checked for the tasks that they will perform. Assistants may be trained individually or in groups.

Most Data Entry Assistants will be college students working part time. Depending on the needs of the project and the schedule of each Assistant, an individual Assistant will not necessarily be trained in all skills; for example; an Assistant might be trained in coding and reviewing coding, but not in data entry or verification.

7.2 Data Entry and Verification

Training materials will include:

- computer with data entry software having appropriately labeled fields
- disks for data and backups
- confidential questionnaires and logsheets
- data entry and verification instruction sheet
- data files ready for data entry, and data files that have been entered but not verified

During training:

- the Supervisor will demonstrate how to open a file, create a new file, find the appropriate fields for data entry, enter data, save a file, and rename a file
- the Assistant will read the instruction sheet, practice data entry, and submit the completed

- file to the Supervisor
- the Supervisor will check the file (using software to compare it with a correct file) and discuss any errors with the Assistant
- if the Assistant has made any errors, s/he will practice until s/he demonstrates 100% accuracy
- to learn verification, the Assistant will read the instruction sheet, practice verifying a file, and submit the completed file to the Supervisor
- the Supervisor will check the file (using software to compare it with a correct file) and discuss any errors with the Assistant
- if the Assistant has made any errors, s/he will practice until s/he demonstrates 100% accuracy
- when the Assistant has demonstrated 100% accuracy in data entry and verification, the Supervisor will log the Assistant into a list of Assistants authorized to do data entry and verification.

8. Quality Assurance Procedures

All personnel will have appropriate training.

Quality assurance procedures used by the Data Entry Company are the following: 1. Double entry techniques to ensure the correct entry of the questionnaires.

2. Use of Range limits on field inputs to reduce opportunity for incorrect entry.

The ranges used by the Data Entry Company will be specified by the Project Data Coordinator and will be used in the CDS tables. If an imported value is outside the stated ranges the Data Entry Assistant will automatically be notified by the *Paradox* software and the Data Entry Assistant will investigate the nature of the error and appropriate action will be taken. See SOP G06 on Problem Management.

To guarantee no complete records have been missed, the Data Entry Assistant will check the number of records in the temporary data file, which contains the new data, and the number of records in the old version of the official database to make sure the sum equals the number of records in the new version of the database. Discrepancies will be checked by looking directly at the initial data file sent by the Data Entry Company to see which records may have been lost or duplicated. If no error is observed the Data Entry Assistant will check to make sure that all old records were correctly incorporated into the new file.

Spot checks of the questionnaires will also be carried out. See SOP D05 for details.

At the conclusion of the study, a comparison of previous versions of the official database with the final version of the official database will be undertaken. Any discrepancies will be checked.

9. References

Borland Paradox Relational Database V5.0 User's Guide and Online Help.

Harvard University/Johns Hopkins University Standard Operating Procedure:

D01 Data Flow Procedures, Rev 1.0

Harvard University/Johns Hopkins University Standard Operating Procedure:

D02 Questionnaire Data Entry and Preparation, Rev 1.0

Harvard University/Johns Hopkins University Standard Operating Procedure:
D03 Lab Results Data Entry and Preparation, Rev 1.0
Harvard University/Johns Hopkins University Standard Operating Procedure:
D05 Exploratory Data Analysis and Summary Statistics, Rev 1.0

**Appendix A - List of Confidential Questionnaire and Logsheet Data files and
Description of Contents**

Initial Data files

ipdat.db - Confidential baseline questions: D1, D5a,b, D10, T5, B44.
iheal.db - Confidential health questions: B21, F8.
imed.db - Confidential medicine questions: F6,F7.
inam.db - Confidential Question D6a, names of individuals living in the household.
ivis.db - Names of technicians and sampling notes (includes name and address).
itlog.db - Data from logsheet for transformation of lab results.
iclog.db - Comments from logsheet.

The CDS files

pdata.db - Confidential baseline questions: D1, D5a,b, D10, T5, B44.
healt.db - Confidential health questions: B21, F8.
medic.db - Confidential medicine questions: F6,F7.
names.db - Names of individuals living in the household: D6a.
visit.db - Visit information (includes name and address).
conqx.db - Comments on confidential questionnaire
logt.db - Data from logsheet for transformation of lab results.
logc.db - Comments from logsheet or on logsheet data entry.

Appendix B - Set of Confidential Questions

Each questionnaire has an initial page which contains the name, address and phone number of the participant. This information is not going to added to the data set. The questions are:

- D1: Address
- D5a,b: Names of Individuals in household
- D10: Telephone number
- B21: Health information.
- B44: Income levels.
- F6: Medications: prescription
- F7: Medications: non-prescription.
- F8: Pregnancy
- T5: Indicate nearest major intersection.