

National Human Exposure Assessment Survey (NHEXAS)

Arizona Study

Quality Systems and Implementation Plan for Human Exposure Assessment

The University of Arizona
Tucson, Arizona 85721

Cooperative Agreement CR 821560

Standard Operating Procedure

SOP-UA-D-25.0

Title: Correcting Electronic Data (Hand Entry & Scanned)

Source: The University of Arizona

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.*

Correcting Electronic Data

1.0 Purpose and Applicability

The purpose of this procedure is to provide a standard method for correcting electronic data errors. The procedure defines (1) when electronic data may be corrected and by whom, (2) the process of correcting the data, and (3) the process of documenting the correction(s) made.

This procedure applies to all corrections made to working and master databases generated by NHEXAS Arizona, with the exception of certain global codes under specific circumstances. The latter are detailed below.

This protocol does not apply to the routine appendage of newly cleaned data processing batches to the master databases.

2.0 Definitions

2.1 DATA: Classified under this word are the following definitions: DATA, ELECTRONIC; DATA, ENTERED; DATA, LOGIC CHECKED; DATA, MISSING; DATA, PHYSICAL; DATA, RANGE CHECKED; DATA, RAW; DATA, VERIFIED; DATA APPENDAGE, DATA CLEANING, DATA CORRECTION, DATA PROCESSING BATCH, DATA PROCESSING ERROR, DATA RECORD, DATA VALIDATION.

2.1.1 DATA, ELECTRONIC: Data stored on some type of magnetic or optical medium (for example: floppy disk, hard disk, bernoulli, tape).

2.1.2 DATA, ENTERED: Electronic data that have been entered for the first time into a computer database. This is the product of "data entry."

2.1.3 DATA, LOGIC CHECKED: Data records that have been checked for and cleared of all apparent logical errors. This is the product of "logic checking."

2.1.4 DATA, MISSING: A datum or data that were applicable to a sample or question, but were not recorded on the physical data form at the time of initial observation; or one or more data records that were applicable to a data processing batch, but were not appended to the working or master database at the time of initial data appendage.

2.1.5 DATA, PHYSICAL: A datum or data written on a physical data form.

2.1.6 DATA, RANGE CHECKED: Data records where each variable in the record was compared against a pre-established valid range, and adjusted if necessary. This is the product of "range checking."

2.1.7 DATA, RAW: Electronic data that were entered and verified, but not yet cleaned. (See DATA, ENTERED and DATA, VERIFIED.)

- 2.1.8 DATA, VERIFIED: Electronic data that were re-entered into the same table and database into which it was originally entered, and compared against the original entered values. This is the product of "data verification."
- 2.1.9 DATA APPENDAGE: The process of adding newly cleaned data processing batches to a master database. This is distinct from "data correction." (See DATA CORRECTION.)
- 2.1.10 DATA CLEANING: This is the the process of locating and correcting data processing errors (see DATA PROCESSING ERROR below). They can be individual level errors in the electronic and physical data, or they can be system level errors in the data collection, packaging, coding, entry, and cleaning procedures themselves. This process is also referred to as "data validation."
- 2.1.11 DATA CORRECTION: A change made to a datum or data within a working or master database. The change can be the modification of the value in one or more data fields, the deletion of one or more data records, or the addition of one or more data records that were missing (see MISSING DATA). This is distinct from "data appendage" (see DATA APPENDAGE above.)
- 2.1.12 DATA PROCESSING BATCH (DP BATCH): A collection of household packets or physical data forms reviewed for quality assurance and ready for data entry. Each DP batch receives a unique numeric or alphanumeric code that is written on all forms in the DP batch and is entered into the database corresponding to that form.
- 2.1.13 DATA PROCESSING ERROR: This is an error that can occur at any level of data processing (see DATA PROCESSING LEVEL below). It is a procedural mistake, such as a duplicate data record, a typographical error, a logical error, or missing information.
- 2.1.14 DATA RECORD: In the context of this SOP, this is a row of data in a database.
- 2.1.15 DATA VALIDATION: See DATA CLEANING.
- 2.2 DATABASE: Classified under this word are the following: **DATABASE, EMPTY; DATABASE, HEALTH; DATABASE, MASTER; DATABASE, WORKING.**
- 2.2.1 DATABASE, EMPTY: A database structure that contains no data records.
- 2.2.2 DATABASE, MASTER: This is the accumulative database generated from validated data processing batches. Newly cleaned batches are appended to the master database. Copies of this database are used in analyses. All corrections made to copies of the master are made to the master database itself. Thus, it is the most complete and accurate database of its kind.

2.2.3 DATABASE, WORKING: A database earmarked for or in the process of cleaning that contains one or more data processing batches. When cleaned, this will be appended to the master database with the same name.

2.3 FIELD: This is an area on a data entry form (ie., screen) where datum from a physical form is entered (see FORM, DATA ENTRY below).

2.4 KEY VARIABLE(S): A variable or set of variables in a data record whose value or combined values make a data record unique from other data records in the same table or file.

2.5 NULL: A data field that contains the database symbol for system missing. This means that no value exists for this field.

3.0 References

Not applicable.

4.0 Discussion

Since there are multiple data personnel working independently, a standard method is necessary to insure a consistent approach to correcting data.

Documenting data corrections is necessary for two reasons. First, it satisfies the necessity for quality control (QC) and quality assurance (QA) pertaining to electronic data. The correction itself, the date the correction was made, and the author of the correction is documented. Furthermore, any correction is given written approval by the Project Data Coordinator. Thus, any data changed erroneously can be detected and fixed. The procedure assures the validity and consistency of corrections.

Second, the procedure meets data security needs. Electronic data can be accidentally damaged or lost at any time. Although a daily data backup system is in place, it does not take effect until late at night. Damage or loss can occur before the daily backup. A record of the daily corrections made facilitates restoration of any lost database.

5.0 Responsibilities

5.1 Project Data Coordinator

5.1.1 Supervision of Data Corrections

The Project Data Coordinator supervises corrections made to a database by a Student Data Assistant. He or she does this by providing solutions regarding unusual or ambiguous data, sometimes in consultation with the principle investigator, the Project Data analyst, and/or the Project Field Coordinator.

5.1.2 Written Approval of Data Corrections

The Project Data Coordinator gives written approval (ie., initials the data correction log form) for each correction made to a database by a Student Data Assistant.

5.2 Project Data Manager

The responsibility of the Project Data analyst is threefold. He or she (1) approves suggested corrections to the master databases, (2) makes any approved corrections to the master databases, and (3) makes corrections to the master databases that were generated from statistical analyses run on one or more copies of the master.

5.3 Student Data Assistant

5.3.1 The Student Data Assistant is responsible to the Project Data Coordinator.

5.3.2 The Student Data Assistant has custody of the data correction log forms for the data processing batch earmarked for or in the process of being cleaned, until they are submitted to the Project Data Coordinator for approval.

5.3.3 The Student Data Assistant is responsible for documenting any corrections made to the physical and/or electronic data on the appropriate data correction log form; and for printing any data correction files generated from range checked data.

5.4 Project Field Coordinator

The Project Field Coordinator is responsible for providing information regarding the standard operating procedures for the preparation of the household packets, data processing batches, and the collection of field data, when it is needed for data correction purposes.

5.5 Student Field Technician

The Student Field Technician is responsible for (1) providing specific information about the household packets on which he or she has participated, and (2) providing specific information about field data that he or she has helped to collect, when it is needed for data correction purposes.

6.0 Materials and Reagents

6.1 Materials

6.1.1 Data Correction Log Forms

(a) General Information

Each document or page must contain corrections for one database

only. This is because the data correction log forms are filed by database name.

(b) "Working Database Change Log Form" (Figure 1)

This form is used to document data corrections that have been made to the working database(s). It includes the database name, file(s) or table(s), key variables, the suggested change, the reason for the change, staff approval, whether a change was made to the source document, the initials of the person making the change, and the change date.

(c) "Master Database Change Log Form" (Figure 2)

This form is used to document data corrections that have been made to the master database(s). It includes the database name, file(s) or table(s), key variables, the suggested change, the reason for the change, staff approval, whether a change was made to the source document, the initials of the person making the change, and the change date.

(d) "Database Change Log Form" (computer printout) (Figure 3: Example)

This form is the equivalent of the working data corrections log form except that it is generated by the data cleaning application "DATA.EXE" while range checking a working database.

(e) Self-Designed Data Correction Form

The standard forms used to document data corrections are listed above. It is not always necessary nor preferable to use them, however. Data staff may document corrections through self-designed printouts or reports using statistical or database software if this is more efficient. If this is done, it is necessary to include the following information on the document:

1. Database name
2. Table name(s) (if R:BASE database)
3. File name (if dBASE database)
4. Key variable(s)
5. Old, incorrect value(s) in data field(s)
6. New, correct value(s) in data field(s)
7. Reason for the correction
8. Initials of the author of the correction(s)
9. Date correction(s) was made
10. Staff approval (initials) of correction(s)
11. Where database correction(s) reflects correction(s) to the source document

6.1.2 Cleaning Applications(a) Data.exe (date = 7/11/90; 5:37 pm; 124,782 bytes)

This executable C program works with all R:BASE databases in research phases three and four. It is used for data entry, data

verification, range checking, and data correction documentation.

When using Data.exe to range check the electronic data, then any value in a data field that is out of range will appear on the screen in reverse video. If a correction is made using the "F5" function key, then Data.exe "documents" the correction in a file called <database>.dup.

Printing the <database>.dup file yields a paper form entitled "Database Change Log Form" (see 6.1.1.(d) above).

6.1.3 Other

- (a) Ink pen for filling out the data correction log forms

6.2 Reagents

Not applicable.

7.0 Procedure

7.1 Preparations

Not applicable.

7.2 Standard Procedure

7.2.1 Corrections to Working Databases

- (a) Data Corrections Made During Range Checks with Data.Exe:

- 1. Corrections to Document

- A). Cases where the value in the data field is changed from a **null value** to any of the following:

- 1). A global code for **refusal**
- 2). A new value (non-global code)
- 3). A code for **temporary** missing datum (distinct from missing datum)

- B). Cases where the value in the data field is changed from a **non-null** value to any of the following:

- 1). A global code for **refusal, non-applicable, or missing**
- 2). A new value (non-global code)
- 3). A code for **temporary** missing datum (distinct from missing datum)

2. Corrections to NOT Document

A). Cases where the value in the data field is changed from a **null value** to any of the following:

- 1). A global code for **non-applicable**
- 2). A global code for **missing**

3. Print the <database>.cor file generated by Data.exe.

4. Initial and date each correction on the printout.

5. Submit the printout to the Project Data Coordinator for approval.

(b) All Other Data Corrections

1. Use the **working** data correction log form (Figure 1).

2. Follow the data correction documentation guidelines under 7.2.1.(a).1 and 7.2.1.(a).2 above.

3. Initial and date the **working** data correction log form.

4. Submit the log form to the Project Data Coordinator for approval.

7.2.2 Corrections to Master Databases

(a) Use a **master** data correction log form (Figure 2).

(b) Follow the data documentation guidelines under 7.2.1.(a).1 and 7.2.1.(a).2 above.

(c) Make corrections to a copy of the master database (referred to as the "new master").

(d) Log all corrections to the "new master" on master corrections log form(s).

(e) Compare the "new master" with the "old master" (the original master database) and resolve any discrepancies.

(f) Once the validity of the "new master" is established, overwrite the "old master" with the "new master."

7.3 Calculations

Not applicable.

7.4 Quality Control

7.4.1 Tolerance Limits

All corrections of data will be documented according to the rules and procedures outlined in this SOP.

7.4.2 Detection Limits

All documented corrections of data are detectable.

7.4.3 Corrective Actions

Any data discovered to be corrected without proper documentation will be brought to the attention of the Project Data Coordinator. If the correction(s) is still able to be documented, then the Data Coordinator will assign the task to the appropriate person.

8.0 Records

8.1 Record Forms (Attached)

Figure 1: "Working Database Change Log Form"

Figure 2: "Master Database Change Log Form"

Figure 3: "Database Change Log Form" (Example of Computer Printout Generated by Data.exe)

8.2 Location of Forms

8.2.1 Blank Correction Log Forms

A supply of blank correction log forms is in the lobby of room 128, where general forms are filed, and/or in the student data office.

8.2.2 Correction Log Forms: Unapproved

Correction log forms that have been filled out but not yet approved are stored in the Project Data Coordinator's office.

8.2.3 Correction Log Forms: Approved

Approved correction log forms are filed in notebooks according to database name.

DP Batch(es): _____

[illegible]

Form ID: UA-P-16.0-2.0

Corrections made to: ☐ Master ☐ Copy of Master **Why?:** ☐ QA Check ☐ Other:

Form ID: UA-D-25.0-1.0

Figure 3: "Database Change Log Form"
(Example of Computer Printout Generated by Data.exe)

DATABASE CHANGE LOG FORM: FIELD4

PASSU-7

HH10	VISIT	PAGE	tablename	var.name	from	to	reason	ok by	ch.date	SD
433	1	1	PASDATA	COMMENTS		38	FIELD BLANK	LJP	ATV080490	✓
433	1	1	PASDATA	COMMENTS		1	REPLICATE			✓
433	1	3	PASSIVE	STRDATE		86/81/91	OMITTED			✓
433	1	3	PASDATA	COMMENTS		19	OMITTED			✓
433	1	3	PASDATA	COMMENTS		19	OMITTED			✓
433	1	3	PASDATA	COMMENTS		19	OMITTED			✓
433	1	4	PASSIVE	STRDATE		86/81/91	OMITTED			✓
433	1	4	PASDATA	COMMENTS		19	OMITTED			✓
433	1	4	PASDATA	COMMENTS		19	OMITTED			✓
529	1	2	PASDATA	COMMENTS		38	FIELD BLANK			✓