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-31.1**

Title: Global Coding for Scanned Forms

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.*

Global Coding for Scanned Forms

1.0 Purpose and Applicability

This procedure defines the strategy for Global Coding of Scanned Forms. This procedure was developed for use in NHEXAS, the Border Study, and other Health and Environment Projects.

2.0 Definitions

2.1 BORDER STUDY : An alias for "Total Human Exposure Arizona: A Comparison of the Border Communities and the State" conducted in Arizona by the University of Arizona/Battelle/Illinois Institute of Technology consortium.

2.2 DATA CLEANING : The process of locating and correcting data processing errors. 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."

2.3 DATA FIELD : An area on a data entry form (i.e., screen) where datum from a physical form is entered.

2.4 DATABASE, MASTER : A cumulative 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 verified and then made to the master database itself. Thus, it is the most complete and accurate database of its kind.

2.5 DATABASE, WORKING : A database earmarked for or in the process of cleaning that contains one or more data processing batches. When cleaned and QA checked, this will be appended to the master database.

2.6 FORM, PHYSICAL [DATA] : The paper or "hard copy" version of a data form.

2.7 GLOBAL CODING : A set of standard codes used in data within all Health and the Environment Projects designating the status of a data field in three cases: (1) datum refused, (2) datum non-applicable, and (3) datum missing. It can also be a standard coding approach that pertains to questions with the same response structure.

2.8 HEALTH AND ENVIRONMENT PROJECTS (or H & E) : An umbrella title for all projects funded to M. D. Lebowitz and/or M.K. O'Rourke (or their designees) which examine purported or real relationships among environmental factors and any aspect of human health.

2.9 NHEXAS Arizona: Acronym for **N**ational **H**uman **EX**posure **A**ssessment **S**urvey,

a research project conducted in Arizona by the University of Arizona/Battelle/Illinois Institute of Technology consortium.

3.0 References

Teleform 5.0, Copyright 1991-1996 by Cardiff Software, Inc., San Marcos, CA.

4.0 Discussion

The scannable forms are designed primarily as "bubble field" forms. These bubble fields are defined within the program and are set to the appropriate global code values.

Hand writing recognition will be employed on the scannable forms. The box fields are programmed as alpha recognition, numeric recognition or either. Numbers are recognized with the greatest resolution. Therefore whenever possible codes are numeric. Some fields require alpha codes. Therefore, additional global codes must be defined. A summary table of questions needing specific codes and coding lists used are found in Table 1.

First, global coding schemes can be divided into two types: field-dependent, and fixed. Field-dependent global coding schemes are used in three situations: (1) subject refusals, (2) non-applicable data, and (3) missing data. Codes in these cases are "field-dependent" because the exact value of the global code depends on the length and type of the data field receiving the code.

Global coding schemes can further be divided by the type of response that is expected, that is, whether a field is defined as numeric, alpha or alpha-numeric. Numeric and alpha-numeric fields are coded as follows: -5 or -55 for refused values, -8 or -88 for non-applicable values and -9 or -99 for missing values with all lengths dependent on the individual field lengths. The same numbering scheme is used with bubble fields, but the number is positive and out of range (055 = refused, 088 = not applicable, and 099 = missing).

Examples:

(1) A subject refused to divulge the total household income. The variable *INCOME*, which is defined as a three-character numeric field, would receive the code *055* for refused.

(2) A person has not had any individuals in the household who smoke. The variable *PEOPSMOK*, which is defined as a three-character numeric field, would receive the default code of *-88*.

Alpha data is handled in much the same way but a series of letters is used instead of numbers. If the alpha-datum is refused by the subject, then the data field receives a code containing one or more X values depending on field length. If the alpha-datum is non-applicable, then the data field receives a code containing one or

more Y values. Finally, if the alpha-datum is missing, then the data field receives a code containing one or more Z values.

Examples:

(1) A subject refused to reveal the name of a child residing in the household. The variable *NAME*, which is defined as a seven-character alpha field, would receive a code of *XXXXXXX* for refused.

(2) A person lives on a dirt road with no street address, only a post box number. The data field for the variable *STRTNME* which is defined as an eight-character alpha-field would receive a code of *ZZZZZZZZ* for missing.

Second, the global coding scheme, whether field-dependent or fixed, follows the principle of mutual exclusivity. This means that fields receive codes that are never within or near the valid range for real data. For example, hand-entry data are coded with negative global codes (e.g., -55, -88, or -99). The use of negative numbers greatly decreases the chance that global codes are inadvertently included in calculations; further, if they were included, then detection of this mistake is very likely. Questionnaire data, in contrast, receive positive global codes. The codes are one to two characters greater than the highest valid response value for the question (e.g., for a question that has a one- or two-character response value, the global codes would be 055, 088, or 099). It is thus very unlikely that the global codes would be confused with real responses.

Technician ID numbers are used on all forms. These ID numbers are never reassigned and, therefore, are considered a type of global code. A current list may be found in Appendix A.

New Technician ID numbers may be created by the Tracking System Administrator using the User Account Maintenance module (UA-D-28.x).

Finally, certain constraints inherent in our primary database software (*TELEFORM version 5.0*) as well as our customized dictionary programs (UA-D-4.x) have shaped our global coding scheme (UA-D-31.x). Teleform will automatically fill blank fields with a predefined default code.

5.0 Responsibilities

The Project Data Coordinator is responsible for (1) ensuring consistency of global coding throughout all working databases used in the projects; (2) modifying global coding protocol if and when necessary, in written form; (3) notifying the Student Data Assistants of any modifications that affect working databases; and (4) notifying the Project Data Analyst of any modifications that affect master databases.

6.0 Materials and Reagents

6.1 Codes are to be written with a black felt tip pen only.

6.2 Questionnaires are put into a batch once they are coded and recorded on the Batch Description and Custody Recorded.

6.3 Those coding lists that are not in the Coding Lists notebook can be found on line in the /rsc53/TrackNHEXAZ/codes/ directory. The coding list that pertains to Global Coding is listed in section 8.0 Records, and includes Table 1.

6.4 Networked Computer Workstation that can access FoxPro.

6.5 Microsoft FoxPro Professional Edition version 2.6, Copyright 1989-1993 Microsoft Corporation.

6.6 Coding Program v1.0, developed in-house using FoxPro 2.6.

7.0 Procedure

7.1 Criteria for Using Field-Dependent Global Codes

7.1.1 Whenever possible bubble fields are used in all forms.
(a) A response of "Yes" is always coded as 001.
(b) A response of "No" is always coded as 002.
(c) Multiple choice responses are coded with zeroes preceding the selection number (e.g. 00x).

7.1.2 When to Code Data Field as Refused

(a) Subject has crossed out question or field technician has indicated that subject refused the question.
(b) Other source(s) indicate(s) that the question, physical form, or questionnaire was refused.

7.1.3 When to Code Data Field as Non-Applicable

(a) Field technician has written "N/A" on the question, physical form, or questionnaire.
(b) Sample cannot be taken due to the subject's particular situation. For example, no street name exists for a residence.

7.1.4 When to Code Data Field as Missing

(a) The sampler, questionnaire, or datum should have been taken, administered, or gathered according to the standard operating procedure, but was not.
(b) The sampler or questionnaire was lost prior to data entry.

(c) The sampling technique or question was determined to be irrevocably flawed.

7.2 Field-Dependent Global Coding Scheme: Follow the coding scheme given for the various data types listed below.

7.2.1 Date Fields

(a) Refusal: -5/-5/-5
(b) Non-applicable: -8/-8/-8
(c) Missing: -9/-9/-9

7.2.2 Time Fields

(a) Refusal
1. Eight-character field: -5:-5:-5
2. Five-character field: -5:-5

(b) Non-applicable
1. Eight-character field: -8:-8:-8
2. Five-character field: -8:-8

(c) Missing
1. Eight-character field: -9:-9:-9
2. Five-character field: -9:-9

7.2.3 Integer and Real Number Fields

The first character of the field is a minus sign; the remaining character(s) is/are filled with fives for refused, eights for non-applicable, or nines for missing.

7.2.4 Alpha-Numeric Fields

The first character of the field is a minus sign; the remaining character(s) is/are filled with fives for refused, eights for non-applicable, or nines for missing.

7.2.5 Technician Identification Numbers

(a) Technician ID numbers are a two-digit numeric field.
(b) A list of these numbers is located in Table 2.

7.4 Quality Control

7.4.1 Tolerance Limits

N/A

7.4.2 Detection Limits

The Project Data Coordinator ensures global coding consistency throughout all project working databases through the quality assurance checks outlined in SOP# UA-D-26.X.

7.4.3 Corrective Actions

Any discovered inconsistencies in global coding will be addressed and resolved by the Project Data Coordinator.

7.5 Coding Lists

7.5.1 Coding lists for each questionnaire are specified in SOP #'s UA-D-6.x, UA-D-7.x, UA-D-10.x, UA-D-11.x, UA-D-13.x, UA-D-31.x, UA-D-35.x, UA-D-37.x, UA-D-43.x.

7.5.2 Adding to Coding Lists that are computerized.

A. Log in to the computer system.
B. For DOS or Windows 3.11 operating Systems:
   1. Enter "net use l: rsc53/TrackNHEXAZ/codes" on a DOS or Win3.11 machine.
   2. Enter L: at the c:\ prompt
   3. Type "cd\codes" and press enter.
   4. Type "fox" and press enter.
   5. Now type "use <filename> order <index name>
   6. Type "browse" and press enter.
   7. Find the next available code number and press the <ESC> key.
   8. Type "append" and press enter.
   9. Enter the code number and it's description in the data entry fields that appear and press enter.
   10. Press the <ESC> key.
   11. Repeat from step 6 or type "quit" to exit the system.
C. Win95 machines
   1. Select "Map Network Drive" from the Network Neighborhood icon.
   2. Select "\\Lonicera\Track" from the Paths menu.
   3. Double click on the drive you have just mounted.
   4. Double click on the "Codes" sub-directory.
   5. Double click on the FoxPro icon.
   6. Follow the above procedure beginning with step 5 to look add new the codes.

7.6 Looking up On-Line Codes

A. For DOS/Win 3.11 Machines
   1. Log in to the network
   2. Type "coding" at the C:\ prompt.

3. Enter the appropriate database name as described in Table 1.
4. Enter the respondent's answer that needs to be coded.
5. The appropriate code will be displayed. If it is not found follow the procedures in section 7.5 to add a new code.

B. For Win95 Machines

1. Log in to the network
2. Double click on the FoxPro icon.
3. Double click on the file "FINDIT.PRG" to begin the coding program
4. Follow the same steps in section 7.6.A beginning with step 3.

8.0 Records

Any updates or exceptions to the global coding scheme will be documented and filed in the "Coding Instructions" notebook located in the Health and Environment offices.

Inclusions

Table 1. Questionnaires Needing Codes & Coding Lists
Table 2. Technician ID Numbers

Table 1. Questionnaires Needing Codes & Coding Lists

Questionnaire Type	Question Number	Coding List Name	Location	SOP # & Table# of Coding List
FOLLOW UP	6C	TYPE OF ANTACID MEDICATION	/rsc53/TrackNHEXAZ/codes/anatacid.dbf	UA-D-11.x / Table 2
FOLLOW UP	6B	TYPE OF CHELATING AGENT	/rsc53/TrackNHEXAZ/codes/chelate.dbf	UA-D-11.x / Table 4
FOLLOW UP	7C	TYPE OF CHROMIUM SUPPLEMENT	/rsc53/TrackNHEXAZ/codes/chromium.dbf	UA-D-11.x / Table 5
FOLLOW UP	7A	TYPE OF CALCIUM SUPPLEMENT	/rsc53/TrackNHEXAZ/codes/calcium.dbf	UA-D-11.x / Table 3
FOLLOW UP	11	TYPE OF DIET	/rsc53/TrackNHEXAZ/codes/diet.dbf	UA-D-11.x / Table 13
FOLLOW UP	6A	DIURETIC MEDICATION	/rsc53/TrackNHEXAZ/codes/diuretic.dbf	UA-D-11.x / Table 6
FOLLOW UP	6A-7D	DOSAGE ACCORDING TO LABELING	/rsc53/TrackNHEXAZ/codes/dosage.dbf	UA-D-11.x / Table 16
FOLLOW UP	1o	OTHER TYPE OF FILTERING DEVICE	/rsc53/TrackNHEXAZ/codes/filter.dbf	UA-D-7.x / Table 14
FOLLOW UP	6D	HORMONE SUPPLEMENT	/rsc53/TrackNHEXAZ/codes/hormone.dbf	UA-D-11.x / Table 7
FOLLOW UP	7D	MULTI VITAMIN SUPPLEMENTS	/rsc53/TrackNHEXAZ/codes/multi.dbf	UA-D-11.x / Table 8
FOLLOW UP	6A-7D	SPECIFIC MEDECINE NAME	/rsc53/TrackNHEXAZ/codes/m_cod.dbf	UA-D-11.x / Table 11
FOLLOW UP	6E	OTHER TYPES OF MEDICATION	/rsc53/TrackNHEXAZ/codes/other.dbf	UA-D-11.x / Table 9
FOLLOW UP	6E	OTHER UNIT OF MEASURE	/rsc53/TrackNHEXAZ/codes/o_unit.dbf	UA-D-11.x / Table 12
FOLLOW UP	7B	SELENIUM SUPPLEMENT	/rsc53/TrackNHEXAZ/codes/selenium.dbf	UA-D-11.x / Table 10
FOLLOW UP	6 & 7	CODING LIST NOTEBOOK - MEDICAL CATEGORY	DATA COORDINATOR'S OFFICE	UA-D-11.x / Table 15
TECHNICIAN QX	11	TYPE OF CARPETING	/rsc53/TrackNHEXAZ/codes/carpet.dbf	UA-D-35.x / Table 10
TECHNICIAN QX	6A	TYPE OF LAND AROUND HOME	/rsc53/TrackNHEXAZ/codes/area.dbf	UA-D-35.x / Table 12
TECHNICIAN QX	11	CLEANING PRODUCT USED	/rsc53/TrackNHEXAZ/codes/cleanmet.dbf	UA-D-35.x / Table 9
TECHNICIAN QX	6G	OTHER DRIPLINE	/rsc53/TrackNHEXAZ/codes/dripline.dbf	UA-D-35.x / Table 2
TECHNICIAN QX	6J	TYPE OF FOUNDATION	/rsc53/TrackNHEXAZ/codes/found.dbf	UA-D-35.x / Table 5
TECHNICIAN QX	6J	YARD MATERIAL	/rsc53/TrackNHEXAZ/codes/material.dbf	UA-D-35.x / Table 1
TECHNICIAN QX	6H	ROOF TYPE	/rsc53/TrackNHEXAZ/codes/roof.dbf	UA-D-35.x / Table 4
TECHNICIAN QX	10C	OTHER TYPE OF SAMPLING	/rsc53/TrackNHEXAZ/codes/o_samp.dbf	UA-D-35.x / Table 8
TECHNICIAN QX	6C	TYPE OF HOUSE SIDING	/rsc53/TrackNHEXAZ/codes/siding.dbf	UA-D-35.x / Table 3
TECHNICIAN QX	9	RELATION	/rsc53/TrackNHEXAZ/codes/relation.dbf	UA-D-35.x / Table 6
BASELINE QX	18B	WHERE TIME SPENT AWAY FROM HOME	/rsc53/TrackNHEXAZ/codes/away.dbf	UA-D-7.x / Table 9
BASELINE QX	37D	WHAT IS BURNED IN THE FIREPLACE	/rsc53/TrackNHEXAZ/codes/burnf.dbf	UA-D-7.x / Table 21
BASELINE QX	36C	WHAT IS BURNED IN THE STOVE	/rsc53/TrackNHEXAZ/codes/burns.dbf	UA-D-7.x / Table 20
BASELINE QX	14F	TYPE OF CLOTHING AT WORK	/rsc53/TrackNHEXAZ/codes/clothing.dbf	UA-D-7.x / Table 5
BASELINE QX	14G	DUST RESPONDENT EXPOSED TO	/rsc53/TrackNHEXAZ/codes/dust.dbf	UA-D-7.x / Table 6
BASELINE QX	31	OTHER FUEL CODES	/rsc53/TrackNHEXAZ/codes/fuel.dbf	UA-D-7.x / Table 19
BASELINE QX	14H	FUMES ENCOUNTERED IN THE WORK PLACE	/rsc53/TrackNHEXAZ/codes/fumes.dbf	UA-D-7.x / Table 7
BASELINE QX	27B	LOCATION OF HOUSE'S GARAGE	/rsc53/TrackNHEXAZ/codes/garage.dbf	UA-D-7.x / Table 17
BASELINE QX	14C	JOB TITLE/CLASSIFICATION	/rsc53/TrackNHEXAZ/codes/job.dbf	UA-D-7.x / Table 3
BASELINE QX	14D	JOB DUTIES	/rsc53/TrackNHEXAZ/codes/jobd.dbf	UA-D-7.x / Table 4
BASELINE QX	14B	BUSINESS	/rsc53/TrackNHEXAZ/codes/jobi.dbf	UA-D-7.x / Table 2
BASELINE QX	38I, 39G	MIX CODES	/rsc53/TrackNHEXAZ/codes/mix.dbf	UA-D-7.x / Table 23
BASELINE QX	30D	TYPE OF COOLER PADS	/rsc53/TrackNHEXAZ/codes/pad.dbf	UA-D-7.x / Table 18
BASELINE QX	43F	FLEA AND TICK PESTICIDES	/rsc53/TrackNHEXAZ/codes/petchem.dbf	UA-D-7.x / Table 24
BASELINE QX	38C	TYPE OF SURFACE TREATED	/rsc53/TrackNHEXAZ/codes/surface.dbf	UA-D-7.x / Table 22
BASELINE QX	19	METHOD OF GETTING TO WORK	/rsc53/TrackNHEXAZ/codes/transport.dbf	UA-D-7.x / Table 7
BASELINE QX	26C,D & E	SOURCE OF WATER	/rsc53/TrackNHEXAZ/codes/wtrsorce.dbf	UA-D-7.x / Table 13
BASELINE QX	26B	MAIN WATER SUPPLIER	/rsc53/TrackNHEXAZ/codes/water.dbf	UA-D-7.x / Table 12
BASELINE QX	H,N,S,V,W	DISEASE CODES NOTEBOOK	DATA COORDINATOR'S OFFICE	UA-D-7.x / Table 11
BASELINE QX	14j,16j,38f,39d	CODING LIST NOTEBOOK - PESTICIDES	DATA COORDINATOR'S OFFICE	UA-D-7.x / Table 8
N/A	N/A	LISTING OF DATABASES (THIS LIST)	/rsc53/TrackNHEXAZ/codes/codelist.dbf	N/A
ALL FORMS HAVING COMMENTS	N/A	COMMENTS MADE BY FIELD TECHS	/rsc53/TrackNHEXAZ/codes/comment.dbf	UA-D-31.x / Table 2
FOOD DIARY FOLLOW UP	12, 14	REASON SOMETHING WAS/WASNT DONE	/rsc53/TrackNHEXAZ/codes/reason.dbf	UA-D-10.x / Table 3
GLOBAL CODE	N/A	RELATION	/rsc53/TrackNHEXAZ/codes/relation.dbf	UA-D-31.x / Table 3
DESCRIPTIVE	P.7	RACE	/rsc53/TrackNHEXAZ/codes/	UA-D-6.x / Table 2
24 HOUR FOOD DIARY CHECK	A-N	DIET DIARY	UA-D-43.x Appendix A	UA-D-13.x / Table 2

Table 2. Technician ID Numbers

TECH ID	INITIALS	LOGIN NAME	USER NAME	SECURITY

The SOP is that all employees participating in the project are assigned a technicican ID, in the number range from 11 to 99. This purpose is to identifying who performs variouse tasks in the the field, and in coding for the NHEXAS Arizona project.