The Arizona Border Study

An Extension of the
Arizona National Human Exposure Assessment Survey (NHEXAS)Study
Sponsored by the Environmental Health Workgroup of the Border XXI Program

Quality Systems and Implementation Plan for Human Exposure Assessment

The University of Arizona
Tucson, Arizona 85721

Cooperative Agreement CR 824719

Standard Operating Procedure **SOP-UA-D-47.0**

Title: Cleaning: Questionnaire Feedback Form

Source: The University of Arizona

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

Title: Cleaning: Questionnaire Feedback Form	
Document No. UA-D-47.0	APPROVALS
☒ Full SOP ☐ Working SOP #pages 6	On Site Principal Investigator:
Issue Date: June 1997	Project QA Director:
Revision No.	Independent Reviewer:

Revision No:	On Site PI:
Revision Date:	Project QA Director:
Revision Made:	Independent Reviewer:
Revision No:	On Site PI:
Revision Date:	Project QA Director:
Revision Made:	Independent Reviewer:

Distributed To:	Revision No. 1	2	3	4	5	6

Form TP-2

Cleaning: Questionnaire Feedback Form

1.0 Purpose and Applicability

The purpose of this procedure is to define the particular steps involved in cleaning the electronic data generated from data entry of the Questionnaire Feedback Form. It applies to electronic data corresponding to the Questionnaire Feedback Form that was scanned and verified by the Data Staff for NHEXAS Arizona, the Border Study, or other Health and the 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 and field technician 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, ELECTRONIC: Data stored on some type of magnetic or optical medium (for example: floppy disk, hard disk).

2.4 DATA, ENTERED: Electronic data scanned into a data file using Teleform scanning software. Entered data are the product of "data entry."

2.5 DATA, VERIFIED: Electronic data that has cleared through the Teleform Verification process. In the Verification process Teleform reviews all of the entered data and displays any possible errors. These potential errors are reviewed by a Data Technician. Once all of the errors are fixed the data is saved to an ASCII file.

2.6 DATA CLEANING BATCH: A collection of electronic data, along with their corresponding physical forms. Data cleaning batches are formed after one or more data processing batches (see DATA PROCESSING BATCH below) are scanned. The data cleaning batches are then cleaned (see DATA CLEANING) and appended to the master database (as described in UA-D-44.X). Each data cleaning batch is assigned a numeric descriptor of the form MMDDYY, where MM is the month the batch was created, DD is the day the batch was created, and YY is the year the batch was created. If more than one batch is created on the same day, each batch after the first is assigned a descriptor of the form MMDDYY_N, where N denotes the batch as being the Nth batch created that day.

2.7 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.

2.8 HRP SITE: The **H**ealth **R**elated **P**rofessions building, located at 1435 North Fremont Avenue; Tucson, AZ 85719. This is an annex of the Arizona Prevention Center and the primary site of NHEXAS Arizona, the Border Study, or other Health and the Environment projects.

2.9 MASS DATA MASSAGE PROCESS (or MDM) = The data processing program used by NHEXAS Arizona, the Border Study, and other Health and Environment projects.

2.10 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

SOP UA-D-44.X (Operation Manual of the Mass Data Massage Program) (This Volume)

4.0 Discussion

A copy of the Questionnaire Feedback Form is given in Appendix A. This copy has each variable labeled for easy reference. Questionnaires and Forms are reviewed for completeness and accuracy prior to being transferred to the data section of the project. These checks are designed to generate flawless forms and questionnaires prior to data entry. Once entered and verified the data is subject to logic checks through the application of Dictionaries and other projects. A copy of the Dictionary for the Questionnaire Feedback Form is given in Appendix B. This dictionary defines the file structure for the Questionnaire Feedback database, as well as the logic and range checks that are performed on each field. For further reference, see SOP# UA-D-4.X.

Sometimes despite all best efforts, errors are found in data. These errors must be corrected to produce a valid database. At this time, all anticipated problems have been addressed. "Cleaning" protocols are developed concurrently with data entry to address the unanticipated problems. Specific corrections will be documented as outlined in SOP # UA-D-38.X.

5.0 Responsibilities

The Project Data Coordinator is responsible for writing the data cleaning instructions for the Questionnaire Feedback Form when data problems are found during the data entry process.

6.0 Materials and Reagents

6.1 Local Area Network
6.2 Purple Pen
6.3 Copy of Data Dictionary for specific form.

7.0 Procedure

7.1 Steps Followed

I. Data Technicians

A. Entering data

1. Scan and verify the form as described in UA-D-34.X.

B. Cleaning Data

1. Log into UNIX environment.
2. Type "data" to run the *MDM*.
3. Select form type:

   a. The user will enter "3" for Questionnaires.
   b. The *MDM* will then prompt the user for a specific questionnaire. The user will then enter "8" for Questionnaire Feedback Qx.

4. The *MDM* shall prompt the user for an action. The user then selects an operation from a given list. Operations include:

   a. Create a new batch
   b. Clean existing batch
   c. QA an existing batch
   d. [Examine Descriptive Master Database]

6. The user will select the Clean existing batch option:

   a. The *MDM* will list all possible batches to clean.
   b. If there are no batches, the program will inform the user, and prompt the user to press the <enter> key, which will return the user to the previous menu.
   c. If there are batches to be cleaned, the user will be prompted for a specific batch.
   d. Once a batch is selected, it is preprocessed (checking for errors) for cleaning, and the user is given information regarding the status of the preprocessing.
   e. If no errors are found within in the data, the *MDM* will send mail to the Data Coordinator informing him of the newly cleaned batch.
   f. If errors are found in the data, the user is informed as to the number of errors found within the batch.

i. For each error found in the batch, the user will be shown the key variables to locate the record containing the error, which is displayed between dashed lines.
ii. The user is then prompted for an action, the action being either the changing the erroneous value, the skipping of the error, a manual change to a different variable in the current record, a manual change to any variable in any record, or the user may quit.

g. If the user wishes to change the erroneous value:

i. Then "C" must be entered.
ii. The variable name is then displayed, and the *MDM* prompts for a new value.
iii. Once a new value is given, the user is given the option to accept the value given, to accept a value formatted by the *MDM*, or to abort the operation.
iv. If the operation is not aborted, the user is prompted for a reason for the change. Once a reason is given, the update is completed.

h. If the user wishes to skip the error:

i. He/she must enter "S".
ii. This will cause the *MDM* to skip the current error, but not the current record.

i. If the user wishes to change the value of a different variable in the current record:

i. He/she must enter "M".
ii. The user will then be prompted for the number of changes to be made.
iii. Once this number is entered, the *MDM* will then prompt for the variable to be changed.
iv. Once a valid variable is specified, the user is prompted for the new value.
v. Once a new value is given, the user is given the following options:

aa. Accept the value given
bb. Accept a value formatted by the *MDM*
cc. Abort the operation.

vi. If the operation is not aborted, the user is prompted for a reason for the change. Once a reason is given, the update is

completed.

j. If the user wishes to change the value of a variable in a different record:

i. He/she must enter "R".
ii. The user will then be prompted for the number of changes to be made.
iii. Once this number is entered, the *MDM* will then prompt for the key variable values of the record to be modified.
iv. Once valid values are given, the *MDM* prompts the user for the variable to be changed.
v. Once a valid variable is specified, the user is prompted for the new value.
vi. Once a new value is given, the user is given the following options:

aa. Accept the value given
bb. Accept a value formatted by the *MDM*,
cc. Abort the operation.

vii. If the operation is not aborted, the user is prompted for a reason for the change.
viii. Once a reason is given, the update is completed.

k. If the user wishes to quit, he/she must hit the "Q" key. The user is then returned to the menu defined in I.B.4.

7.2 Steps which are unique to the Questionnaire Feedback Form.

7.2.1 If the following variables are left blank, they are coded as N/A by the table.awk program: Num_intr, Num_clar, Respattd, Respintr, Quality, and Lang. The missing and refused values must be added during cleaning.

8.0 Records

8.1 All records are automatically generated by the *MDM*.

8.2 Records of all the forms in a cleaning batch are printed out when the batch is created. This list is then attached to the cleaning batch.

8.3 Records of the changes made to the data are located in the following directory tarred with their associated batch: /rsc53/NHEXAZdata/master/qx/qfqx/data.

8.4 Records of the cleaning batches which have been appended to the master data base are located in the following directory: /rsc53/NHEXAZdata/master/qx/qfqx. The list is kept in the file "read.me".

8.5 All changes to the hard copy of the form must be dated and completed in purple or red ink.

Inclusions:

Appendix A: Questionnaire Feedback Form (1 page)

Appendix B: Questionnaire feedback Dictionary (8 pages)

Appendix A: Questionnaire Feedback Form

QUESTIONNAIRE FEEDBACK FORM

National Human Exposure Assessment Survey

Itemnum

Form type: 09

Study: ○ 1. NHEXAS ○ 2. Border ○ 3. Study ○ 5. _____

Stage

Stage #: _____

Collapsed? Collapse (Y / N / 8)

Qx Type

Qx: ○ 1. Descriptive ○ 2. Baseline ○ 3. Follow-Up ○ 4. _____

Admin. by: Tech ID (Init.)

Oth Qx Code: Oth-Qx

HHID HHID F.S. IRN # HHIDFS IRN

Date of Interview: ___ / ___ / ___ (MO / DAY / YR) ~~Date Interv~~ Eventdate

1. Respondent First Name: F_name

2. Number of interruptions: ..... ○ None Num-inter ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ More than 5

3. Number of clarifications offered: ..... ○ None Num-clar ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ More than 5

4. Respondent attitude: ..... ○ Poor Respattd ○ Fair ○ Good ○ Very Good ○ Excellent

5. Respondent interest: ..... ○ Poor Respintr ○ Fair ○ Good ○ Very Good ○ Excellent

6. Overall quality of interview: ..... ○ Poor Quality ○ Fair ○ Good ○ Very Good ○ Excellent

7. Interview conducted in: Lang
   - ○ 1. English ..... *Stop.*
   - ○ 2. Spanish ..... *Go to Q #8 below.*
   - ○ 3. Other ..... *Indicate language (to right) and then go to Q # 8 below.* Language: _____ Lang. Code: Oth-lang

8. Indicate the interpreter : Intrpert
   - ○ 1. NHEXAS Field Team Member ..... *Stop.*
   - ○ 2. Hired Interpreter ..... *Full Name:* _____
   - ○ 3. HH Resident ..... *Go to Q #9 below.*
   - ○ 4. Other ..... *Go to Q #10 below.*
   - ○ 88. Not applicable (Default Code)

9. First name: _____ IRN-intr IRN #: ___

10. First name: _____ Age-intr Age: ___

Comments / Events / Observations: _____

Status

Office Use Only

Status Code: ○ 1 ○ 2 ○ 3 ○ 5 ○ 7 ○ 8 ○ 9

	Tech. ID	MO	DAY	YR
QC: (Init.)	QCBY	QCDATE	/	/
QA: (Init.)	QABY	QADATE	/	/

	Tech. ID	MO	DAY	YR
DE: (Init.)	DEBY	DEDATE	/	/

DP Batch: DPBATCH QFED1 QXV

NHEXAS Form ID: UA-T-1.0-1.0

46052

Appendix B: Questionnaire Feedback Dictionary

```
* Logic checks reviewed by Mary Kay on 05/07/96
* Last updated by Jared on 06/25/97

K HHID (F6.0) ' ', HHIDFS (A1) ' ', IRN (F2.0) ' ', Evntdt_m (F2.0) ' ', Evntdt_d (F2.0) ' ', Evntdt_y
(F2.0) ' ', QxType (F2.0) ' ', Oth_Qx (F2.0) ' '

Q HHID      1      1   I6
T Household Identification

Q HHIDFS    1      7   A1
T Household Family Schism
Y 'A'  First Schism - default value

Q IRN       1      8   I2
T Individual Identification Number
Y -5 Refused
Y -8 not applicable
Y -9 missing
R (01,10)
P Formstat (F3.0) ' '
L X (any(Formstat,01,02,03,05,09) and (range(IRN,01,10) or any(IRN,-5,-8,-9)))

* The following date field has been broken up into mm/dd/yy
* Q Evntdate  1      10   A8
* T Administration Date

Q Evntdt_m  1      10   I2
T Administration Date, Month
R (01,12)
Y -9  Missing

Q Evntdt_d  1      13   I2
T Administration Date, Day
R (01,31)
Y -9  Missing

Q Evntdt_y  1      16   I2
T Administration Date, Year
R (95,99)
Y -9  Missing

Q Itemnum   1      18   I2
T Form Type
Y 09   Default Value

Q Study     1      20   I2
T Study
Y 01   NHEXAS
Y 02   Border
```

```
Y 03  _____
Y 04  _____
Y 05  _____

Q Stage     1     22   I1
T Stage #:
R (1,6)

Q Collapse   1     23   I2
T Collapsed?
Y 01  Y
Y 02  N
Y 08  8
P Stage (F1.0) ' '
L X (any(Stage,1,4,8) and Collapse = 8)
L X (any(Stage,2,3,6) and any(Collapse,1,2))

Q QxType     1     25   I2
T Qx:
Y 01  Descriptive
Y 02  Baseline
Y 03  Follow-Up
Y 04  _____

Q Oth_Qx     1     27   I2
T Oth Qx Code:
Y -5 Refused
Y -8 Not Applicable
Y -9 Missing
R (01,25)
P Qxtype (F2.0) ' '
L X (any(Qxtype,01,02,03) and Oth_Qx=-8)
L X (Qxtype=04 and (Oth_Qx=-9 or range(Oth_Qx,01,25)))

Q TechID     1     29   I2
T Tech ID
Y -9
Y -8
Y 11
Y 12
Y 13
Y 14
Y 17
Y 21
Y 26
Y 28
Y 30
Y 33
Y 34
Y 37
Y 38
Y 39
```

```
Y 42
Y 43
Y 44
Y 45
Y 49
Y 51
Y 52
Y 53
Y 55
Y 56
L X (any(Qxtype,01,02,03) and (TechID=-9 or range(TechID,11,56)))
L X (Qxtype=04 and (range(TechID,11,56) or any(TechID,-8,-9)))

Q F_name    1    31   A15
T Respondent First Name
V q1
P Formstat (F3.0) ' '
L X (Formstat=01 and not(F_name='YYYYYYYYYYYYYYY'))
L X (any(Formstat,02,03,05,09) and (not(F_name='YYYYYYYYYYYYYYY') or
+ F F_name='YYYYYYYYYYYYYYY'))

* If this variable is left blank, the N/A code is added by table.awk.
* The refused and missing values must be added during cleaning.
Q Num_intr   1    46   I2
T Number of interruptions
V q2
Y 00   None
Y 01   1
Y 02   2
Y 03   3
Y 04   4
Y 05   5
Y 06   More than 5
Y 55   refused
Y 88   not applicable
Y 99   missing
P Formstat (F3.0) ' '
L X (Formstat=01 and any(Num_intr,00,01,02,03,04,05,06,55,99))
L X (Formstat=03 and any(Num_intr,00,01,02,03,04,05,06,55,88,99))
L X (any(Formstat,02,05,09) and Num_intr=88)

* If this variable is left blank, the N/A code is added by table.awk.
* The refused and missing values must be added during cleaning.
Q Num_clar   1    48   I2
T Number of clarifications offered:
V q3
Y 00  None
Y 01  1
Y 02  2
Y 03  3
Y 04  4
Y 05  5
```

```
Y 06  More than 5
Y 55   refused
Y 88   not applicable
Y 99   missing
P Formstat (F3.0) ' '
L X (Formstat=01 and any(Num_clar,00,01,02,03,04,05,06,55,99))
L X (Formstat=03 and any(Num_clar,00,01,02,03,04,05,06,55,88,99))
L X (any(Formstat,02,05,09) and Num_clar=88)

* If this variable is left blank, the N/A code is added by table.awk.
* The refused and missing values must be added during cleaning.
Q Respattd   1      50   I2
T Respondent attitude
V q4
Y 01  Poor
Y 02  Fair
Y 03  Good
Y 04  Very Good
Y 05  Excellent
Y 55   refused
Y 88   not applicable
Y 99   missing
P Formstat (F3.0) ' '
L X (Formstat=01 and any(Respattd,01,02,03,04,05,55,99))
L X (Formstat=03 and any(Respattd,01,02,03,04,05,55,88,99))
L X (any(Formstat,02,05,09) and Respattd=88)

* If this variable is left blank, the N/A code is added by table.awk.
* The refused and missing values must be added during cleaning.
Q Respintr   1      52   I2
T Respondent interest
V q5
Y 01  Poor
Y 02  Fair
Y 03  Good
Y 04  Very Good
Y 05  Excellent
Y 55   refused
Y 88   not applicable
Y 99   missing
P Formstat (F3.0) ' '
L X (Formstat=01 and any(Respintr,01,02,03,04,05,55,99))
L X (Formstat=03 and any(Respintr,01,02,03,04,05,55,88,99))
L X (any(Formstat,02,05,09) and Respintr=88)

* If this variable is left blank, the N/A code is added by table.awk.
* The refused and missing values must be added during cleaning.
Q Quality    1      54   I2
T Overall quality of interview
V q6
Y 01  Poor
Y 02  Fair
```

```
Y 03  Good
Y 04  Very Good
Y 05  Excellent
Y 55   refused
Y 88   not applicable
Y 99   missing
P Formstat (F3.0) ' '
L X (Formstat=01 and any(Quality,01,02,03,04,05,55,99))
L X (Formstat=03 and any(Quality,01,02,03,04,05,55,88,99))
L X (any(Formstat,02,05,09) and Quality=88)

* If this variable is left blank, the N/A code is added by table.awk.
* The refused and missing values must be added during cleaning.
Q Lang      1      56   I2
T Interview conducted in:
V q7
Y 01  English
Y 02  Spanish
Y 03  Other
Y 55   refused
Y 88   not applicable
Y 99   missing
P Formstat (F3.0) ' '
L X (Formstat=01 and any(Lang,01,02,03,55,99))
L X (Formstat=03 and any(Lang,01,02,03,55,88,99))
L X (any(Formstat,02,05,09) and Lang=88)

Q Oth_lang   1      58   I2
T Indicate language (to right) and then go to Q#8 below
V q7a
Y -5 refused
Y -8 not applicable
Y -9 missing
R (01,10)
P Lang (F2.0) ' '
L X (any(Lang,01,02,88,99) and Oth_lang=-8)
L X (Lang=03 and range(Oth_lang,01,10))

Q Intrpert   1      60   I3
T Indicate the interpreter
V q8
Y 001    NHEXAS Field Team Member
Y 002    Hired Interpreter
Y 003    HH Resident
Y 004    Other
Y 055   refused
Y 088   Not applicable (Default C
Y 099   Missing
P Lang (F2.0) ' '
L X (any(Lang,01,88,99) and Intrpert=088)
L X (any(Lang,02,03) and any(Intrpert,001,002,003,004,055,099))
```

```
Q IRN_intr   1     63   I2
T IRN #: (of interpreter)
V q9
Y -5 refused
Y -8 not applicable
Y -9 missing
R (01,13)
P Intrpert (F3.0) ' '
L X (Intrpert=003 and range(IRN_intr,01,13))
L X (any(Intrpert,001,002,004,055,088,099) and IRN_intr=-8)

Q Age_intr   1     65   I3
T Age: (of interpreter)
V q10
Y -55 refused
Y -88 not applicable
Y -99 missing
R (01,80)
P Intrpert (F3.0) ' '
L X (Intrpert=004 and range(Age_intr,10,80))
L X (any(Intrpert,001,002,055,088,099) and Age_intr=-88)

Q Formstat   1     68   I2
T Status Code:
Y 01  Cmp
Y 02  N Cmp
Y 03  P Cmp
Y 04  Re-col
Y 05  Ref
Y 07  Dest
Y 08  N/A
Y 09  Miss

Q QCBY       1     70   I2
T QC: Tech ID
Y 11
Y 12
Y 13
Y 14
Y 17
Y 21
Y 26
Y 28
Y 30
Y 33
Y 34
Y 37
Y 38
Y 39
Y 42
Y 43
Y 44
```

Y 45
Y 48
Y 49
Y 51
Y 52
Y 53
Y 55
Y 56

*Q QCDATE 1 72 A8
T QC: Date

Q QCDATE_M 1 72 I2
T QC Date: Month
R (01,12)

Q QCDATE_D 1 75 I2
T QC Date: Day
R (01,31)

Q QCDATE_Y 1 78 I2
T QC Date: Year
R (95,99)

Q QABY 1 80 I2
T QA: Tech ID
Y 11
Y 12
Y 13
Y 14
Y 17
Y 21
Y 26
Y 28
Y 30
Y 33
Y 34
Y 35
Y 37
Y 38
Y 39
Y 42
Y 43
Y 44
Y 45
Y 48
Y 49
Y 51
Y 52
Y 53
Y 55
Y 56

```
*Q QADATE     1     82    A8
T QA: Date

Q QADATE_M    1     82    I2
T QA Date: Month
R (01,12)

Q QADATE_D    1     85    I2
T QA Date: Day
R (01,31)

Q QADATE_Y    1     88    I2
T QA Date: Year
R (95,99)

Q DEBY        1     90    I2
T DE: Tech ID
Y 16
Y 19
Y 32
Y 46
Y 50
Y 60
Y 54
Y 55

*Q DEDATE     1     87    A8
T DE: Date

Q DEDATE_M    1     92    I2
T DE Date: Month
R (01,12)

Q DEDATE_D    1     95    I2
T DE Date: Day
R (01,31)

Q DEDATE_Y    1     98    I2
T DE Date: Year
R (96,99)

Q DPBATCH     1     100   I3
T DP Batch
R (900,999)

Q QXV         1     103   A5
Y 'QFED1' Pre-filled value
```