

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-IIT-A-9.0

Title: Sampling Weight Calculation

Source: The University of Arizona

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

STANDARD OPERATING PROCEDURE FOR SAMPLING WEIGHT CALCULATION

This Standard Operating Procedure (SOP) uses data that have been properly coded and certified with appropriate QA/QC procedures by the University of Arizona NHEXAS team.

Objective

Calculate sampling weights. The sampling weights are needed to obtain weighted statistics of the NHEXAS data.

Introduction

Three types of sampling weights are defined here:

1. **Single-stage Weight:** It is the inverse of the probability of selection of each sampled unit at a particular stage.
2. **Total Weight:** Total weight of a sampled unit at a particular stage is calculate by 2 different ways depending on the following conditions:
 - *If no weight adjustment has been done in the previous stages:* The total weight is the product of the single-stage weights of that sampled unit at that stage and all other previous stages.
 - *If sampling weights in the previous stages have been adjusted:* The total weight is the product of the single-stage weight of that sampled unit at that stage and the adjusted total weight of that unit at the previous stage.
3. **Adjusted Total Weight:** It is the total weight which has been adjusted to best represent the population of interest. This type of weight is the one to be used in the calculation of weighted statistics of data at each stage.

A diagram which explains development of sampling weights in each survey stage of NHEXAS is presented in Table 6-1.

Table 6-1: Sampling Weight Diagram-Survey Stages

STAGE	ACTIVITY	SINGLE-STAGE WEIGHT	TOTAL WEIGHT
Survey 1	Select 50 Tracts from the State of Arizona	$W_{i,1}$	
Survey 2	Select 250 Blocks from the 50 Tracts	$W_{i,2}$	$WT_{i,2} = (W_{i,1})(W_{i,2})$
Survey 3	Select 1225 HHs from the 250 Blocks	$W_{i,3}$	$WT_{i,3} = (WT_{i,2})(W_{i,3})$
Survey 4	Select 1 primary respondent per HH	$W_{i,4}$	$WPDES = (WTDES^1)(W_{i,4})$

Calculation of Single-stage Weights

In random sampling, each sample unit has equal probability of selection. When n units are selected from a stratum which has N units, the probability of selection of each unit is:

$$\pi = \frac{n}{N}$$

(6-1)

When the "Probability Proportional to Size" or PPS design is applied, sample units which are different in size have unequal probability of selection. In NHEXAS, the number of occupied housing units (OHU) are the measure of size. When unit i is selected from a stratum which has N units, the probability of selection of that unit, π_i , is:

$$\pi = \frac{(\text{size of the unit})}{(\text{total size of all N units})}$$

(6-2)

When n units are selected, the probability of selection of each unit is multiplied by n and becomes the frequency of selection, π_i^* :

¹ WTDES is the adjusted total weight of unit i at the survey stage 3, $WT_{i,3}$. Sampling weight adjustment are explained in SOP#10.

$$\pi_i^* = \frac{(n)(\text{size of the } i^{\text{th}} \text{ unit})}{(\text{total size of all } N \text{ units})}$$

(6-3)

The sampling weight of unit i , W_i , is the reciprocal of the probability of selection or frequency of selection of that unit :

$$W_i = 1/\pi_i \quad \text{or} \quad 1/\pi_i^* \quad (6-4)$$

The equations for weight calculations of each survey stage in NHEXAS are as follow:

Stage 1 : Selection of tracts from each county (PPS design)

$$W_{i,1} = \frac{(\text{OHU} / \text{cnty})_i}{(\text{tract selected} / \text{county})_i (\text{OHU} / \text{tract } 2)_i} \quad (6-5)$$

Stage 2 : Selection of blocks from each tract (PPS design)

$$W_{i,2} = \frac{(\text{OHU} / \text{tract } 2)_i}{(\text{block selected} / \text{tract})_i (\text{OHU} / \text{block } 2)_i} \quad (6-6)$$

Stage 3 : Selection of households from each block (random sampling)

$$W_{i,3} = \frac{(\text{OHU} / \text{block } 2)_i}{(\text{OHU selected} / \text{block})_i} \quad (6-7)$$

Stage 4 : Selection of one primary respondent from each household (random sampling)

$$W_{i,4} = \frac{(\text{member} / \text{OHU})_i}{1} = (\text{member} / \text{OHU})_i \quad (6-8)$$

Calculation of Total Weights

In a multi-stage sampling, weighted statistics of the samples at any particular stage can be calculated by using the total weight of each unit at that stage. That total weight is equal to a product of the single-stage weights resulted from that stage and all the proceeding stages. Mathematically:

$$WT_{i,m} = \prod_{s=1}^m W_{i,s} \quad (6-9)$$

where $WT_{i,m}$ = total weight of unit i at stage m , and $W_{i,s}$ = single-stage weight of unit i at stage s .

Equivalently, a total weight of unit i at stage m can be considered as a product of the single-stage weights resulted from that stage and the total weight of that unit at stage $m-1$. Mathematically:

$$WT_{i,m} = (WT_{i,m-1})(W_{i,m}) \quad (6-10)$$

If the total weight at stage $m-1$ is adjusted for nonresponse or adjusted with other methods, the adjusted total weight will be used to calculate the total weight at the next stage. Therefore, the equation becomes:

$$WT_{i,m} = (WT_{i,m-1,adjusted})(W_{i,m}) \quad (6-11)$$

Variable List

Variable	Description
COUNTY	County I.D., according to the Census.
COUNTY 2	County I.D., after some changes are made (see details in Procedure).
TRACT	Tract I.D., according to the Census.
TRACT 2	Tract I.D., after some changes are made (see details in Procedure).
BLOCK	Block I.D., according to the Census.
BLOCK 2	Block I.D., after some changes are made (see details in Procedure).
HHID	Household I.D. , according to the Census.
RESPONSE	Response status (enrolled or refused to answer the Descriptive Questionnaire).
OHU/CNTY	Number of occupied housing unit (OHU) per county.
OHU/TRACT	Number of OHU per tract.
OHU/TRACT 2	Number of OHU per tract, according to TRACT 2 .
HH ENU/BLK	Number of household enumerated by "field truthing" per block
VAC REP/BLK	Number of vacant household per block
OHU/BLOCK	Number of OHU available for contact per block. This is equal to HH ENU/BLK minus VAC REP/BLK .
OHU/BLOCK 2	Number of household available for contact per block, according to BLOCK 2 .

Variable	Description
<i>TRACT SELECTED/CNTY</i>	Number of tracts selected per county, according to <i>TRACT 2</i> .
<i>BLOCK SELECTED/TRACT</i>	Number of blocks selected per tract, according to <i>TRACT 2</i> and <i>BLOCK 2</i> .
<i>OHU SELECTED/BLOCK</i>	Number of OHU selected per block, according to <i>BLOCK 2</i> .
<i>MEMBER/OHU</i>	Number of member per OHU.
<i>W_{i,1}</i>	Single-stage weight of OHU i, resulted from sampling in stage 1.
<i>W_{i,2}</i>	Single-stage weight of OHU i, resulted from sampling in stage 2.
<i>W_{i,3}</i>	Single-stage weight of OHU i, resulted from sampling in stage 3.
<i>W_{i,4}</i>	Single-stage weight of the primary respondent in OHU i, resulted from sampling in stage 4.
<i>WT_{i,3}</i>	Total weight of OHU i at stage 3.

Procedure

1. In SPSS, open ORIGINAL.DOX, delete all variables except the following: **TRACT**, **BLOCK**, **HHID**, and **RESPONSE**. The data will then be saved as a new file called WEIGHT STRUCTURE MAIN.
2. In Excel, open WEIGHT STRUCTURE MAIN, the following variables will be added into the file and their values will be entered: **COUNTY**, **OHU/CNTY**, and **OHU/TRACT**. Data for the 3 variables will be obtained from the 1990 Census data file **C_TRACT.DAT** sent to IIT by the UA research team. Next, the following variables will be added into the file and their values will be entered: **HH ENU/BLK** and **VAC REP/BLK**. Data for the two variables will be obtained from a document called "NHEXAS RECRUITMENT LOG SUMMARY" which was sent to IIT by the UA research team. Also, a variable called **OHU/BLOCK** will be created. Its value, for each household, is equal to the value in **HH ENU/BLK** minus the value in **VAC REP/BLK**.
3. Since there are combination of tracts and blocks, variables for new tract and block i.d. will be created and called **TRACT 2** and **BLOCK 2**, respectively. The combination data is obtained from the document "NHEXAS RECRUITMENT LOG SUMMARY". For **TRACT 2**, all values will be the same as the original tract i.d. except in Santa Cruz county where its two tracts (9962 and 9964) will be combined. They are considered as one tract with a given tract i.d. "99629964". For **BLOCK 2**, all single blocks will have the same block i.d. as in the original block i.d., while all combined blocks will have a new set of i.d. There is a total of 36 groups of combined blocks. Each of them will be given a block combination number. The new block i.d. will then be equal to the block combination number plus 9900. For example, Pima county has 6 selected blocks: 101, 307, 113, 128, 119, and 201. The last two blocks are indicated as combined blocks. This combination was given a combination number equal to 1. Therefore, these two blocks will be considered as one block with an i.d. 9901. Two important criteria for weight calculation purposes are:
 - a) A block cannot be considered as a single block and a component of combined blocks at the same time. For example, tract 4 in Yavapai county which has 4 single blocks: 107, 229, 220, and 231; and a group of combined blocks: 231+311+312 is considered as having 3 single blocks: 107, 229, and 220; and a group of combined blocks: 231+311+312.
 - b) A block cannot be in different groups of combined blocks at the same time. For example, tract 9611 in Navajo county has 2 single blocks 135 and 316; and 3 groups of combined blocks: 147+218, 211+218, and 239+218. Since block 218 appears in all 3 groups, the groups must be combined. As a result, this tract is considered as having 2 single blocks: 135 and 316; and a group of combined blocks 147+211+239+218.

The new tract and block i.d., which are identified by the variables **TRACT 2** and **BLOCK 2** will be used for the rest of the weight calculation procedure.

4. When using the SUDAAN program, it is required that the number of Secondary Sampling Unit (SSU) selected from each Primary Sampling Unit (PSU) is more than

one. In NHEXAS, the PSU is county and the SSU is tract. Therefore, any county which has only one tract selected must be combined with another county. Based on geographical attributes of the counties, combinations of counties were made and the new county i.d. are:

Combined Counties	New I.D. in COUNTY 2
1 and 17	1017
9 and 11	9011
15 and 25	15025
19 and 23	19023

Similarly, the new county i.d. which is identified by **COUNTY 2** will be used for the rest of the weight calculation procedure.

5. In **WEIGHT STRUCTURE MAIN**, a new variable **OHU/TRACT 2** will be created. Its values will be the same as **OHU/TRACT** except in tract 99629963 where the value is the summation of OHU in tract 9962 and 9963. In other words, this variable contains the numbers of OHU per tract which correspond to **TRACT 2**. Next, a new variable **OHU/BLOCK 2** will be created. For single blocks, the variable will have the same values as **OHU/BLOCK**. For groups of combined blocks, each group will have a value which is the summation of OHU in each combined blocks. In other words, this variable contains the numbers of OHU per block which correspond to **BLOCK 2**. The summation process will be done using available functions in Excel.
6. In **WEIGHT STRUCTURE MAIN**, using functions available in Excel, the following variables will be created: **TRACT SELECTED/CNTY**, **BLOCK SELECTED/TRACT**, and **OHU SELECTED/BLOCK**.
7. Finally, in **WEIGHT STRUCTURE MAIN**, calculate the following sampling weights:
 - $W_{i,1}$, calculated by using Eq. 6-5.
 - $W_{i,2}$, calculated by using Eq. 6-6.
 - $W_{i,3}$, calculated by using Eq. 6-7.
 - $W_{i,4}$, calculated by using Eq. 6-8.
 - $WT_{i,3}$, calculated by using Eq. 6-9.

Spreadsheet Format

In **WEIGHT STRUCTURE MAIN**:

Column	Variable
1	COUNTY 2
2	TRACT 2
3	BLOCK 2
4	HHID
5	RESPONSE

Column	Variable
6	OHU/CNTY 2
7	OHU/TRACT 2
8	HH ENU/BLK
9	VAC REP/BLK
10	OHU/BLOCK , calculated from (HH ENU/BLK) – (VAC REP/BLK)
11	OHU/BLOCK 2 , which is the values from OHU/BLOCK adjusted according to BLOCK 2
12	TRACT SELECTED/CNTY
13	BLOCK SELECTED/TRACT
14	OHU SELECTED/BLOCK
15	MEMBER/OHU
16	W_{i,1} , calculated from (OHU/CNTY) / [(OHU/TRACT 2) × (TRACT SELECTED/CNTY)]
17	W_{i,2} , calculated from (OHU/TRACT 2) / [(OHU/BLOCK 2) × (BLOCK SELECTED/TRACT)]
18	W_{i,3} , calculated from (OHU/BLOCK 2) / (OHU SELECTED/BLOCK)
19	W_{i,4} , which is equal to (MEMBER/OHU)
20	WT_{i,3} , calculated from W_{i,1} × W_{i,2} × W_{i,3}