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-IIT-A-10.0**

Title: Sampling Weight Adjustment

Source: The University of Arizona

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

STANDARD OPERATING PROCEDURE FOR SAMPLING WEIGHT ADJUSTMENT

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.

Objectives

Adjust sampling weights in order to:

1. account for nonresponse.
2. obtain estimates that are comparable to the Census estimates.
3. obtain estimates that represent each season equally.
4. obtain the adjusted total weights to be used with data at each level of the questionnaire and environmental sampling, and with the calculated exposure values.

Introduction

There are 3 types of sampling weights discussed in this SOP[1]:

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.

[1] Calculation of sampling weights are explained in details in SOP #9.

A diagram which explains development and adjustment of sampling weights in each stage of NHEXAS is presented in Table 7-1. The section following Table 7-1 contains descriptions of each adjustment method used with NHEXAS data.

Table 7-1: Sampling Weight Diagram

STAGE	ACTIVITY	RESPONSE	OUTCOME	SINGLE -STAGE WEIGHT	TOTAL WEIGHT	ADJUSTED TOTAL WEIGHT[2]	
						DESCRIPTION	CALCULATION
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 & Sampling 1	Select 1225 HHs from the 250 Blocks and ask all HHs to complete DQx	Response = 955 HHs Nonresponse = 270 HHs	DQx-HH Level data	$W_{i,3}$	$WT_{i,3} = (W_{i,1})(W_{i,2})(W_{i,3})$	Weight adjusted for nonresponse using Simple Weighting Class	$SWCWT_{i,3} = (WT_{i,3})(AF_{SW}[3])$
						Weight adjusted by Poststratification	$WTDES = (SWCWT_{i,3})(AF_{PS})$
Survey 4 & Sampling 1	Select 1 primary respondent per HH from the 955 HHs responded	Response = 955 persons	DQx-Person Level data	$W_{i,4}$	$WPDES = (WTDES)(W_{i,4})$		
Sampling 2	Ask the primary respondents in all 955 HHs that completed DQx to complete BQx	Response = 525 HHs Nonresponse = 430 HHs	BQx-HH Level data			Weight adjusted for nonresponse using Logistic Regression	$LRWT_{i,3} = (WTDES)(AF_{LR})$
						Weight adjusted by Poststratification	$WHBAS = (LRWT_{i,3})(AF_{PS})$

[2] Adjusted total weight is the total weight which has been adjusted for nonresponse, and/or adjusted with other methods. It is to be used with data in order to get weighted statistics. For example, to find the weighted mean of family income, a file containing BQx-HH Level data and the adjusted total weight WHBAS will be prepared. The weighted mean will be obtained by using SUDAAN to process the file.

[3] AF = Adjustment Factor. Subscripts SW, PS, LR, AW, and SF stand for Simple Weighting Class, Poststratification, Logistic Regression, Annual Weighting, and Single Factor, respectively.

Table 7-1: Sampling Weight Diagram

STAGE	ACTIVITY	RESPONSE	OUTCOME	SINGLE -STAGE WEIGHT	TOTAL WEIGHT	ADJUSTED TOTAL WEIGHT[2]	
						DESCRIPTION	CALCULATION
			BQx-Person Level data	$W_{i,4}$	WPBAS = (WHBAS)($W_{i,4}$)		
Sampling 3	Ask all 525 HHs that completed BQx to participate in the intensive Qx and environmental sampling	Response = 403 HHs Nonresponse = 122 HHs				Weight adjusted for nonresponse using Logistic Regression	LRWHBAS = (WHBAS)(AF_{LR})
						Weight adjusted by Poststratification	WHCOR = (LRWHBAS)(AF_{PS})

Table 7-1: Sampling Weight Diagram (cont.)

STAGE	ACTIVITY	RESPONSE		OUTCOME	SINGLE -STAGE WEIGHT	TOTAL WEIGHT	ADJUSTED TOTAL WEIGHT	
							DESCRIPTION	CALCULATION
Sampling 4	Randomly[4] select HHs from the 403 HHs for the intensive Qx and environmental sampling	DDQx	297 HHs	DDQx data			For every Qx and media, the starting weight is WHCOR. For each Qx or media, the starting weight is adjusted using Single Factor, and then Annual Weighting. Example: DDQx Weight adjusted using Single Factor adjustment = WQDD1 = (WHCOR)(AF_{SF}) Weight adjusted by Annual Weighting = WQDD2 = (WQDD1)(AF_{AW})	WQDD2 = (WHCOR)(AF_{SF})(AF_{AW})
		FQx	299 HHs	FQx data				WQFU2 = (WHCOR)(AF_{SF})(AF_{AW})
		TAQx	308 HHs	TAQx data				WQTA2 = (WHCOR)(AF_{SF})(AF_{AW})
		TQx	367 HHs	TQx data				WQTC2 = (WHCOR)(AF_{SF})(AF_{AW})
		indoor air	120 HHs	measured concentration				WEIA2 = (WHCOR)(AF_{SF})(AF_{AW})

[4] Due to time constraint and limited fund, not all HHs that agree to the environmental sampling were sampled. The assumption is made that these HHs were randomly selected for samplings and the Single Factor adjustment is used to maintain the sum of the sampling weights.

		dermal wipe	149 HHs	measured concentration				WEDW2 = (WHCOR)(AF_{SF})(AF_{AW})
		floor dust	208 HHs	measured concentration				WEFD2 = (WHCOR)(AF_{SF})(AF_{AW})
		sill wipe	67 HHs	measured concentration				WESW2 = (WHCOR)(AF_{SF})(AF_{AW})
		yard soil	270 HHs	measured concentration				WEYS2 = (WHCOR)(AF_{SF})(AF_{AW})
		food	157 HHs	measured concentration				WEFO2 = (WHCOR)(AF_{SF})(AF_{AW})

Adjustment Methods for Sampling Weights

1. Nonresponse Adjustment by Simple Weighting Class

The procedure for this method begin with selection of one variable, among variables containing information of all samples (respondent and nonrespondents), to organize the samples in several classes. The criteria for an appropriate variable selection are: 1) the classification resulted in at least 20 – 30 respondents in each class, and 2) the respondents and nonrespondents are more alike within classes than between classes with respect to their response to the survey. For each class, a control total is calculated from the summation of the total weight of all samples in that class. The control total of each class divided by the summation of the total weight of all respondents in each class is called an adjustment factor. The product of the respondent's total weight and the adjustment factor is the adjusted total weight. The sum of the adjusted total weights is equal to the sum of the total weights of all respondents before the adjustment.

This adjustment is appropriate when only little information is obtained from the nonrespondents, e.g. when adjusting for the nonrespondents in the Descriptive questionnaire. More detail about this method can be found in many literature papers and books, including Kalton, G. and Maligalig, DS., (1991).

2. Nonresponse Adjustment by Logistic Regression

The method of logistic regression analysis involves the calculation of the predicted probability to response, θ_i , using the following relationship:

$$\theta_i = 1 / \{ 1 + \exp [- (\beta_0 + \Sigma \beta_j X_{ij})] \} \qquad (7\text{-}1)$$

where:

β_0 = the intercept in logistic regression model

β_j = $(\beta_1, \beta_2, \ldots, \beta_k)$ is vector of unknown regression coefficients.

X_{ij} = $(X_{i1}, X_{i2}, X_{i3}, \ldots, X_{ik})$ is the set of values of k predictors X_1, X_2, X_3, … , X_k, specific to person i. These predictor variables contain the data of respondents obtained from questionnaires. X_{ij} equals to 1 if subject i is of categorical variable in predictor variable j, and equal to zero otherwise.

These predicted probabilities are used to adjust the total weights of respondents who participated in Baseline Questionnaire (BQx). The adjusted total weight, W_i^*, is calculated from:

$$W_i^* = (W_i)(Y_i) / \theta_i \qquad (7\text{-}2)$$

where:

W_i = the total weight of respondent i before the logistic adjustment.

Y_i = 1 if respondent i enrolled in BQx.
= 0 otherwise

The adjustment factor in this method is, therefore, Y_i/θ_i. The adjusted total weights of respondents who refused in BQx are, therefore, equal to zero. The sum of the adjusted total weights is equal to the sum of the total weights of all respondents before the logistic adjustment.

This adjustment by logistic regression is appropriate when a certain amount of information about the nonrespondents is known, e.g. when adjusting for the nonrespondents to the Baseline questionnaire and the intensive questionnaires and environmental sampling. More detail about this method can be found in many literature papers and books, including Iannacchione, VG, et.al., (1991).

3. Poststratification

The procedure is used to adjust the total weights of respondents after all the nonresponse adjustments are done. The reason for doing poststratification is so that the estimates would be comparable to the actual statistics of the Census regarding the state of Arizona. This method begin with selection of one variable containing information of all respondents for poststratification of samples. The information (statistics) of this variable must be available at the Census. The Census statistics are the control totals for the poststrata. The control total of each poststratum divide by the summation of the total weight of all respondents in each poststratum is called the adjustment factor. The product of the respondent's total weight and the adjustment factor is the adjusted total weight. The sum of the adjusted total weights in each poststratum is equal to the Census statistics.

The poststratification is normally used after nonresponse adjustments. More detail about this method can be found in many literature papers and books, including D. Holt and T. M. F. Smith, (1979).

4. Annual Weighting

Annual Weighting adjustment is needed in order to represent annual average of the environmental sampling data which are collected during various time of the year. According to the seasonal variation in Arizona, the sampled households will be stratified into 4 strata[5] according to their sampling dates:

[5] A stratum that contain too few samples may be combined with another stratum.

1) Winter stratum, from November 15 - February 21 (99 days).
2) Spring stratum, from February 22 - June 30 (129 days).
3) Summer stratum, from July 1 - September 15 (77 days).
4) Fall stratum, from September 16 – November 14 (60 days).

The proportions of time of winter, spring, summer, and fall season are 99/365, 129/365, 77/365, and 60/365, respectively.
An annual weight and seasonal weights will be calculated. The annual weight is the sum of the sampling weights of all strata, while the seasonal weight is the sum of the sampling weight of a single stratum. Then, for each stratum:

adjusted seasonal weight = proportion of time of the season × annual weight

adjustment factor = adjusted seasonal weight / seasonal weight

The product of the sampling weight and the adjustment factor is the adjusted total weight of each subject.

5. Single Factor Adjustment

There are 403 households responding to stage 4 sampling: the intensive questionnaire and environmental sampling activities. Due to limitations, however, the NHEXAS field team could not include all the respondents in the activities. To maintain the sum of respondents' weights, the weights must be adjusted to account for the weights of the respondents who were not included. The only available option for this situation is to make an assumption that the households were selected randomly to be included in each sampling activity. With the assumption, this adjustment can be done by multiplying the respondents' weights with a single adjustment factor which is a ratio of the sum of the weights of all respondents divided by the sum of the weights of the included respondents.

This single factor adjustment is also used to adjust the sampling weights of the respondents with calculated exposure values. The number of respondents having enough information for exposure calculation is less than 403, so again the assumption that these respondents were randomly selected is made to allow the adjustment by a single factor and, consequently, to obtain sampling weights that are appropriate for the calculation of weighted statistics of exposure.

Variable List

Variable	Description
COUNTY 2	County I.D., after some changes are made (see details in SOP #10).
TRACT 2	Tract I.D., after some changes are made (see details in SOP #10).

Variable	Description
BLOCK 2	Block I.D., after some changes are made (see details in SOP #10).
HHID	Household I.D. , according to the Census.
RESPO_D	Response status (enrolled or refused) for the Descriptive Questionnaire.
$WT_{i,3}$	Total weight of OHU i at stage 3.
AF_{SW}	The adjustment factor for the Simple Weighting Class method.
AF_{PS}	The adjustment factor for the Poststratification method.
AF_{LR}	The adjustment factor for the Logistic Regression method.
AF_{AW}	The adjustment factor for the Annual Weighting method.
AF_{SF}	The adjustment factor for the Single Factor method.
$SWCWT_{i,3}$	The simple weighting class adjusted total weight.
AGE 2	Age of the householders, years.
WTDES	The adjusted total weight to be used with the Descriptive Questionnaire data.
RESPO_B	Response status (enrolled or refused) for the Baseline Questionnaire.
WHBAS	The adjusted total weight to be used with the BQx-household level data.
WPBAS	The adjusted total weight to be used with the BQx-person level data.
WHCOR	The adjusted total weight for the households that agree to participate in intensive questionnaires and environmental sampling.
$LRWT_{i,3}$	The logistic regression adjusted total weight at the sampling stage 2.
LRWHBAS	The logistic regression adjusted total weight at the sampling stage 3.
MEMBER	The number of members in each household.
LRV1, LRV2, ..., LRVN	The set of N variables in questionnaires which are in the final logistic regression model.
LRV1BETA, LRV2BETA, ..., LRVNBETA	The set of N coefficients of variables in questionnaires which are in the final logistic regression model.
THETA	The predicted probability of response.
RESPO_E	Response status (agree or not agree) for the intensive questionnaires and environmental sampling.
WQDD1	The adjusted total weights of the household in the Diet Diary questionnaire after adjusted by using Single Factor method.
WQDD2	The adjusted total weights of the household in the Diet Diary questionnaire after adjusted by using poststratification.
WQDD3	The adjusted total weights of the household in the Diet Diary questionnaire after adjusted by using annual weighting, respectively. This weights is the one to be used with the Diet Diary data.

Procedure: Adjustment At The Descriptive Questionnaire Level

1. In Excel, open WEIGHT STRUCTURE MAIN, select only ***COUNTY 2, TRACT 2, BLOCK 2, HHID, RESPO_D,*** and $WT_{i,3}$. Save them as DQX WEIGHT ADJUSTMENT COMBINED CNTY.
2. In DQX WEIGHT ADJUSTMENT COMBINED CNTY, use the Simple Weighting Class method to obtain the simple weighting class adjusted total weight, $SWCWT_{i,3}$.
3. In DQX WEIGHT ADJUSTMENT COMBINED CNTY, create a variable ***AGE 2***, using the information from the Descriptive Questionnaire. This variable, the age of the householder, will be used for poststratification. The Census statistics for this variable will be obtained from a document in the Census web site (ST-96-20R: Estimates of Housing Units, Households, Households by Age of Householder, and Persons per Household: July 1, 1996). Use the Poststratification method to calculate the poststratification adjusted total weights, ***WTDES***. This weight is the adjusted total weight to be used with data at the Descriptive Questionnaire level.

Procedure: Adjustment At The Baseline Questionnaire-Household Level

1. In SPSS, open DQX REARRANGED OUTPUT WITH CORRECTION IRN1, which contains the primary respondents who complete the Descriptive Questionnaire. Create the following variables: ***COUNTY 2, TRACT 2, BLOCK 2, RESPO_B,*** and ***WTDES***. The first 3 variables will be from WEIGHT STRUCTURE COMBINED CNTY, ***RESPO_B*** will be obtained from Baseline Questionnaire, and ***WTDES*** will be obtained from DQX WEIGHT ADJUSTMENT COMBINED CNTY. Save them as DQX FOR LOGISTIC ADJUSTMENT COMBINED CNTY.
2. To apply the Logistic Regression adjustment method, the file, DQX FOR LOGISTIC ADJUSTMENT COMBINED CNTY, will be use as a data file in the SUDAAN program to obtain vector of unknown regression coefficients (β) in logistic regression model. First, all predictor variables derived from DQx will be used in the model. Then, the final model will be determined by eliminating any predictor variables that are not statistically significant at the significance level (α=5%) by backward elimination procedure. The logistic coefficients (β) of the final model will be used to compute the predicted probability of response (θ) for all primary respondents.
3. In Excel, create a file called BQX LR WEIGHT ADJUSTMENT. Calculate the logistic regression adjusted total weights, $LRWT_{i,3}$, using Eq. 7-1 and 7-2., and the coefficients obtained from SUDAAN.
4. In BQX LR WEIGHT ADJUSTMENT, create a variable ***AGE 2***, using the information from the Descriptive Questionnaire. This variable, the age of the householder, will be used for poststratification. The Census statistics for this variable will be obtained from a document in the Census web site (ST-96-20R: Estimates of Housing Units, Households, Households by Age of Householder, and Persons per Household: July 1, 1996). Use the Poststratification method to calculate the poststratification adjusted total weights, ***WHBAS***. This weight is the adjusted total weight to be used with data at the Baseline Questionnaire-Household level.

Procedure: Adjustment At The Intensive Qx and Environmental Sampling Level

1. In BQX LR WEIGHT ADJUSTMENT, select only ***HHID*** and ***WHBAS***, and save them as SAMPLING OVERALL.
2. In SPSS, open SAMPLING OVERALL. Create the following variables: ***COUNTY 2***, ***TRACT 2***, ***BLOCK 2***, and ***RESPO_E***. The first 3 variables will be from WEIGHT STRUCTURE COMBINED CNTY. ***RESPO_E*** is the value that indicate if the household agree to the intensive Qx and the environmental sampling and will be obtained from HH MATRIX. Then merge this file with BQX IRN1 COMPLETE. Save this file as SAMPLING OVERALL FOR LOGISTIC ADJUSTMENT.
3. To apply the Logistic Regression adjustment method, the file, SAMPLING OVERALL FOR LOGISTIC ADJUSTMENT, is use as a data file in the SUDAAN program to obtain vector of unknown regression coefficients (β) in logistic regression model. First, all predictor variables derived from BQx will be used in the model. Then, the final model is determined by eliminating any predictor variables that are not statistically significant at the significance level (α=5%) by backward elimination procedure. The logistic coefficients (β) of the final model are used to compute the predicted probability of response (θ) for all primary respondents.
4. In Excel, create a file called SAMPLING OVERALL LR WEIGHT ADJUSTMENT. Calculate the logistic regression adjusted total weights, ***LRWHBAS***, using Eq. 7-1 and 7-2., and the coefficients obtained from SUDAAN.
5. In SAMPLING OVERALL LR WEIGHT ADJUSTMENT, create a variable ***AGE 2***, using the information from the Descriptive Questionnaire. This variable, the age of the householder, will be used for poststratification. The Census statistics for this variable will be obtained from a document in the Census web site (ST-96-20R: Estimates of Housing Units, Households, Households by Age of Householder, and Persons per Household: July 1, 1996). Use the Poststratification method to calculate the poststratification adjusted total weights, ***WHCOR***.
6. To obtain the adjusted total weights of each intensive Qx or environmental sampling data in each media, first use the Single Factor adjustment, then apply the adjustment for annual weighting. The results will be the adjusted total weight of the data. See the spreadsheet format for an example on Diet Diary Questionnaire. See also the Weight Diagram in Table 7-1.

Spreadsheet Format

In DQX WEIGHT ADJUSTMENT COMBINED CNTY:

Column	Variable
1	***COUNTY 2***

Column	Variable
2	***TRACT 2***
3	***BLOCK 2***
4	***HHID***
5	***RESPO_D***
6	$WT_{I,3}$
7	AF_{SW}, calculated from the Simple Weighting Class method.
8	$SWCWT_{I,3}$, calculated from $WT_{I,3} \times AF_{SW}$, only for those who responded.
9	***AGE 2***
10	AF_{PS}, calculated from the poststratification method.
11	***WTDES***, calculated from $SWCWT_{I,3} \times AF_{PS}$

In **BQX LR ADJUSTMENT**:

Column	Variable
1	***HHID***
2	***WTDES***
3	***RESPO_B***
4	***LRV1***, obtained from the DQx.
5	***LRV1BETA***, obtained from SUDAAN.
6	***LRV2***, obtained from the DQx.
7	***LRV2BETA***, obtained from SUDAAN.
...	...
n	***LRVN***, obtained from the DQx.
n+1	***LRVNBETA***, obtained from SUDAAN.
n+2	***THETA***, calculated by using Eq. 7-1.
n+3	$LRWT_{I,3}$, calculated by using Eq. 7-2.
n+4	***AGE 2***
n+5	AF_{PS}, calculated from the poststratification method.
n+6	***WHBAS***, calculated from $LRWT_{I,3} \times AF_{PS}$

In **SAMPLING OVERALL LR WEIGHT ADJUSTMENT**:

Column	Variable
1	***HHID***
2	***WHBAS***
3	***RESPO_E***
4	***LRV1***, obtained from the BQx.
5	***LRV1BETA***, obtained from SUDAAN.

Column	Variable
6	***LRV2***, obtained from the BQx.
7	***LRV2BETA***, obtained from SUDAAN.
...	...
n	***LRVN***, obtained from the BQx.
n+1	***LRVNBETA***, obtained from SUDAAN.
n+2	***THETA***, calculated by using Eq. 7-1.
n+3	***LRWHBAS***, calculated by using Eq. 7-2.
n+4	***AGE 2***
n+5	AF_{PS}, calculated from the poststratification method.
n+6	***WHCOR***, calculated from ***LRWHBAS*** × AF_{PS}

Example of the calculation spreadsheet for sampling weights of the Diet Diary questionnaire data:

Column	Variable
1	***HHID***
2	***WHCOR***, obtained from **SAMPLING OVERALL LR WEIGHT ADJUSTMENT**.
3	AF_{SF}, calculated from the Single Factor method.
4	***WQDD1***, calculated from ***WHCOR*** × AF_{SF}
5	AF_{AW}, calculated from the annual weighting method.
6	***WQDD2***, calculated from ***WQDD1*** × AF_{AW}

Reference

D. Holt and T. M. F. Smith, POST STRATIFICATION, J. R. Statist. Soc., 1979.

VG Iannacchione, JG Milne, and RE Folsom, RESPONSE PROBABILITY WEIGHT ADJUSTMENTS USING LOGISTIC REGRESSION, American Statistical Association, Proceedings of the Section of Survey Research Methods, 1991.

G. Kalton and D. S. Maligalig, A COMPARISON OF METHODS OF WEIGHTING ADJUSTMENT FOR NONRESPONSE, Annual Research Conference Proceedings, Bureau of Census, 1991.