

**The Health of Philippine Emigrants Study (HoPES):
Guidance for Using Post-Stratification Weights**

Version 1

September 5, 2019

Last Updated: September 9, 2019

Introduction

The Health of Philippine Emigrants Study (HoPES) is a dual-cohort longitudinal study that started data collection in 2017. HoPES includes a cohort of migrants to the United States recruited in the Philippines prior to migration, and a cohort of non-migrant Philippine residents. HoPES is designed to collect both survey (questionnaire) data and objective health measures beginning pre-migration and continuing for three years. Further details about study design can be found in Gee et al. (2018).

In order to improve the generalizability of analyses conducted using HoPES data, post-stratification weights are provided with the data set. In general, post-stratification weights are used to align a sample with a known target population. For HoPES, the post-stratification weights align the sample with the age, sex and education distribution of recent Filipino immigrants to the United States based on the 2015 American Community Survey (ACS). Thus the purpose of using the weights is to improve the generalizability of results to the population of recent Filipino immigrants to the U.S. This guidance explains how the post-stratification weights were calculated and how to use the weights in analyses with the HoPES data.

This guidance provides the post-stratification weights for use with the full baseline dataset, the 3-month assessment dataset (which included the migrant cohort only) and the one-year assessment dataset. Examples are provided for using the weights in Stata, SAS and SPSS. The Stata code used to generate the weights is also provided.

Post-stratification weights for the baseline samples

Post-stratification weighting involves categorizing observations into mutually exclusive strata based on observed characteristics (e.g., sex) and applying weights to observations within each strata such that the weighted totals equal the known population totals. The purpose of using these weights in analyses is to adjust estimates and inferences such that they are more representative of this target population.

For HoPES, the post-stratification strata are based on age (21-34 or 35-59 years old), sex (female, male) and educational attainment (no college, any college), for a total of 8 strata. Table 1 shows the proportions of individuals in each strata in the target population, which is individuals born in the Philippines and with 2 years or less of US residence based on the 2015 American Community Survey.

Table 1. Distribution of age, sex and educational attainment of individuals who were born in the Philippines and have 2 years or less of residence in the United States based on the 2015 American Community Survey

Age (years)	Sex	College	Population proportion
21-34	Female	None	0.0836
21-34	Female	Any	0.2435
21-34	Male	None	0.0480
21-34	Male	Any	0.1264
35-59	Female	None	0.0938
35-59	Female	Any	0.2237
35-59	Male	None	0.0522
35-59	Male	Any	0.1288

In general, post-stratification weights are created by obtaining the strata proportions for the sample and calculating the weights for each stratum as (population proportion)/(sample proportion). Table 2 provides the calculation of post-stratification weights for the full HoPES migrant and non-migrant samples. Sample proportions are obtained for the migrant and non-migrant samples separately. The post-stratification weights are obtained as (population proportion)/(sample proportion).

Table 2. Full baseline HoPES migrant and non-migrant cohort: calculation of post-stratification weights

Stratum no.	Age (years)	Sex	College	Population proportion (A)	Sample proportion (B)	Baseline dataset post-stratification weight (A/B)
Migrant sample (n = 832)						
1	21-34	Female	None	0.0836	0.0745	1.060
2	21-34	Female	Any	0.2435	0.2969	0.835
3	21-34	Male	None	0.0480	0.0397	1.059
4	21-34	Male	Any	0.1264	0.1154	0.927
5	35-59	Female	None	0.0938	0.0998	1.113
6	35-59	Female	Any	0.2237	0.1875	1.211
7	35-59	Male	None	0.0522	0.0697	0.818
8	35-59	Male	Any	0.1288	0.1166	1.115
Non-migrant sample (n = 803)*						
1	21-34	Female	None	0.0836	0.0708	1.116
2	21-34	Female	Any	0.2435	0.2174	1.141
3	21-34	Male	None	0.0480	0.0323	1.300
4	21-34	Male	Any	0.1264	0.1466	0.723
5	35-59	Female	None	0.0938	0.1789	0.621
6	35-59	Female	Any	0.2237	0.2112	1.075
7	35-59	Male	None	0.0522	0.0472	1.208
8	35-59	Male	Any	0.1288	0.0932	1.395

*The total non-migrant sample is 805. However, two participants are missing age values and were not included in the calculation of the weights.

When the sample proportion in a stratum is less than the population proportion, such individuals are underrepresented in the sample relative to the population; the post-stratification weight will be greater than 1 and the observation will be upweighted in analyses. When the sample proportion in a stratum is more than the population proportion, such individuals are overrepresented in the sample; the weight will be less than 1 and the observation will be downweighted.

In HoPES, the sample proportions and population proportions are similar and the post-stratification weights are generally close to 1. This is because the migrant and non-migrant samples were obtained using quota sampling, with quotas for each stratum that were based on population proportions from the 2011-2013 American Community Survey.

Post-stratification weights for the three-month assessment sample (migrants only)

A 3-month assessment was conducted for migrants only. This assessment was timed to occur after migration to the U.S. These weights are appropriate to use for analyses in which the migrant sample consists of only these 440 respondents. Examples of such analyses include:

- An analysis involving only these 440 migrant cohort members;
- An analyses involving these 440 migrant cohort members AND the baseline non-migrant cohort. Such an analysis might be conducted by merging the baseline and 3-month datasets and dropping the migrants without a 3-month assessment. In this situation, the appropriate weights would be the baseline weights for the non-migrant respondents and the 3-month weights for the migrant respondents. Code is provided in the appendix that shows how to produce this new weight variable.

Table 3. Three-month assessment HoPES migrant sample: calculation of post-stratification weights

Stratum No.	Age (years)	Sex	College	Population proportion (A)	Sample proportion (B)	3-month post-stratification weight (A/B)
1	21-34	Female	None	0.0836	0.0955	0.876
2	21-34	Female	Any	0.2435	0.3500	0.696
3	21-34	Male	None	0.0480	0.0273	1.760
4	21-34	Male	Any	0.1264	0.1250	1.011
5	35-59	Female	None	0.0938	0.0773	1.214
6	35-59	Female	Any	0.2237	0.1864	1.200
7	35-59	Male	None	0.0522	0.0409	1.276
8	35-59	Male	Any	0.1288	0.0977	1.318

Post-stratification weights for the one-year assessment

A one-year assessment was conducted for both the migrant and non-migrant cohorts. These weights are appropriate to use when the analytic dataset consists of only these respondents. Examples of such analyses include:

- A cross-sectional analysis of the one-year assessment data;
- A longitudinal analysis that involves merging the baseline and one-year assessments and analyzing the sample that has one-year assessment data (that is, dropping respondents without a one-year assessment).

Table 4. One-year assessment HoPES migrant and non-migrant samples: calculation of post-stratification weights

Stratum No.	Age (years)	Sex	College	Population proportion (A)	Sample proportion (B)	One-year dataset post-stratification weight (A/B)
Migrants (n = 442)						
1	21-34	Female	None	0.0836	0.0928	0.901
2	21-34	Female	Any	0.2435	0.3575	0.681
3	21-34	Male	None	0.0480	0.0203	2.357
4	21-34	Male	Any	0.1264	0.1312	0.963
5	35-59	Female	None	0.0938	0.0679	1.382
6	35-59	Female	Any	0.2237	0.1787	1.252
7	35-59	Male	None	0.0522	0.0362	1.442
8	35-59	Male	Any	0.1288	0.1154	1.116
Non-migrants (n = 758)						
1	21-34	Female	None	0.0836	0.0673	1.243
2	21-34	Female	Any	0.2435	0.2124	1.146
3	21-34	Male	None	0.0480	0.0330	1.455
4	21-34	Male	Any	0.1264	0.1478	0.856
5	35-59	Female	None	0.0938	0.1873	0.501
6	35-59	Female	Any	0.2237	0.2124	1.053
7	35-59	Male	None	0.0522	0.0462	1.131
8	35-59	Male	Any	0.1288	0.0923	1.395

Reference

Gee GC, de Castro AB, Crespi CM, Wang MC, Llave K, Brindle E, Lee NR, Kabamalan MMM, Hing AK (2018) Health of Philippine Emigrants Study (HoPES): study design and rationale. BMC Public Health 18(1): 77

Frequently Asked Questions

When should I use these weights? Are there situations in which I might not want to use them?

The purpose of using the weights is to improve the generalizability of the results of an analysis to the population of recent Filipino immigrants to the U.S. The weighted sample has the same age, sex and education distribution as the U.S. population of individuals born in the Philippines and with 2 years or less of U.S. residence, based on the 2015 American Community Survey.

The data can also be analyzed without using the weights. In this case, the analyses can be conducted and interpreted much like any analysis. A limitation in this case is that the results may not be as generalizable to the population of recent Filipino immigrants to the U.S.

These weights should not be used if the analysis is conducted on a subset of the data, for example, only respondents with high BMI or only married respondents. In this situation, the weights will not align the sample properly with the population of recent Filipino immigrants to the U.S. An exception is that the weights will still be valid and align the sample properly if the subset consists of several of the strata used to generate the weights, for example, men only or women only.

I am analyzing only the migrant sample (or only the non-migrants). Can I still use these weights?

The migrant and non-migrant samples can be analyzed separately and the weights are still valid to use for such analyses.

I am analyzing only the one-year data cross-sectionally and not earlier waves. Do I still use the one-year weights?

In this situation, the one-year assessment weights can be used; doing so will improve the generalizability of the results to the population of recent Filipino immigrants to the U.S.

Appendix: Using post-stratification weights in software

The post-stratification weights can be used in analyses by using survey data analysis procedures, which are widely available in statistical software.

Stata

In Stata, such analyses are performed by first running a `svyset` command and then using the `svy:` prefix in the command lines.

- * The `svyset` command should be run before analyses.
- * Examples of how to use:

```
svyset _n [pweight=sampwt], strata(migrant)
```

```
svy: prop obese  
svy: prop obese, over(migrant)
```

```
svy: tab obese migrant
```

```
svy: reg bmi age_ male anycollege
```

```
svy: logit obese age_ male anycollege
```

- * To get estimates for a subclass defined by value of some variable,
* use the `subpop(var)` option.
- * Do not use the "if" method of subsetting the data.

```
* This option requests analysis for the specific  
* subpopulation identified by var==1.  
svy, subpop(migrant): prop obese
```

```
* You can also use logical statement to identify the subpopulation:  
svy: prop obese, subpop(if migrant==0)
```

```
* This code shows how you can generate a new weight (new_wt) that uses the  
* baseline weight for non-migrants and the 3-mo weight for migrants. This  
weight could be used if you were merging the baseline and 3-month datasets  
and conducting analyses that included the 803 non-migrants with baseline data  
and the 440 migrants with baseline and 3-month data:
```

```
gen new_wt=sampwt_b1  
replace new_wt=sampwt_3mo if migrant==1
```

SAS

To use the post-stratification weights in SAS, use the `SURVEY` version of your desired procedure and use the `WEIGHT` statement followed by the name of the appropriate weight variable. In SAS, the `STRATA` statement is used to name the variables that form the strata in a stratified sample design. HoPES had a stratified sampling design in that sampling was stratified on migrant and non-migrant. Thus we include the `STRATA migrant;` statement. (Note that sampling strata are

conceptually distinct from the post-stratification strata; sampling strata refer to how the respondents were sampled; post-stratification strata are formed after the data are collected.) Below is sample code:

```
PROC SURVEYMEANS DATA = datasetname;
WEIGHT sampwt;
STRATA migrant;
VAR variablename;
RUN;
```

SPSS

Using post-stratification weights in SPSS requires the creation of an analysis design. Below is code that can be used to account for the sample design from HoPES. The `PLANVARS ANALYSISWEIGHT` subcommand applies the HoPES sampling weights to the dataset. The `SRSESTIMATOR` subcommand accounts for the sample as a simple random sample. This includes a finite population correction. The `DESIGN STRATA` subcommand accounts for how the respondents are sampled. Finally, the `ESTIMATOR TYPE` command accounts for the estimation methods.

```
*CSPLAN ANALYSIS.

/PLAN FILE = "D:\datasetname.csaplan"
/PLANVARS ANALYSISWEIGHT = sampwt
/SRSESTIMATOR TYPE = WOR
/PRINT PLAN
/DESIGN STRATA = migrant
/ESTIMATOR TYPE = WR.
```

After applying the HoPES survey weights to the dataset in SPSS, you will need to use a “CS” command - such as `CSDESCRIPTIVES`, `CSTABULATE`, `CSGLM`, `CSLOGISTIC` - to call your dataset with applied weights.

Technical appendix: Stata code for generating post-stratification weights

The Stata code used to generate the baseline post-stratification weights is provided here for reference. In general, users will not need to run this code in order to use the weights.

```
* Code to generate post-stratification weights for HoPES data
* that align the migrant and non-migrant samples
* with the ACS 2015 distribution of recent (<2 years)
* Filipino immigrants to the US

* Post-stratification weights are formed as ACS (population)
* proportion divided by HoPES (sample) proportion.
* ACS proportions are hard-coded in here.
```



```

* load the dataset:
use hopesbaselinedata.dta, clear

*get proportions in each of the 8 strata, by migrant status
egen x = count(1) if migrant==1, by(post_strat)
egen n_m = count(1) if migrant==1
gen prop_m = x/n_m
drop x
egen x = count(1) if migrant==0, by(post_strat)
egen n_nm = count(1) if migrant==0
gen prop_nm = x/n_nm

*make sampling weights
gen sampwt = .
replace sampwt = .0836/prop_m if post_strat == 1&migrant==1
replace sampwt = .2435/prop_m if post_strat == 2&migrant==1
replace sampwt = .0480/prop_m if post_strat == 3&migrant==1
replace sampwt = .1264/prop_m if post_strat == 4&migrant==1
replace sampwt = .0938/prop_m if post_strat == 5&migrant==1
replace sampwt = .2237/prop_m if post_strat == 6&migrant==1
replace sampwt = .0522/prop_m if post_strat == 7&migrant==1
replace sampwt = .1288/prop_m if post_strat == 8&migrant==1

replace sampwt = .0836/prop_nm if post_strat == 1&migrant==0
replace sampwt = .2435/prop_nm if post_strat == 2&migrant==0
replace sampwt = .0480/prop_nm if post_strat == 3&migrant==0
replace sampwt = .1264/prop_nm if post_strat == 4&migrant==0
replace sampwt = .0938/prop_nm if post_strat == 5&migrant==0
replace sampwt = .2237/prop_nm if post_strat == 6&migrant==0
replace sampwt = .0522/prop_nm if post_strat == 7&migrant==0
replace sampwt = .1288/prop_nm if post_strat == 8&migrant==0

*save the dataset
save hopesbaselinewithsampwts.dta, replace

```