ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

ICPSR 36537

Hispanic Established Populations for the Epidemiologic Study of the Elderly (HEPESE) Wave 7, 2010-2011 [Arizona, California, Colorado, New Mexico, and Texas]

Weight Information

Inter-university Consortium for
Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

Hispanic Established Populations for the Epidemiologic Study of the Elderly (HEPESE) Wave 7, 2010-2011 [Arizona, California, Colorado, New Mexico, and Texas]

Kyriakos Markides
UTMB

Nai-Wei Chen
UTMB

Ronald Angel
UT-Austin

Raymond Palmer
UT-San Antonio, Health Science Center

James Graham
UTMB

Terms of Use

Information about Copyrighted Content

Weights and PSU's.

Baseline (n=3050, 1990 Census). Weights are in the baseline SAS data set public1.

When analyzing data from the HEPESE Baseline (Wave 1), there are two weights to select from:

> FW (final weight, normalized to the 3,050 baseline subjects)
> IW (inflation weight, reflecting approximately 500,000 Mexican Americans residing in the 5 Southwest states in the study –TX, NM, AZ, CA, CO)

For SUDAAN and other programs that adjust for clustering and design effects, use the variable called PSU. For example in SUDAAN:

Proc rlogist data=work.in1 design=wr filetype=SAS est_no=3050;
nest _one_ psu;
Weight iw; *or Weight fw:

Wave 2 (n=2438) Wave 3 (n=1981) Wave 4 (n=1682)

For Waves 2-4, the weights were not adjusted for the change in size of the Mexican American population, so we can no longer use the variables FW and IW. It is still advised to run the programs through SUDAAN and use a weight of 1 for all subjects. (In your SAS code create a variable such as newwgt below: newwgt=1;)

Proc rlogist data=work.in1 design=wr filetype=SAS est_no=3050;
nest _one_ psu;
weight newwgt;

Wave 5 (n=2069, 1167 original subjects, 902 new subjects, 2000 Census)

At Wave 5, new subjects were added and this created some new considerations. First all the census tracts for the surviving original cohort (n=1167) and the 902 new subjects were geographically coded to the 2000 Census. Rake weights were created separately for the original subjects C1RAKE and for the new subjects C2RAKE. There is also a combined weight to use when analyzing the new and the old together: CW5RAKE. A new PSU variable was created based on the 2000 census tracts. That variable is PSUTRACT. This SAS system file is called wave5_wgts_psus.

Proc rlogist data=work.in5 design=wr filetype=SAS est_no=2069;
nest _one_ psutract;
Weight cw5rake; *when combining;

Proc rlogist data=work.in5o design=wr filetype=SAS est_no=1167;
nest _one_ psutract;
Weight c1rake; *when using original cohort only;

Proc rlogist data=work.in5n design=wr filetype=SAS est_no=902;
nest _one_ psutract;
Weight c2rake; *when using new cohort only;

Weights for Wave 5.

The attached file contains weights for wave 5.

c1rake contains rake weights for survivors of the first cohort to wave 5. I calculated it for informational purposes to compare it to the wave 1 weight variable, with which the correlation is 0.85.

c2rake contains rake weights for the new cohort

To calculate these weights I first calculated sampling weights as the inverse of the probability of selection. This was unequal despite the use of PPES sampling, because of the different number of completes in the different PSUs.

Differential completes were strongly correlated with tract percent Mexican (TpM), with relatively ratio of completes to population in low deciles of TpM, and the highest ratio of completes to population in deciles 6 and 7 of TpM. What accounts for this pattern is that (in contrast to the practice at wave 1), we closed PSUs with very high TpM, while continuing to seek subjects in mid-density PSUs (deciles 3-6, then 3-7). Not surprisingly, we were most successful in recruiting subjects in deciles 6 and 7 through this effort—that is, the most dense of the less dense. To account for these differences, I adjusted the weights to population totals from Census 2000 within collapsed deciles of TpM: deciles 1/3,4/5,6,7,8/10. I collapsed deciles to avoid extremely large weights for deciles with few subjects.

These weights were raked to estimated control totals from the Census 2000 PUMS which contained 13,759 cases of older Mexican Americans ages 75+ living in one of the 5 southwestern states. Raking variables were gender, age (75-79,80-84,85+), state (CA,TX,Other), and the crosstabulation of nativity (U.S. vs. other) and schooling (<5,5 < 12, 12+ years). I cross-tabulated these two variables because of the strong relationship between these variables. I also used TpM in the collapsed deciles lised above as a raking variable.

I used Izrael's rakinge.sas macro, documented in Izrael, David, Hoaglin, David C., and Battaglia, Michael P. (2000), "A SAS Macro for Balancing a Weighted Sample." *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference*, Paper 275, and Izrael, David, Hoaglin, David C and Battaglia, Michael P. (2004) To Rake or Not To Rake Is Not the Question Anymore with the Enhanced Raking Macro. *Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference*, Paper 207. Raking converged after 6 iterations.

Weights were normalized to the new cohort sample size, 902.

1) Below is an OLS regression for the weights with respect to raking variables. As you can see, and not surprisingly, low density tracts are associated with large weights, while the 7th decile is associated with smaller weights, for the reason

mentioned in 1c. (I collapsed all other deciles for the regression, as they were not significantly different from one another.) California and being a high school graduate are also associated with larger weights. Other relationships are small. The important adjustments are all exactly as expected in my opinion.

	fw_2c	Coef.	Std. Err.	t	Beta
Vs. others	TpM Decile 1	1.779	0.025	71.62	0.84
	TpM Decile 7	-0.271	0.020	-13.23	-0.15
Vs Ed 5 < 12	Ed < 5	0.097	0.020	4.94	0.07
	Ed 12+	0.396	0.025	15.85	0.20
	Ed unknown	0.116	0.022	5.18	0.07
Vs. Texas	California	0.405	0.017	23.36	0.29
	Other, not Texas	-0.072	0.025	-2.82	-0.03
	Female	0.079	0.016	5.04	0.06
	Age	-0.033	0.010	-3.39	-0.04
	US Born	-0.058	0.017	-3.38	-0.04

cw5rake Combined weights for both samples

For the combined weights, I used as input the final weight (fw) from public1 in the first sample for the first cohort, and w2rake for the second cohort as input weights. I renormalized the wave1 weights for the wave5 survivors to the current cohort size. For TpM assignment for the wave 1 cases, I used the geocoded wave 5 residence tract. If this was not available, I substituted the TpM score from baseline to make the assignment. The combined sample was raked on the same variables as described above.

An analysis of the cw5rake yielded results similar to those reported in the regression table above. cw5rake was correlated 0.95 with each of the input weights. Correlation was near .8 to an alternative weight that was constructed with unweighted input.

Recommended short description of the weights for methods sections:

“Sampling weights for each wave were raked to population totals from Census 2000 5% public-use file and Census summary files 1 and 4 using Izrael’s enhanced raking macro (Izrael, Hoaglin, Battaglia 2000, 2004). Raking variables were age, gender, state of residence, education by immigrant status, and percentage Mexican American in census tract of residence.” (References above)