



TO: HCHS/SOL Steering Committee
CC: Subcommittee Chairs and co-Investigators
FROM: Marston Youngblood, HCHS Coordinating Center
DATE: September 4, 2018
RE: Distribution of HCHS SOL Bootstrap Replicate Sampling Weights

123 W. FRANKLIN STREET
SUITE 450
CAMPUS BOX 8030
CHAPEL HILL, NC 27599-8030

T 919.962.6971
F 919.962.3265
www.csec.unc.edu

The bootstrap replicate weights for both visits (baseline and visit 2) from HCHS/SOL were calculated using a two-step approach that accounts for the complex survey sampling design. First, primary sampling units (PSUs) were sampled with replacement and equal probability from each stratum. To ensure efficiency of the bootstrap estimators and rectify potential biases due to small number of PSUs per stratum, the selected participants from each PSU had their base weights scaled according to the recommendations from Kolenikov (The Stata Journal, 2010). Then, the baseline and the visit 2 scaled weights were further adjusted by their respective response rate. Second, the scaled and non-response adjusted weights were trimmed at the 98th percentile, calibrated and normalized using Census 2010 selected track population counts. This two-step process was independently repeated a hundred times.

Baseline bootstrap weights (HCHS_BOOT_WEIGHTS_FINAL_INV4.sas7bdat)

The final dataset has for each participant three sets of 100 bootstrap replicate sampling weights (expanded, normalized to the overall sample, and normalized to each site). The total number of records in this dataset is 16415, the number of HCHS/SOL participants at baseline. Just like in the HCHS/SOL study, for each bootstrap replicate the sum of the WEIGHT_FINAL_NORM_OVERALL# adds up to the total number of participants (n=16,415).

Visit 2 bootstrap weights (HCHS_BOOT_WEIGHTS_FINAL_V2_INV2.sas7bdat)

The final dataset has for each participant three sets of 100 bootstrap replicate sampling weights (expanded, normalized to the overall sample, and normalized to each site). The total number of records in this dataset is 11,623, the number of HCHS/SOL participants that participated in the second study visit. For each set of bootstrap replicate, the total sum of the V2_WEIGHT_FINAL_NORM_OVERALL# adds up to the total number of participants (n=11,623).

Reference:

Kolenikov, S. (2010) Resampling variance estimation for complex survey data. Stata Journal, 10(2), 165--199.

Please contact us at the Coordinating Center if you have any questions about the contents of this HCHS/SOL data distribution for exclusive use by the study investigators.