

Analytic Note Regarding 2007-2010 Survey Design Changes and Combining Data Across other Survey Cycles

The sample design of NHANES 2007-2010 is different than the sample designs for earlier cycles of the continuous NHANES particularly with regard to oversampling of demographic subgroups of the population. This analytic note provides guidance for analyses that are based on survey data from 2007-2010 combined with data from earlier periods.

For the 1999-2002 and 2003-06 survey periods, the sampling fractions and screener rates were set to over-sample certain groups (i.e. low income persons, adolescents, the elderly, Blacks, and Mexican Americans) to increase the ability to obtain more precise estimates for these groups. In contrast to the oversampling of Mexican Americans, the proportion of non-Mexican American Hispanics in the NHANES sample from these years is much smaller than in the US population and it is generally considered to be too small to produce reliable estimates. NCHS strongly recommends that researchers should not calculate estimates for non-Mexican American Hispanics or for all Hispanics for the 1999-2002 and 2003-2006 survey periods.

Oversampling of population subgroups was changed in the sample design for the 2007-2010 survey period subgroups. Adolescents are no longer oversampled. The most significant change is that all Hispanic persons are targeted for oversampling rather than just Mexican American persons. In addition to allowing estimates for the total group of Hispanics, sampling fractions were set so that the sample size for Mexican Americans would not be very different from previous years and would be sufficient to produce reliable estimates for this group. The methodology for the oversampling of Hispanics does not provide sample sizes that are sufficient for calculating estimates for other Hispanic subgroups.

In general, any two year data cycle in NHANES can be combined with adjacent two year data cycles to create analytic data files based on four or more years of data in order to calculate survey estimates with greater precision (smaller sampling error). The NHANES sample designs for the periods 1999-2002 and 2003-2006 were similar so that combining data cycles from these periods does not present any analytic issues. However, because of the sample design change for

2007-2010, the data user should be aware of the implications if these data are combined with data from earlier survey cycles.

With the availability of data from 2007-2008 data cycle, data users may be interested in combining data from 2005-2006 with 2007-2008 data to create a 2005-2008 data file. As a result of the sample design changes initiated in 2007, users should consider the following when calculating estimates based on data from the four year period 2005-2008. First, because of the small sample numbers and design inefficiency for non-Mexican American Hispanics (and thus for total Hispanics) in 2005-2006, NCHS recommends that data users should not create estimates for total Hispanics for the 2005-2008 data period. Second, for the non-Hispanic white, non-Hispanic black, and Mexican American sample domains, rescaling the sample weights to create four-year weights (as described in the analytic guidelines) should be sufficient for producing combined 4 year estimates for most analytic variables. However, users are cautioned to examine their estimates carefully to see if the four year estimates (and sampling errors) are consistent with each set of 2 year estimates. Among all the variables collected in NHANES, it is possible that aberrant results in either basic estimates or estimates of sampling error may be obtained as a result of the differences in oversampling between the two sample designs.

SUMMARY:

1. When NHANES 2005-2006 and NHANES 2007-2008 are combined to form a 4-year dataset (NHANES 2005-2008): a) adjust the sample weights as described in the analytic guidelines; and b) compare point estimates and standard errors from the 4 year dataset with those seen in both 2-year datasets for consistency.
2. Some changes in appropriate Hispanic analytic subgroups have occurred due to changes in sampling starting in NHANES 2007. The following table summarizes the recommendations for appropriate Hispanic subgroups to be used when analyzing data from NHANES 2003-2006, NHANES 2007-2008 or NHANES 2005-2008.

	Appropriate Hispanic subgroup for analysis			
	NHANES 1999-2002	NHANES 2003-2006	NHANES 2007-2010	NHANES 2005-2008
Mexican Americans	Yes	Yes	Yes	Yes
Hispanics other than Mexican American	No	No	No	No
All Hispanics	No	No	Yes	No