Attachment 14b - Current NHANES 2015-2018 sample design approved by the NCHS Research Ethics Review Board

III.   Study Design

A.   *Sample Design and Study Population*

The sample design of NHANES is based on a continuous on-going annual survey of the non-institutionalized, civilian population of the U.S. Each single year and any combination of consecutive years will comprise a nationally representative sample of the United States population. Data are publicly released every two years to improve analytic stability of estimates and to decrease disclosure risk.

Several subgroups are oversampled to ensure adequate sample sizes for selected population subgroups. For NHANES 2017-2018, African Americans, Hispanics, Asian Americans, older persons, and low-income whites will be oversampled. Table 1 shows the sampling domains for NHANES. These domains represent the analytic subpopulations of interest. For NHANES 2015-2018, there are 87 sampling domains defined by race and Hispanic origin, gender, age, and, for the white/other domains, low income status (i.e., households in which the household income is below 185% of the poverty level).

The sample design for NHANES is planned for four years. The design for NHANES 2015-2018 was planned in 2012 and 2013. After the sampled primary sampling units (PSUs) were selected, no change is made during those four years. Table 1 shows the annual and cumulative estimates of the target sample sizes by analytic subdomain, based on the assumption that two mobile examination center (MEC) teams will be in operation and approximately 5,000 persons will be examined in 15 PSUs per year. The expected sample size is based on past NHANES experience with response rates for each subdomain of interest. The goal for the overall examination response rate for NHANES 2015-2018 was targeted for 70 percent. In NHANES 2014, 2015, and 2016 the examination response rates were 65, 63 and 55 percent, respectively. Experiences prior to 2014 indicated a sample design based on a target of 70 percent was reasonable. Additionally, we have a goal of performing 5,000 examinations per year. When a lower than expected response has occurred in PSUs, more screening is performed to obtain more examinations in an attempt to meet the goal of 5,000 examinations. The ability to screen more is limited by available time in a PSU and available field staff. We are currently determining the NHANES 2019-2020 sample design; and the projected response rates will be more consistent with recent trends.

B.   *Sampling Methods*

As with previous NHANES surveys, the design for the current NHANES is a stratified, multistage probability sample of the civilian non-institutionalized population of the U.S. In hierarchical order, the stages of the sample selection are first: selection of Primary Sampling Units or PSUs (counties or small groups of contiguous counties); second: segments within PSU (a block or group of blocks containing a cluster of households); third: households within segments; and fourth: one or more survey participants within households.

To produce unbiased cross-sectional estimates for the entire civilian, non-institutionalized population of the United States, the sample data is inflated to the level of the population from which the sample is drawn. As in previous NHANES, the sampling weight for each sample person is the product of three factors: the reciprocal of the probabilities of selection (PSU, segment, household, person); an adjustment for non-response; and a post-stratification adjustment factor to make the resulting survey estimates in each age-sex-race-Hispanic origin income category approximately equal to independent control totals from the American Community Survey (ACS) conducted by the U.S. Bureau of the Census.

To analyze multiple year samples, sampling weights can either be averaged over the sampled years used or can be readjusted to population controls for the midpoint of the combined years. Variances for NHANES are estimated using the Taylor series linearization procedure, SUDAAN, available through the Statistical Analysis System. This technique has been used in previous NHANES and the Hispanic HANES (HHANES). Alternatively, jackknife and BRR-type replicate weights are generated to allow other replication methods to compute variance when sample sizes may not be adequate to produce reliable estimates from Taylor's series approximations.

A primary objective of NHANES is to estimate, with acceptable precision, the health and nutritional status of subgroups of the population. For NHANES 2015-2018, the analytic subgroups of interest are defined through 87 sampling domains (sampling flags) by race/Hispanic origin, gender, age, and low income status (i.e., households in which the household income is below 185% of the poverty level) for the white/other domains. These domains can be found in Table 1. A number of domains require oversampling to ensure adequate sample sizes are achieved so selected health conditions can be estimated for minority groups. The over-sampled groups include African Americans, Hispanics, Asian Americans, younger and older persons, and low-income white/other domains. The sampling procedure is implemented using a set of sampling flags that designate for each household which domains are eligible for sampling. This is accomplished by selecting a sample of households with equal probability at a rate equal to the maximum oversampling rate assigned to the domains. Each selected household is then assigned a set of sampling flags corresponding to the 87 sampling domains. The flags are set at the domain sampling rates assigned to achieve the target sample size for each domain. The sampling flags are then pre-loaded into the Computer Assisted Personal Interview (CAPI) system on NHANES laptops prior to the start of data collection. That is, interviewers are not required to carry out any subsampling operation. They are instead instructed by the system on which persons to include in the sample.

Therefore, the probability of selection is related to race/Hispanic origin, gender, age, and low-income status. This differential probability of selection is designed to meet departmental and public health objectives to reduce disparities. The NHANES sample persons, however, are chosen at random.

Table 1: Projected population size, number of sampled persons, and projected response rates for NHANES 2015-2018 in 60 PSUs by age, race and Hispanic origin, income, and gender						
			Projected population average over years 2015-2018[1]	Total sample	Estimated exam response rate[2]	Target number of exams for 2015-2018
Black, non-Hispanic	M&F	< 1 year	726,585	216	85%	184
		1-2 yrs.	1,439,050	370	83%	308
		3-5 yrs.	2,092,808	387	80%	308
	M	6-11 yrs.	2,010,582	383	80%	308
		12-19 yrs.	2,557,093	390	79%	308
		20-39 yrs.	5,531,423	665	74%	492
		40-49 yrs.	2,247,656	343	71%	244
		50-59 yrs.	2,379,097	351	69%	244
		60+ yrs.	2,758,652	804	63%	504
	F	6-11 yrs.	1,953,238	384	80%	308
		12-19 yrs.	2,544,785	401	77%	308
		20-39 yrs.	6,421,630	656	75%	492
		40-49 yrs.	2,752,614	339	72%	244
		50-59 yrs.	2,848,545	327	75%	244
		60+ yrs.	3,836,365	788	64%	504
Total Black, non-Hispanic			42,100,123	6,806	73%	5,000
Hispanic	M&F	<1 year	1,171,535	334	86%	288
		1-2 yrs.	2,316,621	333	87%	288
		3-5 yrs.	3,362,637	353	82%	288
	M	6-11 yrs.	3,225,375	338	86%	292
		12-19 yrs.	3,943,976	356	82%	292
		20-39 yrs.	9,528,110	720	68%	492
		40-49 yrs.	3,804,496	330	75%	248
		50-59 yrs.	2,827,360	364	68%	248
		60+ yrs.	2,659,160	814	61%	496
	F	6-11 yrs.	3,086,758	357	82%	292
		12-19 yrs.	3,783,981	371	79%	292
		20-39 yrs.	8,777,249	613	80%	492
		40-49 yrs.	3,721,308	324	76%	248
		50-59 yrs.	2,865,586	339	73%	248
		60+ yrs.	3,259,775	881	56%	496
Total Hispanic			58,333,927	6,828	73%	5,000

Table 1: Projected population size, number of sampled persons, and projected response rates for NHANES 2015-2018 in 60 PSUs by age, race and Hispanic origin, income, and gender						
Non-Hispanic, non-Black Asian	M&F	< 1 year	273,292	79	66%	52
		1-2 yrs.	546,059	162	62%	100
		3-5 yrs.	804,702	241	66%	160
	M	6-11 yrs.	784,302	237	61%	144
		12-19 yrs.	978,909	292	64%	188
		20-39 yrs.	2,868,204	554	58%	324
		40-49 yrs.	1,325,307	326	50%	164
		50-59 yrs.	1,060,302	319	50%	160
		60+ yrs.	1,318,080	397	46%	184
	F	6-11 yrs.	758,637	223	50%	112
		12-19 yrs.	960,905	288	61%	176
		20-39 yrs.	3,087,674	571	55%	316
		40-49 yrs.	1,519,346	314	51%	160
		50-59 yrs.	1,226,492	287	56%	160
		60+ yrs.	1,707,959	508	39%	200
Total non-Hispanic, non-Black Asian			19,220,170	4,800	54%	2,600
Non-Hispanic White/Other	M&F	< 1 year	475,509	140	97%	136
Low Income		1-2 yrs.	922,061	169	81%	136
		3-5 yrs.	1,346,124	154	91%	140
	M	6-11 yrs.	1,216,699	150	93%	140
		12-19 yrs.	1,469,297	151	93%	140
		20-29 yrs.	2,120,036	141	76%	108
		30-39 yrs.	1,384,933	127	85%	108
		40-49 yrs.	1,253,827	128	84%	108
		50-59 yrs.	1,492,999	149	70%	104
		60-69 yrs.	1,257,791	146	71%	104
		70-79 yrs.	628,636	162	64%	104
		80+ yrs.	401,776	122	52%	64
	F	6-11 yrs.	1,144,717	158	89%	140
		12-19 yrs.	1,467,306	171	82%	140
		20-29 yrs.	3,015,976	123	84%	104
		30-39 yrs.	1,956,563	117	89%	104
		40-49 yrs.	1,457,591	115	90%	104
		50-59 yrs.	1,841,567	124	84%	104
		60-69 yrs.	1,690,869	145	72%	104
		70-79 yrs.	1,354,362	130	80%	104
		80+ yrs.	1,273,741	203	51%	104
Total non-Hispanic White/Other Low Income			29,172,380	3,028	79%	2,400

Table 1: Projected population size, number of sampled persons, and projected response rates for NHANES 2015-2018 in 60 PSUs by age, race and Hispanic origin, income, and gender						
Non-Hispanic White/Other	M&F	< 1 year	1,657,990	395	73%	288
Not Low Income		1-2 yrs.	3,358,271	519	56%	288
		3-5 yrs.	5,020,468	449	64%	288
	M	6-11 yrs.	5,336,202	423	69%	292
		12-19 yrs.	7,829,377	413	71%	292
		20-29 yrs.	10,253,435	351	60%	212
		30-39 yrs.	10,708,541	361	59%	212
		40-49 yrs.	10,911,032	346	61%	212
		50-59 yrs.	13,237,738	364	58%	212
		60-69 yrs.	11,863,100	357	59%	212
		70-79 yrs.	6,916,646	399	53%	212
		80+ yrs.	3,292,269	450	47%	212
	F	6-11 yrs.	5,089,881	401	73%	292
		12-19 yrs.	7,429,312	434	67%	292
		20-29 yrs.	9,355,737	388	55%	212
		30-39 yrs.	10,317,801	336	63%	212
		40-49 yrs.	10,882,958	307	69%	212
		50-59 yrs.	13,380,811	348	61%	212
		60-69 yrs.	12,395,307	351	60%	212
		70-79 yrs.	7,350,344	408	52%	212
		80+ yrs.	4,297,525	513	41%	212
Total non-Hispanic White/Other Not Low Income			170,884,743	8,312	60%	5,000
Total non-Hispanic White/Other			200,057,123	11,339	65%	7,400
TOTAL			319,711,342	29,773	67%	20,000

NOTES: M is male. F is female. Yrs. is years.

[1] Civilian non-institutionalized population in the 50 United States and the District of Columbia.

[2] Unconditional response rates are the domain-level response rates (number of examined persons divided by the number of identified sampled persons) from the 2011-2012 experience, each adjusted downward by the overall screener response rate experienced in those years to result in conservative estimates.