

ICPSR 20240

**Collaborative Psychiatric  
Epidemiology Surveys (CPES),  
2001-2003 [United States]**

Margarita Alegria

*Center for Multicultural Mental Health Research at  
Cambridge Health Alliance*

James S. Jackson

*University of Michigan. Institute for Social Research*

Ronald C. Kessler

*Harvard Medical School. Department of Health Care  
Policy*

David Takeuchi

*University of Washington*

User Guide

Inter-university Consortium for  
Political and Social Research  
P.O. Box 1248  
Ann Arbor, Michigan 48106  
[www.icpsr.umich.edu](http://www.icpsr.umich.edu)

## Table of Contents

Introduction.....	2
National Comorbidity Survey Replication (NCS-R) .....	2
National Survey of American Life (NSAL).....	2
National Latino and Asian American Study (NLAAS).....	3
Sample Design.....	4
National Comorbidity Study Replication (NCS-R) sample design.....	5
National Study of American Life (NSAL) sample design.....	6
National Latino and Asian American Study (NLAAS) Sample Design.....	7
Questionnaire Development.....	10
Survey Management.....	11
Data Collection.....	13
Data Processing.....	15
Data transmission.....	15
Data editing and coding.....	15
Data file creation.....	15
Data processing notes.....	15
Data file documentation.....	15
Weighting.....	16
I. Introduction.....	16
II. CPES Survey Population.....	16
III. Race/Ancestry Populations.....	16
IV. Sample Frame Geographic Domains.....	17
V. Integrated weight for the pooled CPES data set.....	18
VI. Final Weights and Special Analysis Consideration for Weighted Analysis.....	21
VII. Procedures for Sampling Error Estimation in Design-based Analysis of the CPES Data.....	23
VIII. Weights for Study pairs.....	27
IX. CPES Weights Chart.....	39
X. References and Readings.....	39
Replication of Results.....	44

## Introduction

The Collaborative Psychiatric Epidemiology Surveys (CPES) were initiated in recognition of the need for contemporary, comprehensive epidemiological data regarding the distributions, correlates and risk factors of mental disorders among the general population with special emphasis on minority groups. This project joins together three nationally representative surveys: the National Comorbidity Survey Replication (NCS-R), the National Survey of American Life (NSAL), and the National Latino and Asian American Study (NLAAS).

The primary objective of the CPES was to collect data about the prevalence of mental disorders, impairments associated with these disorders, and their treatment patterns from representative samples of majority and minority adult populations in the United States. Secondary goals were to obtain information about language use and ethnic disparities, support systems, discrimination, and assimilation in order to examine whether and how closely various mental health disorders are linked to social and cultural issues.

The CPES surveys were developed under the sponsorship of the National Institute of Mental Health (NIMH), and the data collection was conducted by the Survey Research Center (SRC) of the Institute for Social Research at the University of Michigan from early 2001 through the end of 2003.

The following provides a brief description of each of the three surveys combined in CPES.

### National Comorbidity Survey Replication (NCS-R)

The National Comorbidity Survey Replication (NCS-R) is a probability sample of the United States carried out a decade after the original 1992 NCS (NCS-1) was conducted. The NCS-R repeats many of the questions from the NCS-1 and also expands the questioning to include assessments based on the diagnostic criteria of the American Psychiatric Association as reported in the Diagnostic and Statistical Manual - IV (DSM-IV), 1994. The two major aims of the NCS-R were first, to investigate time trends and their correlates over the decade of the 1990s, and second, to expand the assessment in the baseline NCS-1 in order to address a number of important substantive and methodological issues that were raised by the NCS-1.

The NCS-R survey was administered to a sample of non-institutionalized English-speaking adults aged 18 or older residing in households located in the coterminous United States.

The survey was administered in two parts. Part 1 included a core diagnostic assessment of all 9,282 respondents. Part 2 included questions about risk factors, consequences, other correlates, and additional disorders. Part 2 was administered only to 5,692 of the 9,282 Part 1 respondents.

The interviews took place between February 2001 and April 2003. For the most part, interviews were conducted using laptop computer-assisted personal interview (CAPI) methods in the homes of the respondents.

### National Survey of American Life (NSAL)

The National Survey of American Life (NSAL) is a study designed to explore racial and ethnic differences in mental disorders, psychological distress, and informal and formal service use from within the context of a variety of presumed risk and protective factors in the African-American and Afro-Caribbean populations of the United States as compared with White respondents living in the same communities.

The NSAL survey was administered to a sample of non-institutionalized English-speaking adults aged 18 or older residing in households located in the coterminous United States. The survey identified three target survey populations: African American, Afro-Caribbean, and White American. The African American survey population included only Black adults who did not identify ancestral ties in the Caribbean. The Afro-Caribbean survey population was limited to Black adults who self-identified as being of Caribbean ancestry. The White survey population, drawn from predominantly Black neighborhoods, included all Caucasian adults except persons who self-reported Hispanic ancestry.

The interviews took place between early 2001 and the spring of 2003. For the most part, interviews were conducted using laptop computer-assisted personal interview methods in the homes of respondents (approximately 14 percent of interviews were conducted either partially or entirely by telephone).

The NSAL project yielded 6,199 adult interviews: 3,570 African American, 1,623 Afro-Caribbean, and 1,006 non-Hispanic white, although the actual number of cases in the CPES data file is 6,082 (3,570 African American, 1,621 Afro-Caribbean, and 891 non-Hispanic white). An extremely small sample (n=115) of White adults who were interviewed in households where the White subsample was less than 10% of the African American density stratum were excluded from the final dataset as well as two of the Afro-Caribbean interviews when it was later discovered that they were duplicate cases.

## **National Latino and Asian American Study (NLAAS)**

The National Latino and Asian American Study (NLAAS) is a nationally representative community household survey that estimates the prevalence of mental disorders and rates of mental health service utilization by Latinos and Asian Americans in the United States. The central aims of the NLAAS were threefold. First, to describe the lifetime and 12-month prevalence of psychiatric disorders and the rates of mental health services use for Latino and Asian American populations using nationwide representative samples of these groups. Second, to assess the associations among social position, environmental context, and psychosocial factors with the prevalence of psychiatric disorders and utilization rates of mental health services. Third, to compare the lifetime and 12-month prevalence of psychiatric disorders, and utilization of mental health services of Latinos and Asian Americans with national representative samples of non-Latino whites (drawn from the National Comorbidity Study-Replication (NCS-R) and African Americans (drawn from the National Survey of American Life (NSAL)).

The NLAAS survey was administered to a sample of non-institutionalized Latino and Asian American adults aged 18 or older residing in households located in the coterminous United States. The NLAAS sample identified four Latino target survey populations (Cuban, Mexican, Puerto Rican, and other adults of Latino descent), four Asian American target survey populations (Chinese, Filipino, Vietnamese, and other adults of Asian American descent), and a control group of non-Hispanic, non-Asian white respondents. Recognizing that language barriers may discourage survey participation for minorities who do not speak English, or are not fluent in it, the NLAAS instrument was translated into four languages.

The interviews took place between May 2002 and December 2003. For the most part, interviews were conducted using laptop computer-assisted personal interview methods in the homes of the respondents.

The NLAAS project yielded 4,864 adult interviews: 2,095 Asian respondents, 2,554 Latino respondents, and 215 non-Hispanic, non-Asian white respondents.

Much of the information in this User Guide is derived from Special Issue 2 of the International Journal of Methods in Psychiatric Research, Volume 13, Number 4, 2004 entitled "The NIMH Collaborative Psychiatric Epidemiology Surveys Initiative: Designs, Methods, and Instrumentation" (<http://www3.interscience.wiley.com/cgi-bin/jissue/112542415>). See also Volume 13, Number 2, 2004 of the same journal for additional information (<http://www3.interscience.wiley.com/cgi-bin/jissue/112542416>).

## Sample Design

The NIMH-CPES survey data collections were each based on a multi-stage area probability sample conducted in a total of 252 geographic areas or primary sampling units across the United States. The sample was selected using the sampling frames and selection procedures that are common to the University of Michigan Survey Research Center's (SRC) National Sample design. The national area probability samples for the three individual studies include unique features designed to optimize the cost and error properties of the study-specific samples. The general features of each study sample are summarized in Table 1.

<b>Table 1. Key features of the Collaborative Psychiatric Epidemiology Studies (CPES) sample designs</b>			
<b>Sample design feature</b>	<b>National Comorbidity Survey Replication (NCS-R)</b>	<b>National Survey of American Life (NSAL)</b>	<b>National Latino and Asian American Study (NLAAS)</b>
Survey population	All adults, age 18+ residing in households in the coterminous United States. Exclusions include institutionalized persons, those living on military bases, and non-English speakers.	African-American, Afro-Caribbean, and non-Hispanic white adults, age 18+ residing in households in the coterminous United States. Exclusions include institutionalized persons, those living on military bases and non-English speakers.	Latino and Asian-American adults, age 18+ residing in households in the coterminous United States, Alaska, and Hawaii. Exclusions include institutionalized persons and those living on military bases.
Sample frame	Four-stage national area probability sample.	Four-stage national area probability sample with special supplement for Afro-Caribbean adults.	Four-stage national area probability sample with special supplements for adults of Puerto Rican, Cuban, Chinese, Filipino and Vietnamese national origin.
Sample size	13,054 sample housing units screened for eligible adults. 9,282 completed interviews with eligible respondents.	26,495 sample housing units screened for eligible adults. 6,199 completed interviews with eligible respondents.	27,026 sample housing units screened for eligible adults. 4,649 completed interviews with eligible respondents.
Special features	Selection of two adult respondents in a subsample of households. Special study of main survey nonresponse.	Two-phase sample design to control survey costs in final stages of data collection.	Sample linked to NCS-R for statistical comparisons. Selection of two adult respondents in a subsample of households. Two-phase sample design to control survey costs in final stages of data collection.

## General features common to each CPES sample design

The selection of a probability sample of respondents for each study's interview required a four-step sampling process -- a primary stage sampling of U.S. Metropolitan Statistical Areas (MSAs) and counties, followed by a second stage sampling of area segments, a third stage sampling of housing units within the selected area segments, and concluding with the random selection of eligible respondents from the sample housing units.

The primary stage units (PSUs) of SRC's National Sample are either MSAs, single counties, or a grouping of geographically contiguous counties with small populations. In each CPES sample design, PSUs are assigned to explicit sampling strata based on MSA/non-MSA status, PSU size, geographic location, and population characteristics. Depending on the CPES study sample design, from 12 to 20 of the primary stage strata contain only a single self-representing (SR) metropolitan PSU. Each SR PSU is included in the sample with certainty in the primary stage of selection. The remaining non-self-representing (NSR) primary stage strata in each design contain more than one PSU. From each of these NSR strata, one PSU is sampled with probability proportionate to its size measured in occupied housing unit counts reported at the most recent census.

The designated second-stage sampling units (SSUs) in each CPES sample design are termed area segments. Area segments were formed by linking geographically contiguous census blocks to form units with a minimum number of occupied housing units (typically 50 to 100 based on the needs of the study). Within primary stage units, area segments were stratified at the county level by geographic location and race/ethnicity composition of residents' households. The race/ethnicity stratification of area segments played a particularly important role in the NSAL and NLAAS sample designs where it was used both to improve the sampling precision of the design and as a basis for more cost-effective oversampling in area segments with higher densities of households for targeted race and ethnicity subpopulations. Within each second stage stratum, the actual probability sampling of area segments was performed with probabilities proportionate to census counts of the occupied housing units for the census blocks that comprise the area segment.

The SRC field staff conducted an up-to-date enumeration or 'listing' of all housing units located within the physical boundaries of the selected area segments for each CPES sample design. A third-stage sample of housing units was then selected for screening interviews according to a predetermined sampling rate.

The third stage sampling rate was computed for each selected area segment in the CPES sample design. This rate was then used to select a systematic random sample of actual housing units from the area segment listing. Each sample housing unit was contacted in person by an interviewer. Within each cooperating sample household, the interviewer conducted a short screening interview with a knowledgeable adult to determine if household members met the study eligibility criteria. If the informant reported that one or more eligible adults lived at the sample housing unit address, the interviewer prepared a complete listing of household members and proceeded to randomly select a respondent for the study interview. The random selection of the respondent was performed using a special adaptation of the objective household roster/selection table method developed by Kish (1949).

## National Comorbidity Study Replication (NCS-R) sample design

The survey population for the NCS-R included all U.S. adults aged 18 years and older residing in households located in the coterminous 48 states. Institutionalized persons including individuals in prisons, jails, nursing homes, and long-term medical or dependent care facilities were excluded from the survey population. Military personnel living in civilian housing were eligible for the study but due to security restrictions residents of housing located on a military base or military reservation were excluded. Adults who were not able to conduct the NCS-R interview in English were not eligible for the survey.

The NCS-R was designed to be a cross-sectional replication of the original 1992 National Comorbidity Survey (NCS; Kessler, 1994). To improve the statistical efficiency for cross-time comparison of results from these two surveys, a decision was made early in the NCS-R planning process to maximize the overlap in the primary and secondary stages of the multi-stage sample designs for the two studies. Data from the 2000 U.S. census were not available at the time of the NCS-R sample selection. Therefore, the primary stage design for the NCS-R was carried forward directly from the 1992 NCS multi-stage sample selection with no changes in primary stage strata or PSU definitions and no adjustment to the 1990 census-based measures of size or primary stage selection probabilities. The shared NCS/NCS-R primary stage sample design consisted of a single PSU selection from each of 62 primary stage strata. As shown in Table 2, 16 of these NCS-R PSUs were the largest self-representing MSAs. A total of 31 non-self-representing PSU selections represented the remaining MSAs in the US survey population. More rural non-MSA counties were represented by 15 nonself-representing PSU selections.

Table 2 provides a summary of primary and secondary stage sample allocation for the NCS-R study.

Table 2 . Primary and secondary stage sample allocation for the NCS-R			
Domain of Primary stage strata	Number of NCS-R Primary Stage Units (PSUs)	Number of NCS-R second stage units (SSUs)	
		Total NCS area segments	New listing needed
Total	62	1001	174
SR MSA	16	395	24
NSR MSA	31	402	48
NSR Non-MSA	15	204	102

## National Study of American Life (NSAL) sample design

The NSAL survey populations included all US adults in the three target groups who were age 18 and older and resided in households located in the coterminous 48 states. The African-American survey population included only Black adults who did not identify ancestral ties in the Caribbean. The Afro-Caribbean survey population was limited to Black adults who self-identified as being of Caribbean ancestry. The White survey population included all Caucasian adults except persons of self-reported Hispanic ancestry. Institutionalized persons including individuals in prisons, jails, nursing homes, and long-term medical or dependent care facilities were excluded from the study population. Military personnel living in civilian housing were eligible for the study but residents of housing located on a military base or military reservation were excluded. The NSAL survey populations were restricted to adults who were able to complete the interview in English.

The NSAL multi-stage sample design combines a 'core' national area probability sample of households with a special supplemental sample of households in areas of higher Afro-Caribbean residential density. The NSAL Core national sample is designed to be optimal for a national study of the African-American survey population. The design of the NSAL Core sample closely resembles that used for the 1979-80 National Survey of Black Americans (NSBA) (Hess, 1985; Jackson, 1991). The NSAL Supplement design served solely to augment the sample size from the Afro-Caribbean survey population in a cost and statistically efficient manner and did not contribute to the representative samples of the NSAL's African-American and White survey populations. The NSAL national area probability sample was selected independently of the sample for the NCS-R and the NLAAS although the three designs share many common features such as PSU and area segment definitions and sample selection methods.

The Survey Research Center (SRC) 1990 National Sample of US households (Heeringa et al., 1994) was the starting point for NSAL sample selection. To adapt the sample to be optimal for a national study of the African-American survey population for NSAL, some modification to the primary stage of the basic 1990 SRC National Sample design was needed. The definitions of the primary sampling units in the primary stage frame for the SRC National Sample remained unchanged, but measures of size used in the PPS selection of PSUs were changed from 1990 census counts of total occupied households to African-American occupied households. Some reorganization (combining, splitting) of 1990 'A' National Sample strata (Heeringa and Redmond, 1994) was also required to transform the design from one that was optimal for surveys of all US households to one that emphasized precision for samples of African Americans.

As shown in Table 3, the NSAL Core primary stage design includes 64 PSU selections. The eight largest self-representing MSA PSUs in the 'A' partition of the SRC National Sample remained self-representing (SR) selections in the NSAL primary stage sample. An additional 13 MSA PSUs were designated as self-representing PSUs for the NSAL on the basis of the size of their African-American population, bringing the total number of NSAL SR PSUs to 21. The NSAL primary stage design includes 43 NSR selections, 14 PSUs selected from strata representing the MSA and non-MSA regions of the census Northeast, Midwest and West regions and 29 PSUs selected from MSA and non-MSA strata representing the census South - the region that includes almost 50% of the US African-American population. The primary stage sample allocation for the urban and rural areas of the census South region was deliberately increased to improve sample precision for national estimates derived from the African-American sample. The PPS selection of the 43 NSAL NSR PSUs used a probability sampling method that maximized the overlap of the NSAL primary stage sample with the 1990 SRC National Sample 'A' partition selection for the design stratum. The objective in maximizing the overlap of the NSAL primary stage sample with the 1990 SRC National Sample was, where possible, to take advantage of experienced, trained SRC staff in the National Sample primary stage sample locations.

<b>Domain of Primary Stage Strata</b>	<b>Number of NSAL Primary stage units (PSUs)</b>	<b>Number of NSAL Second Stage Units (SSUs)</b>
Total	64	456
SR MSA	21	198
NSR MSA	27	162
NSR Non-MSA	16	96

The NSAL sample of African-Americans was identified exclusively from the screening of the sample of housing units selected from the 456 NSAL Core area segment listings.

The NSAL sample of Afro-Caribbean households was identified through samples selected from two overlapping area probability sample frames. The first sample source for Afro-Caribbean respondents was from the screening of households in the nationally representative NSAL Core sample. As described above, all sample housing units in this national probability sample were contacted and a screening interview was conducted with each eligible, cooperating household. In total, 266 adult Afro-Caribbeans were successfully interviewed in the NSAL Core national sample. Therefore it was necessary to supplement the NSAL Core sample in order to achieve the original NSAL target sample size of 1,600 Afro-Caribbeans.

Construction of the NSAL Caribbean Supplement sample began with the selection of a stratified sample of eight supplemental PSUs. From these eight PSUs, 86 area segments were selected from the set of qualifying census block groups within the PSUs. To qualify for the Caribbean Supplement, a block group population needed to be at least 10% Afro-Caribbean (based on the 1990 census estimates). Once the primary and secondary stage sampling units were selected, field staff visited each area segment to list housing units.

<b>Table 4 . Primary and secondary stage sample allocation for the NSAL Caribbean Supplement</b>		
<b>Domain of Primary stage strata</b>	<b>Number of NSAL Primary Stage Units (PSUs)</b>	<b>Number of NSAL Second Stage Units (SSUs)</b>
Total	8	86
SR MSA	5	66
NSR MSA	3	20

The NSAL White sample was a stratified, disproportionate sampling of non-Hispanic white adults in the US household population. Although in the strictest sense it may be described as a nationally representative sample of White adults, it is not optimal for descriptive analysis of the U.S. White adult population. Instead, the NSAL White sample was designed to be optimal for comparative descriptive and multivariate analyses in which residential, environmental and socioeconomic characteristics are carefully controlled in the black/white statistical contrasts. As described above, the NSAL sample of White adults was identified by screening the national probability sample of housing units selected for the NSAL Core. The original completed interview target for the NSAL White sample was set at  $n = 1,800$ . Later in the study period, a decision was made to reduce this target to  $n = 1,000$  White adult interviews based on survey costs and updated analysis objectives for the NSAL project. By the nature of its equal probability national sampling of all US households, the NSAL Core screening for eligible African-American and Afro-Caribbean households was projected to identify far more eligible White households than required to meet the sample size target. Therefore, subsampling of eligible White adults at the screening stage was employed to bring the sample of interviews with this group in line with the study targets.

## National Latino and Asian American Study (NLAAS) Sample Design

The survey populations for the NLAAS study included all Latino and Asian American adults who resided in households in the US states and Washington, DC. Latinos were divided into four strata of interest: Mexican, Puerto Rican, Cuban, and all other Latinos. The Asian American survey population was also stratified based on eligible adults' ancestry or national origin: Chinese, Filipino, Vietnamese, and all other Asians. This stratification of the NLAAS survey populations relied on self-reports by household members at the time of the household screening. In cases where a member of the survey population reported belonging to more than one Latino or Asian American target population, the following order of priority was used to assign individuals to a single group for the purpose of the stratified sample selection:

1. Vietnamese;
2. Cuban;
3. Filipino;
4. Puerto Rican;
5. Chinese;
6. Mexican;
7. other Asian; and
8. other Latino.



Institutionalized persons including individuals in prisons, jails, nursing homes, and long-term medical or dependent care facilities were excluded from the study populations. Military personnel living in civilian housing were eligible for the study, but due to security restrictions residents of housing located on a military base or military reservation were excluded.

The NLAAS is based on a stratified probability sample design that includes multiple area probability sample components:

- An NLAAS Core sampling of PSUs, area segments, and housing units that is designed to be nationally representative of all US populations including Latinos and Asians
- The NLAAS High Density (HD) supplemental samples, targeted oversamples of geographic areas with greater than 5% residential density for individual national origin groups of interest in the NLAAS

The NLAAS Core sample is designed to provide a nationally representative sample of Latinos and Asian Americans without regard to geographic residential patterns. The price for the national representation under the NLAAS Core sample design was a high per unit cost of data collection for eligible respondents. This high cost per interviewed case was due to the fact that many area segments in the Core sample had very low density of the populations of interest in NLAAS and there was a need to screen large numbers of households to identify the targeted samples of Latinos and Asians. Even for the more prevalent and widely distributed Mexican or Chinese ancestry groups, it was very costly to screen a general national area probability sample to identify and interview a large nationally representative sample of eligible adults. Survey costs would have been prohibitively high if this method alone had been used to obtain desired numbers of sample observations of less prevalent national origin groups (such as Puerto Ricans, Cubans, Filipinos, and Vietnamese).

To maximize the statistical efficiency of comparisons between the NLAAS survey populations and the larger US adult population, the primary and secondary stages of the NLAAS Core national sample design were completely integrated with the National Comorbidity Survey Replication (NCS-R) national sample design. The NLAAS Core and NCS-R designs shared the same 62 primary areas representing the MSA and non-MSA strata for the 48 coterminous United States (see Tables 2 and 5). Since full representation of Asian ancestry populations was critical to the NLAAS, the Honolulu HI MSA was added to the primary stage sample as a metropolitan self-representing PSU, bringing the total number of NLAAS National Sample PSUs to 63. The second stage of the NLAAS national sample design component was also fully integrated with the second stage of the NCS-R national sample. The two designs did not share exactly the same area segments and housing unit listings; however, each selected NLAAS Core area segment was paired with an NCS-R area segment and the paired segments from the two samples were physically adjacent to one another - maximizing the 'geographical/ecological correlation' of the two samples (Kish, 1987). The decision to introduce geographic 'overlap' with the NCS-R to the NLAAS Core national sample was based on statistical aims for the NLAAS. A primary aim of the NLAAS was to enable comparisons of mental health characteristics both among the NLAAS survey populations of Latinos and Asians and with the larger US population. Full geographic linkage of the NLAAS national sample area segments to the NCS-R maximized the geographic and socio-economic correlation of the two samples. Since both the NCS-R and the NLAAS Core were designed to be nationally representative, this 'correlation of designs' produced no major inefficiencies for stand alone analysis of the NLAAS survey data but significantly reduced the variance of statistical analyses designed to contrast the populations from the two studies.

For the purpose of statistical efficiency in comparing the NLAAS survey population and the larger US adult population, the primary and secondary stages of the NLAAS Core national sample design were integrated with the National Comorbidity Survey Replication (NCS-R) national sample design.

<b>Table 5 . Primary and secondary stage sample allocation for the NLAAS Core National Sample</b>			
<b>Multi-stage sample design units</b>	<b>Primary stage sample domain</b>		
	<b>Total</b>	<b>SR</b>	<b>NSR</b>
NLAAS Core primary stage sample units (PSUs)			
Core PSUs in original NCS-R national sample (plus Honolulu)	63	17	46
Core PSUs not fielded due to near zero expected interviews	25	1	24
Core PSUs fielded in NLAAS household screening	38	16	22

<b>Table 5 . Primary and secondary stage sample allocation for the NLAAS Core National Sample</b>			
<b>Multi-stage sample design units</b>	<b>Primary stage sample domain</b>		
	<b>Total</b>	<b>SR</b>	<b>NSR</b>
NLAAS Core second stage sample units (SSUs)			
Core SSUs matched to NCS-R national sample (plus Honolulu)	474	204	270
Core SSUs not fielded due to near zero expected interviews	157	13	144
Core SSUs fielded in NLAAS household screening	317	191	126
Not high density for special HD oversample populations	26	137	126
High-density Puerto Rican	17	17	0
High-density Cuban	4	4	0
High-density Chinese	15	15	0
High-density Filipino	14	14	0
High-density Vietnamese	4	4	0

As with the NSAL study, the screening and interviewing process of the NLAAS was also conducted through a two-phase method:

<b>Table 6 . Primary and secondary stage sample allocation for the NLAAS high-density (HD) samples</b>			
<b>NLAAS HD supplemental samples:Multi-stage sample design units</b>	<b>Primary stage sample domain</b>		
	<b>Total</b>	<b>SR</b>	<b>NSR</b>
NLAAS HD primary stage sample units (PSUs)			
High-density Puerto Rican PSUs	20	12	8
High-density Cuban PSUs	9	7	2
High-density Chinese PSUs	17	13	4
High-density Filipino PSUs	18	10	8
High-density Vietnamese PSUs	18	12	6
NLAAS-HD second stage sample units (SSUs)			
High-density Puerto Rican SSUs	51	34	17
High-density Cuban SSUs	70	66	4
High-density Chinese SSUs	46	34	12
High-density Filipino SSUs	51	32	19
High-density Vietnamese SSUs	60	43	

## Questionnaire Development

The core CPES questionnaire was based largely on the World Health Organization's (WHO) expanded version of the Composite International Diagnostic Interview (CIDI) developed for the World Mental Health (WMH) Survey Initiative, the WMH-CIDI (Kessler and Üstün, 2004). For additional information, see this video on the development of the CIDI (<http://www.icpsr.umich.edu/CPES/files/Kessler-01.mpg>) .

All three surveys used a modified version of the WMH-CIDI, which had been developed over the course of more than a year by an international group of collaborators, as described by Kessler and Üstün (2004). The design of the WMH-CIDI involved modifications and additions to the existing WHO-CIDI. The CIDI was an expansion of the Diagnostic Interview Schedule (DIS), the first standardized psychiatric diagnostic interview developed for administration by lay interviewers. The CIDI was designed to produce diagnoses based on WHO International Classification of Disease (ICD) criteria, while diagnoses from the DIS could only be made based on American Psychiatric Association (APA) Diagnostic and Statistical Manual (DSM) of Mental Disorders criteria. As summarized by Wittchen (1994), Kessler and Üstün (2004), DIS and CIDI validity studies suggest that there is a significant correlation between diagnoses based on data collected using the DIS and CIDI and diagnoses made by clinicians who reinterview a sample of respondents. Researchers who were involved in the design of the WMH-CIDI were interested in continuing to enhance the quality of survey measurement. As part of the WMH-CIDI development process, debriefing interviews were conducted with CIDI respondents and several methodological improvements were made to the instrument based on findings from the debriefing interviews. See Kessler et al. (2000) and Kessler and Üstün (2004) for a detailed discussion of the content additions and methodological enhancements that were incorporated in the WMH-CIDI instrument. The WMH and CPES questionnaires were administered using computer-assisted interviewing (CAI). The instruments were programmed using Blaise, a CAI software package developed by Statistics Netherlands and used by many government statistical agencies and large survey research organizations worldwide (Blaise Survey Processing System, Version 4.5). Blaise software is specifically designed to accommodate very complicated questionnaire skip patterns and sub-sampling algorithms. Potential drawbacks of using Blaise include its cost and the requirement for highly trained programmers to write the code for complex surveys.

The questionnaire design and testing phase for each project spanned approximately one year. The initial goal was to finalize the NCS-R questionnaire first, particularly the psychiatric disorder sections, in order to make these sections available to NSAL and NLAAS. However, the time required to finalize the NCS-R questionnaire was lengthened because development of the NCS-R questionnaire and the World Mental Health Survey Initiative questionnaire occurred in tandem, resulting in many revisions to the core questionnaire over this period.

# Survey Management

An attempt was made to standardize the interview materials across the studies as much as possible. Table 7 provides an overview of the study materials for the CPES project. Nine hundred and forty-six interviewers were recruited and trained based on specific requirements of the project, such as matching the interviewer and respondent race in NSAL or considering language preference of household for the NLAAS project.

Table 7. Interviewer materials for CPES studies	
Study material	Description
Field interviewer manual	Contents: Study overview (e.g., research goals, SRC organizational structure, funding sources for study) Study-specific interviewing procedures Guidelines on contacting households and encouraging cooperation Eligibility screening procedures Importance and necessity of informed consent and consent procedures Interviewer instructions for administration of the questionnaire Other miscellaneous topics (e.g., working with computer hardware and software, quality control measures, procedures for reporting payroll, expenses and productivity)
Interviewer training workbook	Contents: Training slides and exercises Summary sheets or checklists for complex procedures Copies of respondent letters, consent documents, and incentive payment receipt
Sample information	Detailed maps and listing of addresses within each area to assist in locating households A coversheet for each address contained the following information: Sample household address, unique identification number, interviewer to whom case was assigned Study introduction Screening and respondent selection instructions Informed consent statement (for NLAAS) Space for date, time, and outcome of every call attempt Notes to assist in future calls to the household
Identification badge	Picture identification badge for interviewers that confirmed their employment with SRC
Various letters	Introductory letter and study brochure sent prior to initial contact attempt 'Not at home' letter for households that could not be reached after multiple visits Letter for local law enforcement agencies that notified the agency that interviewing staff would be working in the area Press release for the local news media Letters of support from the director of NIMH Project updates highlighting preliminary findings
Question by question instructions	Instructions or definitions for interviewers that were provided on-screen or accessible by pressing a function key and accessing online help file Instructions were also included in the field interviewer manual
Respondent booklet	Reference for respondents to use during the interview. Booklet contained the following information: Response categories for questions with complex response options Response categories for long lists of options Visual aids to assist respondent's understanding of response options

The 946 interviewers that were trained for the CPES project were distributed across each study as follows: 342 interviewers for NCS-R, 329 interviewers for NSAL, and 275 interviewers for the NLAAS. Study-specific training lasted five to seven days, depending on the study and which components were being covered.

The training sessions consisted of five main components: (1) instruction on household eligibility and respondent selection procedures; (2) questionnaire training, which included a section-by-section review of each module of the questionnaire, followed by question and answer sessions and two-hour practice sessions; (3) computer training and practice sessions; (4) review of interview procedures and study materials; and (5) mock interviews in which interviewing and administrative tasks were integrated to model realistic interviewing experiences. To better convey the content and to engage the training participants, trainers used a variety of formats, including large and small group lectures, round-robin practice sessions, mock interviews and one-on-one help sessions. Participants were given homework assignments, which the trainers reviewed to identify interviewers who were having problems with the computer hardware or software. For later trainings, experienced interviewers served as trainers for the two days of general interviewer training. The new interviewers benefited from the descriptions of the experiences of these interviewers, who were able to provide tested and concrete suggestions on how best to handle all aspects of the job.

All three studies provided training in sensitivity to cultural, racial, and socioeconomic diversity that would be encountered while conducting face-to-face interviews. Additional training was also provided on how to interview on sensitive or potentially embarrassing topics. Finally, because some of the questionnaire topics covered subjects that could reveal information about pending harm to the respondent or others, interviewers were trained on their legal obligations and on how to handle these rare but critical situations.

Determining respondent eligibility, which involved a complex procedure of sorting household members into various categories defined by age, gender, race, and ethnicity, was often a challenge for interviewers. Interviewers were provided with initial and ongoing training on the importance of and techniques for reducing non-response, and a wide variety of tools and procedures was developed at the beginning of each project to maximize respondent participation. Additional refusal conversion strategies and protocols were devised and implemented as the study progressed including ongoing interviewer training, 'tailoring' which involved the use of different approaches in introducing the survey to the respondent, respondent incentives (including \$50 offered for participation), persuasion letters, special respondent recruitment offers, distinctive mailings, a toll-free respondent telephone line, use of travelling interviewers, responsive design approaches, increased respondent incentives (up to \$150), an abbreviated interview schedule, interviewer incentives, subsampling, and institutional review board considerations.

## Data Collection

The organizational structure of the field and central data collection staff for the NCS-R, NSAL, and NLAAS was divided into teams of 6 to 12 interviewers. Each team was supervised by a team leader. Approximately three to four teams formed a workgroup, which was supported by a team leader coordinator. Each workgroup was assigned to a regional field manager, who was responsible for the workgroup's interview production efforts, quality control, and personnel management. Whenever possible, teams were comprised of groups of interviewers from the same region. Every effort was made to assign interviewers to teams prior to training so that interviewers on the same team would be able to work together during training.

An overview of the field organization and structure of the data collection for all three studies is described in Figure 1.

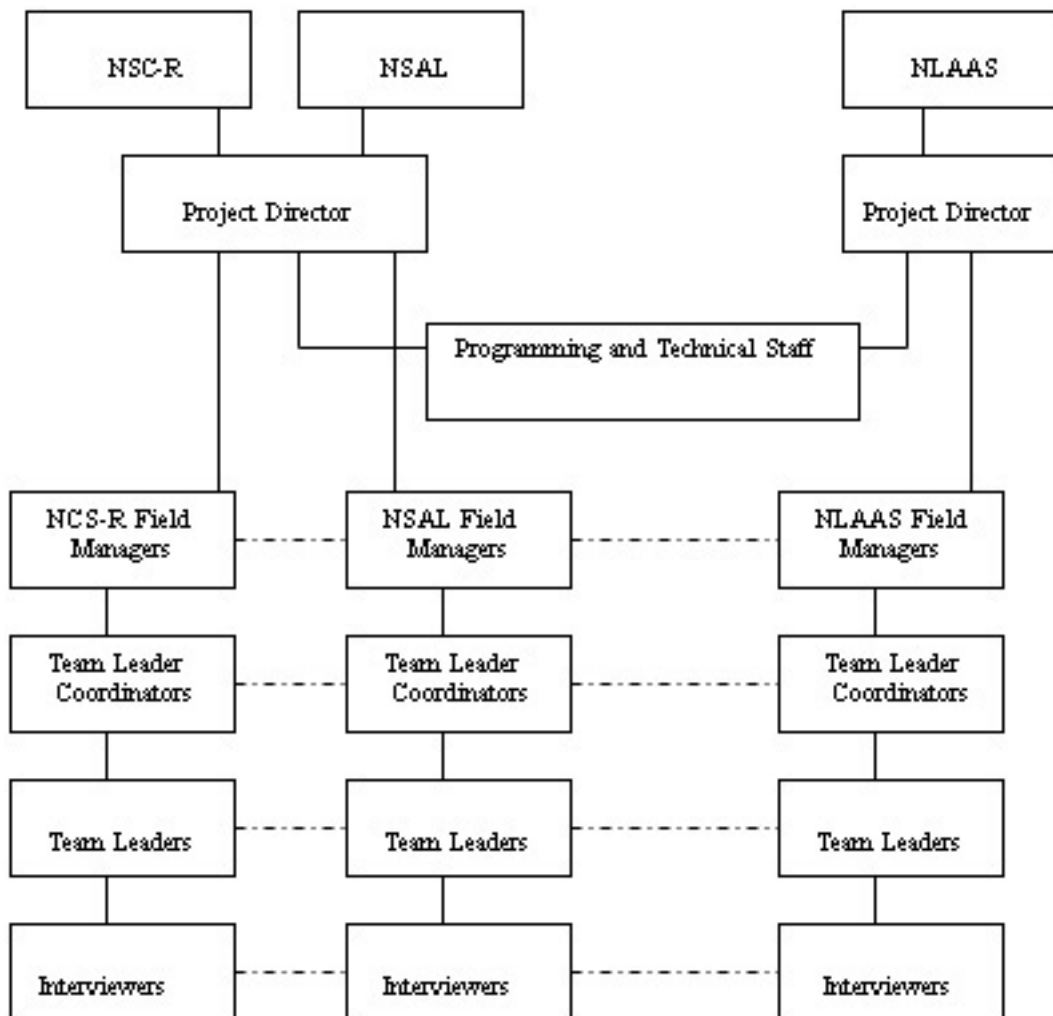


Figure 1. Organizational Structure for CPES data collection

For NCS-R, a sample of 13,054 addresses was fielded. In 98.1% (11,222) of occupied housing units (11,443), interviewers were able to determine whether or not the household was eligible for inclusion in the study. A total of 10,622 addresses yielded an eligible household. As summarized in Table 8 below, 9,282 adult interviews were completed: 7,693 interviews with the main respondent and 1,589 interviews with a second adult in the household. A total of 554 additional interviews (not included in Table 8) were collected using a shortened form of the instrument with a subsample of non-respondents in an effort to assess non-response bias.

<b>Table 8. Interviews, response rate, interview length, and number of contacts for NCS-R, NSAL, NLAAS</b>				
<b>Component</b>	<b>Interviews</b>	<b>Response rate (%)</b>	<b>Averageinterviewlength (mins)</b>	<b>AverageContactsper interview</b>
<b>NCS-R</b>				
Main respondent	7,693	70.9	126	7.1
Second respondent	1,589	80.4	124	4.7
<b>NSAL</b>				
Adult respondent	4,842 (core)1,357 (supplement)	71.5 (core)76.4 (supplement)	145	7.4
<b>NLAAS</b>				
Main respondent	3,620	75.7	161	9.2
Second respondent	1,029	80.3	152	11.6

The final weighted response rate for NCS-R, excluding the Part 1 interviews, was 70.9% for primary respondents (American Association for Public Opinion Research (AAPOR) Response Rate 3; AAPOR, 2004). The response rate for the second respondent was 80.4%. See Kessler et al. (2004) and Heeringa et al. (2004) for a complete discussion of final sample outcomes and response rates.

In contrast to the relatively high sample yield for NCS-R, in NSAL, 11,634 eligible households were identified from 26,495 randomly sampled addresses. A total of 6,199 adult respondents were interviewed as part of NSAL (1,006 White respondents, 1,623 respondents of Caribbean descent, and 3,570 African American respondents). The overall response rate for the core NSAL national sample was 71.5% (AAPOR Response Rate 3; AAPOR, 2004). The Caribbean Supplement sample, which was designed to target areas with high concentrations of persons of Caribbean origin, yielded a weighted response rate of 76.4%. See Jackson et al. (2004) and Heeringa et al. (2004) for additional description of the final distribution of sample dispositions and response rates.

The NLAAS project also involved a significant amount of screening. From a total sample of 27,026 addresses, 4,345 eligible main respondents and 1,234 eligible second adult respondents were identified. For NLAAS, 3,620 main respondent and 1,029 second adult interviews were completed. The weighted response rate for NLAAS was 75.7% among main respondents (77.6% for Latinos, 69.3% for Asians) (AAPOR Response Rate 3; AAPOR, 2004). For second respondents, the final response rate was 80.3% (82.4% for Latinos, 73.7% for Asians). As noted earlier, 215 non-Asian, non-Hispanic white controls were interviewed prior to the elimination of the interviewing of White controls due to budget constraints. More detailed information regarding sample outcomes and calculation of response rates can be found in Heeringa et al. (2004). See Table 8 for a general summary of data collection results across the three studies.

## Data Processing

### Data transmission

Each day interviewers connected their laptops to a telephone line and dialed up an ISP to submit and receive information from the central office. The purpose of the communication was twofold. First, all work performed since the previous communication was transmitted to the central office and the information updated in the master files. As part of the daily communication, interviewers reported the hours that they worked and their expenses. This information transfer allowed for real-time monitoring of interview production, sample disposition, and costs. Second, information from the central office, including program updates, transfers of sample between interviewers, and newly released sample, was transmitted electronically to interviewers.

### Data editing and coding

Many data processing activities that are typically completed when collecting data via paper questionnaires were unnecessary in these studies because the questionnaires were computer administered. For example, making sure that questions were asked in the correct sequence, checking for out-of-range or inconsistent responses, and filling in the appropriate question text based on a respondent's previous answers were all controlled by the interviewing application software. Inconsistent responses that failed the programme's edit checks were brought to the attention of the interviewer who could resolve the inconsistency with the respondent during the interview, improving the quality of data and minimizing the need for back-end editing. Although these software programs automatically performed many of the decisions formerly made by interviewers using paper questionnaires, the data for each study did require some additional editing and coding. Editing operations included processing each interview through a series of programming routines that evaluated question responses and assigned codes to indicate the presence or absence of each mental health disorder assessed by the study. A number of other summary variables based on individual question items were also created in preparation for the project's analysis phase. In addition, each study included several open-ended questions, which were coded.

### Data file creation

Data files were extracted to ASCII and converted to SAS format once data collection began and were updated throughout the data collection period. Early in the data collection phase, these files were used by project managers to identify any problems with administration of the questionnaire and to monitor response trends and patterns. Datasets were produced for the studies' principal investigators on a weekly basis to allow the investigators' staffs to perform preliminary analyses.

### Data processing notes

Additional editing was performed on the files by research and SRO staff for several different purposes, including harmonization, disclosure, consistency, and clarity. For further information, please consult our Data Processing Notes (<http://www.icpsr.umich.edu/CPES/files/processor-notes.pdf>) document, which fully details the changes made to the data files.

### Data file documentation

A codebook and set of companion instructions and study materials were prepared for each study. The codebook provided the information that users need to associate a variable in the data file with the corresponding question on the questionnaire and documented the characteristics of each variable in the data set, such as its format and response codes. The codebook also contained frequencies for nominal and ordinal variables and a set of basic descriptive statistics for continuous variables. For NLAAS, the HTML-compatible codebook included a facility to view each question in any of the five languages.



# Weighting

## **National Institutes of Mental Health (NIMH) Collaborative Psychiatric Epidemiology Survey Program (CPES) Data Set. Integrated Weights and Sampling Error Codes for Design-based Analysis**

**Steven G. Heeringa, Patricia Berglund Statistical Design Group, Survey Research Center, University of Michigan**

June 4, 2007

### **I. Introduction**

Under contract to the National Institutes of Mental Health (NIMH), the Survey Research Center (SRC) has developed an integrated data base for the Collaborative Psychiatric Epidemiology (CPES) surveys: National Comorbidity Survey-Replication (NCS-R), National Survey of American Life (NSAL) and National Latino and Asian American Study (NLAAS). Heeringa, et al (2004) describe the sample designs and sample outcomes for the three CPES surveys. A general description of the survey methodology for the CPES surveys can be found in Pennell, et al. (2004).

This technical report outlines the method for integrating the design-based analysis weights and variance estimation codes for these three studies to permit analysts to approach analysis of the combined dataset as though it were a single, nationally-representative study.

The method of integrating the analysis of these three major survey programs was based on an adaptation of a multiple frame approach to estimation and inference for population characteristics (Hartley, 1962, 1974). There are several features and advantages to the method that are worth noting:

1. It was built on all of the study-specific weight development efforts conducted to date (Kessler et al. , 2004; Heeringa et al. 2004; Heeringa, et al. 2006).
2. It integrated overlapping representation of domains of the CPES survey population in a way that was mathematically transparent and easily understood by analysts of the combined data set. Given the large investments in study-specific weight development, this approach minimized the chance for conceptual or computational errors.
3. It was centered on the assumption that, conditional on the sample domain (e.g., block groups with 10-29.9% African American population) and the race/ethnicity of the respondent (e.g., Mexican-American), each study's sample representation based on the revised weight is proportional to the number of cases it "contributes" to the geographic domain x race/ethnicity cell.

### **II. CPES Survey Population**

The CPES survey population was defined by the union of the survey populations for the three component studies. This included adults age 18 and older, living in households in the 48 coterminous United States (NCS-R, NSAL). The survey population for the Latino and Asian ancestry groups extended to the State of Hawaii as well.

### **III. Race/Ancestry Populations**

CPES analysts are free to define respondent groupings for analysis; however, for purposes of weight development twelve specific race/ancestry groupings were initially specified. These groupings are listed in Table 1. Due to the small number of

persons of Other ancestry interviewed in the NCS-R, those individuals were combined with the White race category for purposes of the CPES weight computation.

<b>Table 1: Race/Ancestry Groupings Required For CPES Weight Development</b>	
<b>CPES Race/Ancestry Population Group</b>	<b>Survey Populations</b>
Vietnamese	NCS-R, NLAAS
Filipino	NCS-R, NLAAS
Chinese	NCS-R, NLAAS
All Other Asian *	NCS-R, NLAAS
Cuban	NCS-R, NLAAS, NSAL
Puerto Rican	NCS-R, NLAAS, NSAL
Mexican	NCS-R, NLAAS
All Other Hispanic *	NCS-R, NLAAS, NSAL
Afro-Caribbean (non-Hispanic)	NCS-R, NSAL
African-American (non-Hispanic)	NCS-R, NSAL
White	NCS-R, NSAL
All Other (Pacific Islander, Native American, etc.)	NCS-R
* Based on NLAAS screening criteria	

The breakdown of the full population into these 12 race/ancestry populations was a direct result of the specific eligibility and oversampling provisions of the NSAL and the NLAAS study designs. As shown in Table 1, NCS-R provided nearly universal coverage of all 12 race/ancestry groups. NSAL and NLAAS provided in-depth coverage of specific populations and with the exception of Afro-Caribbeans from Spanish language countries in the Caribbean (e.g., Cuba, Dominican Republic), the oversampling in each of these two studies did not overlap.

These 12 population groupings form the first dimension of a two-dimensional array that was used to apportion/adjust study-specific weights to create a new weight variable for integrated CPES analyses. These "population" groupings were defined at the respondent level. If individual respondents had multiple race/ancestry, they were assigned to a single category according to the priority order in the NLAAS and NSAL respondent classification rules (e.g. Afro-Caribbean taking preference over African-American, Vietnamese over Chinese). If ancestry for NCS-R cases could not be explicitly established at the level of detail required to map them into the NSAL or NLAAS population categories, they were stochastically assigned to a category based on the prevalence of each population in the Census Block Group in which the respondent's household was located.

## IV. Sample Frame Geographic Domains

The second dimension of the CPES weight computation array was defined based on the geographic domain of the U.S. national sample frame with which individual area segments for the three component samples were associated (see Heeringa et al., 2004). The "domain" groupings were assigned at the area segment level. All respondents from the same segment, regardless of population, were assigned to the same domain. Table 2 defines the 11 domain categories that were used to classify area segments and thereby assign each CPES respondents to a geographic domain.

<b>Table 2. Sample Frame Geographic Domains Required for CPES Weight Development</b>	
<b>CPES Domain</b>	<b>Domain Definition</b>
1	Census Block Group >5% Cuban Population
2	Census Block Group >5% Vietnamese Population
3	Census Block Group >5% Filipino Population
4	Census Block Group >5% Puerto Rican Population

<b>Table 2. Sample Frame Geographic Domains Required for CPES Weight Development</b>	
<b>CPES Domain</b>	<b>Domain Definition</b>
5	Census Block Group >5% Chinese Population
6	Census Block Group >10% Afro-Caribbean (non-Hispanic) (Restricted to NY, NJ, FL, CT, MA, RI and DC)
7	Census Block Group 60-100% African-American
8	Census Block Group 30-59.9% African-American
9	Census Block Group 10-29.9% African-American
10	Census Block Group 0-9.9% African-American
11	Hawaii (NLAAS only)

All segment assignments to geographic domains were performed using Census 2000 data for Block Groups. Like the population assignments for mixed ancestry respondents, area segment domain assignments based on this 11 category classification were not always unique. For example a Census Block Group might have contained a population that was >5% Vietnamese and also >5% Chinese. In cases where a Census Block Group qualified for more than one domain, the corresponding area segment was assigned to the lowest numbered category (e.g. the high density Vietnamese domain for this last example).

## V. Integrated weight for the pooled CPES data set

Case-specific population weights had been developed for each CPES component survey (Kessler et al., 2004; Heeringa, 2004; Heeringa et al., 2005). Each project had carefully developed and refined its weight vector to enable robust probability sampling inference ("design-based") to its chosen survey population. NCS-R was unique among the three component studies in that it required two final analysis weights—one weight for the full sample of cases who participated in the Part 1 interview and a second for the subsample of cases that also completed Part 2 of the NCS-R. Consequently, the CPES combined data set also has two analysis weights—the first for analysis of common data items and the second for analysis of survey items that NCS-R only administered to Part 2 respondents.

The integrated weight development began with the existing final population weights for the NCS-R, NSAL, and NLAAS. The integrated weight development then proceeded according to the following steps:

Step 1. Each NLAAS, NSAL, and NCS-R case was assigned to a race/ancestry category based on the categories and priority order provided in Table 1 (see Section III).

Step 2. Each NLAAS, NSAL, and NCS-R area segment was assigned to a geographic domain based on the definitions and priority order shown in Table 2 (see Section IV). Each NLAAS, NSAL, and NCS-R respondent was assigned to a geographic domain based on its area segment classification.

Step 3. The final population weight values for the three data sets were obtained from the NLAAS, NSAL, and NCS-R investigators. Since the final NCS-R and NSAL weights had been "centered" or "normalized" (mean weight=1.0), they were restored to the original U.S. population scaling based on weighted totals from the March 2002 demographic supplement of the Current Population Survey (CPS).

Step 4. Notation: each case in the CPES pooled data set was indexed as follows:

<b>Table 3. Subscript notation for weight integration expressions</b>		
<b>Index Subscript</b>	<b>Values</b>	<b>Representing</b>
i	1,...,n	Individual sample case subscript
j	1,2,3	Study index, 1=NCS-R; 2=NSAL, 3=NLAAS
k	1-11	Population index (Table 1), collapsing White, Other
l	1-11	Domain index (Table 2)

- Using CPES -

Step 5. From the pooled data set, EXCEL spreadsheets were used to compute the sums of nominal cases for each study by race/ancestry population by geographic domain cell. These counts were then aggregated across the three studies to produce CPES pooled case counts for each population x domain cell:

$$\begin{aligned} n_{+jkl} &= \sum_i I_{ijkl}; \\ n_{++kl} &= \sum_i \sum_j I_{ijkl} \end{aligned}$$

Step 6. The March 2002 CPS data enabled estimation of post-stratification control totals for each race/ancestry group,  $k=1, \dots, 11$ ; however, it did not provide the geographic detail needed to allocate the population total to the  $l=1, \dots, 11$  geographic domains. For this purpose, the weighted population distribution from the CPES study with the most robust estimates of geographic distribution was used. NLAAS was chosen as the basis for allocating the Asian and Hispanic populations to the 11 sample geographic domains. NSAL weighted sample distributions were used to apportion the African-American and Afro-Caribbean populations to the geographic domains. White and Other population totals were allocated to geographic domains based on the empirical distribution of weights in the NCS-R.

$$\hat{N}_{kl} = \sum_{i=1}^{n_{j^*kl}} W_{ij^*kl} \cdot \frac{\hat{N}_{k,CPS}}{\sum_{l=1}^{12} \sum_{i=1}^{n_{j^*kl}} W_{ij^*kl}}$$

where:

$\hat{N}_{kl}$  = the CPES control total for race/ethnicity population  $k$  and domain  $l$ ,

$W_{ij^*kl}$  = the original study-specific weight for case  $i$ , study,

$j^*$  = the CPES study chosen to estimate the domain allocation for population  $k$ , and

$\hat{N}_{k,CPS}$  = the March 2002 CPS population estimate for race/ethnicity category  $k$ .

Table 4 provides the final population controls for the race/ethnicity x domains cells of the CPES weight computation array.

Step 7: The original population weights from each study were post-stratified to the common race/ethnicity x domain population control totals derived from the March 2002 CPS (see Step 6 and Table 4).

- Using CPES -

$$W_{ijkl}^* = W_{ijkl} \cdot \frac{\hat{N}_{kl}}{\sum_{i=1}^{n_{jkl}} \sum_{l=1}^{12} \sum_{k=1}^{11} W_{ijkl}}$$

where:

$W_{ijkl}^*$  = the study specific weight adjusted to 2002 CPS population totals,

$W_{ijkl}$  = the original study-specific population weight for case i, study j,

$\hat{N}_{kl,CPS}$  = the 2002 CPS estimate for race/ethnicity population k allocated to domain l.

Since the original study-specific weights for the major populations of interest had already included some form of population-based control, this rescaling to a common post-stratification standard did not require major adjustments.

Step 8. Since in Step 7 the individual study-specific weights were controlled to exact counts for each race/ancestry x geographic domain cell, the remaining step involved rescaling the study-specific weights to reflect the proportion of nominal cases that each study contributed to the cell in the pooled data set.

$$W_{CPES,i} = W_{ijkl}^* \cdot \frac{n_{+jkl}}{n_{++kl}}$$

where:

$W_{CPES,i}$  = the CPES population weight for case i;

$W_{ijkl}^*$  = the standard population weight for case i, study j, (assigned to population k and domain l).

Conditional on the assigned population (k) and domain classification (l), this rescaling provided a "proportionate to sample size" contribution from each study. It linearly rescaled the weights for each individual study. It did not alter the distribution of the study-specific population weights except to reduce the study specific mean by  $n_{+jkl}/n_{++kl}$  and the variance of the study weights (not relvariance) by a factor of  $(n_{+jkl}/n_{++kl})^2$ .

**Table 4: Standardized Population Control Totals for CPES Weights Based on March 2002 Current Population Survey (Part 1 of 2)**

Race/Ancestry Population Group	Sample Frame Geographic Domain					
	CUBAN >5%	VIET <5%	FILIP >5%	PUERTO RICAN >5%	CHINESE >5%	AFRO-CARIB (see text)
VIETNAMESE	2480	383349	38422	41222	55101	0
FILIPINO	9950	94392	289714	10579	25166	0
CHINESE	48732	54578	240247	132608	541553	0
OTHER ASIAN	35654	19422	221116	6934	247125	0
CUBAN	103587	3643	5000	10334	5000	6447
PUERTO RICAN	48478	19246	39505	1009335	19457	78449
MEXICAN	3750	287255	70701	509666	69584	0
OTHER HISPANIC	202542	110127	24934	1109915	164478	111795
AFRO-CARIBBEAN	3687	2500	1250	126246	0	577382
AFRIC-AMERICAN	102567	68475	242531	1571404	27500	62549
WHITE AND OTHER	1086796	739518	1210972	2026304	2191741	1250
Total	1648223	1782505	2384392	6554547	3346705	837872

**Table 4: Standardized Population Control Totals for CPES Weights Based on March 2002 Current Population Survey (Part 2 of 2)**

Race/Ancestry Population Group	Sample Frame Geographic Domain					
	AFRI-AMER 60-100%	AFRI-AMER 30-59.9%	AFRI-AMER 10-29.9%	AFRI-AMER 0-9.9%	HAWAII	Total
VIETNAMESE	24555	4348	179945	437448	3403	1170273
FILIPINO	11442	27032	389766	674184	421617	1953842
CHINESE	10720	50153	255197	1039660	223468	2596916
OTHER ASIAN	44172	53381	592307	1749290	360977	3330378
CUBAN	16347	11412	149585	805628	0	1116983
PUERTO RICAN	68727	3229	374734	536439	37466	2235065
MEXICAN	282346	67944	2942722	11529509	0	15763477
OTHER HISPANIC	165520	203503	759917	2592100	2945	5447776
AFRO-CARIBBEAN	288119	220638	198343	12119	0	1430284
AFRIC-AMERICAN	10596921	5110387	3545866	721516	0	22049716
WHITE AND OTHER	1943090	6258727	18867372	118404822	0	152730592
Total	13451959	12010754	28255754	138502715	1049876	209825302

## VI. Final Weights and Special Analysis Consideration for Weighted Analysis

### VI.A Final Weights

As noted in Section V above, n=9282 NCS-R survey respondents completed Part 1 of the two-part CIDI-based interview; however, only a subsample of n=5692 NCS-R respondents went on to complete the more in-depth Part 2 questionnaire modules. All n=6082 NSAL and n=4649 NLAAS respondents completed the full interview schedule--the equivalent of NCS-R Parts 1 and 2. To account for the split schedule of NCS-R questionnaire administration, two CPES pooled analysis weights were computed using the calculation sequence described in steps (5) through (8) in Section V above. The first weight, termed the "Part 1" weight, is labeled CPESWGTS in the merged CPES data set. It is the population weight that should be used for analysis involving variables that are included in Part 1 of the NCS-R. The second weight, termed the "Part 2" weight, is stored in the merged data set as the variable, CPESWGTL. It is the population weight that should be used when the CPES analysis includes variables that NCS-R only asked of the Part 2 subsample of respondents.

Table 5 provides a summary of selected distributional statistics for the final CPES Part 1 and Part 2 analysis weights.

<b>Table 5: Distributions of CPES Final Part1 and Part 2 Analysis Weights</b>								
<b>Race/EthnicityPopulationGroup</b>	<b>CPES Final Analysis Weight</b>							
	<b>CPESWGTPS</b>				<b>CPESWGTL</b>			
	<b>n</b>	<b>Mean</b>	<b>Minimum</b>	<b>Maximum</b>	<b>n</b>	<b>Mean</b>	<b>Minimum</b>	<b>Maximum</b>
VIETNAMESE	527	2221	760	11898	526	2225	281	11898
FILIPINO	525	3722	726	15307	520	3757	726	15307
CHINESE	619	4195	754	15820	613	4236	934	15820
OTHER ASIAN	613	5433	587	20950	519	6417	732	27255
CUBAN	625	1787	267	20949	610	6416	731	27225
PUERTO RICAN	654	3417	128	10846	620	3604	75	16372
MEXICAN	1442	10931	781	49382	1214	12984	373	89022
OTHER HISPANIC	899	6060	533	44888	820	6644	591	40582
AFRO-CARIBBEAN	1492	959	111	20796	1476	969	162	21040
AFRIC-AMERICAN	4746	4646	728	18594	4249	5189	978	36257
WHITE AND OTHER	7871	19362	1250	131331	5256	28800	1250	195000
Total Sample	20013	10468	111	131331	16423	12693	75	195000

## VI.B Special Considerations in Weighted Analysis of the CPES Data

The weights have been designed to enable analysts to compute unbiased or nearly unbiased estimates of population statistics and relationships (e.g. bivariate associations, regression relationships) for the larger CPES survey population of U.S. residents. Contemporary statistical software systems such as SAS, Stata, SPSS, and SUDAAN all provide the capability to conduct weighted analysis of the CPES survey data. CPES data analysts are encouraged to consult the user guides and help support for their chosen software package to learn the syntax and program specific features for conducting weighted analysis. The following paragraphs provide guidance on weighted analysis that is specific to the CPES data set.

### VI.B. 1 Part 1 or Part 2 Weight ?

CPES analysts should consult the data documentation to determine if variables of interest in their analysis were obtained in Part 1 or Part 2 on the NCS-R interview. If the analysis includes only Part 1 variables, the CPESWGTS analysis weight should be used. It will include the full sample of NCS-R cases and provide greatest precision for sample estimates of population characteristics or relationships. If the analysis includes one or more variables that NCS-R collected only in Part 2, the appropriate weight for population estimation is CPESWGTL.

In the calculation of the Part 1 and Part 2 weights, the absolute contributions from NSAL and NLAAS to the pooled weight calculation remained unchanged--only NCS-R required changes to the nominal case counts and initial rescaling steps. However, due to the reduced NCS-R sample size for the Part 2 variables, the relative contributions of the NSAL and NLAAS to any give race/ethnicity x sample domain weighting cell did change. Therefore the final CPES Part 1 and Part 2 analysis weights (Step 8 above) differ for NSAL and NLAAS cases as they do for the NCS-R cases.

## **VI.B.2 Subsetting the CPES data by study**

Occasionally, analysts may choose to extract CPES data for only one or two of the three component data sets. The CPES analysis weights will support this type of analysis; however, analysts should recognize that the sum of weights for this special CPES subset may not sum to the population control for that population. For example, consider an analysis which only used Afro-Caribbean data from the NLAAS and NSAL. Since a small number of Afro-Caribbeans interviewed in the NCS-R would be excluded from this analysis, the sum of weights for the combined NSAL and NLAAS cases would no longer match the CPES population control total for the Afro-Caribbean race/ethnicity population. A principle of weighted analysis of data is that population estimates and sampling errors (except for estimates of totals) should be invariant to any linear scaling of the weights (multiplication or division by a constant). Under the procedures used to compute the CPES Part 1 and Part 2 weights, this assumption of linear scaling applies when the data for one or two studies are used independently or are compared.

## **VI.B.3 Subsetting the CPES data based on characteristics or respondents**

In general, CPES analysts can apply the analysis weights for subpopulation analysis (e.g., estimation for women of Mexican-American ancestry). Provided all qualifying cases in the CPES data are included in the subpopulation analysis, the estimates would be unbiased and the sum of the CPES weights would be an unbiased estimator of the 2002 population count for that subset of the larger U.S. population. Experience has shown that due to the sheer numbers of observations and richness of the variable set, data sets such as the CPES generate interest in rare populations or populations for which the original samples were not optimal (e.g. women of Mexican-American ancestry living in the West Census Region and covered by a regional health maintenance organization (HMO) program). CPES analysts who have concerns about the appropriateness of the CPES for subpopulation analysis they are proposing to conduct are encouraged to consult a survey statistician.

## **VI.B.4 Item Missing Data for Analysis Variables**

The original NCS-R, NSAL, and NLAAS analysis weights included adjustments for survey nonresponse. Through the process used to create the combined analysis weights, these adjustments for differential nonresponse are preserved in the CPES Part 1 and Part 2 weights. However, the CPES weights do not include adjustments for item missing data in the CPES data set. With a few special exceptions, most statistical software packages employ "case-wise" deletion as the means to address the problem of missing values for the variables. That is, any case with a missing value on one or more variables (e.g. fitting a multivariate logistic regression model) will cause the case to be dropped from the analysis. If the amount of such case-wise deletion is substantial, the unbiasedness of the weighted estimation may be compromised. Analysts are encouraged to use standard data checking techniques to establish the patterns of missing data in their analysis variables and assess the extent and impact of software-driven case-wise deletion on the integrity of their analysis. If the variables of interest have high rates of item missing data, analysts may consider consulting a survey statistician to consider remediation approaches such as stochastic imputation (Little and Rubin, 2002).

## **VII. Procedures for Sampling Error Estimation in Design-based Analysis of the CPES Data.**

The CPES data set is the product of the merger of three probability samples of the U.S. population and therefore shares the primary stage sample stratification and clustering features of the component sample designs. The NCS-R, NSAL and NLAAS sample designs were very similar in their basic structure to the multi-stage designs used for major survey programs such as the U.S. Health Interview Survey (HIS), the National Survey of Family Growth (NSFG) or the other national scientific surveys. The survey literature refers to these samples as complex designs, a loosely-used term meant to denote the fact that the sample incorporates special design features such as stratification, clustering and differential selection probabilities (i.e., weighting) that analysts must consider in computing sampling errors for sample estimates of descriptive statistics and model parameters. Standard programs in statistical analysis software packages assume simple random sampling (SRS) or independence of observations in computing standard errors for sample estimates. In general, the SRS assumption results in underestimation of variances of survey estimates of descriptive statistics and model parameters. Confidence intervals based on computed variances that assume independence of observations will be biased (generally too narrow) and design-based inferences will be affected accordingly. Likewise, test statistics ( $t$ ,  $X^2$ ,  $F$ ) computed in complex survey data analysis using standard programs will tend to be biased upward and overstate the significance of tests of effects.



This section focuses on sampling error estimation and construction of confidence intervals for survey estimates of descriptive statistics such as means, proportions, ratios, and coefficients for linear and logistic regression models.

## VII.A Sampling Error Computation Methods and Programs

Over the past 50 years, advances in survey sampling theory have guided the development of a number of methods for correctly estimating variances from complex sample data sets. Sampling error programs that implement these complex sample variance estimation methods are available to CPES data analysts. The two most common approaches (Rust, 1985) to the estimation of sampling error for complex sample data are through the use of a Taylor Series linearization of the estimator (and corresponding approximation to its variance) or through the use of resampling variance estimation procedures such as Balanced Repeated Replication (BRR) or Jackknife Repeated Replication (JRR).

## VII.B Taylor Series linearization method:

When survey data are collected using a complex sample design with unequal size clusters, most statistics of interest will not be simple linear functions of the observed data. The linearization approach applies Taylor's method to derive an approximate form of the estimator that is linear in statistics for which variances and covariances can be directly and easily estimated. Stata Release 8 and 9, SAS V8.2/V9.0, SUDAAN Version 9, and the most recent releases of SPSS are commercially available statistical software packages that include procedures that apply the Taylor Series method to sampling error estimation and inference for complex sample data.

Stata (StataCorp, 2005) is a more recent commercial entry to the available software for analysis of complex sample survey data and has a growing body of research users. Stata includes special versions of its standard analysis routines that are designed for the analysis of complex sample survey data. Special survey analysis programs are available for descriptive estimation of means (SVY MEAN), ratios (SVY RATIO), proportions (SVY TAB), and population totals (SVY TOTAL). STATA programs for multivariate analysis of survey data include linear regression (SVY REGRESS), logistic regression (SVY LOGIT) and probit regression (SVY PROBT). STATA program offerings for survey data analysts are constantly being expanded. Information on the STATA analysis software system can be found on the Web at: <http://www.stata.com>.

Programs in SAS Version 9 (SAS, 2003; <http://www.sas.com/>) also use the Taylor Series method to estimate variances of means (PROC Surveymeans), proportions and cross-tabular analysis (PROC SurveyFreq), linear regression (PROC SurveyReg), and logistic regression (PROC SurveyLogistic).

SUDAAN (RTI, 2004) is a commercially available software system developed and marketed by the Research Triangle Institute of Research Triangle Park, North Carolina (USA). SUDAAN was developed as a stand-alone software system with capabilities for the more important methods for descriptive and multivariate analysis of survey data, including: estimation and inference for means, proportions, and rates (PROC DESCRIPT and PROC RATIO); contingency table analysis (PROC CROSSTAB); linear regression (PROC REGRESS); logistic regression (PROC LOGISTIC); log-linear models (PROC CATAN); and survival analysis (PROC SURVIVAL). SUDAAN V9.0 and earlier versions were designed to read directly from ASCII and SAS system data sets. The latest versions of SUDAAN permit procedures to be called directly from the SAS system. Information on SUDAAN is available at the following Web site address: <http://www.rti.org/>.

SPSS Version 14.0 (<http://www.spss.com/>) users can obtain the SPSS Complex Samples module which supports Taylor Series linearization estimation of sampling errors for descriptive statistics (CSDSCRIPTIVES), cross-tabulated data (CSTABULATE), general linear models (CSGLM), and logistic regression (CSLOGISTIC).

## VII.C Resampling methods:

BRR, JRR, and the bootstrap comprise a second class of nonparametric methods for conducting estimation and inference from complex sample data. As suggested by the generic label for this class of methods, BRR, JRR, and the bootstrap utilize replicated subsampling of the sample database to develop sampling variance estimates for linear and nonlinear statistics. WesVar PC (Westat, Inc., 2000) is a software system for personal computers that employs replicated variance estimation methods to conduct the more common types of statistical analysis of complex sample survey data. WesVar PC was developed by Westat, Inc. and is distributed along with documentation to researchers at Westat's Web site: <http://www.westat.com/wesvarpc/>. WesVar PC includes a Windows-based application generator that enables the analyst

to select the form of data input (SAS data file, SPSS for Windows data base, ASCII data set) and the computation method (BRR or JRR methods). Analysis programs contained in WesVar PC provide the capability for basic descriptive (means, proportions, totals, cross tabulations) and regression (linear, logistic) analysis of complex sample survey data. WesVar also provides the best facility for estimating quantiles of continuous variables (e.g., 95%-tile of a cognitive test score) from survey data. WesVar Complex Samples 4.0 is the latest version of WesVar. Researchers who wish to analyze the CPES data using WesVar PC should choose the BRR or JRR (JK2) replication option.

STATA V9 has introduced the option to use JRR or BRR calculation methods as an alternative to the Taylor Series method for all of its svy command options. SUDAAN V9.0 also allows the analysts to select the JRR method for computing sampling variances of survey estimates.

IVEWare is another software option for the JRR estimation of sampling errors for survey statistics. IVEWare has been developed by the Survey Methodology Program of the Survey Research Center and is available free of charge to users at: <http://www.isr.umich.edu/src/smp/ive/>. IVEWare is based on SAS Macros and requires SAS Version 6.12 or higher. The system includes programs for multiple imputation of item missing data as well as programs for variance estimation in descriptive (means, proportions) and multivariate (regression, logistic regression, survival analysis) analysis of complex sample survey data.

These new and updated software packages include an expanded set of user-friendly, well-documented analysis procedures. Difficulties with sample design specification, data preparation, and data input in the earlier generations of survey analysis software created a barrier to use by analysts who were not survey design specialists. The new software enables the user to input data and output results in a variety of common formats, and the latest versions accommodate direct input of data files from the major analysis software systems.

## VII.D Sampling Error Computation Models

Regardless of whether the linearization method or a resampling approach is used, estimation of variances for complex sample survey estimates requires the specification of a sampling error computation model. CPES data analysts who are interested in performing sampling error computations should be aware that the estimation programs identified in the preceding section assume a specific sampling error computation model and will require special sampling error codes. Individual records in the analysis data set must be assigned sampling error codes that identify to the programs the complex structure of the sample (stratification, clustering) and are compatible with the computation algorithms of the various programs. To facilitate the computation of sampling error for statistics based on CPES data, design-specific sampling error codes will be routinely included in all versions of the data set. Although minor recoding may be required to conform to the input requirements of the individual programs, the sampling error codes that are provided should enable analysts to conduct either Taylor Series or Replicated estimation of sampling errors for survey statistics. In programs that use the Taylor Series Linearization method, the sampling error codes (SESTRAT and SECLUSTER) will typically be input as keyword statements (SAS V9.1, SUDAAN V9.0) or as global settings (Stata V9) along with the analysis weight and will be used directly in the computational algorithms. Programs that permit BRR or JRR computations will require the user supplied sampling error codes to construct "replicate weights" that are required for these approaches to variance estimation.

Two sampling error code variables are defined for each case based on the sample design stratum and primary stage unit (PSU) cluster in which the sample respondent resided: Sampling Error Stratum Code (SESTRAT) and Sampling Error Cluster Code (SECLUSTER). The CPES SESTRAT codes were derived directly from a concatenation of the existing sampling error stratum codes for the NCS-R, NSAL and NLAAS sample designs. A total of 180 sampling error strata were defined. These were allocated to the individual contributing samples according to the coding scheme shown in Table 6.

Table 6. CPES Sampling Error Strata	
CPES Component Sample	CPES Sampling Error Strata
NCS-R	1-42
NSAL	43-111
NLAAS	112-180

All original sampling error strata definitions for the NCS-R and NLAAS were preserved unchanged in the mapping to the CPES sampling error stratum code. In general, the assignment of NSAL cases to CPES sampling error strata also followed

the original NSAL coding. The single exception involved a NSAL sampling error stratum that included multiple clusters. This stratum was divided into several pseudo-strata each with a pair of combined clusters. This minor change enables CPES analysts to use any of the sampling error calculation methods (Taylor, BRR or JRR) without having to perform additional recoding of the sampling error variables.

Likewise, with one exception, the values of SECLUSTR for CPES sampling error strata are identical to those in the original NCS-R, NSAL and NLAAS data sets. The exception was the cluster numbering for the one NSAL sampling error stratum with multiple clusters. Clusters in this stratum were randomly grouped into pairs and assigned to pseudo-strata as described in the preceding paragraph. The result is that the CPES SECLUSTR code takes a value of either 1 or 2 and exactly two sampling error clusters are assigned to each sampling error stratum.

## VII.E Syntax for CPES Design-based Variance Estimation Using STATA and SAS

The following two sections provide a short overview of the general syntax and command file structure for computing sampling errors using STATA and SAS programs that have been designed for the analysis of complex sample survey data. Analysts are referred to the user guides and the on-line help facilities of these two software systems for documentation of the individual programs.

### VII.E.1 Stata command syntax

As described above, CPES data analysts who are familiar with the STATA software system can utilize STATA's "svy" commands for the analysis of complex sample survey data. STATA Version 9 syntax for some of the more commonly used analysis programs is illustrated below (shown for the Part 2 weight option) :

```
.svyset seclust [pweight=cpeswgtl], strata(sestrat)
```

This statement defines the sample design variables for the duration of the analysis session. SVY commands issued after this statement will automatically incorporate these design specifications.

To conduct analyses, the following STATA commands and syntax are used (please refer to STATA V9 Reference Manual for specific command syntax and output options):

```
.svy, vce(linearized): mean vars
```

[estimates, standard errors, design effects for means]

```
.svy, vce(linearized): tab v1 v2
```

[estimates, standard errors for proportions of single variable categories, or crosstabulations of two variables with tests of independence]

```
.svy, vce(linearized): regress dep x1 ...
```

[simple linear regression model for a continuous dependent variable]

```
.svy, vce(linearized): logit dep x1...
```

[simple logistic regression model for a binary dependent variable]

To estimate the single statistics or regression models for subpopulations of the survey population in STATA, the following optional syntax is used (illustrated for svytab):

```
.svy, vce(linearized): tab v1 v2, over(var)
```

where var is a categorical variable that defines the subpopulations for which separate estimates are desired (e.g. gender).

## VII.E.2 SAS Version 9 Command Syntax

SAS Version 9 includes four programs for the analysis of complex sample survey data: PROC Surveymeans, SurveyFreq, SurveyReg and SurveyLogistic. The general syntax for specifying the CPES design structure in the SAS system is as follows:

```
PROC SurveyXXXX data=libname.filename;  
  
STRATUM SESTRAT;  
  
CLUSTER SECLUST;  
  
WEIGHT CPESWGTL;  
  
program specific statements here;  
  
RUN;
```

Users are referred to the SAS/STAT(R) 9.1 User's Guide (SAS, 2004) for documentation on program specific statements, keywords and options

## VIII. Weights for Study pairs

Final weights were also developed for analyzing pairs of the CPES studies (NCS-R and NSAL, NCS-R and NLAAS, and NSAL and NLAAS) and for Part I and Part II sub-samples (only for pairs that include NCS-R respondents). This will allow for generating population estimates by analyzing data from study pairs only. Table 7 below summarizes the paired weights and key descriptive statistics for the weight distributions.

Table 7: Descriptive Statistics for Paired Study Weights						
Study Pair	Sample	Variable Name	SampleSize	MeanWeight	StandardDeviationof Weights	Sum ofWeights
NCS-R and NSAL	Short	NCNSWTSH	15364	13588.6116	13251.1943	208775428
NCS-R and NSAL	Long	NCNSWTLG	11774	17731.9026	26960.0142	208775421
NCS-R and NLAAS	Short	NCNLWTSH	13931	15061.7548	11814.4696	209825306
NCS-R and NLAAS	Long	NCNLWTLG	10341	20290.6198	26553.902	209825299
NSAL and NLAAS	---a	NSNLWT	10731	19553.1922	63077.1289	209825306

a. Short and long sub-samples don't not apply as NCS-R sample is not included.

Weights for analysis of CPES study pairs are based on the final CPES 3-study weights and were developed according to the following steps:

Step 1. Each NLAAS, NSAL and NCS-R case was assigned to a race/ancestry category based on the categories and priority order provided in Table 1 (see Section III).

Step 2. Each NLAAS, NSAL and NCS-R area segment was assigned to a geographic domain based on the definitions and priority order shown in Table 2 (see Section IV). Each NLAAS, NSAL and NCS-R respondent was assigned to a geographic domain based on its area segment classification.

Step 3. For each pair of studies, race x domain cell counts were obtained. Due to the lack of over-sampling of certain race groups in specific pairs (Asians in NCS-R and NSAL) some of the CPES race x domain cells had no cases or a small number of cases. Such small cell counts could affect the robustness of post-adjustments and are usually dealt with by collapsing. Collapsing was mainly done over similar race groups (e.g., Vietnamese, Filipino, Chinese and other Asian groups) within a

domain. Similar domains were then collapsed (examples of collapsed domains include Census Block Group > 5% Cuban and include Census Block Group > 5% Puerto Rican) if the cell count was still small (mainly <10) after race group collapsing. The same collapsed groups were used for the long and the short sub samples (whenever applicable). Collapsed groups for each pair of studies are shown in Tables 8-12.

Step 4. CPS 2002 totals were calculated for each collapsed group. Weighted counts using the final CPES weight (short form when dealing with short sample and long form when dealing with long sub-sample) were also generated for each collapsed group.

Step 5. A post-stratification adjustment factor (CPS 2002 total divided by the weighted count using CPES final weight) was calculated and applied to the final CPES weight to generate the paired weights. Respondents' weights for cases belonging to the same race x domain received the same factor.

<b>Table 8: NCS-R and NSAL collapsed cells for the short sample, post-stratification adjustment factor and mean weights</b>					
<b>Collapsed groups for NCSR &amp; NSAL short</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPES weights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnlwtsh)</b>
Asians over all domains	8041944	189	959391	8.382342549	5076.142857
Cuban, PR and other Hisp over all domains	8759413	492	2057662	4.256973692	4182.239837
Mexicans over all domains	15763477	574	6051909	2.604711505	10543.39547
Africarib and AA in Cuban blocks	106254	35	106253	1.000009411	3035.8
Whites in Cuban blocks	1086796	26	611832	1.776298069	23532
Afrocarib and AA in Vit blocks	70975	30	70976	0.999985911	2365.866667
Whites in in Vit blocks	739518	37	691510	1.069424882	18689.45946
Afrocarib and AA in Flip blocks	243781	55	243780	1.000004102	4432.363636
Whites in Flip blocks	1210972	65	944809	1.281710907	14535.52308
Afrocarib in PR blocks	126246	57	126245	1.000007921	2214.824561
AA in PR blocks	1571404	230	1571412	0.999994909	6832.226087
Whites in PR blocks	2026304	130	2014461	1.005878992	15495.85385
Afrocarib and AA in Chinese blocks	27500	22	27500	1	1250
Whites in Chinese blocks	2191741	123	2666727	0.82188428	21680.70732
Afrocarib in Afrocarib blocks	577382	933	577386	0.999993072	618.8488746
AA in Afrocarib blocks	62549	12	62549	1	5212.416667
Whites in Afrocarib and AA 60-100	1944340	267	1944339	1.000000514	7282.168539
Afrocarib in AA 60-100	288119	256	288116	1.000010412	1125.453125
AA in AA 60-100	10596921	2343	10596905	1.00000151	4522.793427

**Table 8: NCS-R and NSAL collapsed cells for the short sample, post-stratification adjustment factor and mean weights**

<b>Collapsed groups for NCSR &amp; NSAL short</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnlwtsh)</b>
Afrocarib in AA 30-59.9	220638	172	220642	0.999981871	1282.802326
AA in AA 30-59.9	5110387	1122	5110400	0.999997456	4554.723708
Whites in AA 30-59.9	6258727	638	6258683	1.00000703	9809.847962
Afrocarib in AA 10-29.9	198343	48	198344	0.999994958	4132.166667
AA in AA 10-29.9	3545866	703	3545843	1.000006486	5043.8734
Whites in AA 10-29.9	18867372	1216	18867235	1.000007261	15515.81826
Afrocarib and AA in AA 0-9.9	733635	220	733649	0.999980917	3334.768182
Whites in AA 0-9.9	118404822	5369	118405824	0.999991538	22053.60849

**Table 9: NCS-R and NSAL collapsed cells for the long sample, post-stratification adjustment factor and mean weights**

<b>Collapsed groups for NCSR &amp; NSAL long</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnswtlg)</b>
Asians over all domains	8041944	83	461432	17.42823211	5559.421687
Cuban, PR and other Hisp over all domains	8759413	364	1524027	5.747544499	4186.887363
Mexican over all domains	15763477	346	4312401	3.655382929	12463.58671
Afrocarib and AA in Cuban blocks	106254	25	106254	1	4250.16
Whites in Cuban blocks	1086796	18	1055336	1.029810411	58629.77778
Afrocarib and AA in Vit blocks	70975	21	70975	1	3379.761905
Whites in Vit blocks	739518	23	739527	0.99998783	32153.34783
Afrocarib and AA in Flip blocks	243781	47	243782	0.999995898	5186.851064
Whites in Flip blocks	1210972	48	1210979	0.99999422	25228.72917
Afrocarib in PR blocks	126246	55	126245	1.000007921	2295.363636
AA in PR blocks	1571404	202	1571417	0.999991727	7779.292079
Whites in PR blocks	2026304	90	2026314	0.999995065	22514.6
Afrocarib and AA in Chinese blocks	27500	14	27500	1	1964.285714
Whites in Chinese blocks	2191741	71	2191747	0.999997262	30869.67606
Afrocarib in Afrocarib blocks	577382	933	577386	0.999993072	618.8488746

**Table 9: NCS-R and NSAL collapsed cells for the long sample, post-stratification adjustment factor and mean weights**

<b>Collapsed groups for NCSR &amp; NSAL long</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnswtlg)</b>
AA in Afrocarib blocks	62549	12	62549	1	5212.416667
Whites in Afrocarib and AA 60-100	1944340	221	1909508	1.018241348	8640.307692
Afrocarib in AA 60-100	288119	253	288116	1.000010412	1138.798419
AA in AA 60-100	10596921	2137	10596881	1.000003775	4958.765091
Afrocarib in AA 30-59.9	220638	170	220642	0.999981871	1297.894118
AA in AA 30-59.9	5110387	1025	5110397	0.999998043	4985.753171
Whites in AA 30-59.9	6258727	509	5852364	1.069435702	11497.76817
Afrocarib in AA 10-29.9	198343	46	198345	0.999989917	4311.847826
AA in AA 10-29.9	3545866	621	3545850	1.000004512	5709.903382
Whites in AA 10-29.9	18867372	891	20300571	0.92940105	22784.0303
Afrocarib and AA in AA 0-9.9	733635	164	733636	0.999998637	4473.390244
Whites in AA 0-9.9	118404822	3385	116088323	1.019954625	34294.92555

**Table 10: NCS-R and NLAAS collapsed cells for the short sample, post-stratification adjustment factor and mean weights**

<b>Collapsed groups for NCSR &amp; NLAAS short</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnlwtsh)</b>
Asians in Cuban blocks	96816	18	96803	1.000134293	5377.944444
Cubans in Cuban blocks	103587	70	103588	0.999990346	1479.828571
PR in Cuban blocks	48478	26	48478	1	1864.538462
Mexican and other His in Cuban blocks	206292	46	201688	1.022827337	4384.521739
Afrocarib and AA in Cuban blocks	106254	19	53552	1.984127577	2818.526316
Whites in Cuban blocks	1086796	25	588300	1.847349992	23532
Vit in Vit blocks	383349	208	383349	1	1843.024038
Flip in Vit blocks	94392	19	94392	1	4968
Chinese and Other Asians in Vit blocks	74000	18	74000	1	4111.111111
Cuban, PR and Mexican in Vit blocks	310144	35	310139	1.000016122	8861.114286
Other Hispanics in Vit blocks	110127	13	110127	1	8471.307692

<b>Table 10: NCS-R and NLAAS collapsed cells for the short sample, post-stratification adjustment factor and mean weights</b>					
<b>Collapsed groups for NCSR &amp; NLAAS short</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnlwtsh)</b>
Afrocarib and AA in Vit blocks	70975	21	48966	1.449475146	2331.714286
Whites in Vit blocks	739518	37	691510	1.069424882	18689.45946
Vit in Flip Blocks	38422	17	38422	1	2260.117647
Flip in Flip Blocks	289714	83	289714	1	3490.53012
Chinese in Flip Blocks	240247	68	240247	1	3533.044118
other Asians in Flip blocks	221116	39	221116	1	5669.641026
Cuban and PR in Flip blocks	44505	13	44505	1	3423.461538
Mexican and other Hisp in Flip block	95635	13	95634	1.000010457	7356.461538
Africarib and AA in Flip blocks	243781	25	110449	2.207181595	4417.96
Whites in Flip blocks	1210972	51	741314	1.633547997	14535.56863
Vit and Flip in PR blocks	51801	14	51801	1	3700.071429
Chinese and other Asians in PR blocks	139542	22	139541	1.000007166	6342.772727
Cuban and PR in PR blocks	1019669	264	980425	1.040027539	3713.731061
Mexican in PR blocks	509666	44	509661	1.00000981	11583.20455
other Hisp in PR blocks	1109915	215	1074917	1.032558793	4999.613953
Afrocarib and AA in PR blocks	1697650	71	420440	4.03779374	5921.690141
Whites in PR block	2026304	110	1704545	1.18876533	15495.86364

<b>Table 10: NCS-R and NLAAS collapsed cells for the short sample, post-stratification adjustment factor and mean weights (Continued)</b>					
<b>Collapsed groups for NCSR &amp; NLAAS short</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnlwtsh)</b>
Vit in Chinese blocks	55101	19	55101	1	2900.052632
Flip in Chinese blocks	25166	10	25165	1.000039738	2516.5
Chinese in Chinese blocks	541553	119	541553	1	4550.865546
other Asians in Chinese block	247125	49	247126	0.999995953	5043.387755
Cuban, PR and Mexican in Chinese blocks	94041	23	94041	1	4088.73913



<b>Table 10: NCS-R and NLAAS collapsed cells for the short sample, post-stratification adjustment factor and mean weights (Continued)</b>					
<b>Collapsed groups for NCSR &amp; NLAAS short</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPES weights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnlwtsh)</b>
Other Hispanics in Chinese blocks	164478	24	164479	0.99999392	6853.291667
Afrocarib and AA in Chinese blocks	27500	22	27500	1	1250
Whites in Chinese blocks	2191741	123	2666727	0.82188428	21680.70732
Asians in Africocarib and AA 60-100 blocks	90889	16	90889	1	5680.5625
Cuban and PR in Africocarib and AA 60-100	169970	28	61807	2.750012135	2207.392857
Mexican in Africocarib and AA 60-100	282346	27	282351	0.999982292	10457.44444
Other His in Africocarib and AA 60-100	277315	19	101448	2.733567936	5339.368421
Whites in Africocarib and AA 60-100	1944340	91	664740	2.924963143	7304.835165
Asians in AA 30-59.9	134914	24	134913	1.000007412	5621.375
Cubans, PR, and Mexicans in AA 30-59.9	82585	31	77710	1.062733239	2506.774194
Other His in AA 30-59.9	203503	30	152627	1.333335517	5087.566667
Whites in AA 30-59.9	6258727	302	2962549	2.112615521	9809.764901
Vit in AA 10-29.9	179945	71	179945	1	2534.43662
Flip in AA 10-29.9	389766	93	389766	1	4191.032258
Chinese in AA 10-29.9	255197	62	255197	1	4116.080645
Other Asian in AA 10-29.9	592307	88	592305	1.000003377	6730.738636
Cubans in AA 10-29.9	149585	81	147761	1.012344259	1824.209877
PR in AA 10-29.9	374734	94	366927	1.021276712	3903.478723
Mexicans in AA 10-29.9	2942722	216	2942725	0.999998981	13623.72685
Other Hisp in AA 10-29.9	759917	99	737569	1.030299538	7450.191919
Whites in AA 10-29.9	18867372	925	14352204	1.314597535	15515.89622
Vit in AA 0-9.9	437448	188	437448	1	2326.851064
Flip in AA 0-9.9	674184	179	674183	1.000001483	3766.385475
Chinese in AA 0-9.9	1039660	259	1039660	1	4014.131274
other asian in AA 0-9.9	1749290	320	1749288	1.000001143	5466.525

**Table 10: NCS-R and NLAAS collapsed cells for the short sample, post-stratification adjustment factor and mean weights (Continued)**

<b>Collapsed groups for NCSR &amp; NLAAS short</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnlwtsh)</b>
Cubans in AA 0-9.9	805628	435	805629	0.999998759	1852.02069
PR in AA 0-9.9	536439	156	536438	1.000001864	3438.705128
Mexican in AA 0-9.9	11529509	1083	11529523	0.999998786	10645.91228
Other His in AA 0-9.9	2592100	340	2576935	1.005884898	7579.220588
Africocarib and AA in Africocarib and AA blocks	21333840	1072	4726615	4.513555684	4409.155784
Whites AA 0-9.9	118404822	5316	117237005	1.009961164	22053.61268
Vit in Hawaiian blocks	3403	2	3403	1	1701.5
Flip in Hawaiian blocks	421617	126	421617	1	3346.166667
Chinese in Hawaiian blocks	223468	69	223469	0.999995525	3238.681159
other Asian in Hawaiian blocks	360977	84	360977	1	4297.345238
Cuban in Hawaiian blocks	0	0	0	0	0
PR in Hawaiian blocks	37466	11	37466	1	3406
Mexican in Hawaiian blocks	0	0	0	0	0
Other Hisp in Hawaiian blocks	2945	1	2945	1	2945
Africocarib in Hawaiian blocks	0	0	0	0	0
AA in Hawaiian blocks	0	0	0	0	0
Whites in Hawaiian blocks	0	0	0	0	0

**Table 11: NCS-R and NLAAS collapsed cells for the long sample, post-stratification adjustment factor and mean weights**

<b>Collapsed groups for NCSR &amp; NLAAS long</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnlwtlg)</b>
Asians in Cuban blocks	96816	14	96802	1.000144625	6914.428571
Cubans in Cuban blocks	103587	63	103587	1	1644.238095
PR in Cuban blocks	48478	23	48478	1	2107.73913
Mexican and other His in Cuban blocks	206292	42	201229	1.025160389	4791.166667
Afrocarib and AA in Cuban blocks	106254	9	37583	2.827182503	4175.888889
Whites in Cuban blocks	1086796	17	994958	1.092303394	58526.94118

<b>Table 11: NCS-R and NLAAS collapsed cells for the long sample, post-stratification adjustment factor and mean weights</b>					
<b>Collapsed groups for NCSR &amp; NLAAS long</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPES weights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnlwtlg)</b>
Vit in Vit blocks	383349	208	383349	1	1843.024038
Flip in Vit blocks	94392	19	94392	1	4968
Chinese and Other Asians in Vit blocks	74000	13	74000	1	5692.307692
Cuban, PR and Mexican in Vit blocks	310144	30	310142	1.000006449	10338.06667
Other Hispanics in Vit blocks	110127	10	110127	1	11012.7
Afrocarib and AA in Vit blocks	70975	12	38540	1.84159315	3211.666667
Whites in Vit blocks	739518	23	739527	0.99998783	32153.34783
Vit in Flip Blocks	38422	17	38422	1	2260.117647
Flip in Flip Blocks	289714	83	289714	1	3490.53012
Chinese in Flip Blocks	240247	68	240247	1	3533.044118
Other Asians in Flip blocks	221116	34	221117	0.999995478	6503.441176
Cuban and PR in Flip blocks	44505	12	44505	1	3708.75
Mexican and other Hisp in Flip block	95635	10	95634	1.000010457	9563.4
Africarib and AA in Flip blocks	243781	17	86858	2.806661447	5109.294118
Whites in Flip blocks	1210972	34	857780	1.411751265	25228.82353
Vit and Flip in PR blocks	51801	14	51801	1	3700.071429
Chinese and other Asians in PR blocks	139542	20	139542	1	6977.1
Cuban and PR in PR blocks	1019669	253	978632	1.041933025	3868.110672
Mexican in PR blocks	509666	41	509665	1.000001962	12430.85366
other Hisp in PR blocks	1109915	209	1073945	1.033493335	5138.492823
Afrocarib and AA in PR blocks	1697650	41	253143	6.70628854	6174.219512
Whites in PR block	2026304	70	1576027	1.285703862	22514.67143
Vit in Chinese blocks	55101	19	55101	1	2900.052632
Flip in Chinese blocks	25166	10	25165	1.000039738	2516.5

<b>Table 11: NCS-R and NLAAS Collapsed Cells for the long sample, post-stratification adjustment factor and mean weights (Continued)</b>					
<b>Collapsed groups for NCSR &amp; NLAAS long</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPES weights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(ncnlwtlg)</b>
Chinese in Chinese blocks	541553	118	541553	1	4589.432203
Other Asians in Chinese block	247125	40	247124	1.000004047	6178.1
Cuban, PR and Mexican in Chinese blocks	94041	20	89041	1.056153907	4452.05
Other Hispanics in Chinese blocks	164478	21	164478	1	7832.285714
Afrocarib and AA in Chinese blocks	27500	14	27500	1	1964.285714
Whites in Chinese blocks	2191741	71	2191747	0.999997262	30869.67606
Asians in Africocarib and AA 60-100 blocks	90889	15	90889	1	6059.266667
Cuban and PR in Africocarib and AA 60-100	169970	25	59887	2.83817857	2395.48
Mexican in Africocarib and AA 60-100	282346	24	282346	1	11764.41667
Other His in Africocarib and AA 60-100	277315	16	94583	2.931975091	5911.4375
Africocarib and AA in Africocarib and AA blocks	21333840	624	3091690	6.900381345	4954.63141
Whites in Africocarib and AA 60-100	1944340	45	390324	4.981348828	8673.866667
Asians in AA 30-59.9	134914	18	134914	1	7495.222222
Cuban, PR and Mexicans in AA 30-59.9	82585	22	77172	1.07014202	3507.818182
other Hisp in AA 30-59.9	203503	24	143650	1.416658545	5985.416667
Whites in AA 30-59.9	6258727	173	1989094	3.146521482	11497.65318
Vit in AA 10-29.9	179945	70	179945	1	2570.642857
Flip in AA 10-29.9	389766	91	389766	1	4283.142857
Chinese in AA 10-29.9	255197	60	255197	1	4253.283333
Other Asian in AA 10-29.9	592307	81	592307	1	7312.432099
Cubans in AA 10-29.9	149585	80	147739	1.012495008	1846.7375
PR in AA 10-29.9	374734	92	366761	1.021738953	3986.532609
Mexicans in AA 10-29.9	2942722	194	2942717	1.000001699	15168.64433

**Table 11: NCS-R and NLAAS Collapsed Cells for the long sample, post-stratification adjustment factor and mean weights (Continued)**

<b>Collapsed groups for NCSR &amp; NLAAS long</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratificationadjustment factor</b>	<b>Mean weight(ncnlwtlg)</b>
Other Hisp in AA 10-29.9	759917	90	735404	1.033332699	8171.155556
Whites in AA 10-29.9	18867372	600	13670503	1.380151996	22784.17167
Vit in AA 0-9.9	437448	188	437448	1	2326.851064
Flip in AA 0-9.9	674184	176	674183	1.000001483	3830.585227
Chinese in AA 0-9.9	1039660	256	1039660	1	4061.171875
Other Asian in AA 0-9.9	1749290	265	1749286	1.000002287	6601.079245
Cubans in AA 0-9.9	805628	432	805629	0.999998759	1864.881944
PR in AA 0-9.9	536439	142	536438	1.000001864	3777.732394

**Table 11: NCS-R and NLAAS collapsed cells for the long sample, post-stratification adjustment factor and mean weights (Continued)**

<b>Collapsed groups for NCSR &amp; NLAAS long</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratificationadjustment factor</b>	<b>Mean weight(ncnlwtlg)</b>
Mexican in AA 0-9.9	11529509	896	11529511	0.999999827	12867.75781
Other His in AA 0-9.9	2592100	298	2574818	1.006711931	8640.328859
Whites AA 0-9.9	118404822	3332	114255079	1.036319987	34290.2398
Vit in Hawaiian blocks	3403	2	3403	1	1701.5
Flip in Hawaiian blocks	421617	126	421617	1	3346.166667
Chinese in Hawaiian blocks	223468	69	223469	0.999995525	3238.681159
other Asian in Hawaiian blocks	360977	84	360977	1	4297.345238
Cuban in Hawaiian blocks	0	0	0	0	0
PR in Hawaiian blocks	37466	11	37466	1	3406
Mexican in Hawaiian blocks	0	0	0	0	0
Other Hisp in Hawaiian blocks	2945	1	2945	1	2945
Africocarib in Hawaiian blocks	0	0	0	0	0
AA in Hawaiian blocks	0	0	0	0	0
Whites in Hawaiian blocks	0	0	0	0	0

**Table 12: NSAL and NLAAS collapsed cells, post-stratification adjustment factor and mean weights**

<b>Collapsed groups for NSAL &amp; NLAAS</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratificationadjustment factor</b>	<b>Mean weight(nsnlwt)</b>
Asian in Cuban blocks	96816	14	76429	1.266744299	5459.214286
Cubans in Cuban blocks	103587	57	84349	1.228076207	1479.807018

<b>Table 12: NSAL and NLAAS collapsed cells, post-stratification adjustment factor and mean weights</b>					
<b>Collapsed groups for NSAL &amp; NLAAS</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(nsnlwt)</b>
PR in Cuban blocks	48478	18	33562	1.444431202	1864.555556
Mexican and other His in Cuban blocks	206292	33	151908	1.358006162	4603.272727
Afrocarib and AA in Cuban blocks	106254	16	52701	2.016166676	3293.8125
Whites in Cuban, PR and Asian blocks	7255331	35	536943	13.51229274	15341.22857
Vit in Vit blocks	383349	207	381506	1.004830855	1843.024155
Flip in Vit blocks	94392	19	94392	1	4968
Chinese and Other Asians in Vit blocks	74000	13	60127	1.230728292	4625.153846
Cuban, PR and Mexican in Vit blocks	310144	17	143960	2.154376216	8468.235294
Other Hispanics in Vit blocks	110127	10	84713	1.30000118	8471.3
Afrocarib and AA in Vit,, Flip, and Chinese blocks	342256	39	155341	2.203256062	3983.102564
Vit in Flip Blocks	38422	17	38422	1	2260.117647
Flip in Flip Blocks	289714	80	279242	1.037501522	3490.525
Chinese in Flip Blocks	240247	67	236714	1.014925184	3533.044776
other Asians in Flip blocks	221116	32	181429	1.218746727	5669.65625
Cuban, PR, Mexican and other Hisp in Flip blocks	140140	17	121287	1.155441226	7134.529412
Vit and Flip in PR blocks	51801	13	48275	1.073039876	3713.461538
Chinese and other Asians in PR blocks	139542	16	125475	1.112109982	7842.1875
Cuban and PR in PR blocks	1019669	249	926172	1.100949932	3719.566265
Mexican in PR blocks	509666	34	393833	1.29411705	11583.32353
other Hisp in PR blocks	1109915	210	1049920	1.057142449	4999.619048
Afrocarib in PR blocks	126246	43	95237	1.325598244	2214.813953
AA in PR blocks	1571404	173	1181981	1.329466379	6832.260116
Vit in Chinese blocks	55101	19	55101	1	2900.052632
Flip in Chinese blocks	25166	9	22649	1.111130734	2516.555556
Chinese in Chinese blocks	541553	117	532451	1.017094531	4550.863248
other Asians in Chinese block	247125	34	171474	1.441180587	5043.352941
Cuban, PR, Mexican, and other Hisp in Chinese blocks	258519	25	144368	1.790694614	5774.72

<b>Table 12: NSAL and NLAAS collapsed cells, post-stratification adjustment factor and mean weights (Continued)</b>					
<b>Collapsed groups for NSAL &amp; NLAAS</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPES weights</b>	<b>Post-stratification adjustment factor</b>	<b>Mean weight(nsnlwt)</b>
Asians in Africocarib and AA 60-100 blocks	90889	13	64385	1.411648676	4952.692308
Cuban and PR in Africocarib blocks	84896	44	84897	0.999988221	1929.477273
Mexican and other Hisp in Africocarib blocks	111795	71	111796	0.999991055	1574.591549
Africocarib in Africocarib blocks	577382	933	577386	0.999993072	618.8488746
AA in Africocarib blocks	62549	12	62549	1	5212.416667
Cuban in AA 60-100	16347	11	14984	1.090963695	1362.181818
PR in AA 60-100	68727	23	58545	1.173917499	2545.434783
Mexican in AA 60-100	282346	14	146402	1.92856655	10457.28571
Other His in AA 60-100	165520	25	133484	1.239998801	5339.36
Afrocarib in AA 60-100	288119	246	276861	1.040663004	1125.45122
AA in AA 60-100	10596921	1840	8321916	1.273375146	4522.780435
Asians in AA 30-59.9	134914	15	76687	1.759281234	5112.466667
Cubans, PR, and Mexicans in AA 30-59.9	82585	17	23486	3.516350166	1381.529412
Other His in AA 30-59.9	203503	28	142451	1.42858246	5087.535714
Afrocarib in AA 30-59.9	220638	168	215511	1.023789969	1282.803571
AA in AA 30-59.9	5110387	908	4135708	1.235674037	4554.744493
Vit in AA 10-29.9	179945	68	172342	1.04411577	2534.441176
Flip in AA 10-29.9	389766	89	373002	1.044943459	4191.033708
Chinese in AA 10-29.9	255197	59	242849	1.050846411	4116.084746
Other Asian in AA 10-29.9	592307	74	498077	1.189187616	6730.77027
Cubans in AA 10-29.9	149585	80	145937	1.024997088	1824.2125
PR in AA 10-29.9	374734	83	323989	1.156625688	3903.481928
Mexicans in AA 10-29.9	2942722	160	2179794	1.350000046	13623.7125
Other Hisp in AA 10-29.9	759917	84	625814	1.214285714	7450.166667
Afrocarib in AA 10-29.9	198343	39	161155	1.230759207	4132.179487
AA in AA 10-29.9	3545866	518	2612729	1.357150321	5043.878378
Vit in AA 0-9.9	437448	185	430467	1.016217271	2326.848649
Flip in AA 0-9.9	674184	172	647819	1.040698096	3766.389535
Chinese in AA 0-9.9	1039660	249	999518	1.040161358	4014.128514
Other Asian in AA 0-9.9	1749290	233	1273702	1.373390322	5466.532189

<b>Table 12: NSAL and NLAAS collapsed cells, post-stratification adjustment factor and mean weights (Continued)</b>					
<b>Collapsed groups for NSAL &amp; NLAAS</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratificationadjustment factor</b>	<b>Mean weight(nsnlwt)</b>
Cubans in AA 0-9.9	805628	425	787108	1.023529173	1852.018824
PR in AA 0-9.9	536439	117	402329	1.333334162	3438.709402

<b>Table 12: NSAL and NLAAS collapsed cells, post-stratification adjustment factor and mean weights (Continued)</b>					
<b>Collapsed groups for NSAL &amp; NLAAS</b>	<b>CPS 2002</b>	<b>Un-weighted Count</b>	<b>Count using CPESweights</b>	<b>Post-stratificationadjustment factor</b>	<b>Mean weight(nsnlwt)</b>
Mexican in AA 0-9.9	11529509	635	6760146	1.705511834	10645.89921
Other His in AA 0-9.9	2592100	238	1803859	1.436974841	7579.239496
Afrocarib and AA in AA 0-9.9	733635	73	243405	3.014050656	3334.315068
Whites in Africocarib and AA blocks	145475261	856	10259583	14.17945164	11985.49416
Vit in Hawaiian blocks	3403	2	3403	1	1701.5
Flip in Hawaiian blocks	421617	126	421617	1	3346.166667
Chinese in Hawaiian blocks	223468	69	223469	0.999995525	3238.681159
Other Asian in Hawaiian blocks	360977	84	360977	1	4297.345238
Cuban in Hawaiian blocks	0	0	0	0	0
PR in Hawaiian blocks	37466	11	37466	1	3406
Mexican in Hawaiian blocks	0	0	0	0	0
Other Hisp in Hawaiian blocks	2945	1	2945	1	2945
Africarib in Hawaiian blocks	0	0	0	0	0
AA in Hawaiian blocks	0	0	0	0	0
Whites in Hawaiian blocks	0	0	0	0	0

## IX. CPES Weights Chart

To better understand weights in CPES, please consult our CPES Weights Chart (</CPES/files/cpes-weights-chart.pdf>) .

## X. References and Readings

Alegria M, Takeuchi D, Canino G, Duan N, Shrout P, Meng X, Vega W, Zane N, Vila D, Woo M, Vera M, Guarnaccia P, Aguilar-Gaxiola S, Sue S, Escobar J, Lin K, Gong F. Considering context, space and culture: the National Latino and Asian American Study. *IJMPR* 2004; 13(4): 208-20.

Alegria M, Vila D, Woo M, Canino G, Takeuchi D, Vera M, Febo V, Guarnaccia P, Aguilar-Gaxiola S, Shrout P. Cultural relevance and equivalence in the NLAAS instrument: integrating etic and emic in the development of cross-cultural measures for a psychiatric epidemiology and services study of Latinos. *IJMPR* 2004; 13(4): 270-88.



- Alegria, M., Takeuchi, D.T., Canino, G., Duan, N., Shrout, P.E., Vega, W., Zave, N., Guarnaccia, P., Aguilar-Gaxiola, Ver, M., Sue, S., Escobar, J., Lin, Keh-Ming, Jang, M. and Gong, F. (2004). " Considering Context, Space and Culture: The National Latino and Asian American Study". *International Journal of Methods in Psychiatric Research*, Vol. 13, No.2, pp. 208-220.
- American Association for Public Opinion Research. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. American Association for Public Opinion Research (AAPOR) standard: <http://www.aapor.org> , 2004.
- Atrostic BK, Bates N, Burt G, Silberstein A. Non-response in US government household surveys: consistent measures, recent trends, and new insights. *Journal of Official Statistics* 2001; 17: 209-26.
- Blaise(r) Survey Processing System: Version 4.5. Statistics Netherlands, 2000-2004.
- Bogen K. The effect of questionnaire length on response rates - a review of the literature. Proceedings of the Section on Survey Research Methods, American Statistical Association 1996: 1020-5.
- Bradburn NM. Respondent burden. Proceedings of the Section on Survey Research Methods. American Statistical Association 1978: 35-40.
- Cannell C, Marquis K, Laurent A. A summary of studies. *Vital Health Stat* 1977; 2: 69.
- Chambers, R.L. and Skinner C.J. (editors). (2003). *Analysis of Survey Data*. JohnWiley and Sons, New York.
- Cheung GQ, Liu Y. Displaying Chinese characters in Blaise. Proceedings of the Eighth International Blaise Users Conference 2003, Copenhagen, Denmark.
- Cochran WG. *Sampling Techniques*, 3 edn. New York: John Wiley & Sons, 1977.
- Cochran, W.G. (1977). *Sampling Techniques*. New York: John Wiley & Sons.
- Corbin J, Morse JM. The unstructured interactive interview: issues of reciprocity and risks when dealing with sensitive topics. *Qualitative Inquiry* 2003; 9(3): 335-54.
- de Leeuw E, de Heer W. Trends in household survey nonresponse: a longitudinal and international comparison. In R Groves, DA Dillman, JL Eltinge and RJA Little (eds) *Survey Non-response*. New York: Wiley, 2002, pp. 41-54.
- Groeneveld R. Using non-Latin alphabets in Blaise. Proceedings of the Eighth International Blaise Users Conference 2003. Copenhagen, Denmark.
- Groves RM, Couper MP. *Non-response in Household Interview Surveys*. New York: Wiley, 1998.
- Groves RM, Fowler FJ, Couper MP, Lepkowski J, Singer E, Tourangeau R. *Survey Methodology*. New York: Wiley, 2004.
- Groves RM. *Survey Errors and Survey Costs*. New York: John Wiley & Sons, 1989.
- Guenzel PJ, Berckmans TR, Cannell CF. *General Interviewing Techniques: A Self-Instructional Workbook for Telephone and Personal Interviewer Training*. Ann Arbor, Michigan: Institute for Social Research, Survey Research Center, 1983.
- Hansen MH, Hurwitz WN, Madow WG. *Sample Survey Methods and Theory, Volumes I and II*. New York: John Wiley & Sons, 1953.
- Hansen MH, Hurwitz WN. The problem of nonresponse in surveys. *Journal of the American Statistical Association* 1946; 41: 517-29.
- Hartley, H.O. (1962). "Multiple frame surveys." *Proceedings of the Socials Science Section of the American Statistical Association Meeting*, Minneapolis, Minnesota.

- Hartley, H.O. (1974). "Multiple Frame Methodology and Selected Applications." *Sankhya, Series C*, 3, pp. 99-118.
- Heeringa SG, Connor J, Darrah D. The 1980 SRC/NORC National Sample. Ann Arbor: Survey Methodology Program, Survey Research Center, University of Michigan, 1984.
- Heeringa SG, Connor J, Redmond G. The 1990 SRC National Sample. Ann Arbor: Survey Methodology Program, Survey Research Center, University of Michigan, 1994.
- Heeringa SG, Groves RM. Responsive Design for Household Surveys. Ann Arbor: Survey Methodology Program, Institute for Social Research, University of Michigan, 2004.
- Heeringa SG, Liu J. Complex sample design effects and inference for mental health survey data. *IJMPR* 1997; 7(1): 56-65.
- Heeringa SG, Wagner J, Torres M, Duan N, Adams T, Berglund P. Sample designs and sampling methods for the Collaborative Psychiatric Epidemiology Studies (CPES) *IJMPR* 2004; 13(4): 221-40.
- Heeringa, S. , (2006). Technical Sample Design Documentation: National Study of American Life (NSAL). Technical Report. Statistical Design Group, Institute for Social Research, University of Michigan, Ann Arbor.
- Heeringa, S., (2004). Technical Sample Design Documentation: 2002-2003 National Latino and Asian American Study (NLAAS). Technical Report. Statistical Design Group, Institute for Social Research, University of Michigan, Ann Arbor.
- Heeringa, S., Wagner, J., Torres, M., Duan, N., Adams, T., Berglund, P. (2004). "Sample designs and sampling methods for the Collaborative Epidemiology Studies (CPES)", *International Journal of Methods in Psychiatric Research*, Vol. 13, No. 4, pp. 221-240.
- Heeringa, S.G. and Liu, J. (1997). "Complex sample design effects and inference for mental health survey data." *International Journal of Methods in Psychiatric Research*, Volume 7, Number 1, 56-65.
- Henderson AS, Jorm AF. Do mental health surveys disturb? *Psychol Med* 1990; 20: 721-4.
- Hess I. Sampling for social research surveys: 1947-1980. Ann Arbor: Institute for Social Research, University of Michigan, 1985.
- Jackson J. Life in Black America. London: Sage Publications, 1991.
- Jackson JS, Torres M, Caldwell CH, Neighbors HW, Nesse R, Taylor RJ, Trierweiler, SJ, Williams DR. The National Survey of American Life: a study of racial, ethnic and cultural influences on mental disorders and mental health. *IJMPR* 2004; 13(4): 196-207.
- Jorm AF, Henderson AS, Scott R, MacKinnon AJ, Korten AE, Christensen H. Do mental health surveys disturb? Further evidence. *Psychol Med* 1994; 24: 233-7.
- Kalton, G. (1977), "Practical methods for estimating survey sampling errors," *Bulletin of the International Statistical Institute*, Vol 47, 3, pp. 495-514.
- Kessler R, Berglund P, Chiu WT, Demler O, Heeringa S, Hiripi E, Jin R, Pennell B, Walters E, Zaslavsky A, Zheng H. The US National Comorbidity Survey Replication (NCS-R): an overview of design and field procedures. *IJMPR* 2004; 13(2): 69-92.
- Kessler R. The National Comorbidity Study of the United States. *International Review of Psychiatry*, 1994; 6: 365-76.
- Kessler RC, Üstün TB. The World Mental Health (WMH) survey initiative version of the World Health Organization Composite International Diagnostic Interview (CIDI). *IJMPR* 2004; 13(2): 93-121.

- Kessler RC, Wittchen H-U, Abelson J, Zhao S. Methodological issues in assessing psychiatric disorders with self-reports. In AA Stone, JS Turkkan, CA Barchrach, JB Jobe, HS Kurtzman and VS Cain (eds) *The Science of Self-Report: Implications for Research and Practice*. Mahwah NJ: Lawrence Erlbaum Associates, 200; 229-55.
- Kessler, R., Berglund, P., Chiu, W.T., Demler, O., Heeringa, S., Hiripi, E., Jin, R., Pennell, B., Walters, E., Zaslavsky, A., Zheng, H. (2004). "The U.S. National Comorbidity Survey Replication (NCS-R): an overview of design and field procedures." *International Journal of Methods in Psychiatric Research*, Vol. 13, No.2, pp. 69-92.
- Kish L. A procedure for the objective selection of the respondent within the household. *Journal of the American Statistical Association* 1949; 44: 380-7.
- Kish L. *Statistical Design for Research*. New York: John Wiley & Sons, 1987.
- Kish L. *Survey Sampling*. New York: John Wiley & Sons, 1965.
- Kish, L. (1965), *Survey Sampling*. New York: John Wiley & Sons, Inc.
- Lessler JT, Kalsbeek WD. *Nonsampling Errors in Surveys*. New York: John Wiley & Sons, 1992.
- Little, R.J.A. and Rubin, D.B. (2003). *Statistical Analysis with Missing Data*, 2nd Edition, John Wiley and Sons, New York.
- Pennell, B.P., Bowers, A. Carr. D., Chardoul, S. Cheung, G-Q, Dinkelmann, K, Gebler, N., Hansen, S.E, Pennell, S. Torres, M. (2004). "The Development and Implementation of the National Comorbidity Survey Replication, the National Survey of American Life, and the National Latino and Asian American Survey", *International Journal of Methods in Psychiatric Research*, Vol. 13, No. 4, pp. 241-269.
- Rao, J.N.K & Wu, C.F.J. (1988.), "Resampling inference with complex sample data," *Journal of the American Statistical Association*, 83, pp. 231-239.
- Research Triangle Institute (2003). *SUDAAN User's Manual*, Release 9.0. Research Triangle Park, NC: Research Triangle Institute.
- Rust, K. (1985). "Variance estimation for complex estimators in sample surveys," *Journal of Official Statistics*, Vol. 1, No. 4.
- SAS Institute, Inc. (2003). *SAS/STAT(R) User's Guide*, Version 9, Cary, NC: SAS Institute, Inc.
- Sharp LM, Frankel J. Respondent burden: a test of some common assumptions. *Public Opinion Quarterly* 1983; 47: 36-53.
- Singer E, Groves RM, Corning AD. Differential incentives: beliefs about practices, perceptions of equity, and effects on survey participation. *Public Opinion Quarterly* 1999; 63: 251-60.
- Singer E, Van Hoewyk J, Gebler N, Raghunathan T, McGonagle K. The effect of incentives in interviewermediated surveys. *Journal of Official Statistics* 1999; 15: 217-30.
- Skinner, C.J., Holt, D., & Smith, T.M.F. (1989). *Analysis of Complex Surveys*. New York: John Wiley & Sons.
- STATA Corp. (2004). *STATA Statistical Software: Release 9.0*. College Station, TX: STATA Corporation.
- Turnbull JE, McLeod JD, Callahan JM, Kessler RC. Who should ask? Ethical interviewing in psychiatric epidemiology studies. *American Journal of Orthopsychiatry* 1988; 58(2): 228-39.
- Westat, Inc. (2000). *WesVar 4.0 User's Guide*. Rockville, MD: Westat, Inc.
- Wittchen H-U. Reliability and validity studies of the WHD Composite International Diagnostic Interview (CIDI): a critical review. *J Psychiatr Res* 1994; 28(1): 57-84.
- Wolter, K.M. (1985 ). *Introduction to Variance Estimation*. New York: Springer-Verlag.

- Using CPES -

World Health Organization. Composite International Diagnostic Interview, Version 1.0. Geneva, Switzerland: World Health Organization, 1990.

World Health Organization. Composite International Diagnostic Interview, Version 2.1. Geneva, Switzerland: World Health Organization, 1997.

## Replication of Results

The master CPES datasets are updated periodically as various, typically small, errors are detected. Diagnostic algorithms are periodically updated as well. This updating will continue in the future as needed and public users will be informed of these updates. It is important for public users to recognize that, because of these changes, it will not always be possible to reproduce results reported in earlier publications.