

6. DATA COLLECTION AND MANAGEMENT

6.1 TELEPHONE SAMPLE

The telephone numbers for the GSPS were obtained from a marketing research vendor and were forwarded to the data collection vendor by CDC on a monthly basis.

6.2 HOUSEHOLD AND RESPONDENT SELECTION

The Computer Assisted Telephone Interview (CATI) was used to conduct the survey. Once eligible household contact was made, the CATI system randomly selected one adult in the household to be interviewed. CATI screener text pertinent to landline and cell phone recruitment is in Appendix A. Table 3 summarizes GSPS eligibility criteria.

Table 3: Eligibility Criteria: Gulf States Population Survey (GSPS), December 2010-December 2011

Household is eligible	Eligible households are within the four state area of the sample and are private residences with at least one adult ≥ 18 years of age who lives within the residence at least 4 months of the year.
Respondent is a member of the household	Household member includes all related adults, unrelated adults, roomers, and domestic workers who consider the household their home AND are ≥ 18 years of age, even though they may not be home at the time of the call. Does not include adult family members who are currently living elsewhere, such as at college, group home, a military base, a nursing home, a correctional facility, and so forth.
Respondent is 18 years of age and older	The first question on the questionnaire regarding household selection asks how many members in the household are 18 years of age or older: If the answer is <i>one</i> and the person answering the telephone is that adult, the interviewer proceeds to the first question on the questionnaire. If the response indicates that there are five or more adults, the interviewer should probe to ensure that they are all 18 years of age or older, that all are currently living in the household, and that the household is not a group home or institution. The interviewer then asks the person answering the telephone the number of adult male members in the household, then asks the number of adult female members in the household. Once this information is entered into CATI, the system will randomly

	select a respondent.
How many are men, and how many are women	<p>The next question asks how many of the adults are men and how many are women. The interviewer enters the appropriate answers.</p> <p>If there is only one adult in the household, the interviewer enters "1" and leaves the other box blank.</p>

6.3 CONFIDENTIALITY

The GSPS maintained respondent confidentiality by adhering to the following criteria:

- No respondent identifiers were retained in the interview records.
- Respondent data were combined; reports cite only aggregate figures.
- No interviews were conducted by interviewers who knew the respondents. Interviewers who inadvertently reached persons known to them rescheduled interviews as appointments, which were reassigned to another interviewer.
- Interviewers were trained using standard protocols for the data collection vendor (i.e. BRFSS standards).
- Interviewer monitoring did not violate respondent confidentiality.
- Specific protocols for sensitive questions were adhered to by interview staff including the provision of toll free numbers to suicide and domestic violence prevention hotlines if respondents needed special assistance.
- The GSPS dataset is devoid of all identifying information. Data elements in the file do not include phone numbers, sequence numbers or any personally identifiable information.

6.4 QUALITY ASSURANCE

The GSPS data collection vendor conducted interviews and sent monthly data files to CDC. Interviews were monitored by the vendor's supervisory staff. Verification callbacks were conducted for approximately 10% of completed interviews. Data were quality checked at CDC by reviewing the vendor's frequency reports, skip patterns, and so forth.

6.5 DISPOSITION CODES

Appendix C presents a complete list of disposition codes that were used for the GSPS and their frequency and percent as a proportion of all phone numbers in the total sample (n=1,092,963). Disposition codes were used to categorize respondents and sample numbers into groups by which eligibility was determined. If the respondents completed all survey questions the interviews were considered **Complete**. **Partial Complete** interviews contained substantive responses to sex and at least three other demographic questions (age, race, Hispanic origin, marital status, employment status, county code, and “Do you have more than one telephone number in your household?”). Responses could not include “Don’t know/Not sure”. GSPS interviewers made a second attempt to complete partial interviews and made up to 15 subsequent call attempts. Data from 38,361 **Complete** and **Partially Completed** interviews are in the GSPS database.

As in all random-digit-dialing surveys, a proportion of the sample did not provide **Complete** or **Partial Complete** data and the non-respondent component of the population must be considered when interpreting survey data. In all, 710,417 phone numbers provided by the vendor were not eligible because they no longer existed, belonged to a business, were dedicated FAX lines, etc. **Complete** or **Partial Complete** interviews could not be obtained from 331,804 potentially eligible numbers. In some of these a person answered the phone and the household was eligible for interview but the interview was not conducted or not completed. In the remainder the selected number provided contact but it was not possible to determine if this was an eligible household.

6.6 CALCULATIONS OF CATEGORIES OF ELIGIBILITY AND STATUS

The GSPS disposition codes (Appendix C) were used to categorize respondents and sample numbers into groups by which eligibility was determined. Table 4 illustrates how disposition codes were used in the GSPS to determine completed interviews, contacts with eligible respondents, contacts with persons of unknown eligibility, contact with ineligible households, contact with households, break offs and terminations, and ineligible phone numbers.

Table 4: Categories of Eligibility and Status by Final Disposition Codes: Gulf States Population Survey (GSPS), December 2010-December 2011

Disposition Code(s)	Symbol for Equation	Category of Eligibility	Comment
110, 120	COIN	Complete	Completed or partial completed interview
110, 120, 220, 210, 230, 240, 250, 260, 270, 280	ELIG	Eligible	Contact is made and enough information obtained to ascertain eligibility of household and/or

			respondent
110, 120, Any 200 level, 305, 310, 315, 370, 335, 340, 405, 437	HH	Households/ Likely households	Household or personal cell phone or likely to be household based on available information
210, 220, 230, 240, 250, 260, 270, 280	ELIGHH	Known household/personal cell phone, not interviewed	Contact is made and enough information is obtained to ascertain eligibility of household and/or respondent but interview is not completed.
310, 315, 320, 325, 332, 330, 345, 350, 360, 370	UNKELIG	Unknown eligibility	No contact is made or contact is made but not enough information is obtained to determine eligibility
210, 220, 270, 280	TERE	Break offs, terminations and refusals	Contact is made and eligibility established, but the interview breaks off prior to completing half of the demographic section
355, 365, Any 400 level	INELIG	Ineligible	Information from individuals or from technological devices is available to ascertain that the phone number, household and/or respondent is not eligible
ELIG/ (ELIG+INELI G)	“e”	Eligibility factor	Estimated proportion of cases of unknown eligibility that are likely eligible

6.7 SURVEY PARTICIPATION RATES

Using the categories of disposition codes in Table 4, rates of cooperation, response, refusal, and interview completion were calculated for the GSPS (Appendix D). The GSPS used formulae adapted from the American Association of Public Opinion Research (AAPOR) [4] and the Council of American Survey Research Organizations (CASRO) [5] to calculate outcomes. AAPOR provides a number of potential methods for calculation of response and two of their standard formulae are provided below. The following formulae were used to calculate rates from eligibility status of respondents, selected respondents, and/or sample phone numbers.

- **Resolution Rate:** The proportion of all telephone numbers in the sample for which their status of the numbers as households with working numbers has been resolved. Records for which household status is unknown are excluded from the numerator.

$$= (\text{ELIG} + \text{INELIG}) / (\text{ELIG} + \text{INELIG} + \text{UNKELIG})$$

- **Cooperation Rate (AAPOR Cooperation Rate 2):** The proportion of all respondents interviewed from known eligible units in which a respondent was selected and contacted. Non-contacts are excluded from the denominator. This rate is based on contacts with households containing an eligible respondent. The denominator of the rate includes completed interviews plus the number of non-interviews that involve the identification of and contact with a selected respondent.

$$= \text{COIN} / (\text{ELIG})$$

- **Interview Completion Rate:** The proportion of selected respondents who successfully completed an interview. This is a type of cooperation rate.

$$= \text{COIN} / (\text{COIN} + \text{TERE})$$

- **Refusal Rate (AAPOR Refusal Rate 2):** The proportion of all eligible respondents who refused to complete an interview or terminated an interview prior to the threshold required to be considered a partial interview.

$$= \text{TERE} / (\text{ELIG} + e*(\text{UNKELIG}))$$

- **CASRO Response Rate (AAPOR Response Rate 4):** The proportion of the number of complete interviews by the number of interviews plus all cases of unknown eligibility that is actually eligible.

$$= \text{COIN} / (\text{COIN} + \text{ELIGHH} + e*(\text{UNKELIG}))$$

- **Response Rate (AAPOR Response Rate 6):** The proportion of the number of complete interviews by the number of interviews.

$$= \text{COIN} / (\text{COIN} + \text{ELIGHH})$$

Analysis and evaluation of GSPS health data to defined populations requires consideration of overall survey participation. Appendix D illustrates rates of outcomes for the GSPS for each state based on categories of final dispositions.

6.8 ACCESSING THE GSPS DATA FILES

The GSPS data files are provided in SAS, SPSS, and ASCII formats and are available at <http://www.cdc.gov/osels/phsipo/dbs/GSPS.HTML>.

Data files: GSPS2012 (in SAS, SPSS, and ASCII)

Format file: GSPS_SASFormats.sas (in SAS)

The following program can be used to load the SAS format and SAS dataset:

```
libname in "C:";
%include "C:\GSPS_SASFormats.sas"; * loading the format;
```

```
data temp;
  set in.GSPS2012;
run;
```

6.9 ACCOUNTING FOR THE COMPLEX SAMPLING DESIGN IN GSPS DATA ANALYSIS

The GSPS dataset was created from a complex survey design. Use of the dataset to most accurately represent populations in the coastal and noncoastal areas must consider the complex survey design. The statistical procedures or modules of statistical software (e.g., SUDAAN, SAS and SPSS) include statements for stratification, clustering and sample weight to account for complex sampling design of survey data. The following are examples for specifying strata, cluster and sample weights when appropriately analyzing the GSPS data:

SAS or SPSS:

```
Strata _STSTR;
Cluster _PSU;
Weight _GSPSWT;
```

SUDAAN:

NEST_STSTR_PSU;
WEIGHT_GSPSWT;

Lack of attention to the complex survey design specifications in analyses will yield biased estimates of population prevalence and of associated confidence intervals. Users should refer to the appropriate SUDAAN, SAS, SPSS, or other software User Manuals for detailed discussion concerning incorporation of complex survey designs before conducting data analyses.

6.10 CALCULATED CATEGORICAL VARIABLES

The GSPS dataset includes variables for all questions asked within the survey questionnaire (See Appendix A). In addition the dataset includes calculated categorical variables to classify depression (PHQ-8), anxiety (GAD-7), Self-Mastery, heavy drinking, binge drinking, and current smoking status (Appendix E).

6.11 GSPS CODEBOOK

The GSPS codebook showing the variable names, location, and frequency of values is provided in Appendix F.