

## **Section B. Collections of Information Employing Statistical Methods**

There are four main objectives for the pilot study: 1) locating the population of interest; 2) determining success in enrolling the population of interest; 3) describing the survey response rates; and 4) performing a comparison of prevalence rates to national surveys. To meet these objectives, this pilot study will use the statistical methods described here.

### **B.1. Respondent Universe and Sampling Methods**

This section will discuss the target and survey population; sampling, reporting, and analytic unit; sampling frame; sample size adjustments; and the general survey design.

#### **Target and Survey Population**

The target population is all people, adults and children, who resided in FEMA-supplied temporary housing units for at least one week in the aftermath of either hurricanes Katrina or Rita. Fortunately, the sampling frame we will use has virtually complete coverage of target population. Consequently, the target population and the survey population are essentially the same.

#### **Sampling, Reporting, and Analytic Unit**

The sampling unit will be the registration identification number provided by FEMA. The registration identification number is a unique identifier for the person who registered for a temporary housing unit. For example, if someone registered for more than one temporary housing unit, each of the temporary housing units will have the same registration identification number. We will refer to the person who registered for the temporary housing unit as the registrant. The registrant will be the reporting unit. That is, the registrant will provide information about all the people, adults and children, who lived in the temporary housing unit. Therefore, the analytic unit will be a person.

#### **Sampling Frame**

The sampling frame is FEMA's financial assistance records and records of occupancy kept since the initial issuance of travel trailers, park homes, and mobile homes as temporary housing in the fall of 2005. These datasets will be populated with contact information on temporary housing unit registrants. The FEMA datasets about occupancy of temporary housing units supplied to ATSDR contain information on approximately 130,000 temporary housing unit registrants. For registration identification numbers that had multiple observations in the database, one observation was selected at random so that each observation in the database represented a unique registration identification number. This resulted in a database that contains 118,684 unique identification numbers, i.e., unique registrants. For the pilot study, this will be restricted to the following counties/parishes list in **Exhibit 1. Feasibility Study Counties/Parishes** which also contains the number of registrants in each county/parish. This restriction will not allow us to generalize to the entire temporary housing unit population but only to the counties/parishes that have been included which include about 75% of the target population. There was a desire to

restrict the counties/parishes to contiguous counties/parishes within a state in order to have a concentrated outreach media campaign informing residents of this study.

**Exhibit 1. Feasibility Study Counties/Parishes**

State	County, State	Registrants
Alabama	Mobile, AL	1,788
Louisiana	Orleans, LA	24,239
Louisiana	Jefferson, LA	19,504
Louisiana	St. Tammany, LA	11,889
Mississippi	Harrison, MS	11,577
Mississippi	Jackson, MS	8,928
Mississippi	Hancock, MS	7,451
Texas	Jefferson, TX	1,604
Texas	Orange, TX	953
Texas	Hardin, TX	522
Texas	Jasper, TX	435
Texas	Tyler, TX	245
Texas	Newton, TX	175

**Sample Size Adjustments**

The analytic sample size is the sample size required to meet the analytic objects, primarily the objective specifying comparison of prevalence rates to national surveys. The actual sample size is the sample size selected in order to attain the analytic sample size after accounting for non-contact, ineligibility, non-cooperation, and attrition. **Exhibit 2. Analytic Sample Size, Sample Size Adjustments, and Actual Sample Size** describes the analytic sample size that was determined by the power calculations (See Degree of Accuracy sub-section in the next section B.2. Procedures for the Collection of Information for power calculations.); the adjustments to the analytic sample size and the expected rates, the sample count, and the actual sample size that will be used for selecting the sample. The actual sample size we will use is 17,000. We rounded up the analytic sample size from 16,525 to 17,000 to account for any uncertainty in the expected rates for the adjustments.

**Exhibit 2. Analytic Sample Size, Sample Size Adjustments, and Actual Sample Size**

Sample	Adjustment	Rate	Count
Actual Sample			<b>16,525</b>
	Retention	0.65	10,741
	Cooperation	0.70	7,519
	Eligibility	0.95	7,143
Analytic Sample	Contact	0.70	<b>5,000</b>

**Exhibit 3. Actual Sample Size Allocation for the Feasibility Study** shows the counties/parishes in feasibility study and the actual sample allocated to these counties/parishes.

**Exhibit 3. Actual Sample Size Allocation for the Feasibility Study**

State	County, State	Registrants	Actual Sample Size
Alabama	Mobile, AL	1,788	340
Louisiana	Orleans, LA	24,239	4,614
Louisiana	Jefferson, LA	19,504	3,713
Louisiana	St. Tammany, LA	11,889	2,263
Mississippi	Harrison, MS	11,577	2,204
Mississippi	Jackson, MS	8,928	1,699
Mississippi	Hancock, MS	7,451	1,418
Texas	Jefferson, TX	1,604	305
Texas	Orange, TX	953	181
Texas	Hardin, TX	522	99
Texas	Jasper, TX	435	83
Texas	Tyler, TX	245	47
Texas	Newton, TX	175	33
Total		89,310	17,000

**General Survey Design**

The survey design is based on probability sampling. The survey design will be stratified simple random sampling with proportional allocation based on the size of the population in each county/parish identified for the pilot. The proportion allocation across the counties/parishes provides us with an equal probability selection method of sample selection. A detailed description of the sample selection will be provided in the Sample Selection Methodology subsection in the next section B.2. Procedures for the Collection of Information.

**B.2. Procedures for the Collection of Information**

This section will discuss the target and survey design in detail; estimation procedures, and power calculations.

**Survey Design in Detail**

We will use a probability sampling design. The sample design will be stratified simple random sampling of unique registration identification numbers with proportional allocation to counties/parishes based on the number of unique registration identification numbers in each county/parish. This is essentially an equal probability selection method. The probability of selection for the unique registration identification number will be the number of unique registration identification numbers selected for the sample in a sampling stratum divided by the total number of unique registration identification numbers in the sampling stratum. That is, the

probability of selection for the  $i^{th}$  unique registration identification number in the  $h^{th}$  sampling stratum is,  $p_{hi}$ , will be

$$p_{hi} = \frac{n_h}{N_h},$$

where  $n_h$  is the number of unique registration identification numbers selected for the sample in the  $h^{th}$  sampling stratum and  $N_h$  is the total number of unique registration identification numbers in the  $h^{th}$  sampling stratum. The design weight for a unique registration identification number will be the inverse of the unique registration identification number probability of selection. That is, the design weight for the for the  $i^{th}$  unique registration identification number in the  $h^{th}$  sampling stratum,  $d_{hi}$ , will be

$$d_{hi} = \frac{1}{p_{hi}}.$$

**Exhibit 4. Actual Sample Size Allocation, Probability of Selection, and Design for the Feasibility Study** shows the counties/parishes in feasibility study, number of registrants, the actual sample allocated, the probability of selection, and design weight for these counties/parishes.

**Exhibit 4. Actual Sample Size Allocation, Probability of Selection, and Design Weight for the Feasibility Study**

State	County, State	Registrants	Actual Sample Size	Probability of Selection	Design Weight
Alabama	Mobile, AL	1,788	340	0.1902	5.2588
Louisiana	Orleans, LA	24,239	4,614	0.1903	5.2535
Louisiana	Jefferson, LA	19,504	3,713	0.1903	5.2535
Louisiana	St. Tammany, LA	11,889	2,263	0.1903	5.2535
Mississippi	Harrison, MS	11,577	2,204	0.1903	5.2535
Mississippi	Jackson, MS	8,928	1,699	0.1903	5.2535
Mississippi	Hancock, MS	7,451	1,418	0.1903	5.2535
Texas	Jefferson, TX	1,604	305	0.1904	5.2523
Texas	Orange, TX	953	181	0.1904	5.2523
Texas	Hardin, TX	522	99	0.1904	5.2523
Texas	Jasper, TX	435	83	0.1904	5.2523
Texas	Tyler, TX	245	47	0.1904	5.2523
Texas	Newton, TX	175	33	0.1904	5.2523
Total		89,310	17,000		

For the analytic file, there will be clustering. The registrant will provide information about all the people, adults and children, who lived in the temporary housing unit. Consequently, the

people on which the registrant reports will be clustered by registrant. Each person for which information is reported will have unique registration identification number design weight assigned to them.

Quality control during sample selection will consist of summing the probabilities of selection in a sampling stratum to ensure that the sum of the probabilities of selection equals the sample size in the stratum. After calculation of the design weights, we will check that the correct number of sampling units have been selected and ensure that the sum the design weights in stratum equals the population size in the stratum.

## **Estimation Procedures**

Inferences from the data will only be to the counties/parishes included in the pilot sample. (Refer to the justification for this in the sampling frame sub-section.) The data will be analyzed using the current version of the SAS survey procedures to appropriately account for the complex survey design, including the clustering, and the weighting which allows for statistically valid inferences. SAS will be used to analyze data, describe demographic characteristics, risk factors and experiences, and perform comparisons of prevalence rates of health symptoms or conditions between the pilot registry and national surveys data (i.e., NHANES and NHIS). The variables used to assess the objectives are:

Health Status: cough (HLTH1, HLTH2), phlegm (HLTH3, HLTH4), wheezing (HLTH5 – HLTH12), dry cough shortness of breath (HLTH13, HLTH14), asthma (HLTH15- HLTH18), sinus problem (HLTH15-HLTH18), chronic bronchitis (HLTH15-HLTH18),

Mental health status (HLTH19, HLTH20)

Socioeconomic: Smoking (SMOKE1-SMOKE4), Alcohol (ALC1-ALC6), Access to Health care (HLTH21, HLTH22), Race and Ethnicity (D8, D9), Marital Status (D10), Education (D11), Employment Status (D12-D14), Income (D15, D16).

## **Power Calculations**

The focus of the power calculations is on prevalence rates for adults. Adults were selected because we know that we will have at least 5,000 of them because the registrant has to be an adult. The analytic objective is to compare prevalence rates for adults from the pilot data to prevalence rates for adults to the most recent National Health and Nutrition Examination Survey (NHANES 2007-2008). The same questions in NHANES, or very similar questions, will be asked of registrants in the pilot study and will be compared to the NHANES estimates. The power calculations were produced using PASS 2008<sup>a</sup> software. A detailed discussion of the power calculations is in Appendix I: Power Calculations for the Katina Pilot Registry. In summary, prevalence rates were calculated for a large number of NHANES questions. The power calculations were based on grouping the prevalence rates to present a reasonable number of power calculations.

Given the respondent sample size of 5,000 expected for the pilot sample, 80 percent power is

---

<sup>a</sup> Hintze, Jerry L. (2008). *PASS 2008*. Utah: Kaysville.

achieved to detect differences of about +/- 1 percentage point for the pilot prevalence rates compared to the NHANES prevalence rate of 2 percent; differences of about +/- 2 percentage points for the pilot prevalence rates compared to the NHANES prevalence rates of 6, 10, and 14 percent; and differences of about +/- 3 percentage points for the pilot prevalence rates compared to the NHANES prevalence estimates of 50 or 75 percent. These minimum detectable differences are acceptable to us for this data collection. The test statistic used is the two-sided Z test with pooled variance. The significance level of the test was targeted at 0.05.

### **B.3. Methods to Maximize Response Rates and Deal with Nonresponse**

#### **Methods to Deal with Nonresponse**

Nonresponse is a challenge in almost all surveys. We will approach this challenge in three sequential phases: monitoring response rates during data collection, nonresponse bias analysis, and post-survey weight adjustments to minimize potential bias. The first phase will be to monitor response rates during data collection. This will allow us to allocate more of the data collection resources to the sampling strata with the lowest response rates in order to get response rates across strata as uniform as possible. Even with this effort, we will not get the same response rate in every stratum. Consequently, we will conduct nonresponse bias analysis.

The nonresponse bias analysis will be conducted in two separate steps. The first step will be to model the contact indicator, and the second step will be to model the cooperation indicator. This will help us identify variables from the frame that are associated with these nonresponse mechanisms. That is, for each of these steps, we will use information from the sampling frame as independent variables in a modeling process. We will use nonparametric tree-based methods to identify the independent variables associated with the appropriate indicator.

Post-survey weighting will be used to minimize the potential bias from nonresponse. Using the information from the nonresponse bias analysis, the tree-based methods will produce separate sets non-contact adjustment cells and non-cooperation adjustment cells that can be used for the nonresponse adjustments, or the variables identified through the tree-based model can be used in generalized exponential models. Either way, the nonresponse adjustments will be implemented sequentially, non-contact then non-cooperation. Using the adjustment cells created from the tree-based models will be our first approach. If for some reason the adjustment cell approach is not satisfactory, e.g. the adjustment cells get too small or adjustment factors get too big, we will use the generalized exponential model implemented in SUDAAN's weight adjust procedure to model the contact and cooperation indicators sequentially. For either approach, the two adjustment factors will be combined to create the overall nonresponse adjustment factor. Quality control for the post-survey weighting will include review of the number of respondents and nonrespondents in the adjustment cells, review of adjustment factors for the adjustment cells, ensuring that the adjusted weights sum to the correct totals at each step of the adjustment process, and monitoring the unequal weighting effect overall and within sampling strata.

### **B.4. Test of Procedures or Methods to Be Undertaken**

A CATI system based on a paper questionnaire will be created. The CATI system will then be used during all interviews to collect data for this pilot registry. CATI systems have several advantages over a paper-and-pencil mode of data collection. First, with a CATI system, survey data is captured electronically, which precludes the need to later key paper data into a database. Theoretically, this reduces transcription errors. CATI systems can also define the type and range of data that can be entered in each field. This can help prevent data entry errors (e.g., entering alphanumeric characters in a social security number field). Finally, a CATI system may also improve the efficiency of an interviewer, resulting in less time being spent writing responses, working through skip patterns in the survey, and a shorter overall interview and respondent burden time.

The CATI data collection instrument will be composed of two parts. The first part will consist of screening questions to determine eligibility for enrollment (Appendix E). The second part—the main questionnaire (Appendix F) will contain contact information of the registrant and other household members, demographics, and health status questions—focusing on respiratory outcomes and mental health.

Health status questions will be identical to those of the 2007-2008 National Health and Nutrition Examination Survey and the 2009 National Health Interview Survey.

The questionnaire was evaluated for ease in administration and comprehension. Skip patterns were checked and instructions to the interviewers for handling various situations that may arise will be developed. In addition, as the CATI is tested, validity checks will be developed to minimize response error.

### **Pilot Testing**

“Cognitive Interviews” were conducted in February 2011 with 9 individuals recruited from an area in which FEMA-supplied temporary housing units were occupied in order to identify any issues related to recall bias; and to determine respondent willingness to provide sensitive information such as social security numbers. RTI conducted 9 cognitive interviews at the Louisiana Public Health Institute (LPHI) in New Orleans. Eight interviews were conducted face-to-face and one was conducted via telephone to simulate a true telephone interview. A cognitive interview protocol was developed to standardize the approach and questions asked during the interview process. This protocol was submitted and approved by RTI’s Institutional Review Board before interviews were conducted. Results from the cognitive interviews will be used to improve the questionnaire, train and monitor the work of interviewers, and to facilitate the interpretation of results.

A timing test of the questionnaire was completed on May 1, 2009. A total of five test interviews were completed in-house by registry staff. The total time needed to answer all the questions ranged from 20 to 25 minutes with an average completion time of 23 minutes. The timing test did not include additional probes. The timing interviews were not video or audio taped.

RTI staff will conduct telephone interviews. Personnel who perform this work will be trained in the purpose of the registry, how to conduct the consent process over the telephone, and how to conduct a telephone interview. Interviewers will also be given training on how to handle

difficult situations that may arise during interviews, such as when respondents react emotionally to questions which remind them of their experiences during or after the hurricanes

Prior to data collection, all staff and contractors will be trained on security and confidentiality policies and procedures. Turnover of interviewers is anticipated to be potentially high, given the intense nature of the material. Accordingly, training will need to be developed that can be stand-alone for each interviewer. In addition, from time to time it is anticipated that some interviewers will need to have method refresher training. For these reasons a CD ROM-based training for interviewers will be developed.

### **Evaluating the Success of the Pilot**

Success of the pilot registry will be measured by the following four factors: locating the population of interest; success in enrolling the population of interest; the survey response rates and; comparison of prevalence rates to national surveys. Documenting self-reported medical conditions will not be included in this OMB request.

**Locating the population of interest.** RTI International, the contractor, was provided the FEMA database of about 130,000 THU occupants. They then designed the sampling plan to sample 17,000 eligible individuals to be traced/located (Appendix H).

**Success in enrolling the populations of interest.** Develop SOPs to track recruitment efforts for each participant including refusals (e.g., number of telephone calls/attempts needed to contact; time expended per dollars spent).

**Survey Response Rates.** A contact rate of >65% would demonstrate the success of the pilot registry. For the calculation of outcome rates for surveys, the standard is the American Association for Public Opinion Research's (AAPOR) *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Survey (AAPOR, 2008)*.<sup>1</sup> This document provides comprehensive methods for calculating outcome rates for surveys conducted by random-digit dialing (RDD) telephone, for personal interviews in a sample of households, and for mail surveys of specifically named persons. While the Katrina Pilot Registry will not neatly fit into one of these three categories, it can be described primarily as a telephone survey of specifically named persons (a combination of all three types listed above). As such, the AAPOR standards serve as the correct guidelines for the calculation of cooperation and contact rates for the Pilot Registry. The cooperation and contact rates will be evaluated to see the success of the Pilot Registry and determine if the full registry might be feasible to complete.

The components of outcomes rates are:

I = Complete interview

P = Partial interview

---

<sup>1</sup> The American Association for Public Opinion Research (2008). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Ann Arbor, Michigan: AAPOR. [http://www.aapor.org/default.asp?page=survey\\_methods/standards\\_and\\_best\\_practices/standard\\_definitions](http://www.aapor.org/default.asp?page=survey_methods/standards_and_best_practices/standard_definitions)



R = Refusal and break-off  
 O = Eligible other non-interview  
 NC = Eligible Non-contacts  
 UH = Unknown if household/occupied household  
 UO = Eligibility unknown, other  
 E = estimated proportion of cases of unknown eligibility that are eligible

Cooperation Rate

A cooperation rate is the proportion of all cases interviewed of all eligible units ever contacted. There are both household-level and respondent-level cooperation rates. The rates here are household-level rates. They are based on contact with households, including respondents, rather than contacts with respondents only. Respondent-level cooperation rates could also be calculated using only contacts with and refusals from known respondents.

$$\text{COOP1} = \frac{\mathbf{I}}{\mathbf{(I + P) + R + O}}$$

Cooperation Rate 1 (COOP1), or the minimum cooperation rate, is the number of complete interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews that involve the identification of and contact with an eligible respondent (refusal and break-off plus other).

A cooperation rate of >70.0% would show the success of the Pilot Registry.

Contact Rate

A contact rate measures the proportion of all cases in which some responsible member of the housing unit was reached by the survey. The rates here are household-level rates. They are based on contact with households, including respondents, rather than contacts with respondents only. Respondent-level contact rates could also be calculated using only contact with and refusals from known respondents.

$$\text{CON1} = \frac{\mathbf{(I + P) + R + O}}{\mathbf{(I + P) + R + O + NC + (UH + UO)}}$$

Contact Rate 1 (CON1) assumes that all cases of indeterminate eligibility are actually eligible.

$$\text{CON2} = \frac{\mathbf{(I + P) + R + O}}{\mathbf{(I + P) + R + O + NC + E(UH + UO)}}$$

Contact Rate 2 (CON2) includes in the base only the estimated eligible cases among the undetermined cases.

**Comparison of prevalence rates to national surveys.** The comparison of prevalence rates obtained through the pilot registry with estimates from national surveys will help determine the utility of conducting a full registry. For example, if all or most health outcomes do not appear to be in excess, the value of a full registry may be questionable.

## **B.5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data**

The Surveillance and Registry Branch (SRB) in ATSDR's Division of Health Studies, is in charge of constructing the Katrina-Rita Pilot Registry.

1. Data will be collected under contract with guidance from branch Epidemiologists. Data will be analyzed in-house by statisticians.
2. Questions regarding this OMB package and data collection procedures should be addressed to Dr. Vinicius Antao at 770-488-0555, [VAntao@cdc.gov](mailto:VAntao@cdc.gov).
3. Questions regarding statistical methods should be addressed to Mr. James Sapp at 770-488-3814, [JSapp@cdc.gov](mailto:JSapp@cdc.gov).
4. Questions regarding IT methods should be addressed to Mr. Timothy Copeland at 770-488-3696, [TCopeland@cdc.gov](mailto:TCopeland@cdc.gov).