

Annex A: Methodology

Sampling Scheme

The RALLY survey of Central City incorporated a classic random cluster sample. Central City was divided into 131 clusters of approximately equal size, and a random (equal probability) selection of 33 clusters was made using a sampling interval and a random number sheet. Within each cluster all inhabited households were included in the sample.

After adjusting for non-response weighting (discussed below), the survey can be generalized to the *population of returned households in Central City*.

Weighting

Two major factors influence probability of any particular residence being included in the survey: probability of selection and probability of response. Design weights and non-response weights, respectively, can be calculated to adjust for differences in these probabilities among residences.

Design weights

Sampled households are weighted according to probability of selection. In the case of the Central City survey, the neighborhood was divided into 131 clusters, and 31 were selected at random. Within each cluster, all residences were selected for participation in the survey, and thus each household in Central City had an equal probability of selection.

As all households had an equal probability of being selected for the survey, *no design weights are needed*.

In the Central City survey, variability in the probability of being included in the survey was influenced mainly by the probability of *responding* to the survey.

Non-response weights

The non-response rate (due to both low availability and refusal to respond) was quite high in the Central City survey (73.6%).

Proxy data on a random sub-sample of non-responding households were collected through ad hoc interviews with neighbors. Variables collected by proxy included the number of residents in the non-responding household, ethnicity, income class and gender.

Logistic regression was used to investigate a number of factors, including small household size, high Katrina flood depth and household ethnicity which seemed to be significantly associated with non-response. Model building revealed that being African American was significantly associated with non-response (Table K). Model 1 indicates that the odds of response decrease by a factor 0.431 among African Americans.

Household size and depth of flooding were also significant factors in some models, but were found to be associated with being African American, not with non-response. Thus, the decision was made to use African American ethnicity to distinguish two non-response classes.

Table K. Logistic Regression Models: Variables potentially associated with survey response.

Variable	Model 1	Model 2	Model 3	Model 4
African American	coeff: -0.841 p: 0.093* Odds ratio: 0.431			coeff: -2.034 p: 0.0484* Odds ratio: 0.131
Household Size < 2		coeff: -0.573 p: 0.064* Odds ratio: 0.564		coeff: -0.027 p: 0.945 Odds ratio: 0.973
Flooded > 3ft			coeff: -1.372 p: 0.000* Odds ratio: 0.254	coeff: 0.649 p: 0.311 Odds ratio: 1.913
<i>Intercept</i>	coeff: -1.046 p: 0.000	coeff: -0.670 p: 0.011	coeff: 2.062 p: 0.000	coeff: -1.913 p: 0.005

*significant at $p < 0.1$

Post-stratification by distinct response class (in this case, African American vs. non-African American) is a common way to apply non-response weights. Probability of response for these classes is generally calculated:

$$R_j = s_j / n_j$$

where

- R_j = Probability of responding to the survey
- s_j = Number of households responding in the j^{th} response class
- n_j = Number of households sampled within the j^{th} response class

As the ethnicity data was not available for non-responders, it was necessary to estimate n_i among African Americans and non-African Americans based on the rate observed in the proxy data. The ethnic make-up of the proxy data was applied to the total number of non-responders sampled (Table L), and response probability was calculated based on the *estimated number* of responses in each class. The probabilities of response are presented in Table M.

Table L: Estimated response totals based on proxy data.

Response	Sample size	Estimated Percent African American*	Estimated Number African American	Estimated non-African American
Non-response	607	92.86%	564	43
Response	218	84.86%	185	33
<i>total</i>	<i>825</i>		<i>749</i>	<i>76</i>

*Percentage in the non-response class is based on proxy data collected from a random subsample.

Table M: Estimated probabilities of response.

Non-response class	Probability of response
African American	0.25
Non-African American	0.43

Overall weighting

Sample weights are thus based on the overall probability of selection for each household in Central City.

$$P_{ij} = (m/M) * (n_i/N_i) * R_j$$

where

P_{ij} = Probability of selection for households in the i^{th} cluster and the j^{th} response class.

m = Number of sample clusters chosen

M = Total number of sample clusters

n_i = Number of sampled households within the i^{th} cluster

N_i = Total number of households in the i^{th} cluster

R_j = Probability of response for households in the j^{th} response

The sample weight is the inverse of the probability of being selected.

$$W_{ij} = 1/P_{ij}$$

where

W_{ij} = Overall weight for households in the i^{th} cluster and j^{th} response class

P_{ij} = Overall probability of selection for households in the i^{th} cluster and j^{th} response class

Analysis

The analysis was done with SPSS 15 using the Complex Samples module, which takes non-SRS survey designs and unequal selection probabilities into account when calculating statistics and estimating standard error. It is important to note that this package does not automatically account for variance in the estimation of the response probability itself, and thus the confidence intervals reported here are probably slightly tighter than they otherwise would be. With this in mind, a strict 95% confidence interval is reported throughout this analysis.

Central City Population Estimate

The general formula for estimating the summer 2006 total population of Central City can be expressed:

$$\text{Total population} = \text{Total number of residences in Central city} * \text{Occupancy rate} * \text{Average household size}$$

Specifically, the RALLY population estimate was calculated utilizing three critical pieces of information:

1. Total number of residences in central city (source: 2000 US census).
2. Occupancy rate within surveyed clusters, stratified by response group and flood depth (source: this survey)
3. Average household size of surveyed residences, stratified by response group (source: this survey).

The RALLY survey of Central City was designed to collect the occupancy status and household size for all residences in each survey cluster. Using this data, the occupancy rate and average household size among successfully surveyed

households was calculated. Separate occupancy rates and average household sizes were estimated for non-responding residences based on proxy data discussed under “Non-response weights” in the “Methodology” section of this report. The occupancy rate among non-responding residences was further stratified by flood depth (above/below one foot of Katrina flooding). These separate occupancy rates were applied proportionally to the 2000 US census estimate of the total number of residences in Central City, yielding an estimate of the total number of occupied residences within each stratum. The number of occupied residences within each stratum was subsequently multiplied by the average household size for the respective stratum, yielding a stratified population estimate. The strata totals were summed to give the total population estimate for Central City, summer 2006.