

## Appendix I

# Katrina and Rita Exposures Registry:

## Feasibility Sample Selection

This document describes the sampling plan for the Katrina and Rita Exposures (KARE) Registry Pilot feasibility study. It contains the following sections: (1) sampling frame development, (2) stratification, (3) sample allocation, (4) applicant selection, (5) snowball sampling, and (6) de-duplication of final data file.

### 1. Sampling Frame Development

The sample will be based on the Federal Emergency Management Agency (FEMA) database provided by Centers for Disease Control and Prevention. The FEMA database is a list of adult applicants for temporary housing units (THU), where each adult represents a household that lived in a THU. Each applicant has a unique registration identification number. 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 contained 118,684 observations. See Appendix A: Distribution of Applicants for a map that shows the density of applicants across counties/parishes. For the feasibility study, sample selection will occur in Alabama, Louisiana, Mississippi, and Texas. The database has 114,292 observations with a geocoded address in Alabama (2,447), Louisiana (70,832), Mississippi (34,482), and Texas (6,531).

### 2. Stratification

For the KARE feasibility study, the explicit stratification will consist of designated counties/parishes. That is, designated counties/parishes will be the sampling strata. There is one county in Alabama, three parishes in Louisiana, three counties in Mississippi, and six counties in Texas designated to be in the feasibility study. In each state the counties/parishes are contiguous. Table 1: Feasibility Study Counties/Parishes lists the counties/parishes that will be included in the feasibility study and the number of applicants in each county/parish.

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

State	County, State	Applicants
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

The counties/parishes represent a mix of rural and urban parishes/counties. See Appendix B: Study Counties/Parishes for a map of study counties/parishes. Within each of these counties/parishes, we will use implicit stratification by Census tract to allocate the sample within the explicit sampling strata.

### 3. Sample Allocation

The sample size for the feasibility study was set at 10,000 applicants. The sample will be allocated proportionally based on the number of applicants across Alabama, Louisiana, Mississippi, and Texas. About 2% of the sample will be allocated to Alabama, about 62% to Louisiana, about 31% to Mississippi, and about 4% to Texas. These percentages represent the approximate population proportions of the applicants based on the applicant counts for Alabama (2%), Louisiana (62%), Mississippi (30%), and Texas (6%). Within each of the states, the sample will be allocated proportionally to the designated counties/parishes within the state. Within each of the designated counties, the sample will be proportionally allocated to the Census tracts. Appendix C: Feasibility Study Counties/Parishes Sample Allocation has a list of feasibility study counties/parishes and the sample allocation for these counties/parishes.

### 4. Applicant Selection

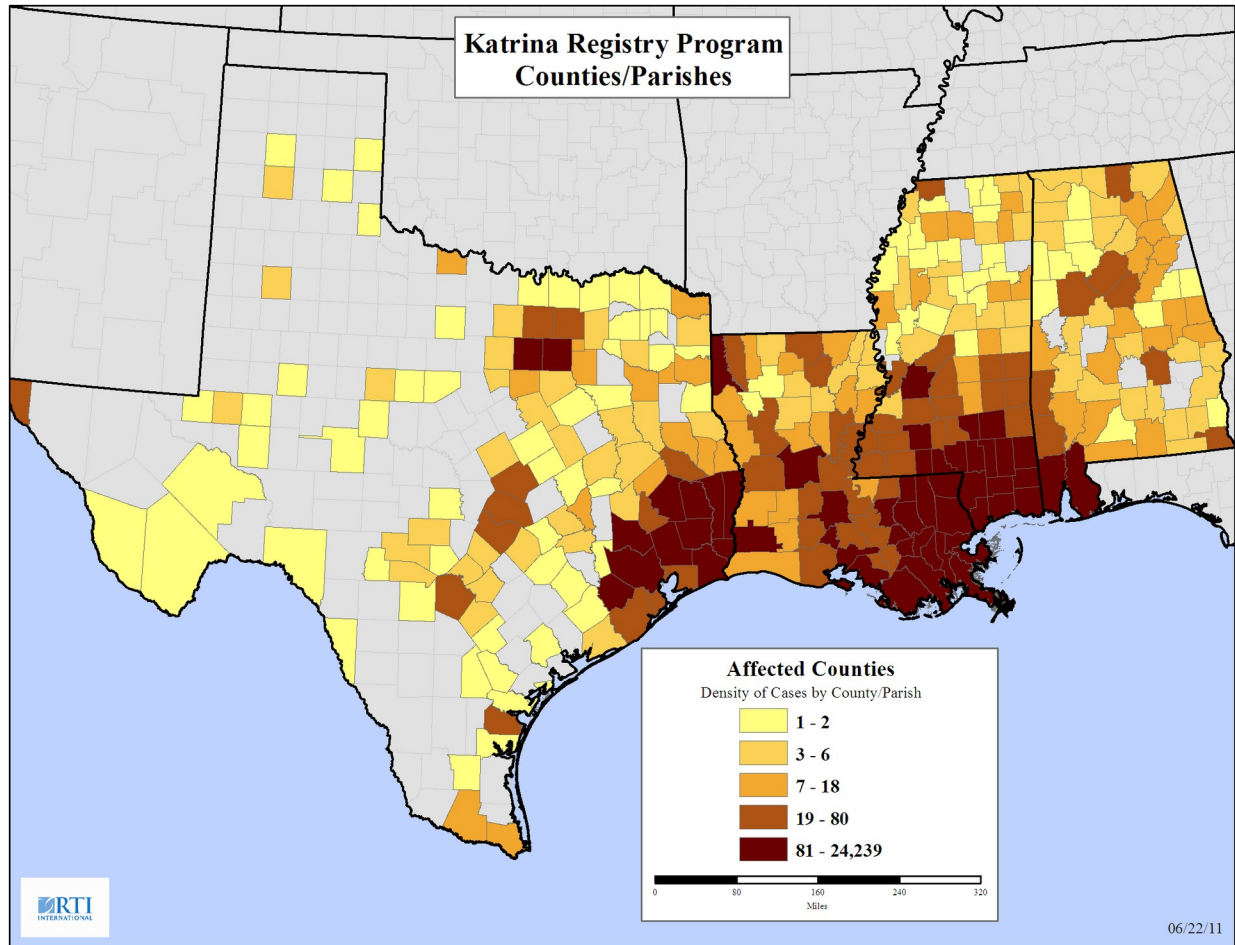
In general, sample selection will be stratified simple random sampling with proportional allocation. The probability of selection for the applicant will be the number of applicants selected for the sample in a sampling stratum divided by the total number of applicants in the sampling stratum. That is, the probability of selection for the  $i^{\text{th}}$  applicant in the  $h^{\text{th}}$  sampling stratum is,  $p_{hi}$ , will be

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

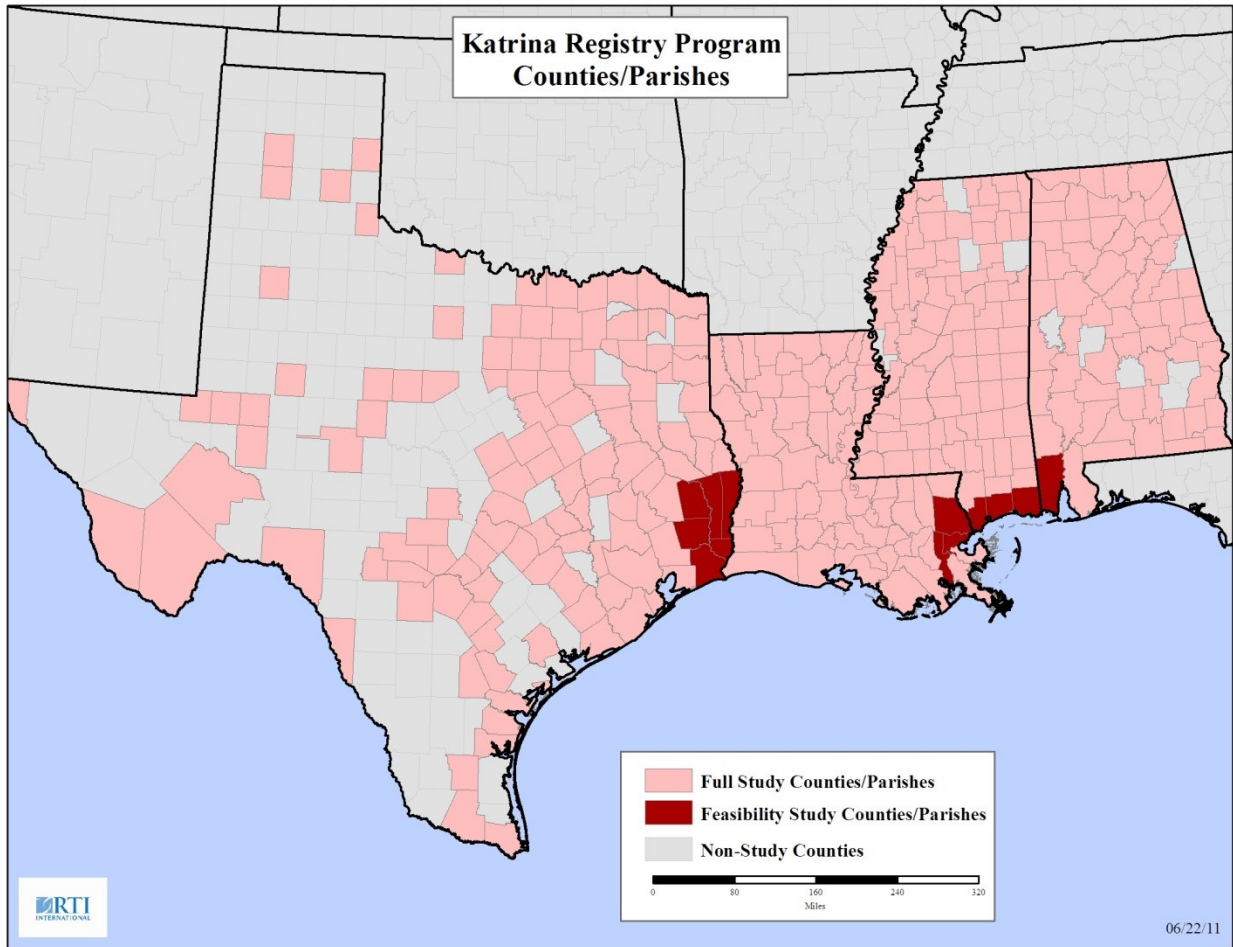
where  $n_h$  is the number of applicants selected for the sample in the  $h^{\text{th}}$  sampling stratum and  $N_h$  is the total number of applicants in the  $h^{\text{th}}$  sampling stratum. The design weight for an applicant will be the inverse of the applicant probability of selection. That is, the design weight for the  $i^{\text{th}}$  applicant in the  $h^{\text{th}}$  sampling stratum,  $d_{hi}$ , will be

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

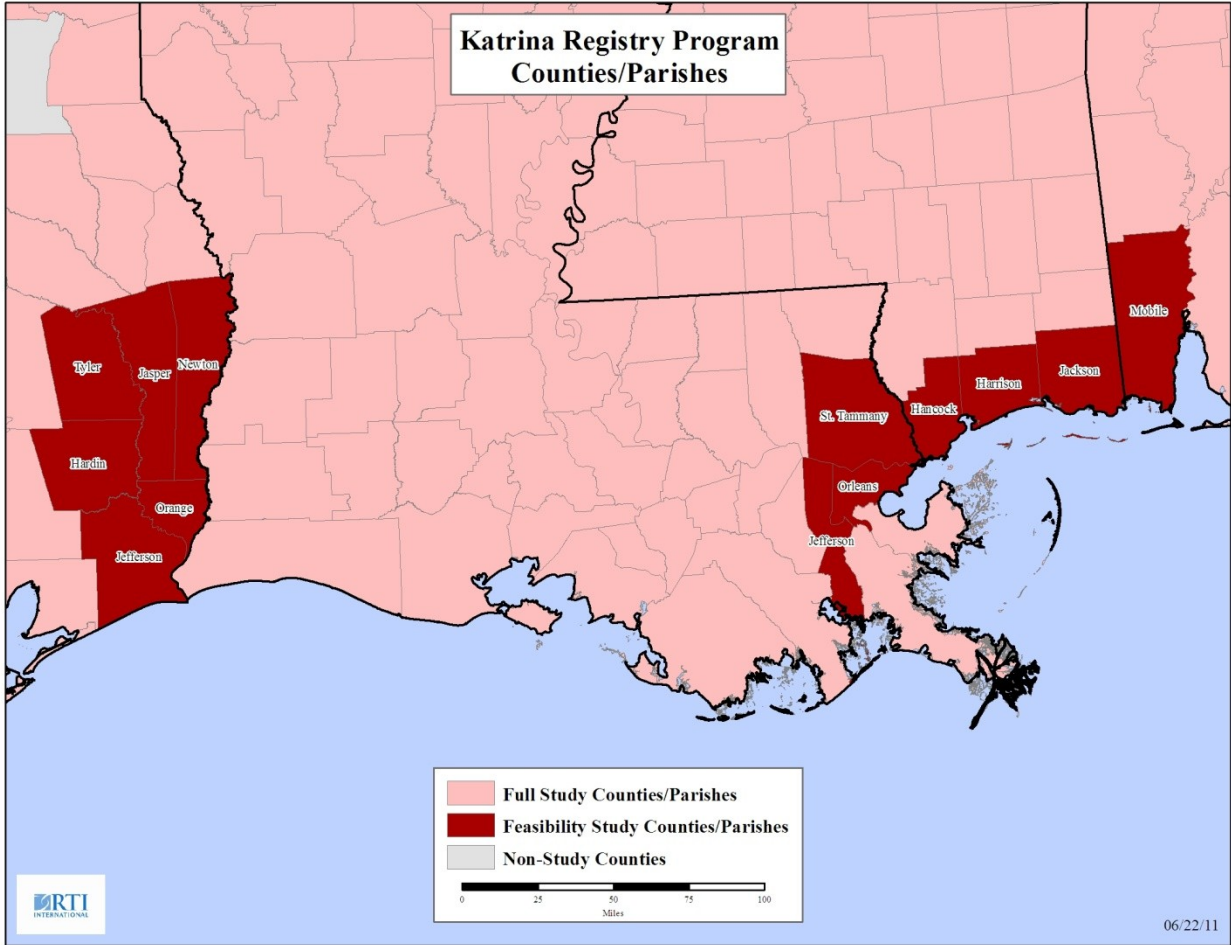
## Appendix A: Distribution of Applicants



Appendix B: Study Counties/Parishes



### Katrina Registry Program Counties/Parishes



**Appendix C: Feasibility Study Counties/Parishes Sample Allocation** (POS is the probability of selection and DW is the design weight.)

stateAlphaCode	County, State	Population Count		Sample Count	POS	DW
AL	Mobile, AL	1,788		200	0.1119	8.9400
AL	Total	1,788	% Sample	0.02002	200	
			Sample Count	200		
TX	Jefferson, TX	1,604		179	0.1118	8.9409
TX	Orange, TX	953		107	0.1118	8.9409
TX	Hardin, TX	522		58	0.1118	8.9409
TX	Jasper, TX	435		49	0.1118	8.9409
TX	Tyler, TX	245		27	0.1118	8.9409
TX	Newton, TX	175		20	0.1118	8.9409
TX	Total	3,934	% Sample	0.0440	440	
			Sample Count	440		
LA	Orleans, LA	24,239		2,714	0.1120	8.9310
LA	Jefferson, LA	19,504		2,184	0.1120	8.9310
LA	St. Tammany, LA	11,889		1,331	0.1120	8.9310
LA	Total	55,632	% Sample	0.6229	6,229	
			Sample Count	6,229		
MS	Harrison, MS	11,577		1,296	0.1120	8.9310
MS	Jackson, MS	8,928		1,000	0.1120	8.9310
MS	Hancock, MS	7,451		834	0.1120	8.9310
MS	Total	27,956	% Sample	0.3130	3,130	
			Sample Count	3,130		
All States	Total	89,310	Sample Size	10,000	9,999	
				Target	Actual	