

Survey to Evaluate the Effectiveness of Mississippi Delta Fish Advisories

0910-NEW

SUPPORTING STATEMENT

B. STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

(i) Respondent Universe

The target population for the survey includes the following four counties within the advisory area: Coahoma, Holmes, Leflore, and Washington. Only the part of Holmes County that is within the advisory area will be included in the survey. Counties were selected by the workgroup to include a mix of rural and non-rural areas and areas with major water resources affected by the advisory (e.g., Roebuck Lake, Moon Lake, Lake Washington, and Bee Lake). Table B.1 provides sociodemographic information on the four counties from the most recently available Census data.

The survey approach includes on-the-bank and household interviews in the four target counties. For the household survey, screening questions will be administered to identify sport and subsistence anglers who harvest wild-caught fish from the Mississippi Delta advisory area and individuals who consume wild-caught fish from the advisory area.

(ii) Sampling Selection

The sampling methods for the on-the-bank and household surveys are described below.

On-the-Bank Survey

The primary advantage of an on-the-bank survey or intercept survey is that it greatly increases the likelihood of reaching individuals fishing in the advisory area who possibly consume the contaminated species. Various approaches are available for conducting intercept surveys. For example, The [Marine Recreational Fisheries Statistics Survey](http://www.st.nmfs.noaa.gov/st1/recreational/overview/overview.html) (MRFSS) (<http://www.st.nmfs.noaa.gov/st1/recreational/overview/overview.html>), developed by the National Oceanic and Atmospheric Administration (NOAA) Fisheries Service to monitor the relative size and impact of saltwater recreational fisheries in the United States, is based on a sample of marinas and other registered sites which anglers use as points of departure and arrival for their boats. We propose a similar approach, albeit on a reduced scale, to capture as much as possible a representative sample of all anglers and their activities in this target area. The design we propose is probability-based and will allow us to draw inferences to the population of all anglers in the four-county survey area. A multi-stage sampling approach is proposed: first fishing access points will be selected, then days and time windows for each access point will be selected, and then anglers present at the selected fishing access points will be selected in the final stage. In some cases we will select one angler to interview in a group which would represent a fifth stage of selection.

Selection of Fishing Access Points

We will develop the sampling frame of fishing access points in the four counties using data available from the Mississippi Department of Wildlife, Fisheries, and Parks (MSDWFP) Web site and GIS data. The MSDWFP Web site identifies over 80 fishing access points in the affected area. For each fishing access point, we will have information on the type of water body (stream versus lake), dock location(s), GIS coordinates of the center point, length (for streams), and area (for lakes). We will work with local organizations to obtain estimates of the number of anglers per unit time (day, week, month) at each site. Approximately 20 fishing access points will be randomly selected to achieve a geographical spread across the four-county target area and to ensure that different types of fishing access points are included in the sample. In order to achieve this goal, we will stratify the population of fishing access points by geography and type water body (lake vs. stream) and proximity to an urban area.

Table B.1 Characteristics of the Four Counties Selected for the Survey

County	Major Cities	Population, 2006	Number of Housing Units, 2005	Land Area, 2000 (sq mi)	Persons per Square Mile, 2000	Race			Median Household Income, 2004
						White	Black	Hispanic	
Coahoma	Clarksdale, Jonestown	28,420	11,587	554	55	26%	73%	1%	\$23,560
Holmes	Durant, Lexington	20,866	8,725	756	29	19%	81%	1%	\$20,295
Leflore	Greenwood	35,752	14,612	592	64	29%	71%	2%	\$22,709
Washington	Greenville	58,007	24,958	724	87	32%	67%	1%	\$25,455

Source: U.S. Census Bureau. State and County Quick Facts. <http://quickfacts.census.gov/qfd/states/00000.html>.

Selection of Anglers

Approximately 400 anglers will be interviewed, 20 in each selected fishing access point. Assuming that (1) 90% of the individuals contacted will be eligible (i.e., aged 18 or older) and of these, 60% will agree to complete the survey, and (2) a reserve sample is needed in case the cooperation rate is lower than anticipated, the estimated starting sample will be 1,000 anglers. Thus, data collection would target a random sample of 50 anglers per fishing access point with the objective of obtaining 20 completed interviews at each access point.

Anglers will be sampled from the selected fishing access points. Access points include parking lots and other entrances to the fishing area where individuals arrive by car to put their boat into the water or to walk to the fishing area. If there are numerous entrances for a particular access point, it may be necessary to randomly sample the entrances for the data collection. The sampling plan will provide coverage across all days of the week, different times of days, and different types of access points. Data collection will require two interviewers. One interviewer will count the number of individuals who arrive and leave each access point during the sampling time window, and the second interviewer will conduct the interview with selected respondents. Interviewers will interview one adult in each group that arrives at an access point, where a “group” is defined as people traveling to the fishing site together. The adult in the group with the most recent birthday will be interviewed. A count of selected individuals who refuse to complete the survey will also be maintained.

Weighting

With the information collected at each access point, coupled with the site selection process, survey weights and inflation factors can be developed to adjust for deviations from sample design, such as variable nonresponse. The inflation factors can be used to infer estimates at the population level with measurable levels of sampling precision.

The average first stage selection probability, P_1 , can be represented as $P_1=20/A$ where A is the total number of fishing access points. The second stage selection probability, P_2 , can be represented as, for example, $P_2=1/10$, in the case that 10% of the days in the data collection period were sampled. The third stage selection probability, P_3 , would be $P_3=1/5$ in the case that 20% of the daily time period was sampled. Finally, the fourth stage selection probability, P_4 , can be represented as $P_4=50/B_i$ where B_i is the estimated number of anglers at access point i . The overall selection probability would be calculated as $P_b=P_1 * P_2 * P_3 * P_4 = [20 * 50] / [A * 10 * 5 * B_i] = 20AB_i$. The corresponding weight would be calculated as $1/P_b = 1/[20AB_i]$. (In some case, there will also be a fifth stage of selection with selection probabilities equal to $P_5=1/n_{ij}$, where n_{ij} is the number of anglers in group j at access point i .)

These selection probability weights will be adjusted to account for nonresponse since not all 50 anglers sampled at each site will respond. Thus, for example, if 50 are sampled and only 20 respond the value of P_4 will be $20/B_i$ instead of $50/B_i$. Furthermore, any population benchmark information will be used to poststratify the on-the-bank sample results to represent the entire population of anglers.

Household Survey

For the household survey, a stratified equal-probability two-stage design will be used employing an addressed-based sample. With this design, census tracts are selected in the first stage, and households within the selected tracts are selected in the second stage. The design will resort to probability-proportional-to-size sampling, which means that first-stage units are selected with probabilities proportional to some measure of size. For example, if the measure of size is total households, a tract with 100 households will have twice the probability of selection as a tract with 50 households. However, in the second stage, a fixed number of households will be selected, leading to overall equal probabilities of selection and evenly distributed interviewer workload over the selected census tracts. With this probability-based design, inferences can be made to the population of all occupied housing units in the four-county survey area. Additionally, the design will allow for comparisons among key sociodemographic variables such as gender and race.

Sample Size

The universe for the household survey consists of all occupied housing units in the four counties of Coahoma, Leflore, Holmes (advisory area only), and Washington. Approximately 600 interviews will be completed. The interviews will be allocated to each county in proportion to the total population for the advisory area (see Table B.2). The combined sample of 600 will result in a 95% confidence interval of (0.44, 0.56) for a proportion estimate of 0.50, assuming a design effect of 2. This level of precision will also result in considerable power for identifying differences between estimates.

Table B.2 Sample Allocation

County	Household Survey
Coahoma	117
Holmes ^a	84
Leflore	150
Washington	249
Total	600

^a For Holmes County, only the portion of the county in the advisory area was used for determining the sample allocation.

To ensure that sample size requirements are met, the required sample sizes will be adjusted upward to account for the anticipated eligibility and response rates. We assume that 75% of the individuals contacted will be eligible and that of these, 60% will agree to complete the survey. Thus, to complete 600 interviews, the starting sample size will be approximately 1,333 households.

First Selection Stage

In the first stage of selection, a total of 30 primary sampling units (PSUs), defined as tracts (i.e., units that contain approximately 500 households) will be sampled from the four

counties. These will be sampled with probabilities proportional to size, where size is an estimate of the total population or total number of households within each county for the advisory area.

Stratification

PSUs will be stratified by geography, access to water bodies of interest, and various demographic and socioeconomic characteristics to ensure a good spread of the sample across all segments of the population. The variables that will play a primary role in stratification will be geographic variables indicating the urban/rural nature of the PSU as well as its proximity to known bodies of water. PSUs will be classified by three levels of urbanicity: center-city, suburbs, rural and the proximity indicator will also have three codes: very distant (more than 10 miles), relatively close (5-10 miles), and very close (less than 5 miles). The combined nine-cell structure will serve as the first level of stratification. Within each of the nine combinations, PSUs will be further sorted by available census neighborhood characteristics such as proportion African American and proportion with a high school education.

Second Selection Stage

Once the 30 PSUs have been selected, a listing of all addresses in the selected PSUs will be made using commercial lists (e.g., Marketing Systems Group or Survey Sampling, Inc.). The list of all addresses in the PSUs is the sampling frame for the second selection stage.

The number of households to sample in each selected PSU will be 20. The households in each PSU will be sorted in geographical sequence and a sample taken, after a random start, of every n th household, where n is the result of dividing the total number of households in the PSU by 20. A simplified version of the “half-open-interval” (HOI) will be used to account for existing households that are not included on the list of addresses. With HOI, the interviewer is given the sampled address and the next address on the list. The interviewer is instructed to ensure that there are no housing units between the two addresses. If there are, a simple algorithm can be used to help decide whether the found housing unit should be interviewed.

Within Household Selection

Within each sampled household, it will be necessary to screen for eligibility and to interview one eligible household member. For a selected household to be eligible for the survey, it must meet the following inclusion criteria: (1) an adult household member has been fishing in the Delta in the past year or (2) a household member has consumed wild-caught fish from the Delta in the past year. If eligible, the interviewer will select an adult within the household to interview. For households that fish, the interviewer will select the adult who fishes most often; for households that consume fish, the interviewer will select the adult who prepares most of the wild-caught fish consumed by household members. For households that fish *and* consume fish, the interviewer will identify the adult who fishes most often and the adult who prepares most of the wild-caught fish, and then use the most recent birthday method to select the person to interview.

Weighting

Although there are unequal probabilities of selection at each stage, the overall two-stage probabilities of selection are equal:

$$p = (30 * MOS_i) / \sum MOS_i * (20 / MOS_i) = 600 / \sum MOS_i$$

where

p = the overall probability of selection, and

MOS = the measure of size for PSU_i (e.g., the total number of households in PSU_i).

However, the measures of size in the first stage are based on best available data, whereas in the second stage they are based on the number of addresses available. To the extent that the two differ, it will be necessary to make adjustments to the second stage sample size or use weighting adjustments. For weighting purposes, adjustments will need to be made for other deviations from the original design, for variable nonresponse, and for final poststratification adjustment to known population benchmarks. SUDAAN, a statistical analysis software tool, will be used to calculate estimates incorporating the complexity of the design, namely, the stratification and clustering. Using the recommended design, the design effect should not exceed 2 for most estimates.

Combining the Results for the On-the-Bank and Household Surveys

The proposed analysis approach is to prepare a combined analysis dataset with responses for both surveys and to develop survey weights for the combined dataset. The following information is needed to develop the combined survey weights: (1) whether the respondent is in the sampling frame for the on-the-bank survey; (2) whether the respondent is in the sampling frame for the household survey; and (3) whether the respondent is in the sampling frame for both surveys. By collecting information from the on-the-bank survey respondents on whether they live in the four-county area and by collecting information from the household survey respondents on whether they fish in the advisory area, we can determine whether the respondent is in one or both frames. With this information, the joint probability of selection can be computed and used to develop weights for the combined dataset. As an example, suppose P_a is the selection probability for the household frame and P_b is the selection probability for the on-the-bank survey, as calculated above. If the respondent is selected in the household frame and does not fish, then P_a would be used and that individual weighted as $1/P_a$. If the respondent is selected for the on-the-bank survey but does not live in the four-county area, then P_b would be used and that individual weighted as $1/P_b$. For respondents who belong to both frames, they would be assigned the selection probability $P_a + P_b - (2 * P_a * P_b)$, namely, the sum of the two probabilities minus twice the probability of belonging to both frames.

Both surveys target essentially the same population except that the household sample will also include non-anglers. These will have probabilities of P_a only since they will not be part of the on-the-bank survey. Each survey will be weighted to stand on its own as a representation of the total population in the case of the household survey and of the angler population in the case of the on-the-bank survey. The survey results will be analyzed separately for non-anglers,

household survey anglers, and on-the-bank survey anglers in order to compare key survey results. Major differences will be identified so these can be taken into consideration when studying the combined results. It is possible that the difference in method of survey administration will induce effects and influence the responses. Identifying and measuring such differences will be one objectives of the analysis plan.

2. Procedures for the Collection of Information

Before conducting the survey, a meeting will be held to brief community leaders about the survey. The purpose of this meeting is to describe the forthcoming survey to community leaders and help establish “buy-in” for the survey. The meeting will be held at Delta State University in Cleveland, Mississippi. FDA’s contractor and individuals from EPA and MDEQ will speak at this meeting.

Before starting full-scale data collection, a pretest of the OMB-approved questionnaire will be conducted following the procedures specified for the full-scale study. To conduct the pretests, 20 participants from the sample population (approximately 10 on-the-bank interviews and 10 household interviews) will be recruited and the questionnaire administered. The questionnaire and interview procedures will be revised, as necessary, based on the pretest findings.

Under the direction of Dr. John Green, Delta State University will conduct the survey. Delta State University will hire and train interviewers who are graduate students or who are local to the area. Using “locals” to conduct the interviews will facilitate access to the interview area and thus help maximize the response rate for the surveys. Additionally, these individuals will be familiar with the Delta culture and water resources. Interviewers will attend a training session before starting data collection. In addition to training on general interviewing methods and refusal/avoidance techniques, interviewers will participate in specialized training on conducting on-the-bank interviews and techniques for collecting data on fish consumption. As part of the training, interviewers will conduct mock interviews to practice administering the questionnaire so that they are familiar with following the skip patterns, selecting the correct response options, and using visual aids properly for some questions.

The procedures for selecting individuals to interview for the on-the-bank and household surveys are described below.

On-the-Bank Survey

No screening questions are necessary for the on-the-bank survey because any individual aged 18 years or older who is observed fishing is eligible for the survey. The interviewer will verify the respondent’s age if the interviewer is not sure if the respondent is at least 18 years old. The respondent selection procedures were described in Section B1(ii).

Using a survey approach similar to that employed by Burger and other researchers (Burger and Waishwell, 2001; Pflugh et al., 1999; and Campbell et al., 2002), interviewers will be assigned to certain water bodies and will be responsible for visiting a certain number of sites each day of the week, including Saturday and Sunday. This approach will help ensure the inclusion of the widest possible cross section of the angler population in the survey. Interviewers

will visit the different sites at least twice each field day to interview anglers who might arrive at different times during the course of the day.

Household Survey

Interviewers will be given a list of the selected households to interview. Within each sampled household it will be necessary to screen for eligibility, and if eligible, identify the household member to interview. The procedures for determining eligibility and selecting a household member to interview were described in Section B1(ii). A follow-up appointment will be scheduled if the selected individual is not available.

The Hispanic population is very small in the selected counties (between 1% and 2%); thus, it is not necessary to translate the questionnaire into Spanish and conduct interviews in Spanish.

3. Methods to Maximize Response Rates

To help ensure that the response rate is as high as possible, FDA's contractors—RTI and Delta State University—will employ all appropriate methods demonstrated in the research literature on survey methodology to maximize the response rate for the survey. These procedures include the following:

- Include community leaders and experts in Delta culture in the development of the questionnaire and survey approach;
- Design a questionnaire that minimizes respondent burden (short in length, written in easy-to-understand language, and uses visual aides to facilitate understanding);
- Test the draft questionnaire using cognitive interviews to ensure that respondents can properly understand the questions and that the response options are robust;
- Test the draft questionnaire in a pretest to ensure that it minimizes burden and refine as appropriate;
- Before data collection, conduct a meeting with community leaders to establish buy-in for the survey;
- Inform respondents that survey results will be available to the public;
- Use “locals” who are familiar with Delta water bodies and the culture to conduct the interviews;
- Train interviewers on refusal/avoidance techniques; and
- Use approaches suggested by experts in Delta culture to facilitate the interviewing process, such as matching interviewers and respondents based on race, having interviewers dress casual, and establishing rapport with respondents before administering the questionnaire.

4. Test of Procedures or Methods

Two sets of cognitive interviews were conducted to test the instrument for the survey. In May 2007, cognitive interviews were conducted with nine individuals. The interviewees

included five male and four female respondents and two Caucasian and seven African American respondents. Based on the cognitive interviews, revisions were made to the instrument, including clarifying difficult questions, deleting redundant questions, revising questions to make them more conversational in nature, and adding picture cards for some questions.

In February 2009, a second set of cognitive interviews was conducted with eight individuals using the revised version of the questionnaire. The interviewees included six male and two female respondents and two Caucasian and six African American respondents. Based on the cognitive interviews, the questionnaire was refined.

5. Individuals Involved in Statistical Consultation and Information Collection

The contractor, RTI, will collect the information and analyze the data on behalf of FDA as a task order under the Quick Turn Around Survey Data Collection contract. Ms. Sheryl Cates is the Project Leader for RTI (919-541-6810) and Dr. Karol Krotki (202-728-2485) is the senior statistician for the study. RTI will subcontract with Delta State University to conduct the interviews. Dr. John Green (662-846-4069) is the primary contact for Delta State University.

REFERENCES

- Burger, J., and L. Waishwell. 2001. "Are We Reaching the Target Audience? Evaluation of a Fish Fact Sheet." *The Science of the Total Environment* 277:77-86.
- Campbell, K.R., R.J. Dickey, R. Sexton, and J. Burger. 2002. "Fishing along the Clinch River Arm of Watts Bar Reservoir Adjacent to the Oak Ridge Reservation, TN: Behavior, Knowledge and Risk Perception." *The Science of the Total Environment* 299:145-161.
- Pflugh, K.K., L. Lurig, L.A. Von Hagen, S. Von Hagen, and J. Burger. 1999. "Urban Anglers' Perception of Risk from Contaminated Fish." *The Science of the Total Environment* 228:203-218.