# Supporting Statement for Continuing Collection RE: Visitor Survey for Cape Hatteras National Seashore
## OMB Control Number 1024-0258

## B. Collections of Information Employing Statistical Methods

The agency should be prepared to justify its decision not to use statistical methods in any case where such methods might reduce burden or improve accuracy of results.

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

The potential respondent universe for the visitor interviews includes all visitors 18 years of age or older using the ocean-side beaches at Cape Hatteras National Seashore (CAHA) from August 2009 through November 2009 (under the emergency clearance) and between December 2009 and July 2010 (under a regular PRA clearance). This includes visitors who use the ocean-side beaches between 6:00 a.m. and 6:00 p.m. Visitor groups will be intercepted based on a sampling plan (described in section B.2) stratified by season, location in the park, day of the week, and time of the day.

Table B1 presents the approximate visitation based on National Park Service (NPS) public use statistics for November 2007 to October 2008. It's assumed that the visitation patterns observed between November 2007 and October 2008 are representative of the visiting behavior expected for the study period. We expect that the large majority of visitors will be on the beach at some point between 6:00 a.m. and 8:00 p.m. For the visitor survey, we expect a response rate of 60%, based on a 74% response rate to a 2002 visitor survey at CAHA that involved a shorter on-site contact, followed by a mailback survey.

**Table B1. Potential Universe for the Visitor Survey.**

| Sampling months | Estimated Size of Respondent Universe | Sample Size | Expected Response Rate | Estimated Responses |
|---|---|---|---|---|
| ALL | 2,112,126 | 2,000 | 60% | 1,200 |

2. Describe the procedures for the collection of information including:
   * Statistical methodology for stratification and sample selection,
   * Estimation procedure,
   * Degree of accuracy needed for the purpose described in the justification,
   * Unusual problems requiring specialized sampling procedures, and
   * Any use of periodic (less frequent than annual) data collection cycles to reduce burden.

Visitors will be sampled on the ocean-side beaches. The sampling plan will be stratified by season, day of week, time of day, and type and location of beach. Scheduled sampling days will be based on data supplied by nine housing rental companies on Hatteras Island for occupancy rates by week for 2007 and counts of visitors on beach sections collected as part

of an on-going effort by RTI International to document baseline vehicle use and the number of visitors on different sections of the beach. The occupancy rates and counting data will be used to estimate the proportion of visitors (including pedestrians) on the beach each week of the year, and selected weeks will be sampled proportionately based on the occupancy data.

Weekends will be sampled at twice the rate of weekdays because information from park staff and from counts of vehicles using the beach indicate that there are about twice as many visitors on the weekend as on weekdays. Sampling days will be divided into three four-hour time periods: from 6 a.m. to 10 a.m., from 10 a.m. to 2 p.m., and from 2 p.m. to 6 p.m. Two four-hour periods will be randomly selected for each day on which the survey is conducted. The beaches will be divided into sections varying in length from $1/16^{th}$ mile to ½ mile (most of the sections will be $1/8^{th}$ mile), based on the expected number of groups on different sections of the beach. GPS coordinates will be available for each segment. The beach segments will be clustered according to location. One cluster of beach segments will be selected for each four-hour interview time period to allow for easy driving or walking between segments. Interviews will be conducted at three beach segments during each four-hour time period. The start time for surveying on each beach segment will be randomized across the four hours.

The visitor survey will be conducted by a team of two interviewers using handheld devices to administer in-person computer-assisted questionnaires. The process is as follows:
- Use GPS receivers to identify the beginning of the section of beach.
- Interviewer 1: Start by walking along the dune line and counting the number of groups of people on the beach and in the water until the end of the segment is reached.
- Based on a pilot test in August 2009, we calculated the number of surveys per hour the interviewers are likely to be able to complete. After counting the groups, the number of groups is entered into a formula in the handheld computer survey device.
- If the number of groups is less than or equal to the number of interviews that are expected to be completed every hour, then every group in the segment will be interviewed.
   i. Interviewer 2: start interviewing at the beginning of the segment
   ii. Interviewer 1: after counting, start interviewing from the end of the segment
- If there are more groups than can be interviewed in the allotted time, then the formula will tell the interviewers to interview every Nth group and randomly select a group between 1 and N to start with. The interviewers will walk in a serpentine pattern, starting at the dune and moving toward the water interviewing every Nth group. When the interviewers reach the water, they will return to the starting point and repeat the selection process until the time limit is reached.
- To interview:
   i. Approach a group;
   ii. Ask who is in the "group," including members who might be in the water;
   iii. Ask to interview the member of the group 18 years of age or older who has the next birthday;
   iv. Conduct the interview

- For safety reasons, do not interview individuals in moving vehicles, but do interview groups sitting near parked vehicles

I

Estimates will be produced using a design-based approach that incorporates the sampling weights to estimate the population parameters. If non-response is in an issue, then appropriate methods, such as weighting-class adjustments, will be used to reduce the potential bias induced by the non-response.

Approximately 25% of the 1/8-mile beach segments will be surveyed (about 60 miles of beach and 30 days of interviewing in four segments each day). We will be sampling the segments with replacement.

The design effect was calculated assuming that six beach segments will be selected at random for each sampling day using a stratified sampling design. It also was assumed that one adult member of every group of beach users in the segment at the time of the survey will be selected at random. This results in a "census of groups of beach users" in the selected beach segment.

The proposed 30 days of interviewing will yield 180 (30 x 6) beach segments (which result in clusters of beach users). The 180 beach segments will be allocated proportionately to the number of segments in the two strata (Stratum 1 = Hatteras and Bodie with 49 beach segments, and Stratum 2 = Ocracoke with 18 beach segments). After combining all this information, the proposed stratified random sample of 180 beach segments and census of groups within selected beach segments results in a design effect of 1.01

Note that a sample of size $n$ drawn using simple random sampling from a population will usually be more efficient (i.e., have smaller sampling errors) than a sample of the same size drawn by means of a sample of pre-existing clusters in the population. Stratification tends to reduce the sampling errors of the estimator; therefore, it tends to increase the efficiency of the sampling design.

The total number of samples is constrained not only by the available budget, but by consideration of the precision that can be expected. Approximate precision estimates may be obtained, assuming the survey design is simple random sampling. Under this assumption, the confidence interval half-width (precision) can be estimated using the following equation:

$$Half - width = Z_{1-\alpha} * \sqrt{\text{estimator variance}}$$

where $Z_{1-\alpha}$ is related to the level of confidence required for the estimate (Cochran, 1987). Given that our proposed survey design considers stratification to account for the different types of beaches, the actual precision estimates are expected to be better (smaller confidence intervals) than those obtained with the equation above.

The visitor survey has two objectives: 1) to estimate the percent of the population that will change visit behavior under a proposed alternative, and 2) to estimate the mean number of days in a trip under a proposed alternative. (Trip length has relevant economic consequences.)

For the first objective, if the true proportion of the population estimated to change its behavior under a proposed alternative is 40%, then to produce estimates within 10% of this true value (95% confidence level, 80% power), we need about 377 participants (Table B2).

**Table B2. Sample Sizes Required to Estimate True Proportions at Specific Power Levels and Confidence Intervals.**

| True Proportion | Precision (Half-width) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 | 0.1 | 0.11 | 0.12 | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 | 0.19 | 0.2 |
| 0.10 | 566 | 393 | 289 | 221 | 175 | 142 | 117 | 99 | 84 | 73 | 63 | 56 | 49 | 44 | 40 | 36 |
| 0.15 | 801 | 556 | 409 | 313 | 248 | 201 | 166 | 139 | 119 | 103 | 89 | 79 | 70 | 62 | 56 | 51 |
| 0.20 | 1005 | 698 | 513 | 393 | 311 | 252 | 208 | 175 | 149 | 129 | 112 | 99 | 87 | 78 | 70 | 63 |
| 0.25 | 1178 | 818 | 601 | 460 | 364 | 295 | 244 | 205 | 175 | 151 | 131 | 115 | 102 | 91 | 82 | 74 |
| 0.30 | 1319 | 916 | 673 | 516 | 407 | 330 | 273 | 229 | 196 | 169 | 147 | 129 | 115 | 102 | 92 | 83 |
| 0.35 | 1429 | 993 | 729 | 559 | 441 | 358 | 296 | 249 | 212 | 183 | 159 | 140 | 124 | 111 | 99 | 90 |
| 0.40 | 1507 | 1047 | 769 | 589 | 466 | **377** | 312 | 262 | 223 | 193 | 168 | 148 | 131 | 117 | 105 | 95 |
| 0.45 | 1555 | 1080 | 793 | 608 | 480 | 389 | 322 | 270 | 230 | 199 | 173 | 152 | 135 | 120 | 108 | 98 |
| 0.50 | 1570 | 1091 | 801 | 614 | 485 | 393 | 325 | 273 | 233 | 201 | 175 | 154 | 136 | 122 | 109 | 99 |

From Cochran W.G. 1987. *Sampling Techniques*. New York. John Wiley & Sons

For the second objective—to estimate the mean number of days in a trip under a proposed alternative—we can use an estimate based on the coefficient of variation (CV), which is the standard deviation divided by the mean, i.e., the standard deviation as a percent of the mean. The smaller the standard deviation as a percent of the mean, the lower the CV. If we expect high variability in responses (i.e., the range in number of days varies substantially in the population), then we should be considering population CVs in the range of 75% to 125%. This results in samples that vary in size from 9 to 1539, depending on the desired precision (Table B3).

At CAHA, the final sample size should be the larger of the two sample sizes that satisfies both the first and second objectives of the visitor survey. This means the sample size for the visitor survey should be closer to 1,539 than to 377.

**Table B3. Sample Sizes Required to Estimate a Parameter at Desired Precision Levels and Coefficients of Variation.**

| Population CV | Percent Relative Error of the Estimated Mean | | | | | |
|---|---|---|---|---|---|---|
| | 5% | 10% | 15% | 25% | 50% | 75% |
| 25% | 98 | 26 | 13 | 6 | 3 | 3 |
| 50% | 387 | 98 | 45 | 18 | 6 | 4 |
| 75% | 867 | 219 | 98 | 37 | 11 | 6 |
| 100% | **1539** | 387 | 173 | 64 | 18 | 9 |
| 125% | 2403 | 603 | 269 | 98 | 26 | 13 |
| 150% | 3460 | 867 | 387 | 141 | 37 | 18 |

Confidence level = 0.95

**3.** Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

Data from the survey will be used to inform the economic analysis of the ORV management alternatives. The analysis will focus on the economic impacts of the alternatives relative to the baseline conditions under the two no-action alternatives (years 2007 and 2008). In order to analyze the new alternatives, we must select parameters to quantify changes in factors that affect visitors (such as ORV restrictions). Because of the uncertainty inherent in forecasting future activity, the economic analysis will estimate scenarios covering a range of possible impacts.

The data from the visitor survey is one source of information that will be used to establish the range of possible values for different parameters. Other sources include publicly available secondary data on visitation and economic conditions. The visitor survey and other data sources will be combined to create reasonable ranges for the parameters in the economic model. We plan to use several sources of information, because no single source provides all the data we need, and each source has positive and negative features. To reduce the concern over the possibility that visitors will answer the survey strategically to influence the outcome of the economic analysis, the questions in the survey and comparisons with other data sources will be used to identify strategic responses.

To maximize response rates to the visitor survey, the instrument was based on input from a number of parties with survey research experience and members of the public and park staff familiar with CAHA visitors. This input also was used to make sure the questions could be understood, the minimum numbers of questions were asked, and the questions collected the data needed for the analysis. The interviewers will be carefully trained in how to sample and make contact with visitors and how to convert initial refusals into completed interviews. Non-response rates will be monitored throughout the year of data collection to identify and correct problems.

We will take several steps to identify non-response bias and correct for bias in our analysis. The interviewer will attempt to ask refusers their Zip code or country of residence and also ask (or observe) if the visitor has ever driven a vehicle on the beaches at CAHA. In addition, any easily observable traits of the visitors (such as group size and presence of children), as well as time of day, day of week, and location of interview will be recorded on a log for each contact for a later comparison of respondents with non-respondents.. This information, along with a code for the interviewer, will be analyzed using a probit or logit model to identify systematic patterns in non-response. In any regression analysis, a two-step selection model (such as a Heckman selection model or propensity model) will be used to help correct regression results for selection bias.

**4.** Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of tests may be submitted for approval separately or in combination with the main collection of information.

The survey instrument was developed with input from contractor staff, including economists, survey research experts, and statisticians. National Park Service staff, including staff from the Social Science Division, and members of the Regulatory Negotiating committee for CAHA, also were consulted about the survey and survey administration. A scoping trip to CAHA took place in October 2008 to identify beaches and beach uses. This information was used to assess the feasibility of different sampling methods. A pre-test took place in December and January involving 9 visitors to evaluate and improve the procedures and the survey instrument.  Finally, a pilot test was undertaken in August 2009 under the emergency clearance to test the final procedures and instrument.

**5.** Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

RTI International, a contractor to NPS, designed the survey and will collect and analyze the data. The project manager is Dr. Carol Mansfield (919-541-8053). Other RTI staff participating in the survey design and data collection are Dr. Brenda Muñoz, Mr. Brian Evans, and Dr. Christine Poulos.