

# **Supporting Statement for Drunk Driver Segmentation Research Plan**

## **Section B**

**December 23, 2015**

**Contents**

B. Collections of Information Employing Statistical Methods.....3

    B1. Describe the potential respondent universe and any sampling or other respondent selection method to be used. Indicate the expected response rates for the collection as a whole.....3

    B2. Describe the procedures for the collection of information.....6

    B3. Describe methods to maximize response rates and to deal with issues of non-response.....8

    B4. Describe any tests of procedures or methods to be undertaken.....8

    B5. 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.....9

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

NHTSA is seeking approval to conduct a Web-based survey to collect information that subsequently will be subjected to factor analysis and cluster analysis methods in order to generate a descriptive set of segments of vehicle drivers (and motorcycle riders) that drive ride while drunk. By identifying the segments, NHTSA will have a pragmatic and cost-effective means to better target and reach intended drunk drivers/riders with messages and communications techniques that are relevant and meaningful. Target market segmentation profiles (such as the type desired by NHTSA) have been proven useful and effective by companies and non-profit organizations throughout the United States for marketing, communications, sales, product/service development, and customer service. This type of a marketing segmentation approach is especially vital, given that NHTSA's communications messages about drunk driving/riding must "compete" for audience attention in the public domain among hundreds of other major marketers, including those in the alcoholic beverage industry.

In short, the objectives guiding this survey are as follows:

1. Gather data on demographics, drinking behaviors, attitudes about drinking and driving, lifestyle characteristics and media use habits of drivers/riders who drive/ride drunk.
2. From the collected data, perform cluster analyses to identify and descriptively define unique similarities and differences in the form of segment profiles. Such segmentation profiles provide NHTSA's Office of Communications and Consumer Information (OCCI) staff a pragmatic and cost-effective means to better target and reach intended communication audiences with messages and communications techniques that are relevant and meaningful to people within the target market.

This study will use an opt-in online panel, and thus a convenience sample. This type of sample is sufficient to identify relationships among variables that occur within a population, which is the core objective of the study. The study will not create population estimates.

### **B1. Describe the potential respondent universe and any sampling or other respondent selection method to be used. Indicate the expected response rates for the collection as a whole.**

NHTSA will utilize a convenience sample managed by Lightspeed GMI, a top-rated marketing research online survey panel provider. The panel for the data collection consists of more than two million members who have previously opted-in to receiving invitations to respond to Web-based research surveys.

*Lightspeed GMI, headquartered in New Jersey, has provided email research samples in the U.S. since 1999. In 2014, Lightspeed panelists completed more than 30 million surveys. The company provides digital data collection services to a client base largely composed of full-service market research firms, marketing communications/ advertising agencies, and research consultancies. To recruit its panel members, Lightspeed uses numerous methodologies to generate interest and opt-in registrations. These recruitment methodologies include e-newsletter campaigns; email lead generation blasts, traditional Internet banner ad placements, social media postings and ad placement, and referrals by existing panel members. To maintain quality control and accuracy of the panel, Lightspeed uses a set of proprietary internal tools to validate quality, prevent fraudulent respondents from joining the panel, remove over-reporters, eliminate duplicates and better ensure panelist engagement. Additionally, they deploy systems for each survey to monitor and disqualify respondents who exhibit questionable response behavior such as respondents who*

*rush through a survey too quickly according to pre-established norms, who rush through grid questions and/or “straight-line” scaling questions, and/or whose responses are illogical, inconsistent or contradictory. Lightspeed’s email invitations and surveys are CAN-SPAM compliant.*

From the Lightspeed panel, survey invitations will be distributed to 220,000 panel members. (Justification regarding 220,000 is provided in the sub-section “Expected Response Rates” starting on page 5). These 220,000 members will be randomly drawn from the total Lightspeed panel in numbers that approximate U.S. Census statistics for demographics. The Census statistic referencing is done to ensure that the 220,000 invitees to the survey will not be demographically skewed in a way that negatively affects expected response rates, qualification rates through the study screener, and survey completion rates. Additionally, this method aids in ensuring the final study sample includes a mix of different demographic groups among the qualified participant population, avoiding demographic skews that may affect interpretation of survey response.

The study’s screening questions will be used to qualify respondents and only those qualified will be presented the full survey. (Note: If not qualified, the respondent will be thanked, and the survey at that point will terminate).

Unique classification logic developed by NHTSA will be used for the screening. The classification logic considers these characteristics:

Participants will be screened to ensure they meet these criteria:

- Age 21 to 54; and
- Currently possess a valid driver’s license for a motor vehicle and/or motorcycle; and
- Drive/ride, on average, 10 or more miles per week; and
- Consume alcoholic beverages multiple times per month; and
- Within the past month, during at least one occasion, consumed four or more drinks (males)/3 or more drinks (females)

NOTE: The age group of 21-54 is of significant importance to NHTSA for this study because this age group constitutes the greatest number of alcohol-related driving/riding fatalities according to USDOT’s Fatality Analysis Reporting System (FARS).

Additionally, the logic considers - -

- Self-reported (within the past month) largest number of alcoholic beverages consumed per hour away from home during a single occasion; accepted respondents must indicate 1.5 or more; and/or
- Particular mode of travel home after that consumption; one of five particular modes must be answered; and/or
- Level of agreement (using a Likert-type scale) to seven particular statements about drinking and driving/riding; and/or
- Past history of being stopped by police for drunk driving/riding; must provide affirmative answer or “prefer not to answer” response.

These criteria will also be applied to ensure - -

- n=1,600 participants who primarily drive a vehicle and n=600 people who primarily ride a motorcycle
- Adequate numbers of low-income (less than \$25,000) respondents and/or respondents of Hispanic origin are included among respondents who are qualified to complete the full survey (goals of 21% and 20%, respectively, regardless of whether a person meets just one or both criteria)

Because of the extent of the screening – plus the method of discernment to identify drunk driving/riding people – the ultimate sample of 2,200 respondents to the full survey will not, in any way, mirror or statistically represent the general U.S. population. The final report’s findings, statistics and comparisons will be relative only to people projected to be drunk drivers/riders.

NHTSA projects that 10% of panel members that receive the invitation will open it (22,000) and 10% of those who open the invitation (2,200) will go through the screener and complete the survey (see Expected Response Rates section for details). The 2,200 respondents will be comprised of 1,600 at risk drivers (of cars, trucks, vans, SUVs, etc.) and an additional 600 at risk motorcycle riders.

While there is no generally accepted rule of thumb for determining appropriate sample size for cluster analyses of this sort, the proposed number of completions is sufficient to identify multiple market segments as found by the earlier GHSA survey referenced in Section A.

Because the survey is conducted using a panel that is a convenience sample, it cannot produce estimates that are projectable to the target population with known levels of confidence. However, the purpose of this project is to assess the relationship of different characteristics related to alcohol-impaired driving/riding in order to identify market segments. In doing so, it is drawing from a massive subject pool expected to contain all pertinent variables found within the target population. The assessment of relationships within a large-scale inclusive sample pool of this type will provide adequate guidance to identify market segments.

#### Expected Response Rates

The panel provider reports that, based on actual field experience for its thousands of surveys, NHTSA can expect approximately 10% of the members receiving the survey invitation will open and start the survey. The percentage of those who start the survey but are ruled ineligible can only be very roughly estimated because there has not previously been a survey administered having the detailed screening criteria that are being used for this survey. In order to generate the estimated percentage of those who enter the screening questions that are ruled eligible to fill out the survey, an initial assumption must be made that the sample entering the screener is similar to the general population age 21 through 54. This is because there are no data available on differential response to survey requests to panel members to use in the computations. Given that assumption, estimates for the percentage of respondents that drop out due to the screener criteria are drawn as best as possible from general population surveys. Perhaps the closest that one can come to accounting for the detailed screening criteria are to use data from NHTSA’s 2008 National Survey of Drinking and Driving Attitudes and Behavior:

<http://www.nhtsa.gov/Driving+Safety/Occupant+Protection/2008+National+Survey+of+Drinking+and+Driving+Attitudes+and+Behaviors> .

That survey provides information on driving frequency, past year drinking frequency, usual number of drinks consumed, and selected behaviors to prevent alcohol-impaired driving. Based on the survey population age 21 through 54, only 42% are likely to indicate that they usually consume alcohol more than once a month, which is one of the screening criteria. A substantial portion of these will subsequently be ruled ineligible based on the largest number of drinks they consumed at one time in the past month, with males needing to exceed three drinks and females needing to exceed two drinks. While there is not an exactly comparable question in the 2008 survey, that survey did ask the usual number of drinks consumed. The response suggests that at least 70% of the 42% that drink more than once a month will drop out from the number of drinks question, leaving about 13%. Additional screening criteria will cause others to drop out: driving on average less than ten miles a week, engaging in activities to avoid impaired driving prior to consuming multiple drinks while away from home, and providing no indication of at-risk attitudes or behaviors. This would likely drop the percentage that is ruled eligible for the survey based on the screening questions to between 5% and 10%. This is NHTSA's best estimate. For purposes of burden calculation, we will assume 10%. Thus NHTSA proposes to send invitations to participate in the survey to 220,000 panel members, 10% of these (22,000) will start the survey screener, and 10% of those (2,200) will be determined eligible based on the screening questions and will proceed to take the full survey.

## **B2. Describe the procedures for the collection of information**

The procedure for the collection of information for this study is as follows:

- Survey will be sent via Internet to 220,000 members of a convenience sample/online panel.
- The screener portion of the questionnaire will take approximately four minutes to complete. All 220,000 panel members who are sent the survey invitation have an equal chance of receiving it. However, only those who open the survey, respond to the screening questions and who subsequently qualify (based on screening criteria) will have the opportunity to complete the main portion of the survey. The screening criteria have been described starting on page 4 of this document.
- Those who meet the screening criteria will be served the main portion of the questionnaire. The respondents who are deemed ineligible via the screening questions will be thanked for their participation at the end of the screening portion, and they will then be terminated from further survey questions.
- As noted in other sections of this document, 2,200 respondents will complete the full survey. Respondents will complete the survey on their own time, and they will have an option to temporarily stop, then resume responding at a later time (within seven days). They will also have the option to contact the panel provider if they experience any technical problems with the survey.
- Included in the above numbers will be a pilot test (or "soft launch") to collect 5% of the total sample (that is, 110 completed surveys) to ensure survey programming, data collection and reporting are performing as intended. With the expectation that this test will show affirmation that the survey is performing properly, the responses collected during this test will be included as part of the total sample of 2,200.

## Data Analysis Plan

The prevailing goal of the study is to identify and describe segments of those deemed “at risk.” To do so, descriptive analyses, factor analysis, and k-means cluster modeling will be applied to the 2,200 “at risk” participants’ responses.

For the basis of the descriptive data analysis, data tabulation sets will be developed, based on a tab plan customized to the final screener and questionnaire. Tabulations will each feature up to twenty-one (21) banner points. This number of banner points, along with their detailed statistical values, is expected to allow for descriptive comparative analyses of responses from sub-groupings of the respondents based on behavioral, attitudinal, and demographic similarities/differences. The individual tabulations will include stubs for all closed-ended data points in the survey, means for Likert and Semantic Differential scale ratings frequency and percentage responses.

Factor analysis will be conducted utilizing SPSS (IBM Statistics) software to examine correlations within sets of variables (across forty-five (45) potential segment defining variables total) reflecting respondents’ self-reported drinking attitudes and behaviors, drinking and driving attitudes and behaviors, social behaviors, and personalities. The identification of correlated variables will aid in selection of variables that are relatively strong in explaining ways in which respondents may cluster. Variables with higher factor loadings from correlated combinations will be considered in cluster modeling.

The cluster modeling will also be conducted utilizing SPSS (IBM Statistics) software using k-means clustering. Cluster analyses will be conducted using variables selected after factor analysis and also informed by descriptive analysis of response to related survey questions. These analyses will vary in the number and selection of variables that will define emerging clusters. It is expected that approximately nine to fifteen (9-15) variables from the set of forty-five (45) potential segment defining variables will be successful in identifying viable cluster models for interpretation. Several cluster modeling outcomes will be examined with four to six (4-6) clusters per model. A cluster model and segmentation approach will be selected and the related clusters will be represented in an additional set of data tabulations for comparative descriptive analysis of responses to survey questions beyond the final defining metrics.

Additionally, thorough quality control reviews will be done for all data files, data tables, cross tabulations and other results used in the study.

### **B3. Describe methods to maximize response rates and to deal with issues of non-response.**

The panel that NHTSA plans to use consists of people who have previously opted-in to receiving and responding to Web-based surveys. Those who choose to participate will be incentivized, as noted in Part A (section A9) of this Supporting Statement. They are attuned and receptive to survey invitations.

Until the project’s sample goals are achieved, a reminder email message will be sent up to three times to each panel member not starting the survey.

### **B4. Describe any tests of procedures or methods to be undertaken.**

NHTSA and its contractors will conduct a test of the online survey to ensure all functions are working properly before the study is sent to potential respondents.

The survey will then experience a ‘soft launch’ period in which a limited number of potential respondents (approximately 5% of the sample) are sent invitations to respond. During this period, we will evaluate the demographic data, completion rate, and time-spent in the survey. Any problems discovered during this test period will be addressed before the survey is sent out to a larger number of panel members.

**B5. 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.**

The company selected as a contractor for this study is The Tombras Group. The Tombras Group is a marketing communications and advertising consultancy for NHTSA’s Office of Communications & Consumer Information. For this study, the contractor has selected W5, Inc., as its research sub-contractor. W5 is a custom marketing research firm with extensive experience in segmentation research and analysis for U.S. companies/manufacturers, brand marketers, advertising agencies and non-profit organizations. This team has extensive experience in both qualitative and quantitative research practices, and has consulted with NHTSA on all aspects of the design. The following individuals have reviewed technical and statistical aspects of procedures that will be used to conduct the NHTSA Drunk Driver Segmentation Study:

Andrew Willard  
Practice Principal  
W5  
3211 Shannon Road, Suite 610  
Durham, NC 27707  
(919) 932-1117

Steve Richardson  
Senior Vice President/Account Planning  
The Tombras Group  
630 Concord Street  
Knoxville, TN 37919  
(865) 524-5376, ext. 1232

Alan Block, MA  
Office of Behavioral Safety Research  
DOT/National Highway Traffic Safety Administration  
400 Seventh Street, SW  
Washington, DC 20590  
(202) 366-6401