Information Collection Request Supporting Statement: Part B
National Survey of Drowsy Driving Knowledge, Attitudes and Behaviors

Abstract:[1] The National Highway Traffic Safety Administration (NHTSA) of the U.S. Department of Transportation is seeking approval to collect information from a random sample of adults (18 years or older) who have driven a motor vehicle in the past month for a one-time voluntary survey to report their knowledge, attitudes, and behaviors associated with drowsy driving. This collection has two parts. The first part is a pilot test for which NHTSA will contact 1,000 households for an expected number of 163 voluntary responses. The second part is the full survey for which NHTSA will contact 81,490 households to achieve a total target of at least 15,000 completed voluntary responses, consisting of 7,000 completed instruments from a nationally representative sample and 2,000 completed instruments from each of four samples representative of States that recently have had drowsy driving law or program activities (Arkansas, Iowa, Massachusetts, and New Jersey). The total estimated burden associated with this collection is 16,323 hours – up to 10,949 hours associated with survey invitations and reminders and up to 5,374 hours associated with completing the survey. NHTSA will summarize the results of the collection using aggregate statistics in a final report to be distributed to NHTSA program and regional offices, State Highway Safety Offices, and other traffic safety stakeholders. This collection will inform the development of countermeasures, particularly in the areas of communications and outreach, for reducing fatalities, injuries and crashes associated with drowsy driving.

B.1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods 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. Response rate means -- of those in your respondent sample, from what percentage do you expect to get the required information (if this is not a mandatory collection). The non-respondents would include those you could not contact, as well as those you contacted but who refused to give the information.

For both the pilot study and data collection, the potential respondent universe includes an address-based sampling frame (ABS) for the entire U.S. population, oversampled for likely young (under age 30) and non-white respondents. Four additional states (Arkansas,

[1] The Abstract must include the following information: (1) whether responding to the collection is mandatory, voluntary, or required to obtain or retain a benefit; (2) a description of the entities who must respond; (3) whether the collection is reporting (indicate if a survey), recordkeeping, and/or disclosure; (4) the frequency of the collection (e.g., bi-annual, annual, monthly, weekly, as needed); (5) a description of the information that would be reported, maintained in records, or disclosed; (6) a description of who would receive the information; (7) the purpose of the collection; and (8) if a revision, a description of the revision and the change in burden.

New Jersey, Massachusetts, and Iowa) that either have existing drowsy driving laws or have or recently had active public information and education campaigns will also be included with additional sampling to ensure representativeness. Based on corporate experience, we anticipate a response rate of 20.4% among households with an eligible respondent.

First, we will generate a sample of residential addresses currently in use using the USPS computerized Delivery Sequence File DSF. Marketing Systems Group (MSG) will generate this sample, as it has for scores of studies for government entities over the past decade. MSG will not only draw the sample, but also augment drop units, simplified addresses and other addresses to be full "city-style," to ensure near complete coverage of all addresses in the United States. The file will contain home and apartment addresses, as well as P.O. boxes (OWGMs, only way to get mail) and all other types of residential addresses. MSG will not sample seasonal or vacant addresses, or non-OWGM P.O. boxes.

The first step in creating this design is to identify households that have a higher likelihood of being Hispanic and African-American households. This is done using census data to target households that live in census block groups which contain 50% or higher density of Hispanic households or 20% or higher density of African-American households. The 50% density Hispanic and 20% density African-American census block groups were chosen based on an estimation of how best to achieve the highest incidence while also minimizing the overall design effect. These strata are represented in the table below as "Geographic Strata" showing the number of households that are "High Hispanic" (from a block group that is 50%+ Hispanic residents), "High African American" (from a block group that is 20%+ African-American residents), or "Residual" (all others). Overall, in the general population, the density of Hispanic persons is 18.2% and of African Americans is 12.1%. Among the high Hispanic strata, 75% are Hispanic, and among the high African American strata, 48% are African American, indicating the increased likelihood of reaching a household with this characteristic.

Second, we leverage listed database information, from Experian and other databases, signaling the presence of persons of specific ages to the file. For the purposes of creating a stratified design, these data are appended to a national database of households aggregated at the census block group level. These strata are represented in the table below as "Age Listed Strata" showing the number of households that are "Listed, Young 18-29" (a listed address with data indicating there is at least one household member age 18-29), "Listed, Old 30+" (a listed address with data indicating that all household members are age 30+), "Listed, no age info" (a listed address, with no data available on the age of the household), and "Not listed" (addresses that are not listed).

Combining region, geographic strata and the age strata together leads to a 4 region x 3 geographic level x 4 age listed design for 48 strata in total that are detailed in Table 1 and 2. These data will be updated based on the latest census information at the time of fielding. For within-household selection, M. Davis and Company, Inc. (MDAC) will follow the recommendation of Dr. Paul J. Lavrakas as outlined in his report, "Within-Household Respondent Selection: How Best to Reduce Total Survey Error?"[2] Of the persons within the household who have driven a motor vehicle within the past month, half of the distributed surveys will ask for the person (18 years or older) with the most recent birthday to complete the survey and half will ask the person (18 years or older) with the next coming birthday to do so.

Table 1: National Sampling Plan for Pilot Study

Region	Geographic Strata[3]	Age Listed Strata[4]	U.S. Households[5]	% of U.S. Households[6]	RSF[7]	Sample[8]	Intervie
Northeast	High Hispanic	Listed, Young 18-29	58,436	0.05%	10.00	0.23%	2
Northeast	High Hispanic	Listed, Old 30+	437,149	0.35%	1.00	0.17%	2
Northeast	High Hispanic	Listed, no age info	459,375	0.37%	3.00	0.55%	6
Northeast	High Hispanic	Not listed	327,019	0.26%	0.75	0.10%	1
Northeast	High African American	Listed, Young 18-29	143,219	0.11%	10.00	0.57%	6
Northeast	High African American	Listed, Old 30+	1,199,945	0.96%	1.00	0.48%	5
Northeast	High African American	Listed, no age info	980,808	0.79%	3.00	1.17%	12
Northeast	High African American	Not listed	563,387	0.45%	0.75	0.17%	2
Northeast	Residual	Listed, Young 18-29	701,612	0.56%	10.00	2.79%	28
Northeast	Residual	Listed, Old 30+	9,053,026	7.25%	1.00	3.60%	36
Northeast	Residual	Listed, no age info	5,533,854	4.43%	3.00	6.60%	66
Northeast	Residual	Not listed	2,187,370	1.75%	0.50	0.43%	4
Midwest	High Hispanic	Listed, Young 18-29	27,470	0.02%	10.00	0.11%	1
Midwest	High Hispanic	Listed, Old 30+	204,652	0.16%	1.00	0.08%	1
Midwest	High Hispanic	Listed, no age info	173,062	0.14%	3.00	0.21%	2
Midwest	High Hispanic	Not listed	105,885	0.08%	0.75	0.03%	0

[2] Lavrakas, P. (2008) Within-Household Respondent Selection: How Best to Reduce Total Survey Error? Prepared for: Media Rating Council, Inc. Retrieved from : http://mediaratingcouncil.org/MRC%20Point%20of%20View%20-%20Within%20HH%20Respondent%20Selection%20Methods.pdf

[3] Geographic Strata indicates the number of households that are in block groups which are either "High Hispanic", "High African American", or all remaining households.

[4] Age Listed Strata indicates the number of households within each Geographic Strata that are identified in the Experian database as having at least one adult age 18 to 29, only adults over 30, listed but with no age information, or does not have listed information from the Experian database.

[5] Total number of households in the United States that falls into each stratum.

[6] Percent of all households that fall into each stratum.

[7] Relative sampling fraction. This is the sampling rate of each stratum relative to the sampling rates of other strata. For example, if stratum 1 has a relative sampling fraction of 3 and stratum 2 has a relative sampling fraction of 4, then stratum 1 is being sampled at ¾ the rate that stratum 2 is being sampled.

[8] Percent of each stratum that will be included in the final sample design.

[9] Estimated number of final interviews obtained from each stratum.

Midwest	High African American	Listed, Young 18-29	183,176	0.15%	10.00	0.73%	7
Midwest	High African American	Listed, Old 30+	1,645,460	1.32%	1.00	0.65%	7
Midwest	High African American	Listed, no age info	1,304,331	1.04%	3.00	1.55%	16
Midwest	High African American	Not listed	585,363	0.47%	0.75	0.17%	2
Midwest	Residual	Listed, Young 18-29	932,860	0.75%	10.00	3.71%	37
Midwest	Residual	Listed, Old 30+	12,438,967	9.96%	1.00	4.94%	49
Midwest	Residual	Listed, no age info	7,343,136	5.88%	3.00	8.75%	88
Midwest	Residual	Not listed	2,327,593	1.86%	0.50	0.46%	5
South	High Hispanic	Listed, Young 18-29	212,519	0.17%	10.00	0.84%	8
South	High Hispanic	Listed, Old 30+	1,530,210	1.23%	1.00	0.61%	6
South	High Hispanic	Listed, no age info	1,665,438	1.33%	3.00	1.99%	20
South	High Hispanic	Not listed	622,073	0.50%	0.75	0.19%	2
South	High African American	Listed, Young 18-29	648,919	0.52%	10.00	2.58%	26
South	High African American	Listed, Old 30+	6,079,636	4.87%	1.00	2.42%	24
South	High African American	Listed, no age info	5,318,464	4.26%	3.00	6.34%	63
South	High African American	Not listed	2,063,630	1.65%	0.75	0.62%	6
South	Residual	Listed, Young 18-29	1,184,836	0.95%	10.00	4.71%	47
South	Residual	Listed, Old 30+	15,051,499	12.05%	1.00	5.98%	60
South	Residual	Listed, no age info	10,790,760	8.64%	3.00	12.86%	129
South	Residual	Not listed	3,561,250	2.85%	0.50	0.71%	7
West	High Hispanic	Listed, Young 18-29	241,238	0.19%	10.00	0.96%	10
West	High Hispanic	Listed, Old 30+	1,778,690	1.42%	1.00	0.71%	7

Table 1: National Sampling Plan for Pilot Study Continued

West	High Hispanic	Listed, no age info	1,836,627	1.47%	3.00	2.19%	22
West	High Hispanic	Not listed	696,992	0.56%	0.75	0.21%	2
West	High African American	Listed, Young 18-29	44,327	0.04%	10.00	0.18%	2
West	High African American	Listed, Old 30+	417,046	0.33%	1.00	0.17%	2
West	High African American	Listed, no age info	404,310	0.32%	3.00	0.48%	5
West	High African American	Not listed	156,954	0.13%	0.75	0.05%	1
West	Residual	Listed, Young 18-29	744,690	0.60%	10.00	2.96%	30
West	Residual	Listed, Old 30+	10,351,863	8.29%	1.00	4.11%	41
West	Residual	Listed, no age info	7,858,700	6.29%	3.00	9.37%	94
West	Residual	Not listed	2,686,327	2.15%	0.50	0.53%	5
			124,864,152	100.00%	2.01539	100.00%	1,00

Table 2: National Sampling Plan for Main Data Collection[10]

Geographic Strata[11]	Age Listed Strata[12]	U.S.	% of U.S.	RSF[15]	Sample[16]	Interviews[17]

[10] The main data collection uses the same sampling proportions as the pilot study, but the total sample will be 7000, instead of the 1000 for the pilot study.

[11] Geographic Strata indicates the number of households that are in block groups which are either "High

Region			Households[13]	Households[14]			
Northeast	High Hispanic	Listed, Young 18-29	58,436	0.05%	10.00	0.23%	16
Northeast	High Hispanic	Listed, Old 30+	437,149	0.35%	1.00	0.17%	12
Northeast	High Hispanic	Listed, no age info	459,375	0.37%	3.00	0.55%	38
Northeast	High Hispanic	Not listed	327,019	0.26%	0.75	0.10%	7
Northeast	High African American	Listed, Young 18-29	143,219	0.11%	10.00	0.57%	40
Northeast	High African American	Listed, Old 30+	1,199,945	0.96%	1.00	0.48%	33
Northeast	High African American	Listed, no age info	980,808	0.79%	3.00	1.17%	82
Northeast	High African American	Not listed	563,387	0.45%	0.75	0.17%	12
Northeast	Residual	Listed, Young 18-29	701,612	0.56%	10.00	2.79%	195
Northeast	Residual	Listed, Old 30+	9,053,026	7.25%	1.00	3.60%	252
Northeast	Residual	Listed, no age info	5,533,854	4.43%	3.00	6.60%	462
Northeast	Residual	Not listed	2,187,370	1.75%	0.50	0.43%	30
Midwest	High Hispanic	Listed, Young 18-29	27,470	0.02%	10.00	0.11%	8
Midwest	High Hispanic	Listed, Old 30+	204,652	0.16%	1.00	0.08%	6
Midwest	High Hispanic	Listed, no age info	173,062	0.14%	3.00	0.21%	14
Midwest	High Hispanic	Not listed	105,885	0.08%	0.75	0.03%	2
Midwest	High African American	Listed, Young 18-29	183,176	0.15%	10.00	0.73%	51
Midwest	High African American	Listed, Old 30+	1,645,460	1.32%	1.00	0.65%	46
Midwest	High African American	Listed, no age info	1,304,331	1.04%	3.00	1.55%	109
Midwest	High African American	Not listed	585,363	0.47%	0.75	0.17%	12
Midwest	Residual	Listed, Young 18-29	932,860	0.75%	10.00	3.71%	259
Midwest	Residual	Listed, Old 30+	12,438,967	9.96%	1.00	4.94%	346
Midwest	Residual	Listed, no age info	7,343,136	5.88%	3.00	8.75%	613
Midwest	Residual	Not listed	2,327,593	1.86%	0.50	0.46%	32
South	High Hispanic	Listed, Young 18-29	212,519	0.17%	10.00	0.84%	59
South	High Hispanic	Listed, Old 30+	1,530,210	1.23%	1.00	0.61%	43
South	High Hispanic	Listed, no age info	1,665,438	1.33%	3.00	1.99%	139
South	High Hispanic	Not listed	622,073	0.50%	0.75	0.19%	13
South	High African American	Listed, Young 18-29	648,919	0.52%	10.00	2.58%	181
South	High African American	Listed, Old 30+	6,079,636	4.87%	1.00	2.42%	169
South	High African American	Listed, no age info	5,318,464	4.26%	3.00	6.34%	444
South	High African American	Not listed	2,063,630	1.65%	0.75	0.62%	43
South	Residual	Listed, Young 18-29	1,184,836	0.95%	10.00	4.71%	330

Hispanic", "High African American", or all remaining households.

[12] Age Listed Strata indicates the number of households within each Geographic Strata that are identified in the Experian database as having at least one adult age 18 to 29, only adults over 30, listed but with no age information, or does not have listed information from the Experian database.

[13] Total number of households in the United States that falls into each stratum.

[17] Estimated number of final interviews obtained from each stratum.

[16] Percent of each stratum that will be included in the final sample design.

[15] Relative sampling fraction. This is the sampling rate of each stratum relative to the sampling rates of other strata. For example, if stratum 1 has a relative sampling fraction of 3 and stratum 2 has a relative sampling fraction of 4, then stratum 1 is being sampled at ¾ the rate that stratum 2 is being sampled.

[14] Percent of all households that fall into each stratum.

South	Residual	Listed, Old 30+	15,051,499	12.05%	1.00	5.98%	419
South	Residual	Listed, no age info	10,790,760	8.64%	3.00	12.86%	900
South	Residual	Not listed	3,561,250	2.85%	0.50	0.71%	50
West	High Hispanic	Listed, Young 18-29	241,238	0.19%	10.00	0.96%	67
West	High Hispanic	Listed, Old 30+	1,778,690	1.42%	1.00	0.71%	49
West	High Hispanic	Listed, no age info	1,836,627	1.47%	3.00	2.19%	153
West	High Hispanic	Not listed	696,992	0.56%	0.75	0.21%	15
West	High African American	Listed, Young 18-29	44,327	0.04%	10.00	0.18%	12
West	High African American	Listed, Old 30+	417,046	0.33%	1.00	0.17%	12
West	High African American	Listed, no age info	404,310	0.32%	3.00	0.48%	34
West	High African American	Not listed	156,954	0.13%	0.75	0.05%	3
West	Residual	Listed, Young 18-29	744,690	0.60%	10.00	2.96%	207
West	Residual	Listed, Old 30+	10,351,863	8.29%	1.00	4.11%	288
West	Residual	Listed, no age info	7,858,700	6.29%	3.00	9.37%	656
West	Residual	Not listed	2,686,327	2.15%	0.50	0.53%	37
			124,864,152	100.00%	2.01539	100.00%	7,000

* Relative Sampling Fraction

This stratified design attempts to address the issue of possible ABS under-representation of younger respondents and respondents of color. To illustrate the plan, consider the first strata, Northeast, High Hispanic, Listed Young 18-29. This is a highly-valued stratum given it will include households most likely to contain a young Hispanic person. As such, while about 0.05 percent of households reside in this stratum, we expect to attain 0.23 percent of interviews from it. Conversely, the Residual strata are under-sampled since those strata will disproportionately attain respondents who tend to be more likely to participate in surveys, i.e., older Caucasians. Without implementing such a disproportionate stratification, ABS will tend to substantially under-represent younger respondents and respondents of color.

Similar sampling plans would be executed in the State samples, as detailed below:

Table 3: Arkansas Sampling Plan

Geographic Strata	Age Listed Strata	AR Households	% of AR Households	RSF*	Sample	Interviews
High Hispanic	Listed, Young 18-29	864	0.07%	10.00	0.35%	7

High Hispanic	Listed, Old 30+	5,999	0.48%	1.00	0.24%	5
High Hispanic	Listed, no age info	5,453	0.44%	3.00	0.67%	14
High Hispanic	Not listed	3,502	0.28%	0.75	0.11%	2
High African American	Listed, Young 18-29	14,846	1.20%	10.00	6.06%	121
High African American	Listed, Old 30+	136,503	11.01%	1.00	5.57%	112
High African American	Listed, no age info	95,825	7.73%	3.00	11.73%	234
High African American	Not listed	61,664	4.97%	0.75	1.89%	38
Residual	Listed, Young 18-29	43,311	3.49%	10.00	17.67%	353
Residual	Listed, Old 30+	453,647	36.59%	1.00	18.51%	370
Residual	Listed, no age info	281,153	22.68%	3.00	34.41%	688
Residual	Not listed	137,110	11.06%	0.50	2.80%	56
		1,239,877	100.00%	2.02	100.00%	2,000

Table 4: New Jersey Sampling Plan

Geographic Strata	Age Listed Strata	NJ. Households	% of NJ Households	RSF*	Sample	Interviews
High Hispanic	Listed, Young 18-29	12,689	0.37%	10.00	1.86%	37
High Hispanic	Listed, Old 30+	101,991	2.97%	1.00	1.50%	30
High Hispanic	Listed, no age info	96,127	2.80%	3.00	4.23%	85
High Hispanic	Not listed	90,492	2.64%	0.75	1.00%	20
High African American	Listed, Young 18-29	25,645	0.75%	10.00	3.76%	75
High African American	Listed, Old 30+	238,623	6.96%	1.00	3.50%	70
High African American	Listed, no age info	197,530	5.76%	3.00	8.69%	174
High African American	Not listed	112,878	3.29%	0.75	1.24%	25
Residual	Listed, Young 18-29	99,036	2.89%	10.00	14.52%	291
Residual	Listed, Old 30+	1,282,585	37.41%	1.00	18.81%	376
Residual	Listed, no age info	881,256	25.70%	3.00	38.77%	776
Residual	Not listed	289,903	8.46%	0.50	2.13%	43
		3,428,755	100.00%	1.98	100.00%	2,000

Table 5: Massachusetts Sampling Plan

Geographic Strata	Age Listed Strata	MA. Households	% of MA Households	RSF*	Sample	Interviews
High Hispanic	Listed, Young 18-29	7,603	0.28%	10.00	1.4%	28
High Hispanic	Listed, Old 30+	39,948	1.46%	1.00	0.7%	15
High Hispanic	Listed, no age info	28,762	1.05%	3.00	1.6%	32
High Hispanic	Not listed	27,536	1.01%	0.75	0.4%	8
High African American	Listed, Young 18-29	14,243	0.52%	10.00	2.6%	52
High African American	Listed, Old 30+	93,454	3.41%	1.00	1.7%	34
High African American	Listed, no age info	54,409	1.99%	3.00	3.0%	60
High African American	Not listed	35,388	1.29%	0.75	0.5%	10
Residual	Listed, Young 18-29	145,296	5.31%	10.00	26.7%	533
Residual	Listed, Old 30+	1,356,330	49.56%	1.00	24.9%	498
Residual	Listed, no age info	609,065	22.25%	3.00	33.5%	671
Residual	Not listed	324,849	11.87%	0.50	3.0%	60
		2,736,883	100.00%	1.99	100.0%	2,000

Table 6: Iowa Sampling Plan

Geographic Strata	Age Listed Strata	IA. Households	% of IA Households	RSF*	Sample	Interviews
High Hispanic	Listed, Young 18-29	365	0.03%	10.00	0.1%	3
High Hispanic	Listed, Old 30+	2641	0.21%	1.00	0.1%	2
High Hispanic	Listed, no age info	2797	0.22%	3.00	0.3%	7
High Hispanic	Not listed	1312	0.10%	0.75	0.0%	1
High African American	Listed, Young 18-29	2,040	0.16%	10.00	0.8%	17
High African American	Listed, Old 30+	15,049	1.20%	1.00	0.6%	12
High African American	Listed, no age info	14,984	1.19%	3.00	1.8%	37
High African American	Not listed	7,726	0.61%	0.75	0.2%	5
Residual	Listed, Young 18-29	52,672	4.19%	10.00	21.5%	430
Residual	Listed, Old 30+	676,709	53.78%	1.00	27.6%	553
Residual	Listed, no age info	360,831	28.68%	3.00	44.2%	884
Residual	Not listed	121,216	9.63%	0.50	2.5%	50
		1,258,342	100.00%	1.95	100.0%	2,000

B.2. Describe the procedures for the collection of information, including:

Statistical methodology for stratification and sample selection.
The study will be executed using an address-based sampling frame (ABS). ABS is an excellent alternative to telephone Random Digit Dialing (RDD), given near universal coverage and higher response rates. However, ABS has systematic nonresponse that leads to respondents that are often older and disproportionately Caucasian (Rapoport, Sherr, & Dutwin, 2012, 2014).[18,19] The MDAC team has in the past executed a stratified sample design that is designed to combat these tendencies. This design will allow for a sample which is highly representative of key segments of the population, including those under the age of 30.

Disproportionate stratification and sequestering of random subsets of sample are strategies that are easily corrected by weighting the sample to reflect a random probability sample with no bias from the sample manipulations. Overall, this stratified design would be applied to a national sample as well as distinct samples of New Jersey, Arkansas, Massachusetts, and Iowa residents.

First, MDAC will generate a sample of residential addresses currently in use by employing the USPS computerized Delivery Sequence File (DSF). Marketing Systems Group (MSG) will generate this sample, as it has for scores of studies for government entities over the past decade. MSG will not only draw the sample, but also augment drop units, simplified addresses and other addresses to be full "city-style", to ensure near complete coverage of all addresses in the United States. The file will contain home and apartment addresses, as well as P.O. boxes (OWGMs, only way to get mail) and all other types of residential addresses. We will not sample seasonal or vacant addresses, or non-OWGM P.O. boxes.

For within household selection, MDAC will follow the recommendation of Dr. Paul J. Lavrakas as outlined in his report, "Within-Household Respondent Selection: How Best to Reduce Total Survey Error?"[20] MDAC will use the modified birthday method so that half of the distributed surveys will ask for the person (18 years or older) with the most recent birthday to complete the survey and half will ask the person (18 years or older) with the next coming birthday to do so.

Estimation procedure.

Estimation procedures will follow standard methods of weighting the data to national (and state) parameters. We anticipate using the American Community Survey for point

[18] Rapoport, R., Sherr, S., & Dutwin, D. (2012). Does Ethnically Stratified Address-based Sample Result in Both Ethnic and Class Diversity; Case Studies in Oregon and Houston. Presented at the annual conference of the American Association of Public Opinion Research in Orlando, FL; May 2012.
[19] Rapoport, R., Sherr, S., & Dutwin, D.(2014) Address Based Samples: Key Factors in Refining this Research Methodology. SSRS Whitepaper Archive
[20] Lavrakas, P. (2008) Within-Household Respondent Selection: How Best to Reduce Total Survey Error? Prepared for: Media Rating Council, Inc. Retrieved from : http://mediaratingcouncil.org/MRC%20Point%20of%20View%20-%20Within%20HH%20Respondent%20Selection%20Methods.pdf

estimates and weighting to common parameters such as age, educational attainment, race/ethnicity, gender, and region. We will also explore other potential nonresponse patterns and, if possible, weight to correct for them. For example, this might include address-based sample frame parameters of Census Planning Database parameters.

Prior to calibration (raking) to the above parameters, a base weight will be developed to ensure equal probability of selection. This will minimally include corrections for the disproportionate sample stratification and likely number of drivers/adults in the household.

Once weighted, estimation would utilize a complex samples procedure (Taylor Series Linearization)[21] to ensure proper standard errors due to the complex sample design.

Degree of accuracy needed for the purpose described in the justification.

The power of a test is closely linked to two other metrics: the sample size and the value of the proportion under the alternative hypothesis (HA) against which the value of the proportion under the null hypothesis (H0) is tested.[22] Rejection of the null hypothesis occurs when the observed data is very unlikely to have come from a population for which H0 is true. In this case, rejecting the null hypothesis means that the true proportion of drivers that have fallen asleep while driving is not equal to the proportion we are testing against. At a fixed HA value, the choice of the desired power level automatically yields the minimal sample size needed to achieve that power. Conversely, for any pre-selected sample size, the power of the test can be determined. Hence, it is possible to simulate the relationship between sample size and power for different choices of the HA value. Two aspects of this relationship are noteworthy. First, for any given HA value of the proportion, greater power can be achieved simply by increasing the sample size. However, given the cost of sampling, the minimal implied sample size (MISS) for achieving a particular level of power should suffice, without any need for increasing the sample size indefinitely. Second, for any given level of power, the sample size needs to become larger as the distance narrows between the values of the proportion under H0 and HA. For example, once the desired power level is selected, a larger sample size is required to distinguish between the H0 and HA values when they are "close together" as opposed to when they are "farther apart." The greater the ability to distinguish between the H0 and HA values, the more likely we are to have made the correct decision about the rejection of the null hypothesis based on the results of a given statistical test.

The disproportionate stratified sample we propose, as well as weighting adjustments, all contribute to increased variance in the estimate. The design effect (calculated not just for the variance of the stratification but all weighting adjustments) is a measure of the loss in statistical efficiency caused by complex sample designs and disproportionate nonresponse, the square root of which is the inflation factor on the margin of error. A sample design with a high design effect, which indicates low efficiency, needs a greater

[21] Lavrakas, P. J. (2008). *Encyclopedia of Survey Research Methods* (Vols. 1-0). Thousand Oaks, CA: Sage Publications, Inc. doi: 10.4135/9781412963947.
[22] TIBCO Software Inc. (2020). Data Science Textbook. *See* tibco.com/products/data-science

amount of sample to achieve the same given performance level as a sample design with a lower design effect. The design effect can be computed as a function of the weights applied to the sample to adjust for the disproportionate sample and differential non-response.[23] For the purposes of this power analysis, we factor in an assumed design effect of 1.6 which is the design effect associated with the disproportionate sample design.

The graphs below show the power (on the y-axis) for various sample sizes and various proportions under HA (on the x-axis). The two null hypothesis values for the proportion of interest (proportion of drivers that have fallen asleep while driving) shown are 0.37,[24] and 0.41.[25] These two null hypothesis proportions were chosen as examples to illustrate the power of different sample sizes because they were the reported proportion of interest (proportion of drivers that have fallen asleep while driving) from two previous studies. Comparing the graphs (below) with the same null hypotheses shows that a larger sample size provides higher power for a greater number of alternative hypotheses.

As illustrated, larger samples give more reliable results. Large samples are justified and appropriate when the differences sought are small and the population variance large. Established statistical procedures help ensure appropriate sample sizes so that we reject the null hypothesis not only because of statistical significance, but also because of practical importance. These procedures must consider the size of the type I and type II errors as well as the population variance and the size of the effect. The probability of committing a type I error is the same as our level of significance, 0.05, called alpha, and represents the possibility of rejecting a true null hypothesis (false positive). The probability of committing a type II error or beta (ß) represents failing to reject a false null hypothesis (false negative).

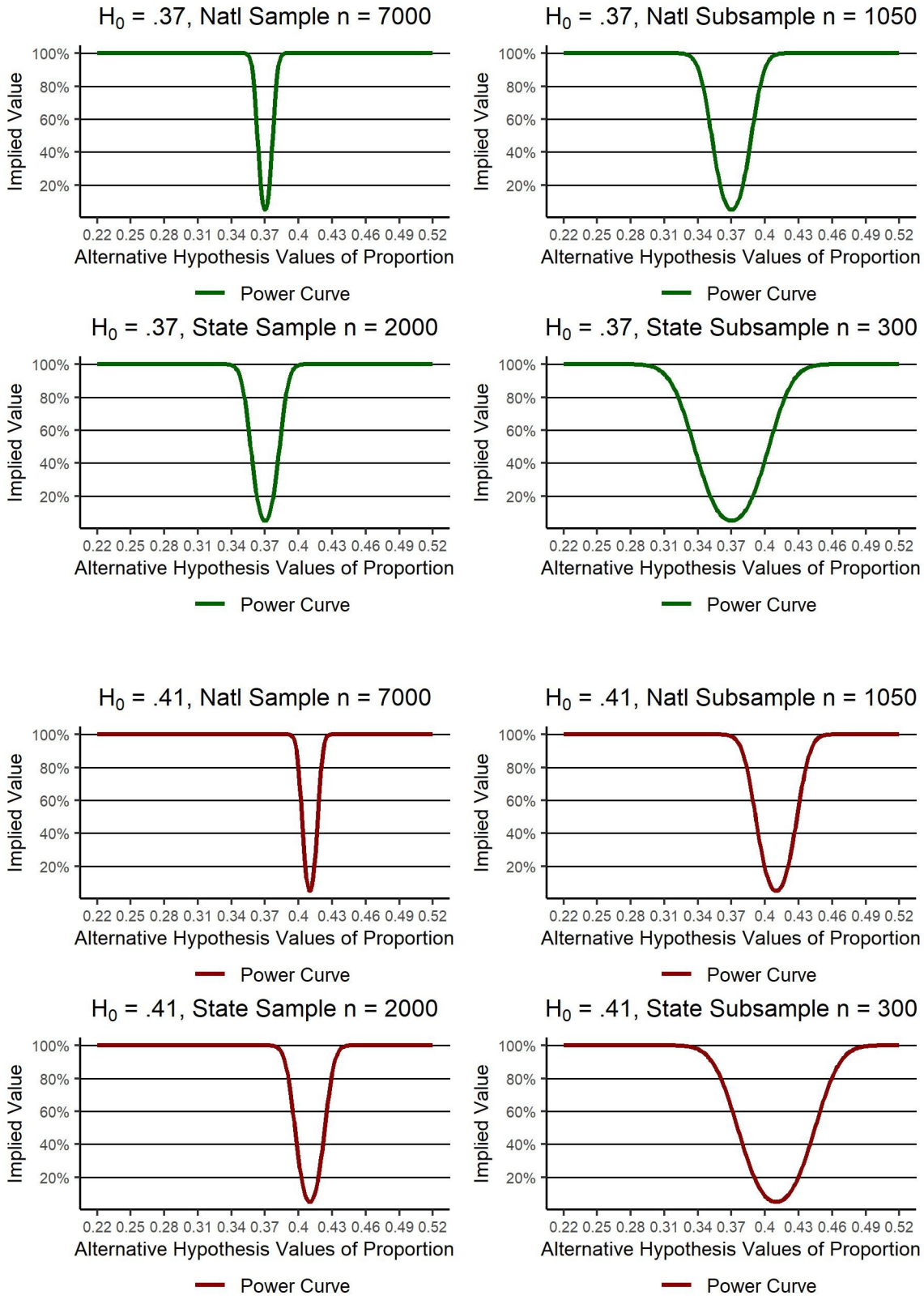
Type I and Type II errors are minimized as the sample size increases. The power of any test is 1 - ß, since rejecting the false null hypothesis is the goal.[26]

[23] The design effect can be computed as the sample size multiplied by the sum of the squared weights divided by the sum of the weights squared.

[24] Royal, A. (2003). *Volume 1: Findings, National Survey of Distracted and Drowsy Driving Attitudes and Behaviors: 2002* (Report No. DOT HS 809 566). Washington DC: National Highway Traffic Safety Administration (NHTSA). Retrieved from https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/hs809566v1.pdf.

[25] American Automobile Association Foundation for Traffic Safety, 2010. *Asleep at the Wheel: the Prevalence and Impact of Drowsy Driving*. Retrieved from https://aaafoundation.org/wp-content/uploads/2018/02/2010DrowsyDrivingReport.pdf.

[26] For additional explanation, *see* Slaughter, Chris. 2008. *Power and Sample Size*. GI Research Conference. Retrieved from http://biostat.mc.vanderbilt.edu/wiki/pub/Main/ChrisSlaughter/powergi.pdf.


H0 = .37, Natl Sample n = 7000
H0 = .37, Natl Subsample n = 1050
H0 = .37, State Sample n = 2000
H0 = .37, State Subsample n = 300
H0 = .41, Natl Sample n = 7000
H0 = .41, Natl Subsample n = 1050
H0 = .41, State Sample n = 2000
H0 = .41, State Subsample n = 300
Implied Value
100%
80%
60%
40%
20%
0.22 0.25 0.28 0.31 0.34 0.37 0.4 0.43 0.46 0.49 0.52
Alternative Hypothesis Values of Proportion
Power Curve

Unusual problems requiring specialized sampling procedures, Paperwork Reduction Act (PRA) Guide 4/27/2011

N/A

Any use of periodic (less frequent than annual) data collection cycles to reduce burden.

This is a one-time survey effort, and data will only be collected one time.

If you are selecting a uniform respondent universe, you may be using a simple random numbers table to select a sample.

We are using a stratified sample, as discussed above.

Stratified sampling is often used when the sampling population can be split into non-overlapping strata that individually are more homogeneous than the population as a whole (e.g., gender and age groups). If there are no obvious "dividing lines", grid lines can be used to divide the population. Random samples are taken from each stratum (or class) and the results are combined to estimate a population mean. Stratified sampling is most successful when the variance within each stratum is less than the overall variance of the population.

Please see the tables above for the details on the stratification approach and sampling process.

B.3 Describe methods to maximize response rates and to deal with issues of nonresponse. 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.

Although the expected response rate of this study is 20.4%, the stratified sampling plan described above will contribute to the collection of reliable data that can be generalized to the U.S. population. Non-responses will be analyzed to determine if differential rates of response contribute to biased estimates from data collection.

Any aspect of your plan that makes it easier and more attractive to comply with the request for information would tend to maximize response rate. This would include:

- **Steps such as prenotification and various types of follow-up with those who did not respond at the first opportunity (give details, e.g., intervals for follow-up, types of follow-up, how many times you will follow up).** As recommended by the Dillman, *Tailored Design Method*,[27] there are five total mailings to the selected household for the data collection. The first is a letter requesting the respondent complete the survey online. This letter will be on NHTSA letterhead and describe the purpose of the study in a clear and relatable way. Included in this first letter will be a non-contingent incentive of a $2 bill. Based on the social exchange theory, it is believed that this will engender good will with recipients, and they may be more likely to complete the survey. The second is a postcard requesting the respondent complete the survey online. The third is a letter with a mailed paper survey. The paper format of the questionnaire will be similar in design to the web layout, adhering to usability principles with easy to read fonts and a pleasant visual layout. In addition, a Business Reply Envelope (BRE) will be included, so as not to add the burden of the cost of return mail to the potential survey respondents. The fourth is a postcard reminder. The fifth outreach is a second mailing of the paper survey with a BRE. Each of the mailings will occur approximately two weeks after the previous mailing. A process involving verification of completion of the survey will occur after each mailing so as to reduce the number of mailings that respondents receive.

- **Making the questions as simple and brief as possible.**

  The research team conducted cognitive interviews with eight respondents. These respondents went through a draft questionnaire with a trained moderator who probed the respondents for their impressions of the questions and ability to answer them efficiently and accurately. Respondents were encouraged to provide feedback about the length of the interview overall as well as the length of each question. They were asked to “think aloud” as they

[27] Dillman, D. A., Smyth, J. D., Christian, L. M., & Dillman, D. A. (2009). *Internet, Mail, and Mixed-mode Surveys: The Tailored Design Method.* Hoboken, N.J: Wiley & Sons.

considered their responses to the questions, thereby providing insight as to how well they understood the questions, whether the language was too complex and cumbersome, and whether they were losing patience with the length of the questions or the survey overall. This feedback provided valuable information about how best to create a survey that is manageable for respondents who will be engaging in a self-administered survey procedure.

After interviews were completed, the moderator provided a written report that contained question-by-question summaries on respondents' reactions to the questions and suggestions for how to improve questions in a way that will create a more positive and efficient survey experience.

- **Already having a good working relationship with this group or the group's perception that actions based on the information collected would be helpful to them.**

  Results from the study will be publicly disseminated in a peer-reviewed journal. In addition, a dissemination plan will be created to reach a broader audience, which includes stakeholders and the general public.

B.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.

Usability testing will be conducted with up to nine participants on the finalized versions of the online and paper surveys. After obtaining informed consent, the facilitator will ask the participant to use the questionnaire as they would in real life, making note of any confusion, reluctance, or inability to proceed, but intervening as little as possible. Then the facilitator and participant will walk through the questionnaire a second time, discussing each question and probing on issues that arose during the first pass.

Due to the sensitive nature of the subject matter, sessions will be timed, but not recorded. The facilitator will take notes on a laptop. Most of the usability data will be qualitative, consisting of observations, probe questions, and participant quotes. To obtain a quantitative measure of the questionnaire's usability, the 10-question System Usability Scale (SUS)[28] questionnaire will be used. The SUS is a public-domain, extensively validated, generic instrument that produces a numeric score with a maximum of 100.

The questionnaire will be tested via web browser on a PC, web browser on a tablet or mobile phone, and via paper.

[28] The System Usability Scale. Retrieved from https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html.

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

M. Davis and Company, Kim Dorazio and Dr. Tom Sexton, (215-790-8900)
Econometrica, Dr. Fred Bellemore, (240-333-4808)
SSRS, Jonathan Best (484-840-4310).

Appendix: Question Justification

Part I: Questions A and B
Screening questions to determine eligibility for the survey. Study sample is composed of adults who have driven in the past month.

Part II.: Questions 1-24
Measurement of Ajzen's Attitudes, Subjective Norms, and Perceived Behavioral Control. We hypothesize that strong attitudes against drowsy driving, high perceived susceptibility and risk to/of drowsy driving, strong perceived subjective norms against drowsy driving, strong motivation to comply with subjective norms, and perceived ability to change situations that influence drowsy driving are associated with less frequent self-report drowsy driving.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50(2),* 179–211. doi:10.1016/0749-5978(91)90020-T

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Addison-Wesley.

Part III.: Questions 25-44
Questions from NHTSA's 2002 distracted and drowsy driving survey, the last nationwide data collection effort to assess drowsy driving. These questions will allow comparisons about the incidence of drowsy driving in two different time periods to determine if drowsy driving prevalence, and its consequences, have declined, stayed the same, or increased.

Royal, D. (2003). *Volume 1: Findings, National Survey of Distracted and Drowsy Driving Attitudes and Behaviors: 2002* (Report No. DOT HS 809 566). National Highway Traffic Safety Administration. https://rosap.ntl.bts.gov/view/dot/1725

Part IV.: Questions 45-59
Questions to determine whether residents of States with drowsy driving laws or active public information and education campaigns are more aware of the dangers of drowsy driving than States without such programs.

Part V. Questions 60-68
Questions from NHTSA's 2002 distracted and drowsy driving survey. These questions will allow comparisons about the actions taken when people are drowsy and driving and assess people's behaviors in two different time periods.

See Royal (2003) above.

Part VI. Questions 69-80

Because there is not a definitive definition of drowsy driving, comparison in certain studies is difficult. This set of questions relaxes the stricter definition of drowsy driving (falling asleep or nodding off while driving) to assess the occurrence, predecessors, and consequences associated with a more colloquial definition of drowsy driving (driving while feeling tired or sleepy enough that you thought you might fall asleep or nod off) to more easily allow comparisons with other studies.

Arnold, L. S. & Tefft, B. C. (2015). Prevalence of Self-Reported Drowsy Driving, *AAA Foundation for Traffic Safety.* https://aaafoundation.org/prevalence-self-reported-drowsy-driving-united-states-2015/

Goncalves, M., Amici, R., Lucas, R., Åkerstedt, T., Cirignotta, F., Horne, J., Léger, D., McNicholas, W. T., Partinen, M., Téran-Santos, J., Peigneux, P., Grote, L., Mallin W., Verbraecken, J., Dogas, Z., Sõõru, E., Arnulf, I., Penzel, T., Schiza, S. E.,… Aksu, M. (2015). Sleepiness at the wheel across Europe: A survey of 19 countries. *Journal of Sleep Research, 24(3),* 242–53. doi: http://10.1111/jsr.12267

Swanson, L. M., Drake, C., & Arnedt, J. T. (2012). Employment and drowsy driving: A survey of American workers. *Behavioral Sleep Medicine, 10(4),* 250–257. doi: http://10.1080/15402002.2011.624231

Wheaton, A. G., Chapman, D. P., Presley-Cantrell, L. R., Croft, J. B., & Roehler, D. R. (2013). Drowsy driving – 19 States and the District of Columbia, 2009-2010. *Centers for Disease Control and Prevention, Morbidity and Mortality Weekly Report, 61(51),* 1033–1037. https://www.cdc.gov/mmwr/pdf/wk/mm6151.pdf

Part VII. Questions 81-98

Questions to provide information regarding factors found to be associated with an increased risk of drowsy driving and drowsy-driving-related crashes including: environmental factors such as roadway or driving conditions; sleep deprivation; sleep disorders such as sleep apnea or narcolepsy; medical conditions; medications; work schedules; circadian rhythm; alcohol use; and lifestyle.

National Highway Traffic Safety Administration (NHTSA). (1998). *Drowsy Driving and Automobile Crashes: Report and Recommendations*. NHTSA. https://rosap.ntl.bts.gov/view/dot/1661

Zanier, N., Eby, D. W., Arnedt, J. T., Molnar, L. J., Shelgikar, A., St. Louis, R., Antonucci, T., Jackson, J. S., Nelson, J., Ryan, L., & Smith, J. (2010). *Drowsy Driving among Older Adults: A Literature Review*. (Report No. M-CASTL 2010-04). Michigan Center for Advancing Safe Transportation throughout the Lifespan. http://www.umtri.umich.edu/our-results/publications/drowsy-driving-among-older-adults-literature-review

Part VIII. Questions 99-107

Questions to assess how work schedules affect propensity to drive while drowsy and alternative transportation options.

Lee, M. L., Howard, M. E., Horrey, W. J., Liang, Y., Anderson, C., Shreeve, M. S., O'Brien, C. S., & Czeisler, C. A. (2015). High risk of near-crash driving events following night-shift work. *Proceedings of the National Academy of Sciences, 13(1),* 176–181. doi: http://10.1073/pnas.1510383112

Scott, L. D., Hwang, W., Rogers, A. E., Nysse, T., Dean, G. E., & Dinges, D. F. (2007). The relationship between nurse work schedules, sleep duration, and drowsy driving. *Sleep, 30(12),* 1801–1807. doi: http://10.1093/sleep/30.12.1801

Part IX. Question 108

Question regarding ride share services to assess whether respondents have access to alternative forms of transportation.

Part X. Questions 109-114

Demographic questions to help ascertain who are drowsy drivers and are they more prevalent for specific demographic groups. These questions are intended to allow for targeted countermeasure development.

Wheaton, A. G., Olsen, E. O., Miller, G. F., & Croft, J. B. (2016). Sleep duration and injury-related risk behaviors among high school students – United States, 2007 – 2013. *Morbidity and Mortality Weekly Report, 65(13),* 337–341. doi: http://dx.doi.org/10.15585/mmwr.mm6513a1