SECTION B

INFORMATION COLLECTION SUPPORTING STATEMENT

Psychological Constructs Related to Seat Belt Use

B) Collections of Information Employing Statistical Methods

The proposed survey, Psychological Constructs Related to Seat Belt Use, will employ statistical methods to sample and analyze the information collected from respondents. The following sections describe the procedures for respondent sampling and data tabulation. The survey will be administered using the GfK KnowledgePanel, a probability-based web panel that has been in existence since 1999. All KnowledgePanel surveys are completed online. To improve representation, panelists who do not have internet access are provided netbooks for the duration of their panel participation. The panel allows for easily obtained representative samples for studies, and the probability-based nature of the design allows for weights and variances to be calculated using standard, accepted statistical techniques. The KnowledgePanel recruitment and empanelment process is designed to comply with CAN-SPAM[1] and Council of American Survey Research Organizations (CASRO) guidelines as well as the participant treatment protocols outlined by the federal Office of Management and Budget (OMB) that follow guidelines from the Belmont Report.[2] The following sections describe the procedures for respondent sampling and data tabulation for the present study.

B.1. Describe the potential respondent universe and any sampling or other respondent selection method to be used.

a. Respondent Universe

The potential respondent universe is U.S. residents aged 16 years or older who have driven or ridden in a motor vehicle (defined as a "car, van, truck, taxi or ride-sharing service") within the past year. Eligible respondents may only ride as passengers in motor vehicles, only drive motor vehicles, or both ride and drive depending on the situation. In the screener (Form 1365), we will verify that the respondent has driven and/or ridden in a motor vehicle in the past year.

[1] The CAN-SPAM Act is a law that sets the rules for commercial email.

[2] National Commission. (1978). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research (National Commission for the Protection of Human Subjects of biomedical or behavioural research.)*. Washington DC: Department of Health, Education and Welfare.

b. Respondent Sampling and Estimated Response Rates

To reach the target population—U.S. residents aged 16 years or older who have driven or ridden in a motor vehicle within the past year—survey participants will be recruited via the KnowledgePanel, which allows for probability-based samples of U.S. residents. Panelists are currently recruited onto the KnowledgePanel using address-based sampling (ABS), which is based on the U.S. Postal Service's computerized delivery sequence file. Before 2009, panelists were also recruited via random digit-dialing (RDD) techniques. ABS is used to obtain a probability-based sample of addresses. Individuals residing in sampled addresses are invited to join the KnowledgePanel via a series of mailings, and follow-up invitations are conducted with nonrespondents via phone (for addresses that can be matched to a phone number). Full-panel weights are computed for all panelists. These weights reflect selection probabilities and incorporate a calibration adjustment to ensure that key demographic distributions align with the most recent data from the U.S. Census Bureau's Current Population Survey (CPS).

More precisely, let w_i be the calibrated weight for panelist $i \in P$, where P represents the set of panelists. Then the calibration estimator is of the form $\hat{X}_{cal} = \sum_{i \in P} w_i X_i = X$, where $X = \sum_U X_i$ is a vector of population totals used for calibration over the universe U; that is, the estimated vector of demographic totals $\hat{X}_{cal}$ is equal to the population totals used in calibration.

The next step involves oversampling individuals who have a lower probability of being in the panel, so that each member of the population has the same chance of being sampled for NHTSA's survey, prior to screening. A study-specific sample of invitees will be drawn from the panel through a probability-proportional-to-size (PPS) sampling procedure, using the full-panel weights as a measure of size to achieve an epsem (i.e., equal probability of selection method) sample. This method generally results in a higher effective sample size and greater statistical power than would otherwise be obtained. PPS sampling will be conducted independently for each of four NHTSA-defined regions,[3] without replacement; the number of invitees—but not necessarily completes—by region will be proportional to the regional population size. This will not affect the epsem properties of the sample but will ensure that the sample of invitees is balanced by region before nonresponse and subsampling in the screening phase. More precisely, the probability of study-specific selection (prior to screening) for individual i within stratum h is $\pi_{hi} = n_h \frac{w_{hi}}{\sum_i w_{hi}}$, where:

[3] **Region 1:** AK, CA, HI, OR, WA; **Region 2:** AZ, CO, IA, ID, KS, MN, MO, MT, ND, NE, NM, NV, SD, UT, WY; **Region 3:** AL, AR, DC, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV; **Region 4:** CT, IL, IN, MA, ME, MI, NH, NJ, NY, OH, PA, RI, VT, WI

- h refers to the NHTSA-defined region, for $h \in \{1,2,3,4\}$;
- w_{hi} is the full-panel weight for panelist i within stratum h;
- $\sum_i w_{hi}$ is the sum of the weights of all panelists within stratum h; and
- n_h is the desired number of survey invitees in stratum h, computed as $n_h \propto N_h$; that is, the desired stratum sample size is proportional to the stratum population size N_h.[4]

The above sample design results in an adjusted design weight of

$$w'_{hi} = \frac{w_{hi}}{\pi_{hi}} = w_{hi}\left(n_h \frac{w_{hi}}{\sum_i w_{hi}}\right)^{-1} = \frac{\sum_i w_{hi}}{n_h}$$ for an individual $i \in S$, where S is the set of invitees.

This figure is constant within a stratum, yielding properties resembling that of an epsem sample. Further, since $n_h \propto N_h$, we have that $w'_{hi} = \frac{\sum_i w_{hi}}{n_h} \propto \frac{\sum_i w_{hi}}{N_h} \approx 1$, given that the total weight within stratum is approximately equal to the stratum population size. As a result, the overall sample is approximately epsem, and the stratification ensures a balanced sample of invitees with respect to the NHTSA-defined region.

Responding panelists will be screened to ensure that they are in the target population and to classify them as “always users” versus “not-always users” of seat belts.[5] This classification will be used when implementing an additional subsampling phase to oversample “not-always users” (relative to “always users”), which will improve precision for domain estimates of “not-always users.” An individual’s probability of

---

[4] For sample allocation purposes, this stratum population size can reasonably be estimated using the full panel weights, that is, $\hat{N}_h = \sum_i w_{hi}$; alternatively, it could be obtained from external benchmarks (e.g., from the U.S. Census Bureau’s American Community Survey). Both designs are measurable and should produce approximately equivalent results.

[5] We explored stratification as a possibility for increasing the expected proportion of not always belt users in the sample. We used the public use data from the *2013 National Survey of Drug Use and Health* to conduct a Chi-square Automatic Interaction Detector (CHAID) analyses to explore possible stratification variables. The outcome variable was dichotomous seatbelt usage (always as driver and passenger versus not always), and the predictor variables were age, educational attainment, race/ethnicity, sex and marital status. The analysis suggested a four-strata design producing subpopulation estimates of not always users ranging from 5% (college-educated females) to 21% (young people without a college education) with a population estimate of 13%. We then examined the effect of different allocations on the design in terms of the effective sample sizes for always users and not always users. The results of this analysis suggested that the stratification produced marginal increases in the effective sample size for not always users and significant decreases in the effective sample size for always users, which is due to the fact that not always users were not sufficiently concentrated within the strata. The use of the additional stratification would also significantly increase the complexity of the sample design because it already stratifies for four geographic regions to account for expected differences in belief structures. Based upon these factors, the final design opts for an oversample of not always users based upon a screener rather than using additional stratification.

selection in this subsampling phase will be retained and incorporated into the subsequent computation of study-specific survey weights.

Let r_j be the subsampling rate for group j, where $j=1$ for not-always users and $j=2$ for always users of seatbelts. Then the adjusted design weight from above can be modified as $w''_{hij}=w'_{hij} I_{hij} \delta_{hij} (r_j)^{-1} + w'_{hij}(1-I_{hij})$ where:

- I_{hij} is an indicator variable that is equal to 1 if individual i from within group j and stratum h completed the screener and 0 otherwise; and
- δ_{hij} is an indicator variable that is equal to 1 if the given individual was sampled following the screening phase and 0 otherwise.

In other words, the weight is divided by r_j for those who were subsampled; the weight is removed for those who completed the screener but were not subsampled; and the weight is not modified for those who did not complete the screener (although the weights of nonrespondents would need to be subsequently removed and accounted for during nonresponse and/or calibration adjustments). The subsampling rates will be determined in a fashion that ensures sufficient precision for not-always users, while also aiming to avoid creating too much of a design effect due to weighting.[6] For example, assuming a population incidence of 13% for non/part-time users, a subsampling rate of 100% for not-always users and subsampling rate of 60% for always users would result in approximately 1,196 survey completes among not-always users and a design effect from weighting of 1.3 among those completing the screener, assuming no differential nonresponse patterns.[7]

Information regarding the anticipated precision for the final set of respondents, overall and for key domains, is provided in Section B.2 (b).

In general, response rates can be affected by many factors, such as mode, sponsor, topic, questionnaire length, use of incentives, and frequency and intensity of follow-up efforts for nonrespondents (see, for example, Groves & Couper 2012; Massey &

---

[6] The design effect due to weighting is a correction factor (1 + *L*) for weighting variability and its effect on precision, where $L=n^{-1}\sum_s \frac{(w_i-\overline{w})^2}{\overline{w}^2}$ and is defined as the relative variance of the sample weight w_i. This 1 + *L*, termed the relative loss due to weighting by Kish (1992) and commonly referred to as the unequal weighting effect (UWE), can be a reasonable approximation for the design effect (DEFF) when the weights are unrelated to the outcome of interest (e.g., see Spencer, 2000). In practice, note that design effects are specific to a particular estimator, and thus can vary depending on the type of analysis and the specific parameter being estimated.

[7] For instance, assuming a 13% incidence and 95% survey completion rate among subsampled individuals who completed the screener, then 9,687 individuals completing the screener would result in $9687\times.13\times.95=1196$ interviews of not-always users (due to subsampling all such users), versus $9687\times(1-.13)\times.6\times.95=4804$ interviews of always users (due to subsampling at 60%), for a total of $6{,}000$ interviews.

Tourangeau 2013; Plewes & Tourangeau 2013).[8] Although survey outcome rates can vary greatly by study, we can estimate rates based on the results of numerous studies that have been conducted with the GfK KnowledgePanel. Based on this experience, we anticipate the following rates for this study: Recruitment rate (RECR) of 15%; profile rate (PROR) of 64%; screener completion rate (SCRR) of 50%; completion rate (COMR) of 95%;[9] and cumulative response rate (CUMRR) of 5%. Note that RECR and PROR are due to nonresponse at the panel-level, whereas SCRR and COMR are from study-specific nonresponse. Thus, were we to focus solely on survey participation among recruited and profiled panelists, we would attain a study participation rate of approximately 48% (i.e., the product of SCRR and COMR), although this is not an official response rate, but rather, a measure of study-specific response among panelists.

Table 1 shows the anticipated sample counts, based on this information. We assume that roughly 95% of sample members will be eligible to complete the study (i.e., have driven or ridden in a car in the past year). In practice, it is not clear exactly what this rate will be, although we believe that it will be very high given that 91% of households are estimated to own at least one vehicle[10] and 97% of individuals are estimated to be in a household with at least one driver.[11] We also note that anticipated response rate calculations and burden calculations are fairly insensitive to this rate. Further, our sample size calculations reflect that an estimated 13% of U.S. residents aged 16 or older do not always use seatbelts.[12]

Table 1. Overview of Anticipated Sample Counts

	Estimate
Total Panelists Contacted	20,394
Screener Completion Rate (SCRR)	0.50
Estimated Eligibility Rate	0.95
Estimated Rate of Not-Always Users of Seatbelts (Among Eligibles)	0.13

[8] Groves, R. M., & Couper, M. P. (2012). *Nonresponse in household interview surveys*. John Wiley & Sons.
Massey, D. S., & Tourangeau, R. (2013). Where do we go from here? Nonresponse and social measurement. *The Annals of the American Academy of Political and Social Science, 645*(1), 222-236.
Plewes, T. J., & Tourangeau, R. (Eds.). (2013). *Nonresponse in social science surveys: a research agenda*.
[9] GfK provided the recruitment rate (RECR) of 15%; profile rate (PROR) of 64%; screener completion rate (SCRR) of 50%; and completion rate (COMR) of 95%. These figures are based on the response rates observed in their KnowledgePanel over the past 17 years.
[10] U.S. Census Bureau (2014). *American Community Survey.* Retrieved December 9, 2016, from https://www.census.gov/programs-surveys/acs/data/data-via-ftp.html.
[11] U.S. Department of Transportation Federal Highway Administration (2009). *National Household Travel Survey.* Retrieved December 9, 2016, from http://nhts.ornl.gov/download.shtml.
[12] Substance Abuse and Mental Health Services Administration, *Results from the 2013 National Survey on Drug Use and Health: Mental Health Findings*, NSDUH Series H-49, HHS Publication No. (SMA) 14-4887. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2014.

Subsampling Rate of Always Users of Seatbelts	0.60
Survey Completion Rate (COMR)	0.95
Complete Screeners (20,394 * .50)	10,197
Eligible Persons (10,197 * .95)	9,687
Not-Always Users (9,687 * .13)	1,259
Eligible Persons: Always Users (9,687 * .87)	8,428
Ineligible Persons (10,197 * .05)	510
Subsampling	
Not-Always Users (9,687 * .13)	1,259
Always Users (8,428 * .60)	5,057
Subsampled Persons (Total)	6,316
Survey Completion	
Not-Always Users (1,259 * .95)	1,196
Always Users (5,057 * .95)	4,804
Complete, Eligible Surveys (Total)	6,000

SCRR = (Complete screeners) / (Sampled persons) = 10197/20394 = .5

COMR = (Complete, eligible surveys) / (Subsampled persons) = 6000/6316 = .95

CUMRR = PROR * RECR * SCRR * COMR = .15 * .64 * .5 * .95 = .05

B.2. Describe the procedures for the collection of information.

a. Procedures for Collection of Information

KnowledgePanel participants will be recruited for this survey via email (Appendix A). Respect for the principles of voluntariness and informed consent is central to GfK's procedures in building and maintaining the KnowledgePanel. Participation in research is voluntary at the time that respondents are asked to join the panel, at the time they are asked to participate in any particular survey, and at the time they answer any given question in a survey.

For each panel member selected to complete the screener for this study, an email (Appendix A) will be sent to their password-protected email account that will notify them that a survey is available for completion. Surveys will be self-administered and accessible at any time during the designated fielding period. Participants will be able to complete the survey only once. Participants who click on the link provided in the recruitment email will be administered a five-item screener (Form 1365). Responses to these items will serve two purposes:

1. Respondents who report neither driving nor riding as a passenger in a motor vehicle within the past year (i.e., individuals who are not members of the target population) will be screened out.
2. The selection of participants into the full survey will be informed by their self-reported status as an "always" versus a "not-always" user. Respondents will first be classified into one of the two groups based on their response to Q3. If a self-reported "always" user then reports not wearing a seat belt within the past year in Q4 and/or Q5, he or she will be reclassified as a "not-always" user for oversampling purposes.

Participants selected for completing the full survey (Form 1366) based on their responses to the screener will be immediately redirected to the full survey, so as to reduce the likelihood of attrition as compared to a two-phase administration approach.

All personally identifiable information (PII; e.g., names, addresses, emails, etc.) will be kept secured in a separate office in the Information Technology section of the main GfK offices in Palo Alto, CA, and all data transfers from web-enabled devices (PCs and laptops used for survey administration) to the main servers will pass through a firewall. All PII will be retained by GfK.

Moreover, all electronic survey-specific data records will be stored in a separate secured database that does not contain PII. GfK staff members who have access to the PII, which is limited to the Panel Management staff members, do not have access to the survey response data, and vice versa, with the exception of IT administrators who must have access to maintain the computer systems. The secured database contains field-specific permissions that restrict access to the data by type of user, as described above, thus preventing unauthorized access.

b. Precision of Sample Estimates

The ability to detect differences in attitudes and behaviors based on respondent characteristics can be affected by several factors, including the type of analysis, population-level differences between groups, and the study's statistical design. Table 2 displays the margin of error for key domains, under some simplifying assumptions.[13]

[13] When using a disproportionate sampling allocation, the margin of error for a 95% confidence interval of +/-5% for a proportion p is $MOE \approx 1.96 \cdot \sqrt{\sum_{h=1}^{H} \left(\frac{N_h}{N}\right)^2 (1-f_h) \frac{p_h(1-p_h)}{n_h - 1}}$, where N_h is the stratum population size, N is the total population size, $1-f_h = 1-\frac{n_h}{N_h}$ is the stratum-level finite population correction, n_h is the stratum sample size, and p_h is the estimated proportion. In order to produce more conservative confidence intervals, we assume a proportion of 50% and an ignorable finite population correction. We assume an overall design effect due to weighting of 1.34, a number of interviews by region that is proportional to estimates from the 2014 American Community Survey one-year-estimates, a

These margin of error calculations include an adjustment for an anticipated loss in precision due to weighting of 1.34 for total estimates and estimates by region. For domain estimates of always seatbelt users and not-always seatbelt users, we assume a design effect from weighting of no greater than 1.3, based on experience with general population studies from the GfK KnowledgePanel. Although we have an approximately epsem design prior to nonresponse and oversampling of not-always users of seatbelts, an assumed design effect of 1.3 would allow for losses in precision due to the effects of weighting for nonresponse and calibration. This design effect is further adjusted upward by a factor of 1.3 for region and overall estimates, to account for oversampling of not always seatbelt users.

Table 2. Margin of Error (MOE) Estimates

Population	**MOE**
Total	1.5%
Region 1	3.6%
Region 2	3.9%
Region 3	2.4%
Region 4	2.6%
Not-always users	3.2%
Always users	1.6%

The sample size of 6,000 ensures an overall MOE of 1.5%, adequate precision for regional estimates and estimates based on seatbelt usage, as well as no greater than a 4.0% MOE for domain estimates for any subpopulations with at least a 13.5% population incidence (assuming a 1.34 design effect, which would result if there is minimal covariance between subpopulation membership and the weights). Although we have planned our sample with our reporting requirements in mind, we note that the results of the data analysis may suggest the utility for estimates of domains that have not been pre-specified. As such, our sample size of 6,000 will allow adequate precision not only for estimates of our key planned domains (i.e., by region and seatbelt usage), but also for any unplanned domains of at least moderate size.

c. Power Analysis

We furthermore confirmed our sample size requirement of 6,000 to be sufficient to detect effects of the size that we expect in the data using a power analysis. Specifically, we used a simulation approach (e.g., Muthén & Muthén, 2002)[14] based on a model similar to the most complex model we propose to test. We simulated input data with bivariate effect sizes in the small-to-medium range (i.e., .2), generated several thousand datasets with different sample sizes, estimated models on those data, and recorded the

population incidence of 13% for not-always users, and a subsampling rate of 60% for always users.

[14] Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*, 599–620.

proportion of model estimates that were statistically significant at a hypothesis rejection/alpha probability of .05. After compiling the observed proportions, we fit the several thousand simulation results to a linear model using the sample size used as a predictor, assuming the power function was approximately linear. Using the results from the linear model, we solved for a sample size that would obtain statistically significant effects 80% of the time given the estimated models. Because our analysis will incorporate complex survey design (i.e., survey weights and strata), we adjusted the resulting sample size estimate by our estimated design effect of 1.34 to obtain the sample size needed to obtain sufficient statistical power given the increase in standard error estimates that accompany weighting and stratification. Our results show that a sample size of 4,500 was near the 80% point without the design adjustment. With the 1.34 design adjustment, we accepted the value of 6,000 as a sample size that will provide sufficient statistical power to answer our research questions.

d. Cumulative Response Rate

The American Association for Public Opinion Research's Standard Definitions[15] provides clear guidance for calculating cumulative response rates for probability-based internet panels, such as the one being used for the current data collection. The approach is discussed in more detail by Callegaro and DiSogra (2008).[16] AAPOR indicates that the cumulative response rate (CUMRR) can typically be computed as the product of panel recruitment rate (RECR), profile rate (PROR), and completion rate (COMR). AAPOR furthermore indicates that the use of screening questions to determine study eligibility requires an additional step, which in our case involves the use of a screening completion rate (SCRR). The computations are also affected by the treatment of partial interviews, which in our case will be excluded from the numerators of the relevant rates. For our study, the components of the cumulative response rate can be computed as follows:

- Recruitment Rate (RECR): For a given recruitment cohort, RECR is computed as the number of households providing initial consent to join the panel divided by the number of consents plus refusals, non-contacts, others, and the product of unknowns and the estimated proportion of cases of unknown eligibility that are eligible. In essence, this is AAPOR Response Rate 3 (RR3), as applied to the panel recruitment process. Note that panelists are recruited over a series of recruitment cohorts. Thus, RECR is computed separately for each cohort, after which the study-specific RECR is computed as a weighted average of the cohort-specific rates for each sample member.
- Profile Rate (PROR): For a given cohort, PROR is computed as the number of households completing the profile survey divided by the sum of the numbers of

[15] The American Association for Public Opinion Research (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.9th edition*. AAPOR.
[16] Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly, 72*(5), 1008–1032.

profile completes, profile partials, refusals, non-contacts, and others. For this study, PROR is computed separately for each cohort, after which the study-specific PROR is computed as a weighted average of the cohort-specific rates for each sample member.

- Screening Completion Rate (SCRR): The number of sampled panel members who complete the survey screener divided by the total number of sample members invited to participate in the study.
- Completion Rate (COMR): The number of sampled panel members who complete the full survey divided by the number determined to be eligible for the study based on the applicable screener questions.

Following data collection, the cumulative response rate will be calculated following American Association for Public Opinion Research (AAPOR) guidelines. GfK's survey tracking process will ensure that all sample members are accounted for and given the proper disposition code in line with AAPOR and Council of American Survey Research Organizations (CASRO) guidelines. This will allow us to appropriately calculate response rates for the probability-based panel sample and track any problems with the survey effort. The tracking of disposition codes will allow for the creation of appropriate weights for eligible respondents that are adjusted to account for sample members with unknown eligibility and eligible nonrespondents (see B.2(c)(d)).

e. Non-response Bias Analysis

There are several methodologies that are widely used for implementing nonresponse bias analyses. For example, a nonresponse bias analysis could be a benchmark study in which survey estimates are compared with external estimates; it could be a nonresponse bias follow-up study in which a higher level-of-effort is used to gain additional respondents, in order to compare late versus early respondents; it could involve comparing respondents with nonrespondents based on auxiliary variables from the sampling frame and/or external data; or it could even involve constructing alternate sets of weights in order to assess the impact of nonresponse on various estimators. The appropriateness and utility of particular nonresponse bias analysis methods is affected by factors such as characteristics of the sampling frame, availability of high-quality external sources of information about the population, and cost.

For this survey, we expect to employ benchmarking methods and/or a comparison between respondents and nonrespondents using auxiliary variables available for all sample members. Benchmarking methods would simply entail comparing survey estimates for the given population with external benchmarks for the same population from a source such as the U.S. Census Bureau's Current Population Survey (CPS) or American Community Survey (ACS). An advantage of this method is that it would encompass not only nonresponse error but also coverage error. The main disadvantage is that such an analysis would not be possible for all estimators, as benchmarks may only be available for key sociodemographic variables. We also note that a calibration weighting approach will ensure conformity with respect to key demographic characteristics, therefore removing any differences with respect to the adjustment categories.

Alternatively, we may compare respondents with nonrespondents using auxiliary variables in a two-step process.

1. First, logistic regression methods would be used to estimate response propensity using auxiliary variables in the sampling frame (e.g., age, gender, and race/ethnicity) that are available for both respondents and nonrespondents. This analysis would be restricted to the final stage of nonresponse among KnowledgePanel sample members, and would examine participation among panelists sampled for the study, given that this will allow for the use of profile information collected for all impaneled households.
2. Then, for those demographic characteristics found to be significant predictors of response status, additional logistic regression analyses would be conducted to determine whether these characteristics are also significantly related to key outcome variables of interest. For characteristics related to both response propensity and survey outcomes, unadjusted estimates would be subject to bias, and therefore this is normally taken into account in computing survey weights. It is important to note that nonresponse bias is specific to a statistic, so separate assessments are often needed for different estimates.

f. Sample Weighting

Using proven procedures, we will develop full-sample weights to reflect the study design and mitigate the risks of sampling error, nonresponse error, and/or coverage error. We expect to compute weights as follows:

1. We will begin by computing weights for all KnowledgePanel members that will account for the panel design. These weights will be calibrated on several key demographic characteristics to external benchmarks from the U.S. Census Bureau's Current Population Survey (CPS) or American Community Survey (ACS) to mitigate possible coverage or nonresponse error.

2. The weights from Step 1 will then be adjusted to account for study-specific sampling from the KnowledgePanel, prior to screener-based subsampling. By multiplying the full-panel weights by the inverse of the probability of selection for the corresponding sampling stage, the weights of panelists who were not invited to participate in this study can be redistributed to panelists who were. Because the study-specific selection probability for invitees is set to be proportional to the full-panel weights, this results in behavior that is similar to an equal probability of selection method, which improves survey precision in many situations.

3. Next, the weights of qualified respondents from Step 2 will be adjusted to account for subsampling during the screening phase (i.e., oversampling of "not-always users" of seat belts). The inverse of individuals' selection probabilities from the corresponding sampling stage will be applied as an adjustment to redistribute the weights of non-subsampled individuals to subsampled individuals.

4. The weights of qualified respondents will then be calibrated to relevant benchmarks using a raking technique (i.e., iterative proportional fitting). This adjustment will ensure that the calibrated weights conform to external benchmarks taken from a high-quality source of population information, such as the U.S. Census Bureau's Current Population Survey (CPS) or American Community Survey (ACS). The adjustment categories will be based on key sociodemographic variables such as region, gender, age, race/ethnicity, education, household income, and/or metropolitan area. In conducting calibration, we will take special care to ensure comparability between the set of individuals entering the calibration stage and the set of benchmark totals.

5. Following calibration adjustments, we will examine the distribution of the resulting weights and, if necessary, trim outliers at the extreme upper and lower tails of the weight distribution to reduce weight variation. The final calibrated and trimmed weights will then be scaled, so that they sum to the total number of eligible respondents.

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

The contractor is taking a number of steps to boost the survey's response rate. The study's sponsorship by NHTSA will be emphasized in the recruitment email in order to attract potential participants. During the cognitive testing of the survey instrument, participants reported that the explicit mention of NHTSA as the survey's sponsor lent credibility to the study and made them more likely to want to take it. The existing relationship between GfK and its panelists, meanwhile, is expected to also encourage respondents to accept the invitation to participate in this particular survey.

Furthermore, GfK also operates a modest incentive program—primarily through the use of a point system—to encourage participation. Incentives fall into two categories: general and survey-specific. General incentives are provided for each completed survey. Those who use their own computer and internet connection (i.e., an internet household) are awarded 1,000 loyalty points for completing each survey. One thousand points is roughly equivalent to $1.00. Those who did not have a computer and internet connection at the time of recruitment (i.e., a non-internet household) are provided one at no cost. They are allowed to keep and use the computer on an unrestricted basis for the duration of their tenure on the panel. For surveys longer than 15 minutes, such as the present one, an additional incentive is offered in the form of an entry into a sweepstakes. KnowledgePanel has an existing sweepstakes in place for its panel that has already been vetted from a legal standpoint to ensure compliance in all 50 states.

The survey's administration will also include multiple features to decrease the likelihood that respondents become frustrated and terminate their participation prior to submission of a completed questionnaire. For example, the survey interface will provide easy navigation from page to page and furnish the capability for respondents to pause and leave the system and then re-enter at the departure point without losing any previously inserted information. In addition, a live telephone help desk will be operated from early morning through late evening, 7 days a week, for the duration of the survey's fielding period.

Furthermore, the survey instrument has already undergone cognitive testing, and will additionally undergo usability testing, once programmed. Both of these activities will help to ensure that the survey is clearly understood and easy for respondents to complete.

B.4. Describe any tests of procedures or methods to be undertaken.

As part of the study design, the contractor conducted nine cognitive interviews to test and refine questions for the survey. All interview participants were drawn from the study's target population.

Data collection consisted of two phases (Phase I n = 5, Phase II n = 4), with a one-day break scheduled in between to allow for the revision of the survey instrument prior to starting Phase II. The entire survey instrument was tested. In each session, the participant and the moderator sat in a room together while other members of the team observed from another room. At the start of the interview, the moderator discussed the purpose of the interview and provided detailed instructions to the participant. Although the actual survey will be web-based, cognitive testing was conducted with a paper version for ease of administration and discussion. Notes were taken in real time for analysis and no audio or video recordings were made.

Participants were directed to complete the survey and to share aloud their thoughts and decision-making process as they answered each question. The purpose of this "think aloud" protocol was to gain a deeper understanding of not just the *what* but also the *why* of participants' reactions and responses to the survey. Frequently, the moderator would ask follow-up questions to probe specifics of the participant's evaluation process. There was also a set of standardized probes included in the Moderator's Guide. Furthermore, as time permitted, additional probes were supplied to the moderator from the observing team over a messenger/chat application the moderator passively monitored.

Across both phases, both the pre-determined and ad-hoc probes were generally designed to:

1. Ensure that participants understood the survey items as intended.
2. Assess the language and clarifying definitions included.
3. Assess the appropriateness of the response options/anchors used for the survey items.
4. Assess the ordering of the survey items.

Notes from the cognitive interviews were captured by two researchers—the moderator conducting the interview and an observer. The moderator was also able to capture non-verbal cues such as facial expressions to help identify areas where participants might be getting stuck and/or disliked the phrasing of a question. Following each interview, the team debriefed and compared notes on the session, occasionally modifying, adding, or removing probes. In most cases, there was full agreement among the team on the items that gave participants difficulty and might require revision. The team discussed any points of disagreement until consensus was reached.

B.5. Provide the name and telephone number of individuals consulted on statistical aspects of the design.

The following individuals have reviewed technical and statistical aspects of procedures that will be used to conduct this survey:

William Walton, PhD
Director, Personnel & Development
Fors Marsh Group
1010 N. Glebe Rd., Suite 510
Arlington, VA 22201
(571) 858-3794

Anna Sheveland, PhD
Senior Researcher, Personnel & Development
Fors Marsh Group
1010 N. Glebe Rd., Suite 510
Arlington, VA 22201(571) 858-3815

Jonathan Mendelson, MS
Senior Scientist, Scientific Techniques and Analysis Team
Fors Marsh Group
1010 N. Glebe Rd., Suite 510
Arlington, VA 22201
(571) 858-3753

Joseph Luchman, PhD
Senior Scientist, Scientific Techniques and Analysis Team

Fors Marsh Group
1010 N. Glebe Rd., Suite 510
Arlington, VA 22201
(319) 333-1443

Mansour Fahimi, PhD
Senior Vice President, Chief Statistician
GfK
1250 I St NW #330
Washington, DC 20005
(202) 747-1820

Alan Block, MA
Office of Behavioral Safety Research
National Highway Traffic Safety Administration
1200 New Jersey Ave, SE
Washington, DC 20590
(202) 366-6401

Mary T. Byrd, MA
Office of Behavioral Safety Research
National Highway Traffic Safety Administration
1200 New Jersey Ave, SE
Washington, DC 20590
(202) 366-5595