SUPPORTING STATEMENT
U.S. Department of Commerce
National Oceanic & Atmospheric Administration
Using Quick Response Surveys to Build a Public Perception and Response Database
OMB Control No. 0648-0805

SUPPORTING STATEMENT PART B

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

The target population for the Quick Response Survey (QRS) is adults over 18 who have experienced a specific flash flood, tornado/high wind, or winter weather event. Fourteen National Weather Service (NWS) Weather Forecast Offices (WFOs) participated in Phase I and will continue to do so in Phase II. Phase II engages an additional 36 offices (Table 1) reaching a potential of 188,065,658 people. Surveys will be targeted to people residing in the counties impacted by a specific weather hazard. Respondents will be surveyed 1 day to 4 weeks after a select severe or winter weather event occurs to limit recall bias by the respondent and provide the NWS with rapid results.

Table 1. Potential Respondent Universe

City	State	WFO/CWA	Region	Population
Albuquerque	New Mexico	ABQ	Southern	1,621,716
Amarillo*	Texas	AMA	Southern	427,566
Austin/San Antonio	Texas	EWX	Southern	5,426,844
Birmingham	Alabama	BMX	Southern	2,788,108
Blacksburg	Virginia	RNK	Eastern	1,764,086
Burlington*	Vermont	BTV	Eastern	832,929
Chicago	Illinois	LOT	Central	10,082,145
Denver	Colorado	BOU	Central	4,125,237

Fort Worth*	Texas	FWD	Southern	9,557,527
Honolulu	Hawaii	HFO	Pacific	1,445,592
Houston/Galveston	Texas	HGX	Southern	7,890,842
Jackson	Mississippi	JAN	Southern	1,743,191
Kansas City/Pleasant Hill	Missouri	EAX	Central	2,698,866
Las Vegas	Nevada	VEF	Western	2,535,401
Little Rock*	Arkansas	LZK	Southern	1,680,735
Mobile	Alabama	MOB	Southern	1,670,748
Mount Holly*	New Jersey	PHI	Eastern	12,365,092
Nashville*	Tennessee	OHX	Southern	2,891,417
Peachtree City*	Georgia	FFC	Southern	8,601,769
Phoenix*	Arizona	PSR	Western	4,895,596
Pittsburgh*	Pennsylvania	PBZ	Eastern	3,705,637
Portland*	Oregon	PQR	Western	3,837,661
Raleigh*	North Carolina	RAH	Eastern	4,917,044
Seattle	Washington	SEW	Western	5,396,736
Shreveport*	Louisiana	SHV	Southern	2,079,655
Springfield	Missouri	SGF	Central	1,381,595
St Louis	Missouri	LSX	Central	3,669,734
Taunton*	Massachusetts	BOX	Eastern	7,959,719
Twin Cities/Chanhassen	Minnesota	MPX	Central	4,926,276
Upton*	New York	OKX	Eastern	16,877,835

Year 1 Population - 30 WFOs				**139,797,299**
10 Additional WFOs Avg. Pop	TBD			2,413,418
Year 2 Population - 40 WFOs				**24,134,179**
10 Additional WFOs Avg. Pop	TBD			2,413,418
Year 3 Population - 50 WFOs				**24,134,179**
Total Population				**188,065,658**
* Offices that disseminated at least one survey during Phase 1				

There will be two sampling methods used. The first method is a convenience sample. Surveys will be disseminated on participating WFOs social media pages. In addition, each WFO will enlist core partners (local weather, Emergency Managers (EMs), community groups) to share the survey links on their social media pages and email listservs to reach broader segments of the population. The goal is to create a high quality, geodemographically stratified convenience sample, called a quota sample.

Table 2 estimates surveys completed by the participating offices in Phase 2 based on results from Phase 1. In Phase 1, 15,226 surveys were completed. For Phase 2, we estimate that offices will field 1 - 4 surveys per year.

Table 2. Convenience Sample Estimated Respondents Phase II.

	Avg. Numb. WFOs (1)	Surveys per WFO (2)	Responses per Survey (3)	Total Surveys
Year 1	22	1.5	500	16,500
Year 2	35	1.5	550	28,875
Year 3	45	2	600	54,000

Total				**99,375**

1. *Targets are: 30 WFOs by end of Year 1, 40 WFOs by end of Year 2, and 50 active WFOs by project end. Numbers in Table 2 reflect average annual WFOs that are expected to participate.*
2. *Phase I average surveys per WFO was 1 (range 1 - 4); Phase II projected average surveys per WFO is 1.5-2.0.*
3. *Phase I responses per survey was 412; projections of 500 - 600 responses per survey reflect planned improvements in response.*

The second sampling method will be an internet-based probability sample. Although a probability-based sample has historically been the gold standard for this type of survey, it would be cost prohibitive to conduct the multiple surveys required for this effort. Therefore, an important topic for this study is how to collect and weight the quota sample that addresses bias issues that are common to convenience samples (see question 2), and to what extent this approach is appropriate for making statements about the respondents and for understanding causal factors. A sampling experiment will be conducted in Year 1 to compare U.S. Census-based benchmarks, a probability-based sample, and a convenience sample for 1 – 3 events (see question 3). The probability-based sample is being purchased from Ipsos Knowledge Panel. Ipsos uses probability sampling to recruit a representative U.S. online panel. Panel members are recruited using probability selection algorithms for both random-digit dial (RDD) telephone and address-based sampling (ABS) methodologies. The Ipsos response rate is expected to be 55-60% (Ipsos 2015). Table 3 estimates the number of expected respondents to be 1650.

Table 3. Probability Sample Estimated Respondents

Information Collection	Population or Potential Respondents Universe (a)	Number of Respondents Selected (b)	Maximum Ipsos surveys (c)	Expected Completion Rate (d)	Expected Number of Respondents (e) =(b) x (c) x (d)
Ipsos KnowledgePanel	All of the population within the selected Yr 1 NWS WFO County Warning Areas = **139,797,299**	600	5	55%	1,650
Total					**1,650**

References

OMB 2021. Available from https://omb.report/icr/202105-0990-004/doc/111804401

Pew Research 2020. Available from https://www.pewforum.org/2020/01/22/methodology-31/

Pew Research 2021. Available from https://www.pewresearch.org/internet/2021/04/07/social-media-U.S.e-in-2021/

Ipsos 2015. Documentation for Human Subject Review Committees: Ipsos company information, past external review, confidentiality, and privacy protections for panelists. Available from https://www.Ipsos.com/sites/default/files/Documentation%20for%20IRBs.pdf

We Are Social 2021. Digital 2021: Global October Snapshot Report. Available from https://wearesocial.com/blog/2021/10/social-media-users-pass-the-4-5-billion-mark/

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

2.1 Statistical methodology for stratification and sample selection

This research will use a demographically stratified, convenience sample, also called quota sample. The intent is to create a high-quality convenience sample that can mitigate the selection bias issues that are common with convenience samples, and therefore enable use of a standard statistical inference approach as an approximation, as suggested in Vehovar et al. 2016.

Probability-like sample properties will be introduced to the convenience sample using the following steps. First, collecting a convenience sample from multiple social media sites and email lists has been demonstrated to introduce some degree of randomness into the data because different sites address different populations. Second, using quota sampling helps to create a stratified sample which can be matched to American Community Survey (ACS) census data through sample matching. Third, sample weights will be constructed from the matched ACS sample with known population parameters using raking ratio estimation and, perhaps, propensity score matching to make the sample representative of the populations within each WFO. Lastly, these methods will be evaluated by comparing the quota sample results to the results from the Ipsos probability sample (See Question 3).

2.2 Estimation Procedure

The data collected in the survey will primarily be categorical as survey respondents select specific actions or perceptions through multiple choice questions, or they indicate the importance of varied factors using a Likert scale. Data from surveys will be aggregated at the hazard level. The primary outcome variables are: **1)** actions taken before the event (for example, Did you cancel a planned trip based on predicted flood risk?) and **2)** actions taken during the event (for example, Did you drive on a flooded roadway?). The independent variables include environmental, cognitive, situational, and demographic factors. These data will be used for two primary purposes:

1. Descriptive Statistics and tests of association. The data will be used to examine measures of central tendency, standard deviation, and distributions. It will also be used to conduct Chi-square analyses to test the association between the outcome variables and the various factors. This information will be generated on an automated basis soon after a data collection effort is complete for evaluation by NWS forecasters to understand its usefulness.
2. Logistic and Multinomial regression. Since the outcome variables are categorical or binary, non-parametric regression will be used to analyze how the different factors influence the odds of taking protective actions before or during the event. Once a model is estimated, the internal validity of the model will be evaluated using Receiver Operating Curve (ROC) analysis, looking at the area under the receiver operating curve (AUC) and K-fold cross-validation AUC scores indicate the ability of the model to correctly predict outcomes in data while avoiding false negative and false positive predictions. K-fold cross-validation divides a sample into *k* number of subsamples, each of which is tested against the larger remaining sample for similarity between analysis results. Coupling this method with AUC scoring provides a dual approach to ensuring internal validity (Kohavi 1995). In addition, once the internal validity of the model is

demonstrated, marginal probabilities will be computed for taking protective action given a specific factor.

3. In addition, statistical techniques such as latent class analysis, mediating variables will be used to understand if there are relevant groups within the sample.

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

The analysis will focus on understanding how different factors influence protective action decisions. The degree of accuracy needed to analyze different groups will be addressed through power analysis. Power analysis evaluates the ability to detect differences between two groups in a given dataset. Put another way, it enables researchers to correctly reject the null hypothesis that two samples are taken from the same distribution. In current practice, most data collection efforts look for 0.8 power, meaning there is an 80% probability that the difference between the two groups will be detected, if it exists. Power is a function of sample size, the level of precision desired, the anticipated prevalence of the group in the sample, and effect size. Effect size measures the marginal distance between groups and is usually obtained empirically through prior studies or a pilot project (Durlak, 2009). This logic can be applied to any subgroup that might be analyzed on its own, and the proportion can be changed to suit other hazards.

2.4 Unusual Problems and Use of Less Frequent Data Collection Cycles.

There are no unusual problems requiring specialized sampling procedures. Because of the number of WFOs involved, a less than annual data collection will not be a factor. In this study, surveys will be fielded in the first year and from then, the expectation is to increase the number of participating offices until a decision is made about whether to expand the surveying to all offices. The collection depends on the frequency of weather events. In case of an unusually active weather season, only the most impactful events are surveyed.

References

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., ... & Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. Journal of survey statistics and methodology, 1(2), 90-143.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of pediatric psychology*, *34*(9), 917-928.

Ender, P. B. (2011). STATA: Power Logistic Regression (Powerlog): Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.

Hays, R. D., Liu, H., & Kapteyn, A. (2015). Use of Internet panels to conduct surveys. Behavior research methods, 47(3), 685-690.

Kim, J. K., & Wang, Z. (2019). Sampling techniques for big data analysis. International Statistical Review, 87, S177-S191.

Kohavi, R. (1995, AugU.S.t). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai (Vol. 14, No. 2, pp. 1137-1145).

MacInnis, B., Krosnick, J. A., Ho, A. S., & Cho, M. J. (2018). The accuracy of measurements with probability and nonprobability survey samples: Replication and extension. Public Opinion Quarterly, 82(4), 707-744.

Madigan, D., Stang, P. E., Berlin, J. A., Schuemie, M., Overhage, J. M., Suchard, M. A., ... & Ryan, P. B. (2014). A systematic statistical approach to evaluating evidence from observational studies. Annual Review of Statistics and Its Application, 1, 11-39.

NOAA (2021). Storm Events Database. Available from https://www.ncdc.noaa.gov/stormevents/

Vehovar, V., Toepoel, V., & Steinmetz, S. (2016). Non-probability sampling. The Sage handbook of survey methods, 329-345.

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

Response Rates

Through Qualtrics and a Survey Dashboard that will be created as part of this project, we will monitor the demographic characteristics of the survey responses in real time compared to benchmark American Community Survey statistics to ensure that we are getting the demographic representation needed for increased generalizability and number of responses required for statistical power. As the research progresses, we will evaluate the quality of the response and adjust frequency of reposting links, and the types of organizations that post the links to the survey. This approach is based on several studies (see Antoun et al., 2016; Perrotta et al., 2021; Vehovar et al., 2016) that found quota sampling through Facebook is a systematic strategy that can be used to obtain a stratified sample approximating American Community survey results when considering multiple sampling frames and randomization.

After we develop the process for monitoring survey response, we will ask WFOs to track the time it takes for forecasters to initiate and monitor data collection to understand the burden it will place on WFOs.

Sampling Experiment

In addition to ensuring representativeness in the data collection process as outlined in the section above, the quality and accuracy of the quota sample will be assessed through a sampling experiment where we compare the quota sample results and Ipsos sample results for the same weather event. For the QRS project sampling experiment, we will field identical surveys to an Ipsos sample and a QRS convenience sample for counties in a Metropolitan Statistical Area (MSA) that has experienced a widespread hazard. We will first compare the American Community Survey basic demographics (a benchmark) to the Ipsos survey to calibrate its accuracy. We assume good accuracy based on previous studies and Ipsos' panel management. We will then compare demographic and non-demographic variables in the Ipsos surveys to the QRS convenience sample. We will compare point estimates, overall measures of error across key variables, and compare multivariate regressions for each sample. We will evaluate demographic and non-demographic weighting strategies. In previous studies that have compared probability and nonprobability samples, the probability samples were generally more accurate; however, there was a wide-range of quality for the non-probability samples, with some nonprobability samples having accuracy that was very close to the probability samples. In addition, accuracy can be determined for different demographic or attitudinal segments. The sampling experiment will provide valuable information on biases that may exist in the quota sample and the source of the bias that might lead new weighting strategies or data collection strategies; the sampling experiment can also reveal population groups that might not be adequately represented in the quota sample; however, it may also show that the probability sample does not have sufficient power for certain kinds of analyses. Based on the results of this test, we can begin to assess the potential for inference for this data collection by the quota sample effort and what kinds of caveats should be associated with the data.

References

Antoun, C., Zhang, C., Conrad, F. G., & Schober, M. F. (2016). Comparisons of online recruitment strategies for convenience samples: Craigslist, Google AdWords, Facebook, and Amazon Mechanical Turk. Field methods, 28(3), 231-246.

Callegaro, M., Baker, R. P., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (Eds.). (2014). Online panel research: A data quality perspective. John Wiley & Sons.

CFI Group (2018a). National Weather Service Webmonitor Results Q3 FY2018 (April – June 2018). Retrieved with permission from the NWS Performance Management Web Portal.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus. the Internet: Comparing sample representativeness and response quality. Public Opinion Quarterly, 73(4), 641-678.

Dutwin, D., & Buskirk, T. D. (2017). Apples to oranges or gala versus golden delicious? Comparing data quality of nonprobability internet samples to low response rate probability samples. Public Opinion Quarterly, 81(S1), 213-239.

Perrotta, D., Grow, A., Rampazzo, F., Cimentada, J., Del Fava, E., Gil-Clavel, S., & Zagheni, E. (2021). Behaviours and attitudes in response to the COVID-19 pandemic: insights from a cross-national Facebook survey. EPJ data science, 10(1), 1-13.

Malhotra, N., & Krosnick, J. A. (2007). The effect of survey mode and sampling on inferences about political attitudes and behavior: Comparing the 2000 and 2004 ANES to Internet surveys with nonprobability samples. Political Analysis, 15(3), 286-323.

NOAA 2021. Storm Events Database. Available from https://www.ncdc.gov/stormevents.html

Nunley, C., & Sherman-Morris, K. (2020). What people know about the weather. Bulletin of the American Meteorological Society, 101(7), E1225-E1240.

Rivers, D. (2007, AugU.S.t). Sampling for web surveys. In Joint Statistical Meetings (p. 4).

Vehovar, V., Motl, A., Mihelič, L., Berčič, B., & Petrovčič, A. (2012). Zaznava sovražnega govora na slovenskem spletu. Teorija in praksa, 49(1), 171-189.

Yeager, David S., et al. "Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples." Public opinion quarterly 75.4 (2011): 709-747.

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

The QRS were tested in Phase I and no additional testing is needed.

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

Quick Response Surveys contacts who were consulted on statistical aspects of the design and will collect and analyze the survey data for the agency include:

Brenda Philips, (413) 577-2213; bphilips@engin.umass.edu

Cedar League, (406)202-8167; cedarleague@gmail.com

Nathan Meyers, (734)771-7323; npmeyers@soc.umass.edu

David Westbrook, (413) 522-1409; westy@cs.umass.edu