Understanding people's evolving risk assessments and decisions during tropical cyclone threats: Design and implementation of a novel, longitudinal, real-time survey methodology

Julie L. Demuth[1], Rebecca E. Morss[1], Gabrielle Wong-Parodi[2], Andrea B. Schumacher[1,4],
Joshua J. Alland[3], Dakota Smith[4], Natalie Herbert[2], Hugh D. Walpole[1]

[1] Weather Risks and Decisions in Society Research Group
Mesoscale and Microscale Meteorology Lab
National Center for Atmospheric Research

[2] Woods Institute for the Environment
Department of Earth System Science
Stanford University

[3] University Corporation for Atmospheric Research

[4] Cooperative Institute for Research in the Atmosphere
Colorado State University

Final report to NOAA/OAR/WPO and NOAA/NWS
June 2023

Suggested citation for this report

Demuth, J. L., R. E. Morss, G. Wong-Parodi, A. B. Schumacher, J. J. Alland, D. Smith, N. Herbert, and H. Walpole. (2023). *Understanding people's evolving risk assessments and decisions during tropical cyclone threats: Design and implementation of a novel longitudinal, real-time survey methodology.* Final report to NOAA, 79 pp. https://doi.org/10.25923/wk8q-k529

Suggested citation for the survey instruments for 2020 Hurricanes Laura and Marco

Demuth, J., R. Morss, G. Wong-Parodi. (2023) "2020 Hurricanes Laura and Marco predictive and post-storm survey instruments", in *Public longitudinal panel surveys collected during and after hazardous weather threats: Hurricanes*. DesignSafe-CI. https://doi.org/10.17603/ds2-j9hc-xy24

Suggested citation for the survey instruments for 2021 Hurricane Henri

Demuth, J., R. Morss, G. Wong-Parodi, A. Schumacher, N. Herbert, H. Walpole. (2023) "2021 Hurricane Henri predictive and post-storm survey instruments", in *Public longitudinal panel surveys collected during and after hazardous weather threats: Hurricanes.* DesignSafe-CI. https://doi.org/10.17603/ds2-05s8-ah21

Table of Contents

List of Figures

List of Tables

Executive Summary

Meteorological and computational advances are making more and better information more frequently available when a tropical cyclone (TC) threatens, including as it and the risks it poses evolve. Yet, very little is known about whether and how people are obtaining evolving TC risk information or how their interpretations and responses are changing (or not) with the changing risks. For NOAA and its partners to fully assess the value of evolving TC forecasts and related information they provide and to identify where improvements are needed, it is essential to understand people's risk information obtained, risk perceptions, and behavioral responses and how these evolve for an active, real-world TC as it evolves. To meet this need, the goals of the project described herein are to:

- Goal 1: Develop and demonstrate a novel, rapidly deployable longitudinal panel survey methodology for collecting real-time data from at-risk members of the public during a TC that is threatening the mainland United States.
- Goal 2: Develop research-guided recommendations to NOAA on modifying and expanding this methodology to collect these types of data for future TC threats.

This report documents our design and implementation of this novel methodology that conducts multiple surveys that ask the same questions repeatedly of the same individuals (i.e., a longitudinal panel survey) over multiple days for an active, real-world TC as it evolves. This systematic collection of perishable social science observational data allows us to measure whether, when, and how people get TC risk information, perceive the risks, and respond. We report on the implementation of our longitudinal panel survey approach twice, in two different geographic areas and over two hurricane seasons: first in 2020 for TCs Laura and Marco in the Gulf and second in 2021 for TC Henri in New York and New England.

Here, we summarize the findings and recommendations to NOAA for future work, all pertaining to the methodology. Additional supporting content for each of the findings and recommendations is provided in the main body of the report.

FINDING 1. Successfully designing and implementing a novel, rapidly deployed, event-specific longitudinal panel survey during the multi-day predictive phase of a real-world TC required (a) identifying the methodological challenges in detail, (b) assessing the feasibility of overcoming those challenges, (c) developing a detailed but flexible research approach, and (d) working with "a cross-sector village" of committed research team members and external collaborators who contribute the diverse, needed forms of expertise.

FINDING 2. The proof-of-concept methodology yielded convenience samples that are older, more White, more retired, more educated, and well-resourced with adequate access to basic needs. Nevertheless, these samples provide a basis for exploring modifications to the fielding approach and incentive structures to get more diverse or more representative samples.

FINDING 3. Curating the meteorological data and evacuation orders for integration and analysis with our survey data (a) is a valuable methodological contribution in its own right, (b) but was time-consuming and challenging to do, even with the meteorological expertise of our team members.

FINDING 4. Our planned survey fielding approach was that it would take 6 days from the fielding decision to fully field 3 predictive survey waves—i.e., 24 hours to put the survey in the field, 24-hour fielding periods for each predictive survey wave, and 24-hour intervals between waves. In practice, some aspects took less time, and we explored spending less time in some ways. This suggests there is feasibility to getting a sufficient sample for three predictive survey waves for TCs for which there is less lead time before landfall, but these advantages must be weighed against potential disadvantages.

RECOMMENDATION 1. The longitudinal panel survey methodology designed and implemented here that is event-specific and rapidly deployed during the predictive phase of a hazardous weather threat yielded collection of novel, perishable social science observational data for real-world TCs as they evolved. Collection of such data for additional TCs should be prioritized to develop more comprehensive datasets that will facilitate more robust understanding of people's perceptions and behaviors in response to forecast and other risk information provided by NOAA and its partners when TCs threaten. This actionable social science in turn will help identify where improvements are most needed in NWS's forecast product suite content and in dissemination of information across the forecast and response system.

> **RECOMMENDATION 1a. Different sampling approaches should be developed to acquire more diverse survey samples, including targeted efforts toward (a) more socioeconomic diversity, including more vulnerable populations, and (b) samples that are representative of the population in the geographic areas at risk from different TCs.**

RECOMMENDATION 2. A mechanism could be developed for analyzing in near real time the rapidly deployed, event-specific social science observational survey data to identify critical misperceptions and/or lack of awareness about TC risks and to provide near real-time, actionable input to NWS to guide TC forecast messaging interventions—in other words, to operationalize "incident" TC risk communication alongside operational TC forecasting. NOAA should invest in research and development to explore developing this capability.

RECOMMENDATION 3. It should be explored how to expand the longitudinal panel survey methodology designed and implemented here (event-specific, rapidly deployed, during the predictive phase) to other types of hazardous weather threats, beyond TCs. For example, this methodology could be used for weather threats that tend to be longer-fused and spatially broad (e.g., winter storms, atmospheric rivers, heat) and to threats that tend to be shorter-fused and spatially localized (e.g., severe convective storms, fire weather).

RECOMMENDATION 4. NOAA/NWS meteorological data and products as well as associated emergency response orders should be made more easily accessible to a broad range of researchers and other users, with consistent data formats, clear and long archival periods, and standardized units (when possible).

RECOMMENDATION 5. A dashboard or other web-based platform should be developed to make publicly available the longitudinal social science observational data as well as the

corresponding meteorological and evacuation data. These data should be accompanied by detailed metadata about survey development, data quality control, data treatment, and data source.

1. Introduction

Over the last few decades, advances in atmospheric science and technology have led to dramatic improvements in tropical cyclone (TC) prediction. This has enabled the National Weather Service (NWS) to provide the nation with skillful TC forecasts and warnings at longer lead times and with more localized forecasts for specific TC-related hazards when a TC threatens. In turn, this has allowed NWS's partners in emergency management and broadcast meteorology to provide emergency preparedness and response and additional forecast information to members of the public. At the same time, advances in information and communication technology have made a growing volume of TC risk information available and have transformed capabilities for people to access, share, and use it (Morss et al. 2017).

In other words, meteorological and computational advances are making more and better information more frequently available when a TC threatens, including as it and the risks it poses evolve. *Yet, very little is known about whether and how people are obtaining evolving TC risk information or how their interpretations and responses are changing (or not) with the changing risks.*

A wealth of empirical research has been conducted to study what TC information members of the public obtain, how they understand it, and how they make decisions (see, e.g., reviews by Baker 1991, Dash and Gladwin 2007, Lindell 2012, Lazo et al. 2015, Huang et al. 2016, Tanim et al. 2022), including studies that members of our own research team have conducted (Morss and Hayden 2010; Demuth et al. 2012, 2016; Lazrus et al. 2012; Bostrom et al. 2016, 2018; Cuite et al. 2017, Morss et al. 2016; Wong-Parodi and Feygina 2018; Wong-Parodi et al. 2018). This body of work has developed valuable knowledge about how people's perceptions of hurricane risks and their protective decisions are influenced by a variety of factors, ranging from sociodemographic characteristics to situational factors to risk messages. However, this research utilizes cross-sectional data (collected at one point in time) through surveys, survey-based experiments, or interviews to understand people's perspectives at a specific point in time or integrated over time. Further, the data are collected retrospectively after a TC, for hypothetical events, or about TC risks in general.

Yet, real-world TC risks are dynamic, and the forecast and preparedness information that NOAA and its partners provide about TCs evolve as the risks evolve over the course of several days or longer (Morss et al. 2017). Although some studies have explored how people get information and manage evolving TC risks using retrospective interviews (Gladwin et al. 2001; Taylor et al. 2009; Morss and Hayden 2010) or simulations (Meyer et al. 2013; Wu et al. 2015a,b), these retrospective and hypothetical studies may not capture important details about what people are doing, thinking, and feeling at different times during an actual threat, when risks and the need to make decisions are real and when stress and uncertainty can be high. Meyer et al. (2014) began to fill this gap by gathering real-time data from at-risk members of the public during Hurricanes Isaac and Sandy using multiple telephone surveys during the same hurricane threat, but their data were gathered from different people at different times. Anderson et al. (2016) and Demuth et al. (2018) approached analyzing data from the same people over time through analysis of Twitter narratives from people residing in mandatory evacuation zones during Hurricane Sandy, but tweets offer unstructured data about what people want to share and when.

For NOAA and its partners to more fully assess the value of the evolving TC forecast and related information they provide and to identify where improvements are needed, it is essential to understand people's risk information obtained, risk perceptions, and behavioral responses and how these evolve with the TC. Yet, NOAA currently has no means available for empirically and systematically assessing *in real time* how a broad sample of people who are at risk from a TC respond to the forecast information that NWS provides *as TC risks evolve*. Operational implications of this lack of knowledge are that NWS forecasters must decide how to adapt their TC communication and messaging strategies in the moment based on evidence about how people are responding that may be anecdotal or highly localized from partners (e.g., shelter reports). Research implications of this lack of knowledge are that NOAA personnel must decide how to invest in TC product suite improvements with limited understanding about how members of the public are consuming and processing the evolving, uncertain TC risk information available. These operational and research implications are exacerbated by variations across different TCs in their predictability, evolution, hazards, and impacts. In other words, NOAA not only lacks knowledge about how people interpret and respond to a given TC as it evolves but they also do not know whether and how the patterns of people's interpretations and responses are consistent or different from TC to TC.

To fill this gap, *a longitudinal panel survey—which collects multiple surveys (i.e., waves) asking the same questions repeatedly over time of the same individuals—is needed to develop knowledge about whether, when, and how people get TC risk information, perceive the risks, and respond for an active, real-world TC as it evolves.* However, there are multiple challenges with systematically collecting this type of data, including:

a) Lead time – Identifying a TC risk to the mainland United States with sufficient lead time to conduct multiple survey waves during the *predictive phase*, i.e., while the TC is threatening but before it makes landfall or causes impacts in the area being sampled, given predictive uncertainties;

b) Sampling – Identifying a target population to sample from that is over a large enough geographic area to yield a sufficiently large sample for data analysis of respondents who will respond to multiple surveys over a short time period (see item c) accounting for attrition (see item d) and that captures people who are and will continue to be at risk as the TC threat evolves, given predictive uncertainties;

c) Repeated responses from an individual – Getting an individual to respond to repeated surveys over a time period of only a few days, and developing an approach to confirm that the same person is responding to each survey wave;

d) Attrition – Losing respondents from wave to wave due to the novel nature of this longitudinal panel survey in which people are asked to respond to multiple surveys over only a few days and are facing an active TC threat that they may need to prepare for and/or evacuate from;

e) Fielding – Fielding each survey wave for long enough to get a sufficiently large sample size while also considering time constraints of having intervals between waves (see item f) and wanting to field three survey waves during the predictive phase of a TC threat;

f) Wave intervals – Allowing a long enough interval between survey waves to be able to detect changes, if they occur, in people's responses, while also considering time constraints of fielding three survey waves during the predictive phase of a TC threat (see item e);

g) Survey design – Designing the survey such that it can validly measure key variables of interest and potential changes in those variables over a short period of time given wave interval length;

h) TC predictability limitations – Managing uncertainty about where the TC may make landfall and otherwise pose hazards at the multi-day lead time when the decision must be made to field the survey and for what geographic areas, all of which means deciding on-the-fly where to survey, which in turn presents challenges for the survey company;

i) Survey company – Identifying a survey company that has the necessary potential respondent panel, technological capacity, and flexible survey deployment capability to field such a longitudinal panel survey, and that is willing to work closely with the research team to design a feasible fielding strategy and successfully implement it, given the other challenges and the uncertainties associated with developing this new method; and

j) Cost – Implementing the survey at a reasonable cost level, given all of the other challenges.

Recognizing the knowledge gap and challenges described above, there are two primary goals of the project described herein.

- Goal 1: Develop and demonstrate a novel, rapidly deployable longitudinal panel survey methodology for collecting real-time data from at-risk members of the public during a TC that is threatening the mainland United States.
- Goal 2: Develop research-guided recommendations to NOAA on modifying and expanding this methodology to collect these type of data for future TC threats.

Achieving these goals will further build NOAA's knowledge base and methodological toolkit—particularly its capacity to collect event-specific, perishable[1] social science observations (NOAA 2021)—to answer questions about how members of the public are behaving in response to TC forecasts and other risk information that is being provided. The knowledge gained by implementing this methodology over time and over multiple TCs will enhance NOAA's ability to prioritize and evaluate TC product suite improvements, and improve the effectiveness of its forecast and warning communication to provide relevant information that helps people understand and manage approaching TC threats, thereby enhancing human health and safety and reducing economic disruptions.

This report details the methodology developed and demonstrated for two longitudinal panel surveys: for TCs Laura and Marco in 2020, which jointly threatened the Gulf Coast where Laura made landfall and Marco nearly made landfall, and for TC Henri in 2021, which made

[1] Perishable data are highly transient data that may degrade in quality, be irrevocably altered, or be permanently lost if not collected soon after such data is generated. Perishable data collection most often occurs immediately before, during, or in the direct aftermath of a disaster (CONVERGE 2023).

landfall in New England. Initially, we had funding to implement the longitudinal panel survey for only one TC event in 2020, and thus we designed the methodology with this in mind. NOAA later provided funding to implement the survey for two more TC events, one of which occurred in 2021 and the other of which is pending. We therefore structured this report to foreground the description of the methodology for Laura and Marco in 2020, with changes or specifics for Henri in 2021 described separately in most sections. We describe the development of the longitudinal panel survey research approach (Section 2), development of the survey design (Section 3), fielding of the two longitudinal panel surveys (Section 4), the survey samples (Section 5), access and integration of meteorological data with the survey data for enhanced data analysis (Section 6), and a summary of our findings and recommendations for future work (Section 7).

2. Developing the Longitudinal Panel Survey Approach: Survey Parameters, Sample, Programming and Fielding, Incentives, and Costs

As described above, there are many challenges with fielding a longitudinal panel survey for an active, real-world TC, given the aim of collecting multiple waves of survey data during the predictive phase to evaluate how people assess and manage evolving TC risks. Accordingly, the primary goal of this project was to develop and pilot a methodology for collecting such data. In other words, the goal was to determine whether and how this method could work in a way that yielded a viable dataset for analysis.

Our first step in designing a research approach was to work with a survey company to determine the feasibility of successfully conducting a longitudinal panel survey, particularly during the predictive phase of a TC, given the challenges noted above. We contacted multiple survey companies with our initial ideas of parameters for survey fielding and sampling, and we discussed initial details with them.

Ultimately, only one survey company, YouGov[2] (https://today.yougov.com/), indicated that our idea was feasible, challenges notwithstanding, and was willing to work with us to more fully develop an approach. We developed a close, collaborative relationship with YouGov in which we interacted regularly, which was essential to the success of the surveys. YouGov was highly responsive, flexible, and accommodating while also being rigorous and honest about what they could and could not promise. We worked closely with them to develop our survey parameters, sampling approach, and fielding approach. We also developed scenarios for the survey costs, including different incentive structures depending on the re-contact rates. A copy of the document that NCAR provided to YouGov when establishing our contract with them (for the original longitudinal hurricane survey in 2020) that details these elements is provided in Appendix A. Below, we summarize the key elements.

Through initial discussions with YouGov, we decided that, to maximize feasibility for the proof-of-concept, the surveys would be web-based (but mobile-friendly), English-speaking only, and with respondents from YouGov's existing survey panel. We planned to field two to three survey waves during the predictive phase, when a TC is approaching and threatening the mainland United States, and a final survey wave post-storm. With input from YouGov, we decided to try fielding each predictive wave for 24 hours and have 24-hour intervals between

[2] YouGov is a reputable international survey research company that has a large existing panel of members for surveys such as ours.

waves. With three predictive waves, this timeline translates to 120 hours (5 days) needed to field. To enable YouGov to field the survey as quickly as possible, most of the survey programming was to be completed well in advance (as further discussed below) and potential sampling areas would be predefined. Still, YouGov indicated they would need approximately 24 hours (1 day) after we made the decision to field the survey for finalizing logistics, specifically so they could program storm-specific wording (e.g., the TC name) into the survey, program the geographic areas to survey for, and prepare the recruitment emails. Thus, the total planned time needed to implement the final survey logistics, field up to three predictive waves, and have intervals between the predictive waves was 144 hours (6 days) (Figure 1). In other words, our original approach required that we needed to make a decision to field the longitudinal panel survey for a threatening TC at least six days before it could make landfall in the mainland United States.

To be clear, we did not know whether allotting only 24 hours for each predictive survey wave to be in the field would yield our desired sample size (discussed below), nor whether 24-hour intervals between predictive survey waves would be a sufficient amount of time for people to change their perceptions and behaviors. However, allowing more time for each of these steps would translate to having to make a decision to field at an even longer lead time, which seemed unrealistic for most TC scenarios given predictability limitations. Again, this was a pilot effort of methodology, so we considered this a reasonable attempt, knowing that we would learn valuable lessons regardless of the outcome.

Following the predictive survey waves, we planned to field a post-storm wave approximately 7 to 14 days after the TC made landfall (if it did) or dissipated (Figure 1). If the TC we studied caused extensive and/or considerably severe impacts, our ethic was to field the post-storm survey later to give people time for immediate response and short-term recovery before asking them to complete a survey. We planned to field the post-storm survey for approximately 10 days to give people ample time to respond and to maximize the number of responses, although we discussed with YouGov the potential of leaving the post-storm survey open longer, if the initial response rate was low.

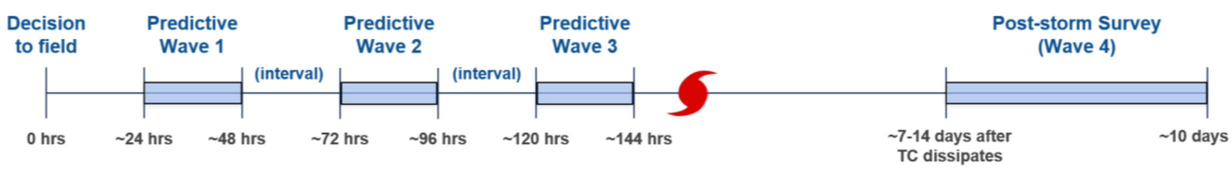


Figure 1. Conceptual depiction of the timeline needed for deciding to field the survey, fielding three predictive waves with wave intervals, and fielding the post-storm survey.

We planned to design the survey so that the same core set of questions would be asked in each of the predictive survey waves. These repeated measures are, in part, what allow for statistical analysis of whether, when, and how people's assessments, perceptions, and responses change over time. The post-storm survey wave would include many analogue questions to measure what people thought retrospectively and what behaviors they did or did not engage in. Furthermore, we asked several additional questions one time on either Wave 1 or Wave 4, to gather data on additional concepts and covariates of interest that we anticipated would not change during the time period of interest, as further explained in Section 3.

For the longitudinal data analysis, we wanted the final dataset for analysis to be *balanced*, meaning we have data from a given respondent for all survey waves (i.e., on all measurement occasions). To ensure it was the same person responding to all waves, YouGov asked questions on each survey wave about birth year and gender, and respondents were either matched based on these two characteristics or were eliminated from the sample.

We wanted a final, balanced sample of at least n=700 people who responded to all survey waves based on power analysis estimates. To determine whether this was feasible (and to develop a feasible sampling strategy), we had to consider the number of people available on YouGov's existing panel in different geographical areas, and we had to consider attrition rates from wave to wave. With the help of NCAR's GIS program, we developed a list of all coastal zip codes that were within 50, 75, and 100 miles from the coast[3] in the Gulf of Mexico and the Atlantic Ocean, i.e., from Texas to Maine, stratified by state. Florida was the only state we subdivided (by bisecting the peninsula), to give us the option to field in the western portion of Florida for a Gulf TC or in the eastern portion for an Atlantic TC. We provided this listing of zip codes to YouGov who then provided us with their panel feasibility numbers by the three distances from the coast and by state (Appendix A, Table A1).

To develop a sampling strategy, we needed to consider potential attrition rates (how many respondents might drop out) from wave to wave in conjunction with our desired final sample size. Because attrition rates were unknown for a rapid-turnaround longitudinal survey during an active hazard threat, YouGov developed two scenarios: one with a conservative, higher attrition rate across waves and one with an optimistic, lower attrition rate. We also offered monetary incentives to respondents, to incentivize them to complete all survey waves. YouGov's scenarios therefore included a higher incentive amount for the post-storm wave if attrition rates were high (in other words, if we were losing more respondents between predictive waves and wanted to motivate more respondents to complete the final wave) and consistent incentive amounts for all waves if attrition rates were lower. Details of these two scenarios are provided in Appendix A and Tables A2 and A3, respectively.

Based on the conservative scenario, we estimated we would need a sample of at least n=1277 completed respondents at Wave 1, which in turn meant that (based on YouGov's existing panel) we would have to field the survey over multiple states. Fortunately, this approach to get a sufficient sample—fielding over a broad area along the coast—corresponded with the large geographic area that could be at risk of TC impacts at a six-day lead time when we had to make the decision to initiate fielding the surveys.[4] We therefore planned to field over a large geographic area, which would be determined by the areas at risk from the TC, and to continue surveying the same people as the TC risks evolved, which likely would include areas where the risks decreased and areas where they increased. Additionally, having predetermined zip codes that were within 50, 75, and 100 miles inland from the coast meant we could choose to field farther inland, thereby also fielding over a larger geographic area, if it seemed that a given TC would pose substantial risks inland.

---

[3] We included zip codes that had any part of its area within these distances, meaning we did not require the full zip code area or some threshold proportion of the area to be within these distances.

[4] We recognized that, due to predictability limitations, a TC might dissipate or recurve away from the mainland United States. Accordingly, we built into the methodology the possibility that we could not continue fielding for all later survey waves.

In order to be ready to field the survey when there was a TC we wanted to study, we did all of the work on the survey instrument development, web programming, sampling, and other planning well in advance of the June 1 beginning of the Atlantic hurricane season. We spent several months in Winter 2020 developing the predictive and post-storm surveys (discussed in Section 3). In Spring 2020, we provided the final surveys to YouGov who programmed them, and then we tested the survey links and YouGov accordingly made corrections and visualization and formatting changes. We also pretested the survey (discussed more in Section 3) and made another round of changes accordingly. Also, prior to hurricane season, we predetermined the different sampling and incentive strategies, and we developed a communication plan among the research team to monitor TCs (see Section 4a) and to communicate with YouGov and with our NOAA collaborators. By the beginning of the tropical season, our surveys were finalized and ready for fielding. The only outstanding specifics to provide to YouGov were the TC name, which was programmed into survey because we structured the survey questions to specifically refer to the storm, and which states we wanted to field for and how far inland from the coast. To facilitate these final steps, if we were monitoring and considering fielding for a TC, we contacted YouGov with a "heads-up" email with the TC name and geographic areas we were targeting.

a. Changes to approach for 2021 surveys (of TC Henri)

We implemented two substantial changes to our longitudinal panel survey approach in 2021. One change was that we planned for fielding each predictive wave for 18 hours instead of 24 hours. This was based on our analysis from the 2020 survey, which showed that 92–100% of responses were received within 18 hours. The other change was that during the predictive phase of surveying, we allowed for *unbalanced* data, meaning that participants could miss responding to a wave (i.e., data not on all measurement occasions).[5] We required responses for Wave 1, the first predictive survey, and for Wave 4, the post-storm survey. However, respondents could skip responding to either (but not both of) predictive Waves 2 or 3. Thus, the three possible response patterns for the 2021 survey were:

- Balanced responses to Wave 1, Wave 2, Wave 3, and Wave 4
- Unbalanced responses to Wave 1, Wave 2, and Wave 4
- Unbalanced responses to Wave 1, Wave 3, and Wave 4.

3. Survey Instrument Design

Because TCs pose risks and because NWS's mission to protect life and property accordingly involves providing forecast information to reduce risks, we designed the survey to focus on measuring multiple dimensions of a few core risk-related concepts: *risk information, risk perception, efficacy beliefs,* and *behavioral responses.* By gathering data about the same measures multiple times, we are able to evaluate whether, when, and how different aspects of people's risk assessments and decisions changed as TC risks and associated forecasts evolved. Below, we provide details of how we specifically measured different dimensions of these concepts, along with other data we collected as covariates (to control for their effects on the dependent variables of interest) and as moderators (to examine whether and how relationships

[5] The statistical analysis for longitudinal multilevel modeling does not require having data from respondents on all measurement occasions, and thus it can be done with unbalanced data. However, later when doing more detailed analysis of our data, we determined that it is highly beneficial to have measures over at least three time points, i.e., waves, for more robust analysis and interpretation.

between two variables depend on the level of a third variable). Due to predictability limitations at the approximately six-day lead time when the survey fielding decision had to be made (see Section 2), we designed the survey so that it was not required that a TC make landfall for the data to be useful. Rather, the survey questions were developed to gather data from people who might be tracking the TC or consider themselves at risk at some point during the threat, regardless of how the TC manifested.

Table 1 provides a high-level summary of all survey concepts and dimensions included in the survey instruments, with the predictive survey measures in the left-hand column and with the corresponding post-storm survey measures in the right-hand column. The predictive surveys measured what people were thinking and doing about the TC risks in real time as the TC was threatening, and the post-storm survey asked questions about what people experienced and thought after the fact. Full copies of the predictive and post-storm surveys are provided in Appendices B and C, respectively.

The predictive survey began with informed consent, followed by questions about birth year and gender, which were used to match respondents across waves, as discussed above. Following that, the initial TC-specific survey question asked, “Are you aware that Tropical Cyclone [name] is currently in the [Atlantic / Gulf of Mexico / Caribbean Sea]?”, where the fields denoted in brackets were populated with the storm-specific information. If people said they were not aware of the TC, they were asked a few open-ended questions to gather general data about where they typically get TC forecast information, their perceptions of the likelihood of the United States and the area where they live being affected by a TC in the next two weeks, and their perceptions of what the impacts would be. If people said they were aware of the TC, they were asked a set of questions to measure the core risk-related concepts of *risk information*, *risk perception, efficacy beliefs,* and *behavioral responses*, described further below.

We designed the TC awareness question assuming that the survey would pertain to only one TC. However, as described in Section 4b, TCs Laura and Marco were threatening the same geographic area for the first longitudinal survey we fielded in 2020. We therefore worked with YouGov (who were very accommodating) to quickly modify this question in the hours before fielding to account for two TCs by asking whether respondents were aware of either named storm. For people who indicated they were aware, we also added a question to ask them to indicate which storm was more relevant to where they live. The storm name of the TC they selected as more relevant (Laura or Marco) was populated into the storm name field in the questions about the core risk-related concepts, and the rest of the survey proceeded as originally designed.

For the post-storm survey, the initial TC-specific question asked, “Were you aware of [Hurricane / Tropical Storm] [name], which recently occurred?”, where the fields denoted in brackets were populated with storm-specific information. For the 2020 longitudinal survey, this question was restructured to ask about Laura and Marco. If people were aware of either or both of these storms, they were asked which was more relevant to where they live. If they were not aware, they were asked other questions not specific to the TCs, further described below.

Risk information includes any information about the potential threat, severity, consequences, or recommended actions associated with a risk (NRC 1989). This includes forecast information, information about recommended preparedness or protective actions, environmental cues, and social cues, all of which can be obtained from many different sources via many different channels. We drew on past research by members of our research team and by other scholars to develop risk information measures for this survey (e.g., Morss and Hayden 2010; Demuth et al. 2012; Meyer et al. 2013, 2014; Lazo et al. 2015; Demuth et al. 2016; Trumbo et al. 2016; Morss et al. 2016, 2018; Cuite et al. 2017; Bostrom et al. 2018; Wong-Parodi and Feygina 2018; Wong-Parodi et al. 2018).

For the predictive surveys specifically, we developed questions pertaining to TC [name] to measure respondents': (a) frequency of getting information about the storm from different sources (e.g., NWS, local TV meteorologist) and channels (e.g., website, TV, social media), including from environmental cues; (b) channels used to obtain NWS information, for those who indicated they did; (c) frequency of information seeking overall; (d) frequency of seeing or hearing about other people doing things related to the storm, i.e., of social cues; (e) perceived importance of different kinds of forecast information (e.g., storm track, wind speeds); and (f) perceived information of other kinds of risk information (e.g., potential impacts, how to prepare). All of these questions were asked with a specific time referent of the prior 24 hours, for instance, how often respondents got information from different sources during the last 24 hours.

For the post-storm survey, we asked two open-ended questions about risk information, one about what forecast information about TC [name] respondents wished they had had but did not and a second similar question about information about preparing for TC [name]. Then, to match the questions from the predictive survey, we asked questions to measure respondents' perceived usefulness of (a) different sources and channels for obtaining forecast or preparedness information; (b) different types of forecast information; and (c) other kinds of risk information (e.g., potential impacts, how to prepare).

Risk perception includes a person's beliefs, attitudes, judgments, and feelings about events, situations, or activities that could lead to negative consequences (Pidgeon 1992; Renn 2008), which are made through the process of collecting and interpreting signals about uncertain impacts of events (Wachinger et al. 2013). Risk perception has been studied for decades, in many different ways, a synthesis of which is beyond the scope of this report. To design the risk perception measures for these surveys, we drew on past research by members of our research team and by other scholars (Lazo et al. 2015; Demuth et al. 2016; Trumbo et al. 2016; Demuth 2018; Morss et al. 2016, 2018; Wong-Parodi and Feygina 2018; Wong-Parodi et al. 2018; Walpole and Wilson 2020), and we extended this research to include a variety of risk perception measures relevant to the context of an evolving TC. This includes measures related to exposure (the chance of being affected by a hazard), susceptibility (the chance of negative impacts due to a hazard), severity (how bad negative impacts could be), and negative affect (bad feelings about the hazard) (e.g., Slovic et al. 2004; Trumbo et al. 2016; Walpole and Wilson 2020).

For the predictive surveys specifically, we developed questions to measure respondents' risk perceptions of two risk targets: the mainland United States and the respondent themselves (Sjoberg 2000). For the United States as a risk target, we measured respondents' perceptions of the mainland United States' (a) exposure to the TC overall; (b) TC intensity at landfall; (c) time until TC effects begin; (d) exposure to different hazards that might be caused by the TC (e.g., strong winds, flooding due to storm surge); (e) susceptibility to negative impact overall; and (f) susceptibility to different types of negative impacts (e.g., power outages, people injured). For the respondent personally as a risk target, we measured their perceived (a) exposure to the TC overall; (b) probability of being exposed to tropical storm-force and hurricane-force winds; (c) exposure based on being in different forecast areas (e.g., in the cone of uncertainty, in a hurricane watch or warning) and being under an evacuation order; (d) exposure to different hazards that might be caused by the TC; (e) susceptibility to negative impacts overall; (f) susceptibility to different types of negative impacts; (g) severity of negative impacts overall; (h) severity of negative impacts due to different hazards; (i) worry; and (j) feeling of safety in their home.

For the post-storm survey, we asked questions pertaining to the respondent personally as the risk target. To match the questions from the predictive survey, we asked: (a) whether the respondent's home was ever in different forecast areas (e.g., in the cone of uncertainty, in a hurricane watch or warning) and under an evacuation order; (b) whether the respondent's home was ever affected by different hazards caused by the TC; (c) respondent's assessment of the severity of negative impacts overall; (d) respondent's assessment of the severity of negative impacts due to different hazards; and (e) whether the respondent experienced different types of negative impacts, with similar types of forecast areas, TC hazards, and negative impacts as in the corresponding predictive survey questions.

Efficacy is a person's beliefs about the ability to produce a desired or intended outcome. More specifically, *response efficacy* is a person's belief that engaging in a recommended action will reduce harm, and *self efficacy* is a person's belief that they have the capacity to perform the recommended action (Bandura 1977, Rogers 1983, Witte 1992). For the predictive survey, we measured respondents' response efficacy related to taking five different types of recommended actions (e.g., evacuating, getting emergency supplies) for this storm. We measured respondents' perceived self efficacy as how well they think they could take protective actions to reduce their negative impacts from this storm, and how confident they are in their ability to do so. For the post-storm survey, we asked how well the respondent was able to take protective actions to reduce their negative impacts from the storm.

Behavioral responses can include mitigative and protective behaviors as well as informational behaviors that a person engages in (Griffin et al. 1999, Lindell and Perry, 2012). For the predictive survey, we measured whether or not people had, at the time of completing that survey wave, taken several types of protective actions for TC [name]: (a) evacuated; (b) made evacuation arrangements; (c) got emergency supplies; (d) boarded up windows and doors or put up storm shelters; (e) followed the latest weather forecasts; (f) moved indoor furniture or valuables to a safer location; (g) gassed up vehicles; (h) did other home preparation; and (i) engaged in any other action. For the post-storm survey, we asked whether or not respondents ever took this same set of protective actions when [name] threatened and approached. We also asked an evaluative question about whether respondents would make a different decision next time as well as what, if anything, they would do differently next time.

In addition to the set of questions about the core risk-related concepts, we asked questions to measure a number of other concepts and factors that are known to influence people's hazard risk assessments and decision-making that we want to incorporate in our data analyses either as variables to control for or as additional independent variables of interest (covariates and moderators in Table 1). Questions about some of these concepts were asked on the first predictive survey (Wave 1), before respondents were potentially hit by the TC, and questions about other concepts (more stable traits that are not likely to change across survey waves) were asked on the post-storm wave (Wave 4).

On Wave 1, we measured respondents' perceived residence in a hurricane evacuation zone; past hurricane experiences (Demuth et al. 2016); social influence and social support (Wong-Parodi and Feygina 2018); and perceptions of hazard responsibility (Wong-Parodi et al. 2017). Because the first longitudinal survey was conducted in Summer 2020 within a few months of the beginning of the COVID-19 pandemic, we also measured how respondents have been affected by COVID-19 and their perceptions about managing hurricane risks given COVID-19.

On the post-storm wave, we measured respondents' numeracy abilities (Schwartz et al. 1997) and cultural worldviews (Morss et al. 2016, 2020; Johnson and Swedlow, 2020). We also measured sociodemographic characteristics that have been shown to relate to hurricane risk perceptions and behaviors (e.g., residence type, homeownership). YouGov maintains data for all of its panel members about other sociodemographic attributes—such as race, education, income, and marital status—which they provided to us. Finally, following Harlan et al. (2019), we measured whether or not people had access in the past year to each of five basic needs: enough money for housing, enough money for electricity and other utility bills, enough money for health care and medications, enough of the kinds of food wanted for the household, and health insurance.

Table 1. Summary of all questions asked on the 2020 predictive (left-hand column) and post-storm (right-hand column) surveys. The post-storm questions are aligned by row with the predictive question that it matches. Changes for the 2021 survey are discussed in Section 3b.

Predictive survey waves: Outline of sections, concepts, dimensions	Post-storm survey wave: Outline of sections, concepts, dimensions
Screening question to match respondents	***Screening question to match respondents***
• Age	• Age
• Gender	• Gender
Awareness and relevance of TC threat(s)	***Awareness and relevance of TC threats***
Q-0. Aware of TC threat	Q-0. Aware of TC threat
Q-00. Relevance of Marco and/or Laura	Q-00. Relevance of Marco and/or Laura
Questions asked only IF NOT aware of TC threat (if response to Q-0 is "no")	
Q-A. Typical information obtained	
Q-B. Perceived likelihood of TC affecting U.S. in next 2 weeks (exposure, U.S.)	
Q-C. Perceived impacts if a TC occurs in the next 2 weeks (severity, U.S.)	

Q-D. Perceived likelihood of TC affecting respondent in next 2 weeks (exposure, personal)	
Risk information	***Risk information***
	1. Forecast information wanted
	2. Preparedness information wanted
1. Frequency of info obtained, from different sources a. Channel used for NWS sources	3. Usefulness of different sources for getting info
2. Frequency of info obtained, from different channels	4. Usefulness of different channels for getting info
3. Frequency of info seeking overall	
4. Frequency of social cues	
5. Importance of different forecast info	5. Usefulness of different forecast info
6. Importance of different risk info (forecast, prep, impacts)	6. Usefulness of different of risk info (forecast, prep, impacts)
7. Knowledge of storm current intensity	
Risk perceptions	***Risk perceptions***
8. Susceptibility of U.S., aggregate	
9. Exposure of U.S., aggregate	
10. Susceptibility of U.S., specific geographic areas	
11. Exposure of U.S., of different hazards	
12. Susceptibility of U.S., of different impacts	
13. Exposure of U.S., timing	
14. Exposure of U.S., intensity	
15. Exposure personal, tropical storm-force winds	
16. Exposure personal, hurricane-force winds	

17. Exposure personal, relative to forecast and evacuation areas	7. Whether their home was exposed to forecast and evacuation areas
18. Exposure personal, aggregate	
19. Susceptibility personal, aggregate	
20. Severity personal, aggregate	8. Assessment of severity of negative impacts (personal), aggregate
21. Worry personal, aggregate	
22. Safety personal, aggregate	
23. Exposure personal, of different hazards a. Severity personal, of different hazards	9. Assessment of personal exposure / effects, of different hazards a. Assessment of severity of negative impacts (personal), due to different hazards
24. Susceptibility personal, of different impacts	10. Assessment of personal exposure / effects, of different impacts
Efficacy	***Efficacy***
25. Response efficacy	
26. Self efficacy -- how well	11. Assessment of how well they could take the steps needed
27. Self efficacy -- how confident	
Responses	***Responses***
28. Protective action taken	12. Whether or not they took protective actions
	13. Assessment of protective action decisions
	14. Assessment of protective action decisions -- whether or not they'd do anything differently next time
Covariates and moderators	
29. Reside in evacuation zone	
30. Past experience, direct evacuation-related	15. Experience -- whether evacuation order was issued for area where they live a. If yes, whether mandatory or voluntary

31. Past experiences, direct and indirect	
32. Social influence	
33. Social support	
34. Social support	
35. Perceptions of responsibility	
36. COVID-19, personal effects	16. COVID-19, personal effects for this storm
37. COVID-19, perceptions	
	17. Residence type
	18. Homeownership
	19. Year home/apartment built
	20. Length of residence in county
	21. Home-proofing
	22. Access to basic needs
	23. Numeracy
	24. Numeracy
	25. Numeracy
	26. Cultural worldviews

a. Changes to survey content for 2021 survey (of TC Henri)

Based on initial analysis of the data from the 2020 survey (of TCs Laura and Marco), we added two questions to the core set of risk-related questions on the predictive surveys. Pertaining to *risk information*, we added a question to measure frequency of information *sharing* overall, as a counterpart to our measure of frequency of information *seeking* overall. Pertaining to *efficacy*, we added a question to measure self efficacy in a more elaborate way, as how easy or hard it would be for the respondent to take each of the same recommended actions asked about in the response efficacy question. Along with adding these new questions, we changed the response options for all questions about frequency of *risk information* to more easily convert the responses to a ratio-level count.

We also added a few new questions to measure additional covariates of interest. On Wave 1, we added questions to measure injunctive and subjective normative behaviors (Cialdini et al. 1998, Wong-Parodi et al. 2019), the respondent's self-assessment of their physical and mental health, and the respondent's subjective attribution of hurricanes to climate change in general (Wong-Parodi and Garfin 2022). On the post-storm survey (Wave 4), we added questions to measure the respondent's subjective attribution of the TC they just experienced to climate change (as a counterpart to the question asked on Wave 1), and a question to measure respondents' climate change-related adaptation and behavioral intentions (Wong-Parodi and Garfin 2022, Wong-Parodi and Rubin 2022).

4. Survey Fielding

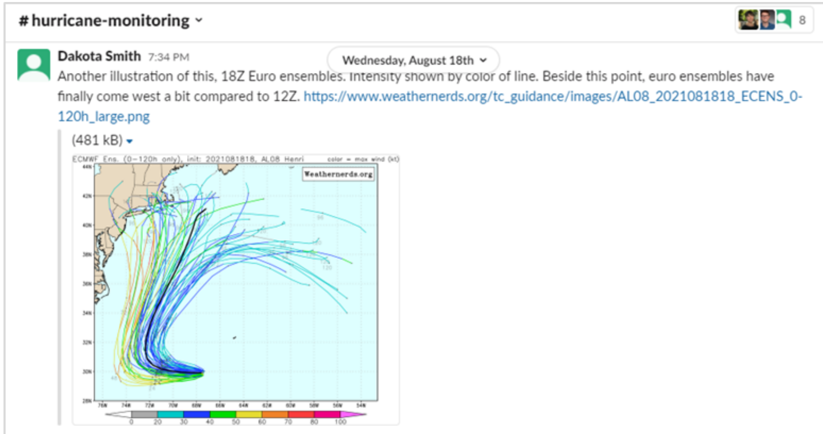
a. TC Monitoring

The first step in fielding the survey was monitoring the tropics for candidate TCs. Given that our primary goal for the first fielding of the longitudinal surveys (in 2020) was to test the methodology, we had three criteria:

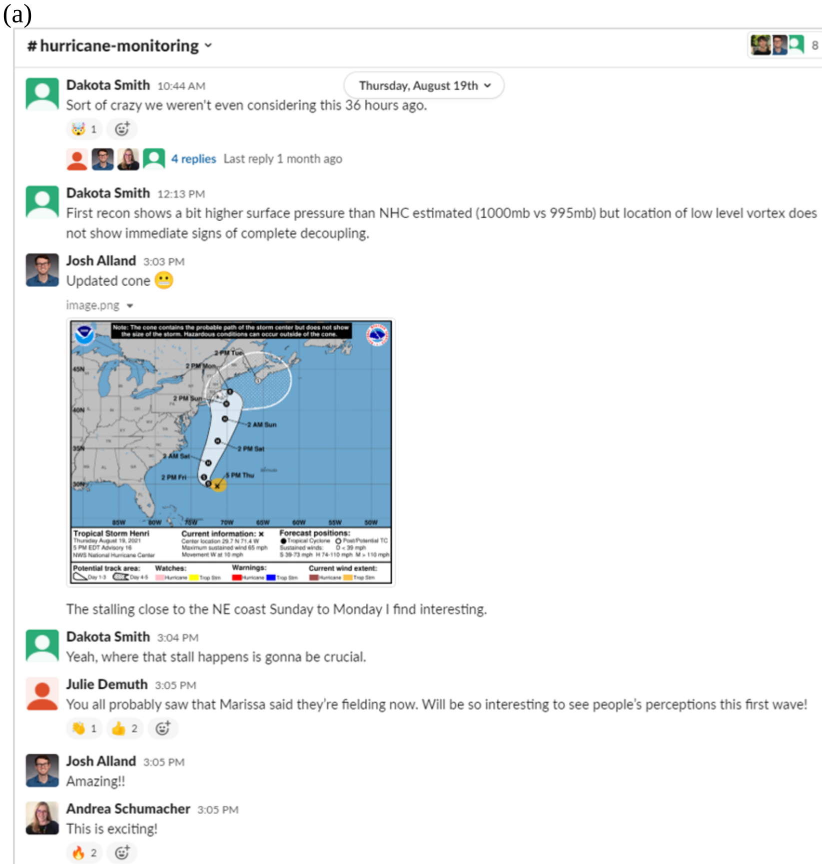
- a TC that was tropical-storm strength or was forecast to be so by the time Wave 1 was fielded, so we could refer to the TC by its name in the survey;
- a TC that had a reasonable chance of making landfall in the mainland United States (while also recognizing that the storm could dissipate or track such that it did not make landfall, e.g., if it was in the Atlantic Ocean and it curved away from land); and
- a TC that had at least a six-day lead time before anticipated landfall, which, given the timeline for fielding the survey, is when the decision to field the survey had to have been made (see Section 2).

To monitor the tropics and decide when to field the survey, we developed a multifaceted, cross-institutional approach incorporating multiple areas of expertise. Three members of our research team with expertise in tropical meteorology led the monitoring efforts: Josh Alland, Andrea Schumacher, and Dakota Smith. For regular monitoring, they utilized multiple resources, including the NHC operational forecast product suite; NOAA Hurricane Research Division (HRD) live forecast discussions; and operational and experimental numerical weather prediction (NWP) and statistical guidance products and observational data from NWS and the research community. We set up a team Slack channel (#hurricane-monitoring), which the monitoring team used to cue the rest of the team to a TC (or potential TC) of interest. Whenever a storm of potential interest was identified, we held discussions on Slack as new information emerged as well as real-time team forecast and survey fielding discussions immediately following the HRD forecast discussions. When our research team identified a TC that met or might soon meet our criteria and that we were considering fielding for, we communicated with our NOAA hurricane supplemental team collaborators—including our funding collaborators in WPO and our NHC operational forecaster collaborators—whose expert input was crucial to our decision-making about whether or not to field.

Figure 2 provides two snapshots from our team Slack channel, one that shows ensemble guidance the team was considering when identifying a TC of interest (Figure 2a) and another that shows communication among the team after we had made a decision to field for the 2021 TC Henri and were continuing to monitor the forecasts (Figure 2b).



(a)



(b)

Figure 2. Screenshots of our research team's #hurricane-monitoring Slack channel, illustrating (a) ensemble track and intensity guidance the team was considering when deciding whether to select this TC to field for, and (b) communication among team members after we had decided to field and were continuing to monitor the storm.

b. 2020 Fielding for TCs Laura and Marco

On August 18, 2020, we were monitoring two invest systems—Invest 97L in the central Caribbean Sea and Invest 98L in the Atlantic Ocean east of the Lesser Antilles—with the idea that one could be a possible candidate for our longitudinal survey. We corresponded with our NHC collaborators via email and decided to watch both systems given the substantial uncertainty in whether and when each would form into a tropical depression. Invest 98L became Tropical Depression (TD) 13 late on August 19, and Invest 97L became TD14 midmorning on August 20. The forecast storm tracks suggested that both TCs had a good chance of making landfall in the mainland United States, and we had enough lead time to field three predictive survey waves if we made the decision soon whether or not to field. We scheduled a call for early the next day, August 21, with our team of NOAA collaborators and, based on their expert input, we made the decision to field.

The images in Figure 3 show the cones of uncertainty for Tropical Storm (TS) Laura and for TD14—which would strengthen to a TS and be named Marco by late that evening—for the approximate day and time that we made our decision to field. Based on these and other forecast products (e.g., ensemble track forecasts, wind speed probabilities), we decided to field in all coastal zip codes up to 50 miles inland for the entire Gulf Coast, including Texas, Louisiana, Mississippi, Alabama, and western Florida (i.e., all coastal zip codes west of the southern tip of the Florida peninsula, including the Florida Keys).

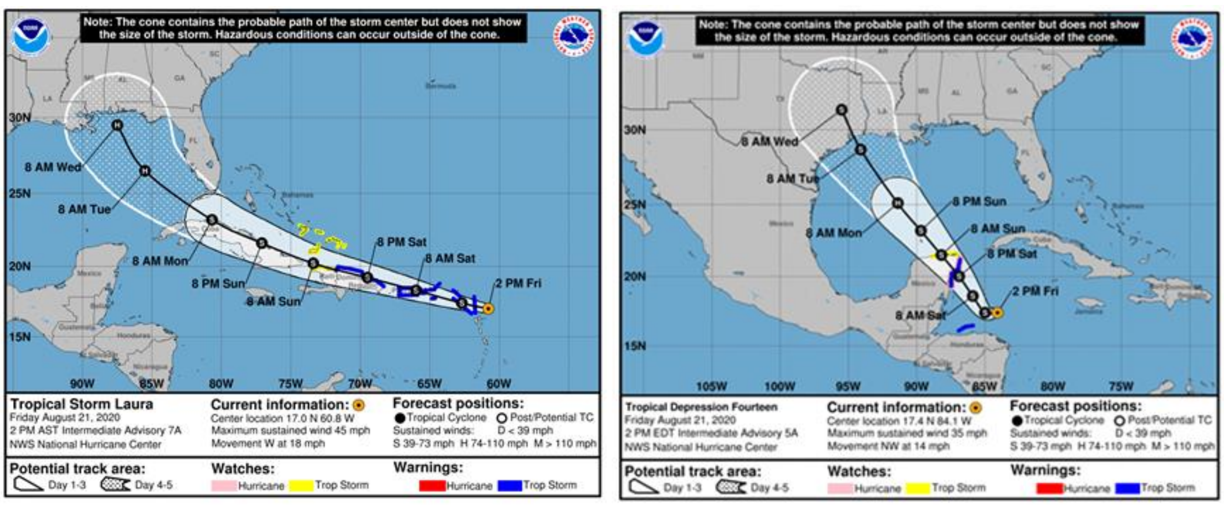
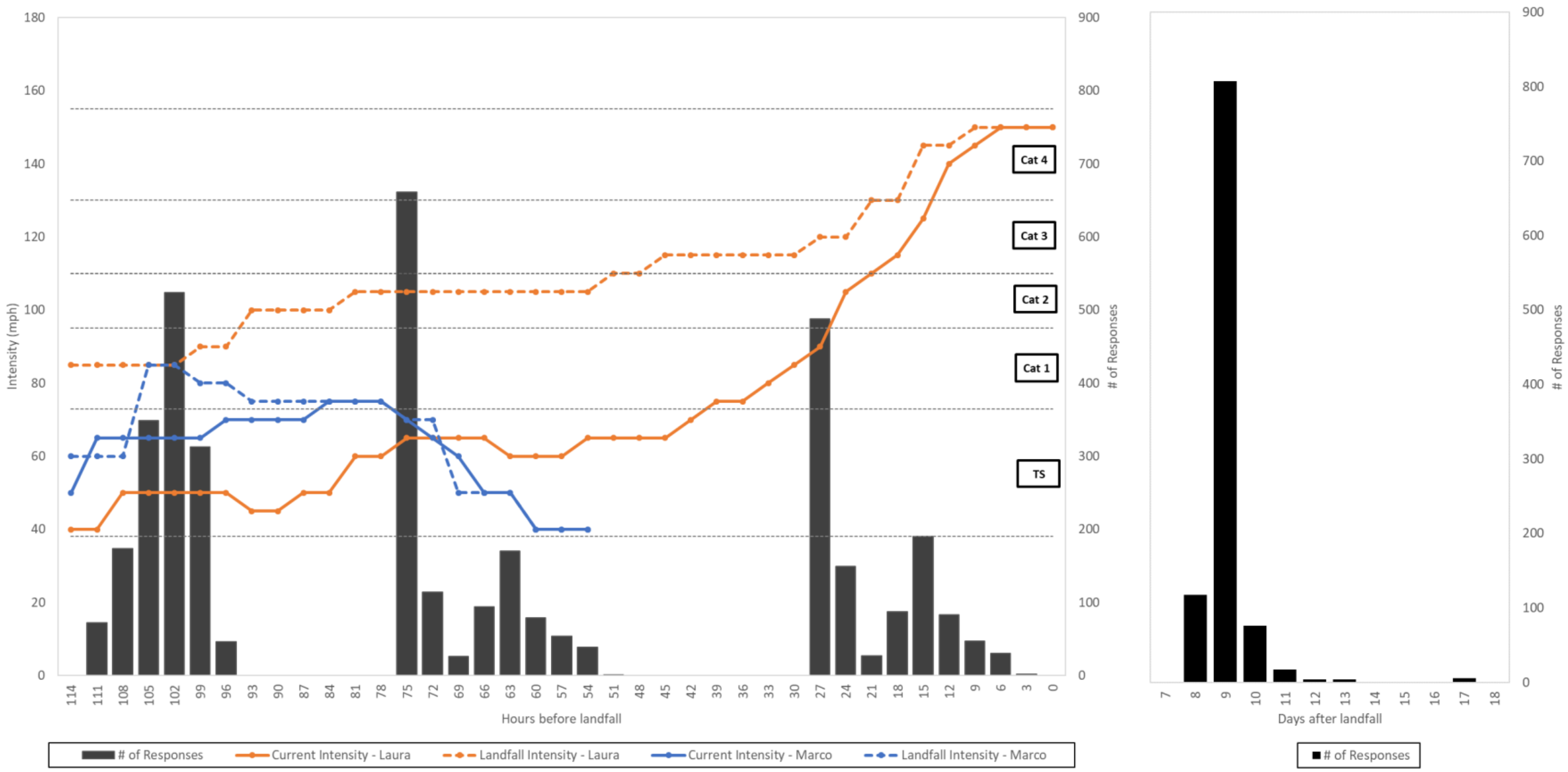


Figure 3. Cones of uncertainty for TS Laura and for TD14, which became TS Marco later that day, that were in effect at the approximate time that our research team made the decision to field the longitudinal hurricane survey for these TCs.

Because YouGov needed approximately 24 hours to finalize the fielding logistics, we fielded Wave 1 beginning early the following day, on Saturday, August 22. We outperformed the targeted sample of n=1277 for Wave 1, but nevertheless we continued to collect responses until we had approximately another 200 completed surveys (in case attrition rates were higher than anticipated), and then we closed Wave 1 after 18 hours. Based on the evolving forecast tracks and translational speeds, we fielded Wave 2 after only an 18-hour interval, to increase the chance of getting two predictive waves before Marco made landfall and three predictive waves before Laura made landfall. Then, we defaulted to the fielding cadence that we originally envisioned (see Section 2); Wave 2 was fielded for 24 hours, there was a 24-hour interval between Waves 2 and 3, and Wave 3 was fielded for 24 hours.

Figure 4 provides a timeline of all longitudinal survey waves with a histogram of the number of survey responses in three-hour bins for the predictive survey waves (Figure 4a) and the post-storm wave (Figure 4b). Overlaid with the timeline of the predictive waves are solid lines that show the *current* TC intensity for Laura (orange) and Marco (blue) and dashed lines that show the *forecast* intensity at landfall for each storm at that time (Figure 4a). Marco was a TS during Wave 1, became a weak Category 1 hurricane shortly before Wave 2, weakened throughout the Wave 2 period, and ultimately its center passed just south of the mouth of the Mississippi River around 6 pm CT on August 24 as a weak TS during Wave 2, a couple hours before Wave 2 closed. Laura was a TS throughout Waves 1 and 2 but was forecast during those waves to make landfall as a Category 1 or 2 hurricane. Shortly before Wave 3 was fielded, Laura began rapidly intensifying. During the 24-hour period that Wave 3 was fielded, the TC intensified from a Category 1 hurricane with 85 mph maximum sustained winds (MSW) to a Category 4 hurricane with 145 mph MSW, and the forecasted intensity at landfall increased from 115 to 150 mph. Laura made landfall on August 27 around 1 a.m. CT in western Louisiana as a Category 4 hurricane, approximately five hours after Wave 3 closed.

We fielded the post-storm survey beginning on September 3, 8 days after the Wave 3 survey closed, until September 13.



(a) (b)

Figure 4. Timeline of all longitudinal survey waves for Laura/Marco, with a histogram of the number of survey responses in three-hour bins for the (a) predictive survey waves and (b) post-storm wave. Solid lines are the current TC intensity for Laura (orange) and Marco (blue). Dashed lines are the forecast TC intensity at landfall for Laura (orange) and Marco (blue).

Figure 5 provides cumulative distributions of the percentage of survey responses over the number of hours after each wave started, for each predictive survey wave. Wave 1, which was fielded in the morning, accumulated survey responses at a slower rate for the first 6 hours, from 7 a.m.–1 p.m. local time. Survey responses then began to accumulate at a faster rate from 7–14 hours after fielding, from 2–9 p.m. local time. Survey responses then tapered off slowly during the next 3 hours, until 1 a.m. local time when Wave 1 closed. Waves 2 and 3 were fielded in the evening, at 7 p.m. and 8 p.m. local time, respectively. They have very similar response curve patterns, with high survey response rates for the first 3 hours during the evening hours in local time, much slower rates from 4–10 hours after fielding, during the overnight hours local time, and then moderate but steady response rates during the second half of the fielding periods.

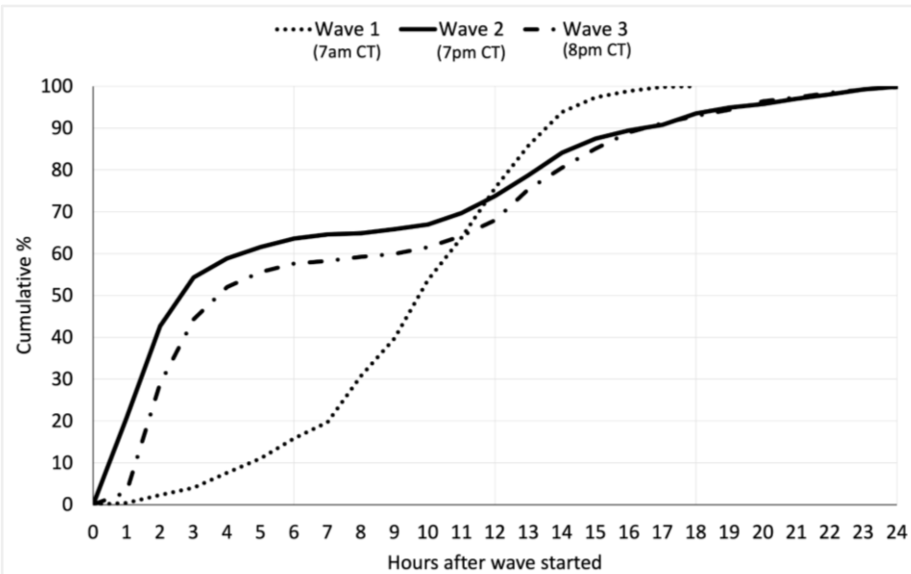


Figure 5. Cumulative distributions of survey responses over time for each predictive survey wave, for the Laura/Marco surveys.

The number of completed surveys and the correspondent attrition and re-contact rates from wave to wave are provided in Table 2. The attrition rates from wave to wave were low, especially for Waves 3 and 4, yielding a final, balanced sample of n=1034. The overall attrition rate from Wave 1 to Wave 4 was 29.6%, calculated as (1 - 1034/1469).

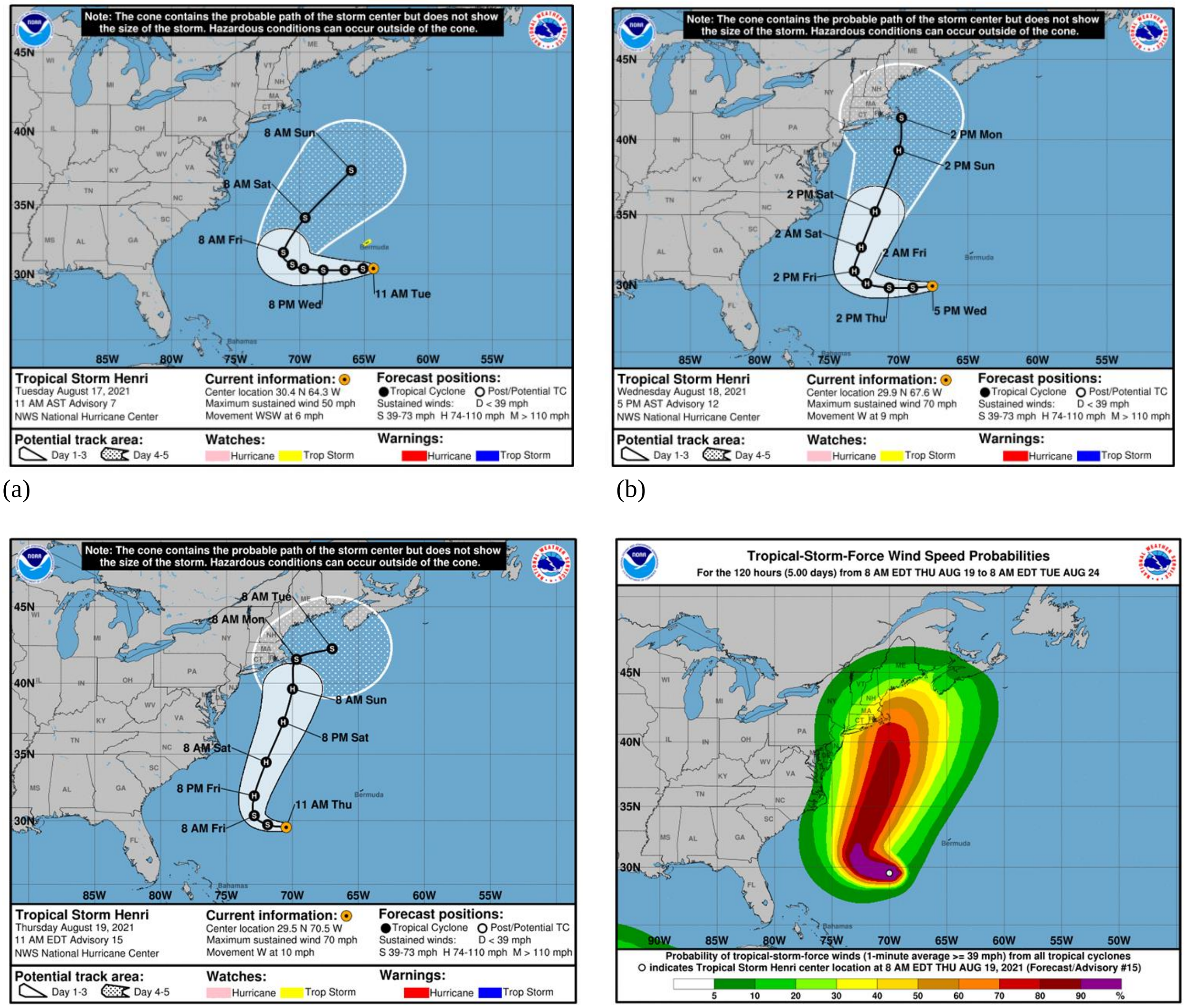
Table 2. Summary of survey fielding dates and times, sample sizes, and corresponding attrition and re-contact rates, for the Laura/Marco surveys.

Wave number	Wave fielding: Dates and (total hours)	Predictive wave interval number of hours	Number of completed surveys	Attrition rate from wave to wave	Re-contact rate from wave to wave
Wave 1 (Laura & Marco)	~7 a.m. CT Saturday, August 22 to ~1 a.m. CT Sunday, August 23 (18 hours)	N/A	1469	N/A	N/A
Wave 2 (Laura & Marco)	~7 p.m. CT Sunday, August 23 to ~7 p.m. CT Monday, August 24 (24 hours)	18	1237	15.8%	84.2%
Wave 3 (Laura only)	~8 p.m. CT Tuesday, August 25 to ~8 p.m. CT Wednesday, August 26 (24 hours)	24	1102	10.9%	89.1%
Wave 4 (Laura & Marco)	September 3 to September 13	N/A	1034	6.2%	93.8%

c. 2021 Fielding for TC Henri

On August 15, 2021, while we were monitoring TS Grace, which at that time was on a similar path as Laura in 2020, TD8 formed in the Atlantic Ocean northeast of Bermuda. On August 16, Grace's track was trending south and west and was forecast to make landfall in Mexico rather than U.S. Gulf states, and thus we considered it unlikely as a candidate TC to study. That day, TD8 strengthened into TS Henri. On August 17, Grace's forecast track continued to be for a Mexico landfall, and thus we ruled it out. TS Henri was forecast to make a clockwise turn and remain out at sea (Figure 6a), and thus we considered it unlikely as a candidate. However, on August 18, Henri's official forecast track shifted west and included the storm strengthening to a hurricane (Figure 6b), thereby posing a chance of a landfalling hurricane in New England. Furthermore, operational and experimental NWP guidance suggested potential for additional westward track shifts. We reached out to our NOAA collaborators that day to ask for their input, and they suggested it was a good, interesting storm to consider because it was a rare opportunity to study a TC threatening this area and it was coincident with the 30-year anniversary of Hurricane Bob (1991), the last hurricane to hit New England.

On August 19, based on the continued track and intensity forecasts, we made the decision to field. Based on the areas covered by the cone of uncertainty (Figure 6c) and the 30% tropical-storm-force wind speed probabilities (Figure 6d), we decided to field in all coastal zip codes up to 50 miles inland in New York, Connecticut, Rhode Island, Massachusetts, Vermont, New Hampshire, and Maine.



(a) (b) (c) (d)

Figure 6. Cone of uncertainty forecast for Henri on (a) August 17 and (b) August 18, when the research team was considering whether or not to field for this TC, and the (c) cone of uncertainty and (d) tropical-storm-force wind speed probabilities for Henri on August 19 when a decision was made to survey this TC.

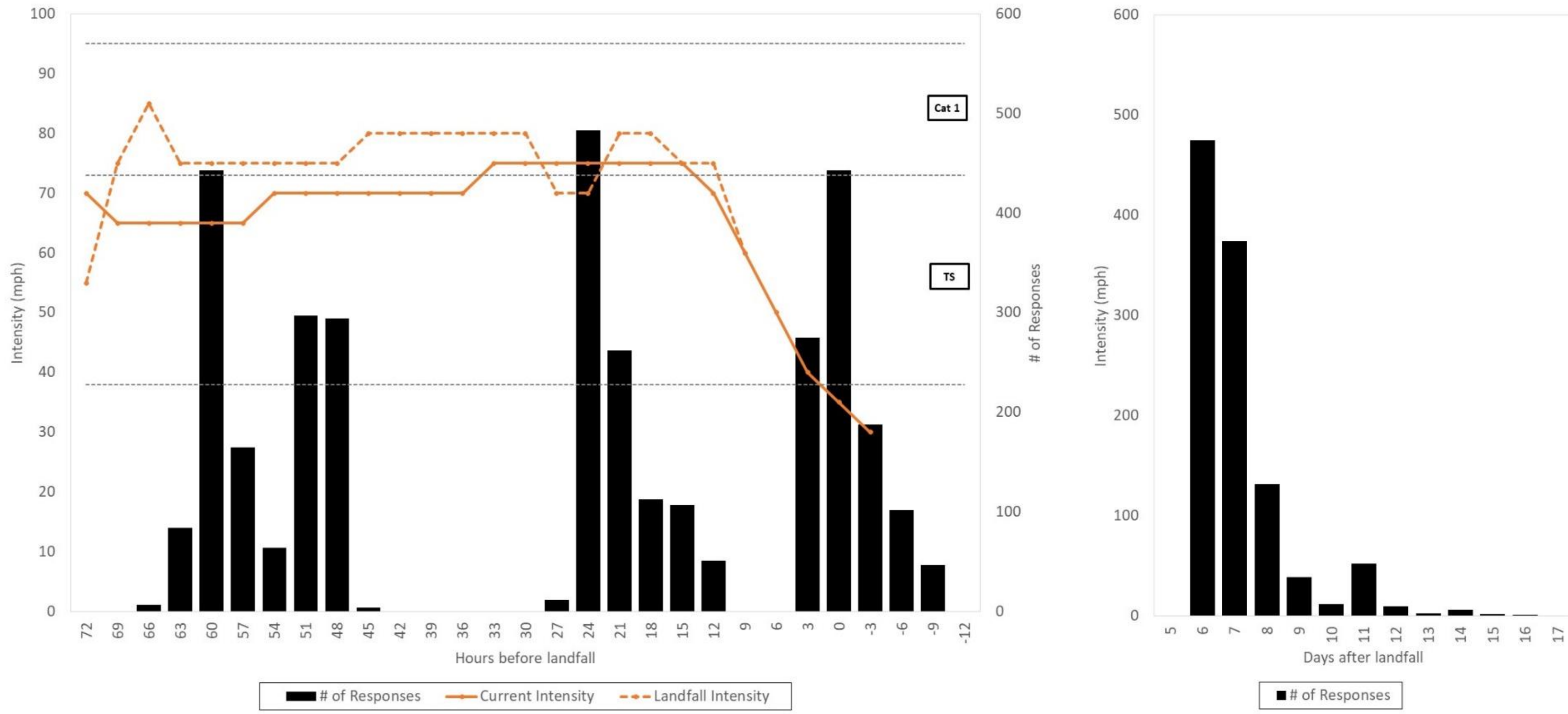
YouGov kindly worked with us to field the survey later that same day (i.e., in less than the 24-hour period they preferred to have to finalize fielding logistics), so we could launch Wave 1 *before* watches and warnings were in effect for the areas, which functionally served as the basis for a natural experiment in which we could examine the effects of these products being in effect versus not.

In the spirit of testing different aspects of our methodology to inform future related work, we opted to try doing two things differently from the 2020 survey. First, we tried fielding for shorter periods with shorter wave intervals, given that our experience with Laura/Marco suggested this was feasible, especially for a high-population area like that sampled in Henri (with a larger number of potential respondents in the YouGov panel than more sparsely populated areas). Second, we tried allowing for unbalanced responses across the four survey waves; we still required responses to the first predictive survey wave (Wave 1) and the post-storm wave (Wave 4), but we allowed people to skip responding during either Wave 2 or Wave 3 (see Section 2a).

We fielded Wave 1 beginning early in the evening on Thursday, August 19. We again outperformed our targeted sample, and collected n=1358 responses for Wave 1, and we closed it after 19 hours. We fielded Wave 2 after a 19-hour interval. Although we fielded Wave 2 for only 15 hours, we got n=1026 completed responses. Based on Henri's evolving track and translational speed, we fielded Wave 3 after only an 8-hour interval, and fielded it for only 13 hours. Henri weakened shortly before landfall and made landfall as a TS partway through Wave 3.

Figure 7 provides a timeline of all longitudinal survey waves for Henri with a histogram of the number of survey responses in three-hour bins for the predictive survey waves (Figure 7a) and the post-storm wave (Figure 7b). Overlaid with the timeline of the predictive waves are solid orange lines that show the *current* TC intensity for Henri and dashed orange lines that show the *forecast* intensity at landfall at that time (Figure 7a). Henri was a TS during Wave 1, became a weak Category 1 hurricane shortly before Wave 2 and remained that strength during most of Wave 2, weakened quickly before and during Wave 3, and ultimately made landfall on August 22 around 2 p.m. ET as a weak TS, about 6 hours into the Wave 3 field period.

We fielded the post-storm survey for Henri beginning on August 27, 5 days after the Wave 3 survey closed, until September 7.



(a) (b)

Figure 7. Timeline of all longitudinal survey waves for Henri, with a histogram of the number of survey responses in three-hour bins for the (a) predictive survey waves and (b) post-storm wave. The solid orange line is the current TC intensity for Henri, and the dashed line is the forecast TC intensity at landfall.

Figure 8 provides cumulative distributions of the percentage of survey responses over the number of hours after each wave started, for each predictive survey wave. Wave 1, which was fielded in the evening, accumulated survey responses at a slower rate for the first 0 to 4 hours, from 5–9 p.m. local time. Survey response rates then increased for a couple of hours (between 9–11 p.m. local time), then flattened for several hours during the overnight period, followed by an increase 13–17 hours after fielding (6–10 a.m. local time) and then tapering off until Wave 1 closed. Waves 2 and 3 were fielded starting early in the morning, at 7 a.m. and 6 a.m. local time, respectively. They have similar response curve patterns, with high survey response rates for the first 6 hours, and then more moderate but steady response rates during the second half of the fielding periods.

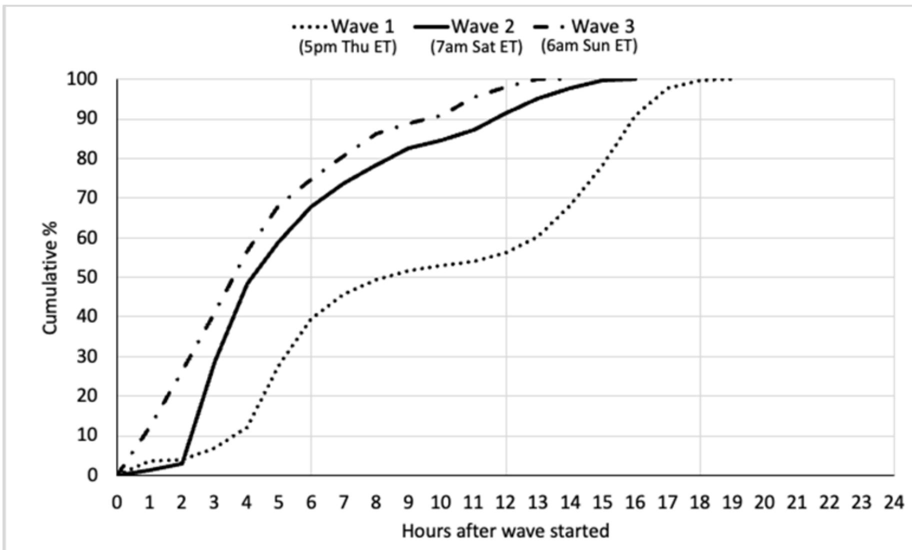


Figure 8. Cumulative distributions of survey responses over time for each predictive survey wave, for the Henri surveys.

The number of completed surveys are provided in Table 3, with data provided for both the *unbalanced* samples, in which people could skip either Wave 2 or Wave 3, and for the *balanced* sample across all waves. *Balanced* sample attrition and re-contact rates from wave to wave also are provided. The *balanced* sample was n=839, which is an overall attrition rate from Wave 1 to Wave 4 of 38.2%, calculated as (1 - 839/1358).

Table 3. Summary of survey fielding dates and times, unbalanced and balanced sample sizes, and balanced attrition and re-contact rates, for the Henri surveys.

Wave number	Wave fielding: Dates and (total hours)	Predictive wave interval number of hours	Number of completed surveys	Balanced attrition rate from wave to wave	Balanced re-contact rate from wave to wave
Wave 1	~5 p.m. ET Thursday, August 19 to ~12 p.m. ET Friday, August 20 (19 hours)	N/A	1358	N/A	N/A
Wave 2	~7 a.m. ET Saturday, August 21 to ~10 p.m. ET Saturday, August 21 (15 hours)	19	1026	24.4%	75.6%
Wave 3	~6 a.m. ET Sunday, August 22 to ~7 p.m. ET Sunday, August 22 (13 hours)	8	1052 - 886 completed W2	13.6%	86.4%
Wave 4	August 27 to September 7	N/A	1106 - 839 completed W2 and W3 (balanced) - 957 completed W2 but skipped W3 - 985 completed W3 but skipped W2	5.3%	94.7%

d. Ethical considerations

There are ethical guidelines that must be adhered to for any data collection with humans, but there are additional considerations of collecting data during a real-world threat for at-risk members of the public who may be evaluating and responding to the risk in the very ways we aim to measure. A chief ethic in this regard is that participants are not required or coerced into responding to any survey wave, and there is no penalty if they begin taking a survey but do not complete it. Additionally, as discussed above, for Henri we allowed for *unbalanced* data during the predictive phase, so that participants could miss responding to a wave and still be invited and have the option to respond to subsequent waves. This more flexible data collection approach recognizes that people may be evacuating, protecting their homes, gathering emergency supplies, and/or helping others, all of which may make them unable to respond to a given survey wave but still allows for their experiences to be represented in the dataset (Demuth et al. 2023). Another important ethical consideration is to ensure the predictive survey is short and focused on measuring the most important variables so that respondents can complete it in 10–12 minutes on average. As discussed in our recommendations (Section 7), analysis of the data will elucidate which variables are most important, so that future work can prioritize streamlining the survey accordingly as well as focus on refining statistical power analyses to sample most effectively.

Our research received approval from NCAR’s Human Subjects Committee and from Stanford’s Institutional Review Board.

5. Survey Sample

a. Demographic Characteristics: TCs Laura and Marco vs. Henri

We fielded the 2020 longitudinal survey of TCs Laura and Marco in all coastal zip codes up to 50 miles inland for the entire Gulf Coast, from Texas through the western side of Florida (Section 4b). A map showing the zip codes of survey respondents from the balanced sample (n=1034) is provided in Figure 9.

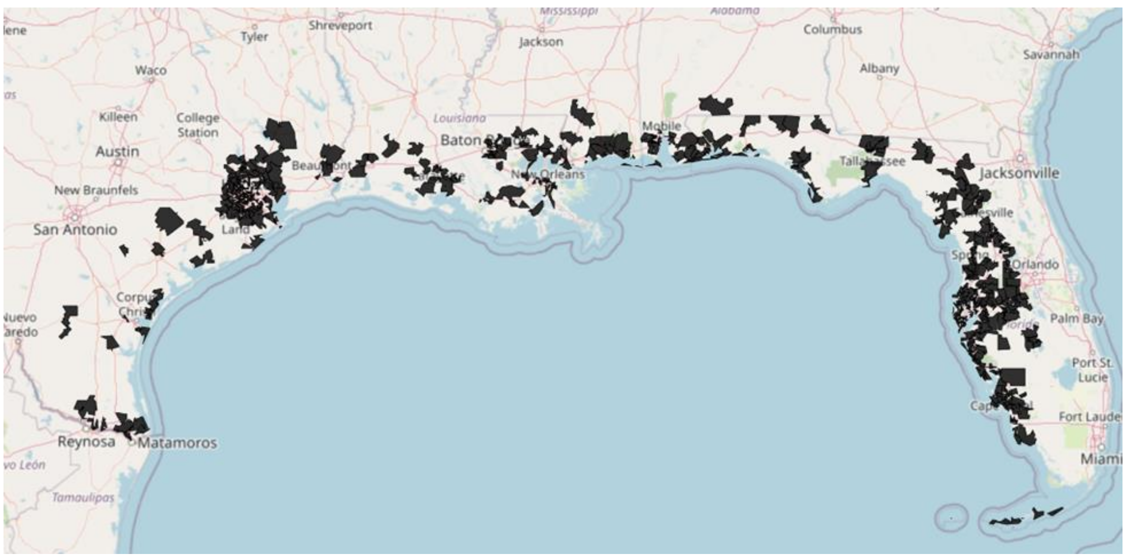


Figure 9. Map showing the zip codes of survey respondents’ locations for the 2020 longitudinal survey of TCs Laura and Marco.

We fielded the 2021 longitudinal survey of TC Henri in all coastal zip codes up to 50 miles inland from New York to Maine (Section 4c). A map showing the zip codes of survey respondents from the balanced sample (n=839) is provided in Figure 10.

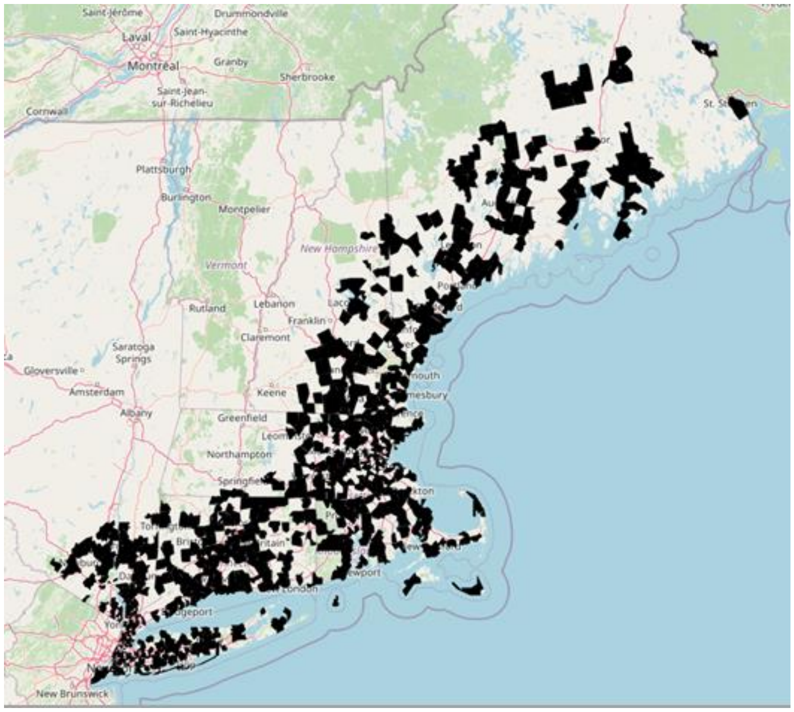


Figure 10. Map showing the zip codes of survey respondents' locations for the 2021 longitudinal survey of TC Henri.

In Table 4, we provide a summary of the demographic characteristics for the balanced sample of respondents (n=1034) from the Gulf Coast for the 2020 survey of Laura and Marco alongside those for the balanced sample of respondents (n=839) from New York and New England for the 2021 survey sample of Henri.

Both samples are comprised primarily of older adults (mean age 63.1–64.6) who are long-term residents of their county. A majority of both samples are homeowners (69.2–78.7%), live in a one-family detached home (56.8–73.4%), are White (86.9–88.7%), and have a 4-year or higher degree (53.1–61.2%). A plurality of both samples are retired (43.9–55.1%). The income quartiles for the Laura and Marco sample indicate that approximately one-quarter have a family income of less than $40,000, one-quarter have between $40,000–70,000, one-quarter have between $70,000–120,000, and the rest make more than $120,000 or prefer not to say. The income distribution for the Henri sample is slightly wealthier, with smaller proportions in the lower income bins and higher proportions in the higher income bins. As discussed in Section 3, we measured whether or not people had access in the past year to each of five basic needs; the vast majority of our samples had access to all five (83.1–83.7%) whereas a small fraction had access to none of the five (1.2–1.4%). The basic need that was missing most commonly in both samples was enough money for health care and medications (7.3–10.0%), followed by enough kinds of food for the household (6.2–7.6%).

Based on these characteristics, the samples for both surveys are not representative of the general population. It is important to remember, however, that our goal with these surveys was to develop a proof-of-concept, to determine whether we could field longitudinal surveys with three predictive waves while a TC was threatening and get an adequate number of people responding to multiple survey waves over a several-day time period. Given the funding available for this proof-of-concept, it was not feasible to attempt to get a representative sample; accordingly, both are convenience samples. In Section 7, we discuss recommendations for building on our methods to explore options for pursuing samples that are more diverse if not representative of the population.

Table 4. Sociodemographic characteristics of the samples from Laura and Marco (2020) and from Henri (2021).

Characteristic	Laura & Marco (2020)	Henri (2021)
Age: mean (SD)	64.6 (13.3)	63.1 (13.3)
Length of residence in county: mean (SD)	21.0 (17.9)	30.7 (21.0)
Gender: % Female	48.2%	50.1%
Dwelling: % One-family detached home	73.4%	56.8%
Dwelling: % One-family home attached to one or more houses	4.9%	8.5%
Dwelling: % Building with two or more apartments or dorm	12.9%	32.7%
Dwelling: % Manufactured or mobile home	8.4%	1.7%
Dwelling: % Boat, RV, van, etc.	0.4%	0.1%
Homeownership: % Own	78.7%	69.2%
Race: % White	86.9%	88.7%
Race: % Black	4.8%	4.1%
Race: % Hispanic	3.4%	1.9%
Race: % Asian, Native American, two or more race, or other	4.8%	5.4%
Education: No high school or high school graduate	12.8%	13.2%
Education: Some college or 2-year college	34.2%	25.6%
Education: % 4-year college	30.4%	31.7%
Education: % Post-graduate	22.7%	29.5%
Employment: % Retired	55.1%	43.9%
Employment: % Full-time	20.3%	29.4%
Employment: % Permanently disabled	5.5%	5.1%
Income: % less than $40,000	25.1%	19.3%
Income: % between $40,000–$70,000	25.6%	20.3%
Income: % between $70,000–$120,000	25.1%	26.7%
Income: % from $120,000 to greater than $500,000	13.8%	19.1%
Income: % prefer not to say	10.3%	14.6%
% With access to all 5 basic needs ($ for housing, $ for utilities, $ for health care and meds, adequate food, health insurance)	83.1%	83.7%
% Without access to all 5 basic needs ($ for housing, $ for utilities, $ for health care and meds, adequate food, health insurance)	1.4%	1.2%
% Structural home improvements for TCs made	25.1%	7.6%

b. *Experiences with the TC Surveyed for based on Post-storm Survey (Wave 4): TCs Laura and Marco vs. Henri*

In addition to the sociodemographic data summarized above, we can further characterize our samples based on their experiences with Laura and Marco (n=1034) or with Henri (n=839), as reported on the post-storm survey (Wave 4). Recall that Marco and Henri ended up weakening before landfall and not causing hazardous TC conditions over a large area, and Laura ended up rapidly intensifying before landfall and causing major impacts but over a less populated region (i.e., and therefore affected only a small fraction of our sample). Thus, most of our sample ended up not having significant negative experiences for these storms, which is to be expected for some storms.

Figure 11 shows the percentage of respondents who indicated their home was ever in each of the forecast or evacuation areas inquired about. Approximately half of respondents from Laura and Marco and from Henri reported they were in the cone of uncertainty at some point during the TCs. Approximately 25–30% more respondents indicated they were in a hurricane or tropical storm watch and warning for Henri than for Laura and Marco, respectively. About 10% more respondents also reported being in a storm surge watch or warning for Henri compared to Laura and Marco. Although small proportions of people reported they were in an area under a mandatory or voluntary evacuation order in each of the TC surveys, a larger proportion reported being in these areas for Laura and Marco than for Henri.

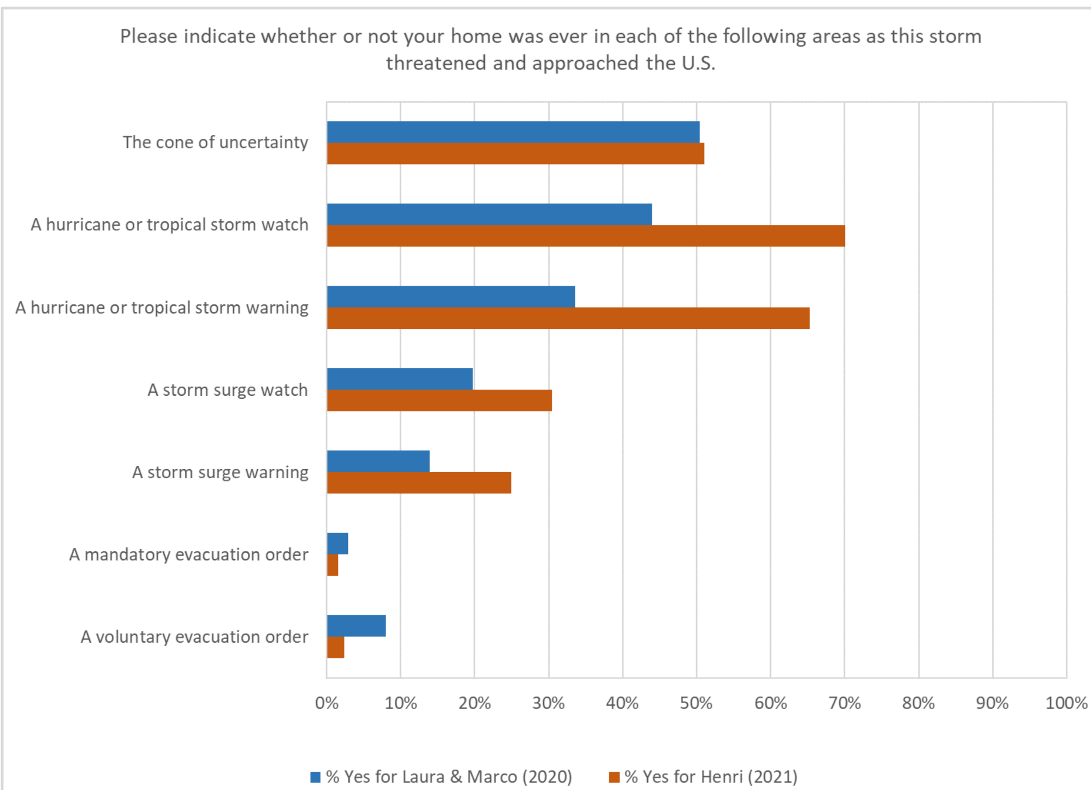


Figure 11. Percentage of respondents who indicated that, yes, their home was ever in each of the forecast products or evacuation orders inquired about for Laura and Marco (blue bars) or for Henri (orange bars).

Figure 12 shows the percentage of respondents who indicated they took different informational and protective behavioral actions as Laura and Marco or Henri threatened. The action taken most, by the vast majority of respondents for Laura and Marco and for Henri, was following the latest weather forecasts. More than half of respondents for Laura and Marco and for Henri reported they gassed up vehicles. The actions most commonly taken next for both TCs were, in descending order, getting emergency supplies, home preparations such as trimming trees and securing loose objects in the yard, and moving indoor furniture or other valuables to a safe location. Small proportions of people reported that they made evacuation arrangements, boarded up windows and doors or put up storm shutters, or evacuated.

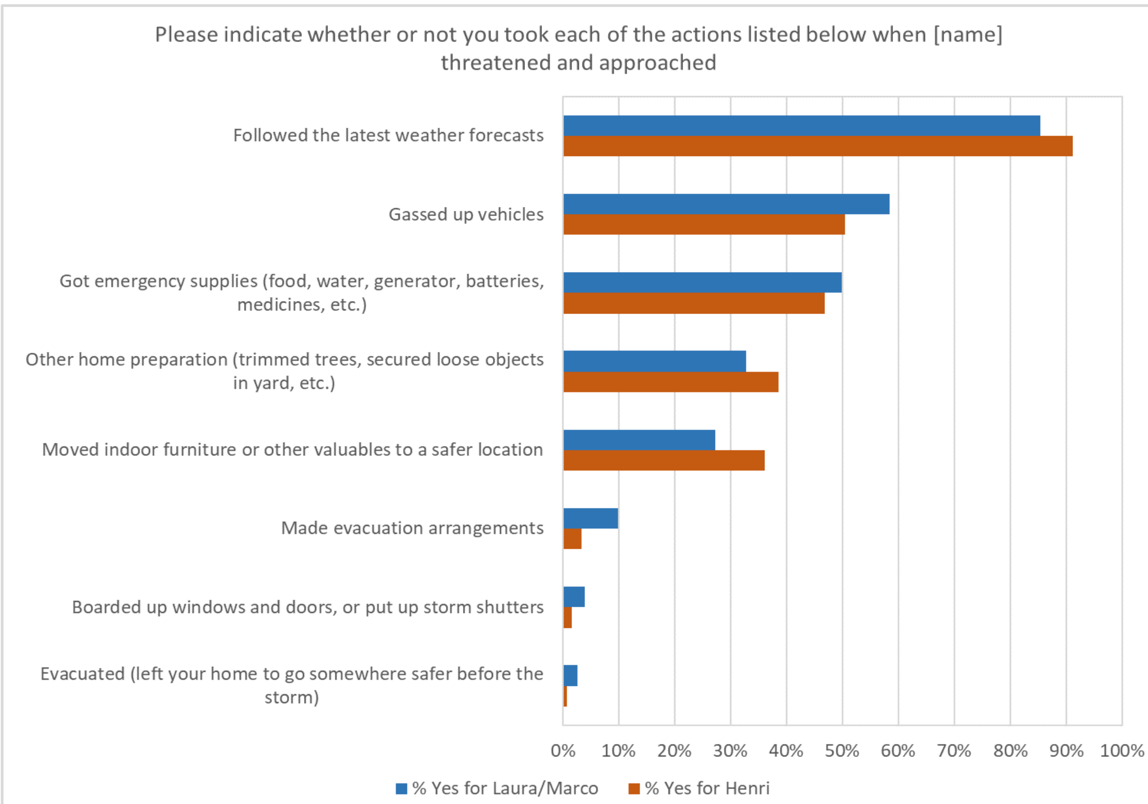


Figure 12. Percentage of respondents who indicated that, yes, they took each of the actions listed for Laura and Marco (blue bars) or for Henri (orange bars).

Figure 13 shows the percentage of respondents who indicated they had experienced different types of negative impacts from Laura and Marco or from Henri. Overall, the proportion of respondents who had negative experiences was low for each storm. Approximately an equal proportion of respondents reported experiencing emotional impacts for both Laura and Marco and for Henri. More respondents reported experiencing road closures for Henri than for Laura and Marco. On the other hand, more respondents reported experiencing power outages for Laura and Marco than for Henri. Fewer than 5% of respondents for either storm reported all other negative experiences: damage to home or property, financial losses, harm to personal livelihood, injury to friend/family/other loved one, or injury to themselves.

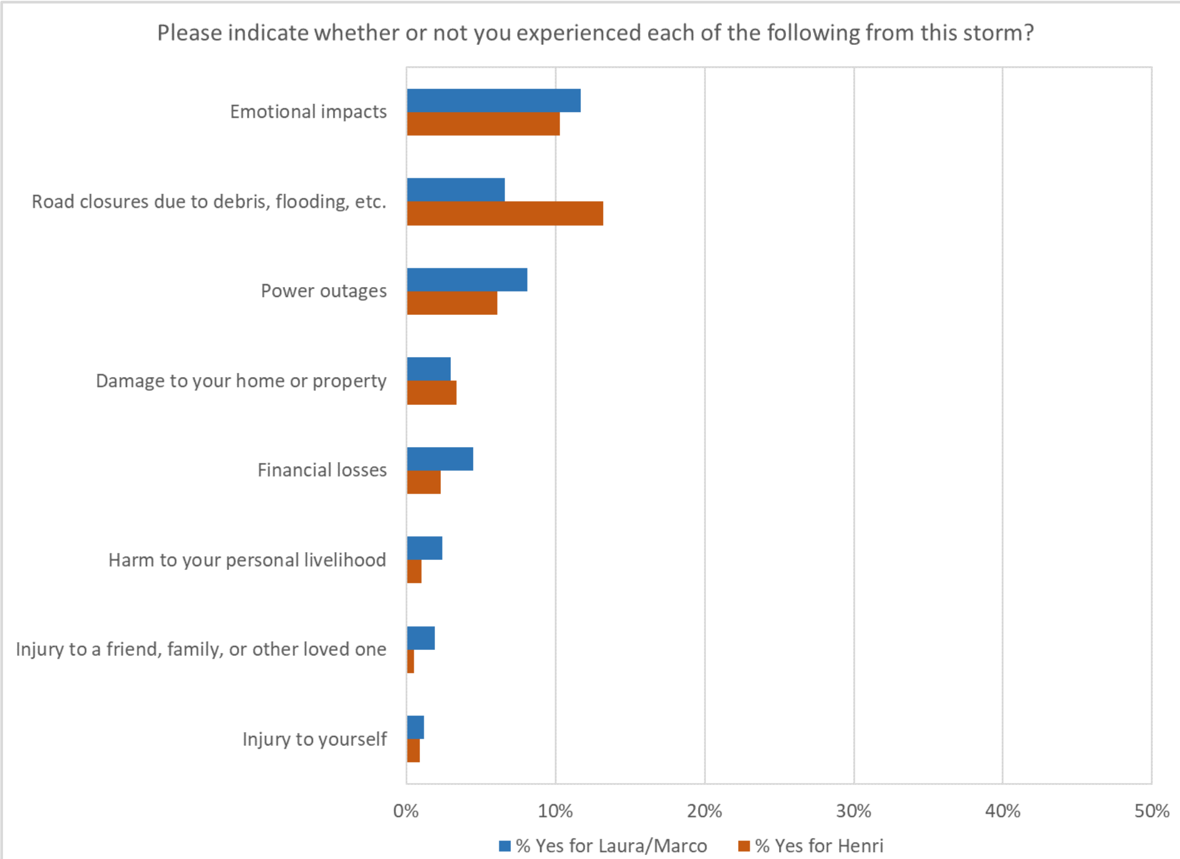


Figure 13. Percentage of respondents who indicated that, yes, they experienced each item listed for Laura and Marco (blue bars) or for Henri (orange bars). Note that the x-axis only goes to 50% in this figure.

6. Integrating Meteorological Data with Longitudinal Survey Data

An exciting aspect of collecting data about people's perceptions of and responses to TC forecast information for a real-world TC is that the social science data can be integrated with and compared to different types of corresponding meteorological data, allowing for richer and novel data analyses. Our planned data analyses required integrating meteorological data with survey data in two primary ways.

First, we wanted to compare people's perceived exposure to different aspects of TC risks—e.g., exposure to different TC hazards, timing of the TC, TC intensity, different forecast products—to the official forecast information that was being provided at the time they took the survey. A list of the exposure risk perception questions asked on the predictive surveys and the corresponding meteorological data we collected for analysis can be found in Table 5.

Table 5. Survey questions to be compared with meteorological data and the relevant NOAA datasets needed.

Predictive wave: Sections, concepts, dimensions	Relevant NOAA datasets
Risk information	
7. Knowledge of storm current intensity	NHC official track and intensity forecasts
Risk perceptions	
11. Exposure of U.S., of different hazards a. Strong winds b. Flooding due to storm surge c. Flooding due to rain d. Tornadoes	a. Wind speed probabilities, TS/hurricane watches and warnings b. Storm surge watches and warnings c. Excessive rainfall outlooks, flood watches and warnings d. Convective outlook tornado probabilities, tornado watches and warnings
13. Exposure of U.S., timing	NHC official track and intensity forecasts
14. Exposure of U.S., intensity	NHC official track and intensity forecasts
15. Exposure personal, tropical storm-force winds	Wind speed probabilities (34-kt)
16. Exposure personal, hurricane-force winds	Wind speed probabilities (64-kt)
17. Exposure personal, relative to forecast and evacuation areas a. Cone of uncertainty b. TS/hurricane watch c. TS/hurricane warning d. Storm surge watch e. Storm surge warning f. Mandatory evacuation g. Voluntary evacuation	a. Cone of uncertainty b. TS/hurricane watches c. TS/hurricane warnings d. Storm surge watches e. Storm surge warnings f. Mandatory evacuations g. Voluntary evacuations
23. Exposure personal, of different hazards a. Severity personal, of different hazards	a. Wind speed probabilities, TS/hurricane watches and warnings b. Storm surge watches and warnings c. Excessive rainfall outlooks, flood watches and warnings d. Convective outlook tornado probabilities, tornado watches and warnings

Second, as described above, we fielded the longitudinal surveys over broad geographic areas, which corresponded to the large areas that were at risk of TC impacts at the approximately six-day lead time when we had to make the decision to field the surveys. Furthermore, we surveyed this same broad area for all survey waves. For Laura and Marco in 2020, we fielded and got responses to the survey along the entire Gulf Coast (see Figure 9). For Henri in 2021, we fielded and got responses throughout coastal New York and New England (see Figure 10). These TCs evolved during the period between when we fielded Wave 1 and when the TC made landfall. The threat areas became more refined, and the types and magnitude of specific hazards evolved. Correspondingly, the TC risks that people faced evolved, with risks increasing in some areas and decreasing in others. We wanted to analyze risk information obtained, risk perceptions, and responses for respondents who were in areas *where TC risks increased* versus *where the risks decreased*. To do this, however, we need to define and characterize mutually exclusive TC risk areas so we could accordingly stratify respondents as inside or outside of them. We defined these risk areas as TC exposed versus non-TC exposed areas based on official forecast products and evacuation orders that were in effect during each survey wave.

In this section, we provide details of the multiple meteorological datasets and evacuation orders that were curated for analysis (Section 6a), how these data were matched to survey respondents (Section 6b), and how we defined TC exposure for more nuanced data analysis (Section 6c).

a. Meteorological datasets and evacuation orders

Here, we provide brief descriptions of the various meteorological datasets listed in Table 5 along with our methods for collecting and formatting the data. When possible, meteorological data were collected in GIS format to facilitate spatial matching with survey respondent zip codes. A summary of the entirety of the information is provided in Table 7 at the end of this section.

i. NHC official track and intensity estimates and forecasts

The National Hurricane Center (NHC) issues official forecast advisories for active Atlantic tropical cyclones every 6 hours at 0300, 0900, 1500, and 2100 UTC. In some situations, NHC also issues intermediate forecast advisories at 0000, 0600, 1200, and 1800 UTC. Information from these advisories, which include current and forecast track and intensity, were obtained from the NHC GIS archive (NHC 2022a). Track and intensity forecasts were available every 12 hours from forecast valid time t=0 to 72 hours, and every 24 hours from forecast valid time t=72 to 120 hours. Current position and intensity estimates (based on available TC observations) were obtained directly from these files (i.e., the t=0 forecast). Landfall intensity forecasts are not explicitly provided and thus were calculated as the greater of (a) the intensity at the last forecast point over water or (b) the intensity at the first forecast point over land (Figure 14).

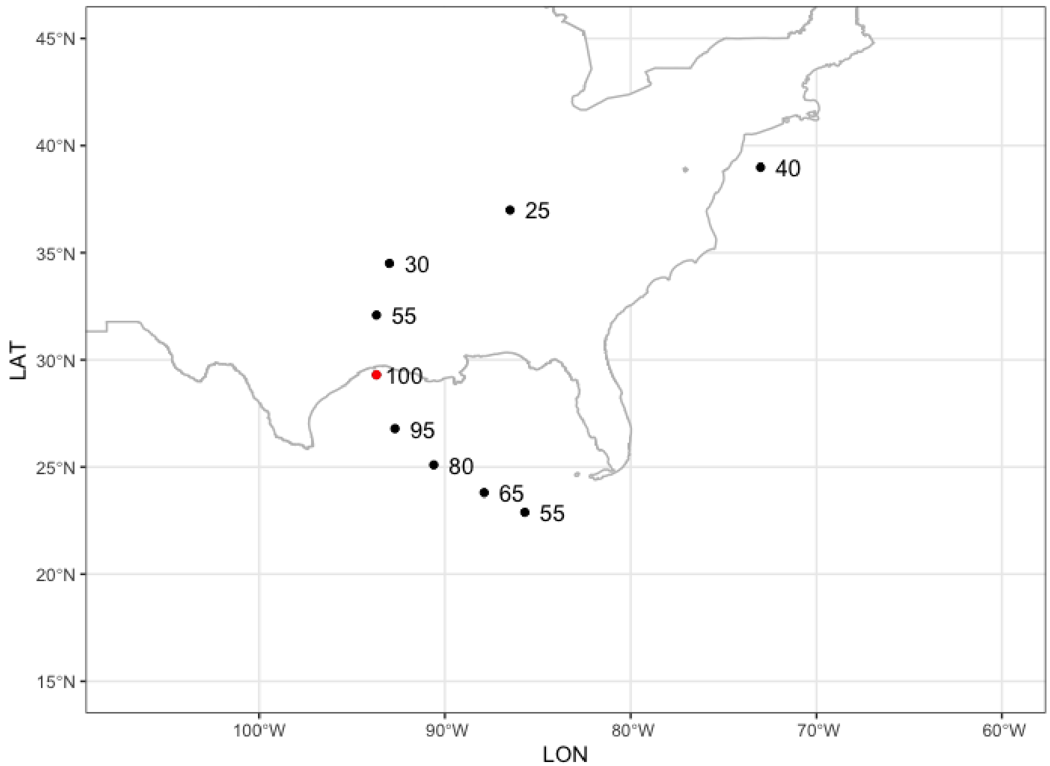


Figure 14. Forecast locations and intensities (in knots) for Hurricane Laura Advisory #22. The red dot indicates the forecast landfall intensity determined by our methodology. Data source: https://www.nhc.noaa.gov/gis/.

ii. Track forecast cone (also known as the cone of uncertainty)

The NHC also issues a graphical Tropical Cyclone Track Forecast Cone (Figure 15) product with each forecast advisory. The Track Forecast Cone, also commonly referred to as the Cone of Uncertainty, is constructed from the past 5 years of official track forecast errors, and it represents the area within which the 5-day path of the center of the tropical cyclone should remain 60–70% of the time (NHC 2022b). GIS data for the Cone of Uncertainty were obtained from the NHC GIS archive (NHC 2022a).

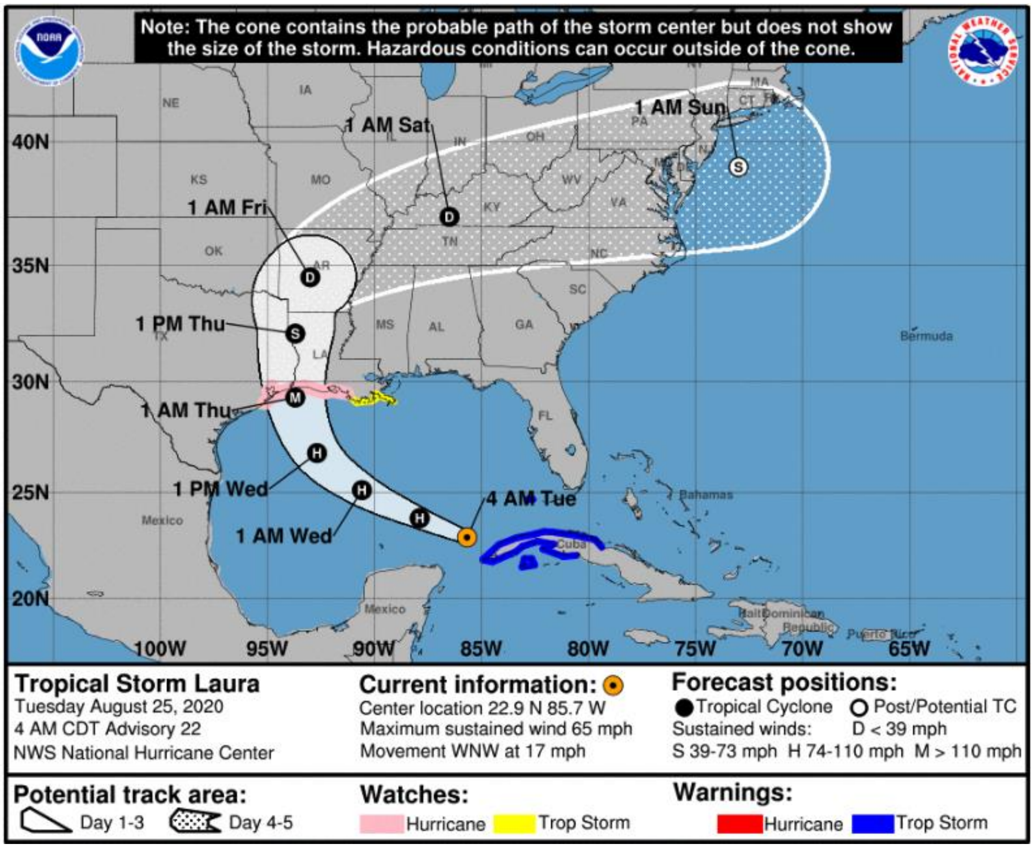


Figure 15. The NHC 5-day cone of uncertainty graphic for Hurricane Laura Advisory #22. Source: https://www.nhc.noaa.gov/archive/2020/LAURA_graphics.php (cone, 5-day no line). Graphic shows the official NHC forecast positions (black dots), tropical cyclone type (letters in dots), potential track area (white-fill and white-hatched cones), and tropical storm and hurricane watches and warnings (colored lines along coast).

iii. Tropical storm and hurricane watches and warnings

Tropical storm and hurricane watches and warnings (TSHWW) are issued by two NWS entities: the NHC and local Weather Forecast Offices (WFOs). The NHC issues coastal TSHWW, which are displayed on the Track Forecast Cone (Cone of Uncertainty) and Tropical Cyclone Surface Wind Field graphics, with colored lines representing coastal areas under a hurricane warning (red), hurricane watch (pink), tropical storm warning (blue) and tropical storm watch (yellow). Definitions of the different types of TSHWW can be found in the NHC Glossary (NHC 2022c). TSHWW valid over inland areas are issued by WFOs.

Our first inclination was to use the NHC-issued TSHWWs because they are displayed on the popular NHC Tropical Cyclone Track Forecast graphic and thus have more public visibility. However, because these TSHWWs are only valid along the coastline and our responses came from zip codes up to 50 miles inland, we opted to use the NWS WFO-based TSHWWs. NHC and WFOs coordinate regularly during tropical cyclone landfalls, so it is reasonable to expect that the NHC and WFO-based TSHWWs are relatively consistent.

WFO-issued TSHWWs were obtained in GIS format from Iowa State University's Iowa Environmental Mesonet (IEM 2022a). The IEM archive allowed us to filter data by date and time, region, and other variables. The parameters used to filter the data for each type of watch and warning are shown in Table 6.

Table 6. Valid Time Event Code (VTEC)variable values used to subset shapefiles by tropical cyclone-related hazard type. Shapefiles obtained from IEM Archived NWS Watch, Warnings, Advisories webpage: (https://mesonet.agron.iastate.edu/request/gis/watchwarn.phtml). Lookup table can be found at https://github.com/akrherz/pyIEM/blob/main/src/pyiem/nws/vtec.py.

Variable Names	VTEC Phenomenon	VTEC Significance
Tropical storm watch	Tropical storm (“TR”)	Watch (“A”)
Tropical storm warning	Tropical storm (“TR”)	Warning (“W”)
Hurricane watch	Hurricane (“HU”)	Watch (“A”)
Hurricane warning	Hurricane (“HU”)	Warning (“W”)
Flood watch	Flood (“FA”, “FL”) Flash Flood (“FF”)	Watch (“A”)
Flood warning	Flood (“FA”, “FL”) Flash Flood (“FF”)	Warning (“W”)
Tornado watch	Tornado (“TO”)	Watch (“A”)
Tornado warning	Tornado (“TO”)	Warning (“W”)

iv. Tropical cyclone wind speed probabilities

The NHC provides probabilities of 34-kt (tropical storm-force), 50-kt, and 64-kt (hurricane-force) winds from forecast valid time t=0 to 120 hours. Tropical cyclone wind speed probabilities (WSPs) are updated every 6 hours at 0000, 0600, 1200, and 1800 UTC. NHC displays WSPs on their main webpage as color contour graphics (Figure 16) and as a text product that lists WSP values at specific locations, e.g., major cities and/or predetermined coastal (NHC 2022d). NHC wind speed probability data were obtained in shapefile format from the NHC GIS archive (NHC 2022a). The shapefiles provide geometry information for the WSPs as thresholded on the NHC WSP graphic (0–10%, 10–20%, …, >90%).

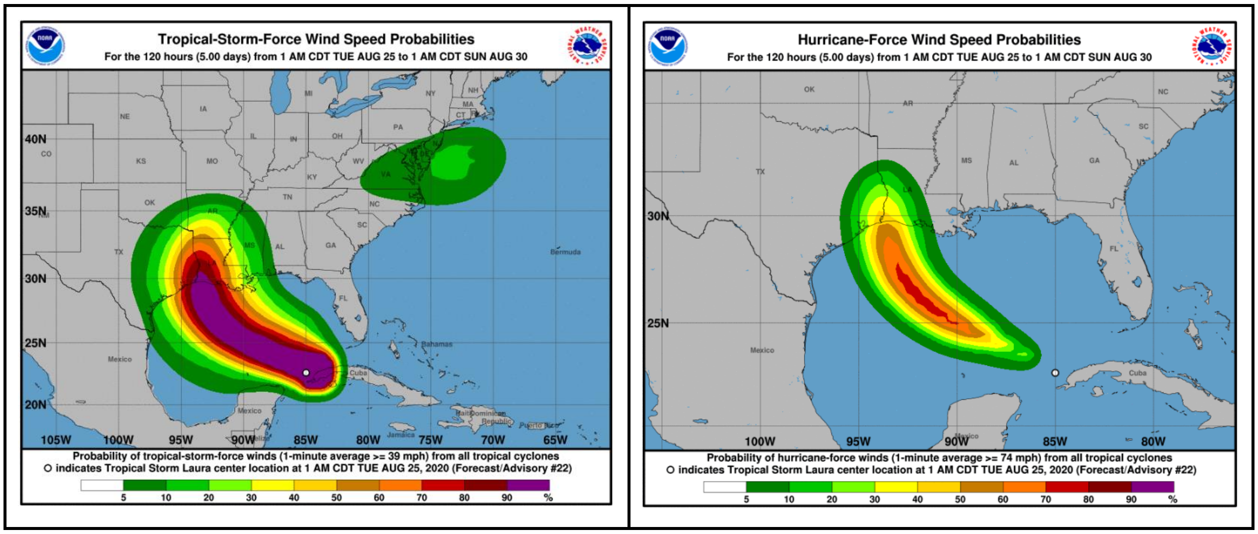


Figure 16. The cumulative 5-day tropical-storm-force (34-kt, left) and hurricane-force (64-kt, right) wind speed probabilities for Hurricane Laura on 8/25/20 0600 UTC (Forecast Advisory #22).

v. Storm surge watches and warnings

Storm surge watches and warnings are issued for active tropical cyclones whenever life-threatening inundation from storm surge is possible along any portion of the Gulf or Atlantic coasts of the continental United States within the next 48 hours of TC landfall (NHC 2022b). A storm surge watch suggests the possibility of life-threatening inundation from rising water moving inland from the shoreline somewhere within the specified area, generally within 48 hours, in association with an ongoing or potential tropical cyclone, a subtropical cyclone, or a post-tropical cyclone (NHC 2022c). A storm surge warning indicates the danger of life-threatening inundation from rising water moving inland from the shoreline somewhere within the specified area, generally within 36 hours, in association with an ongoing or potential tropical cyclone, a subtropical cyclone, or a post-tropical cyclone (NHC 2022c). Storm surge watches and warnings are issued every six hours at 0300, 0900, 1500, and 2100 UTC. Storm surge watches and warnings were obtained in shapefile format from the NHC GIS archive (NHC 2022a). An example of the storm surge watches and warnings issued for Hurricane Laura is shown in Figure 17.

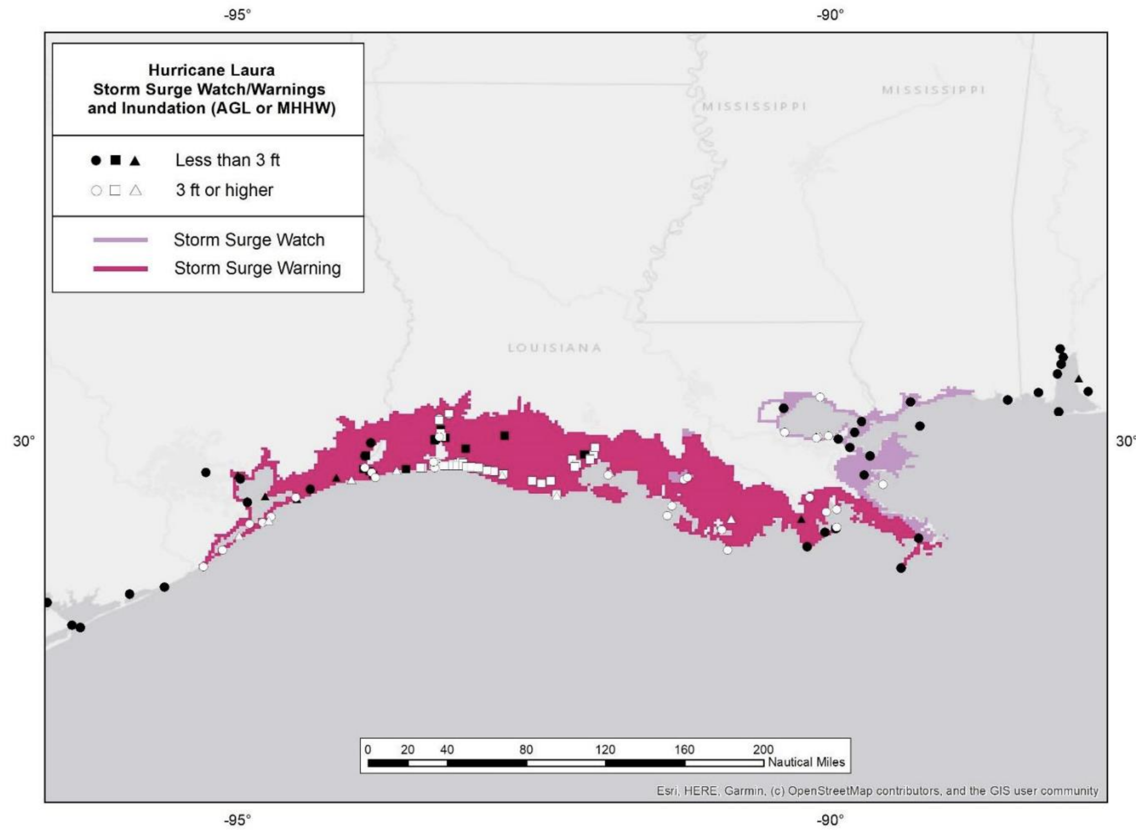


Figure 17. Maximum water levels measured during Hurricane Laura from tide gauges (circles), USGS water level sensors (triangles) and surveyed high water marks (squares), as well as areas covered by storm surge watches (lavender) and warnings (magenta). (Source: Pasch et al. 2021, https://www.nhc.noaa.gov/data/tcr/AL132020_Laura.pdf)

vi. Excessive rainfall outlooks

The NWS Weather Prediction Center (WPC) is responsible for issuing Excessive Rainfall Outlooks (EROs), which are forecasts of the probability that rainfall will exceed flash flood guidance (FFG) within 40 kilometers (25 miles) of a point. Gridded FFG is provided by the twelve NWS River Forecast Centers (RFCs) whose service areas cover the lower 48 states. The Day 1 ERO is issued at 0830 UTC and is valid from 1200 UTC Day 1 (current day) to 1200 UTC Day 2 (the next day). The Day 1 ERO is updated at 1600 UTC Day 1 and 0100 UTC Day 2. The Day 2 ERO is issued at 0830 UTC and is valid from 1200 UTC Day 2 to 1200 Day 3. The Day 3 ERO is issued at 0830 UTC and is valid from 1200 UTC Day 3 to 1200 Day 4. Both the Day 2 and Day 3 EROs are updated at 2030 UTC. Examples of the ERO forecasts for Hurricane Laura issued on August 25, 2020, are shown in Figure 18.

ERO data are available in shapefile format at https://ftp.wpc.ncep.noaa.gov/shapefiles/qpf/excessive/. However, the ERO data are only archived on this site for approximately two months. For the 2020 survey analysis (Laura/Marco), we were able to obtain data more than two months after the storm (past the archival period) in GRIB format from the WPC Science and Operations Officer, Mark Klein. GRIB is a gridded format, and required additional interpolation and processing to be used for our purposes. We then set up scripts to automatically collect the real-time GIS format data in real time going forward, which we used for the 2021 survey analysis.

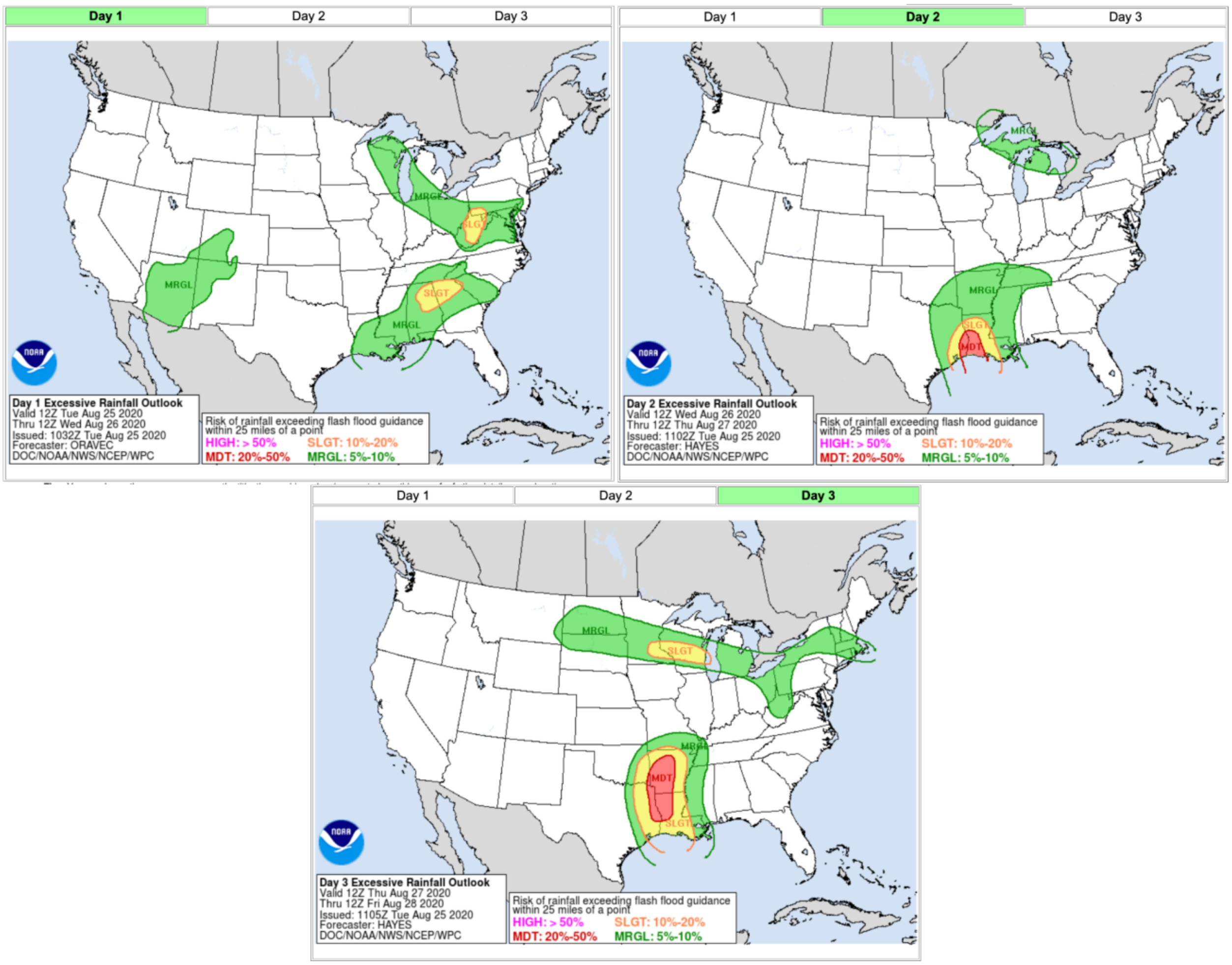


Figure 18. The Day 1 (upper left), Day 2 (upper right), and Day 3 (bottom) Excessive Rainfall Outlooks (ERO) for Hurricane Laura issued on 8/25/20 at 0900 UTC.

vii. Flood watches and warnings

Flood watches and warnings and flash flood warnings are issued by local WFOs. Flood watches are issued when conditions are favorable for flooding. Flood (flash flood) warnings are issued when flooding (flash flooding) is imminent or occurring (NWS 2022). Flood watches and warnings were obtained in GIS format from IEM (IEM 2022a). To capture all the various types of flood watches and warnings issued by the NWS, we included any phenomena code mentioning the term "flood" in the IEM VTEC lookup table (https://github.com/akrherz/pyIEM/blob/main/src/pyiem/nws/vtec.py). The parameters used to filter the data for each type of watch and warning are shown in Table 6.

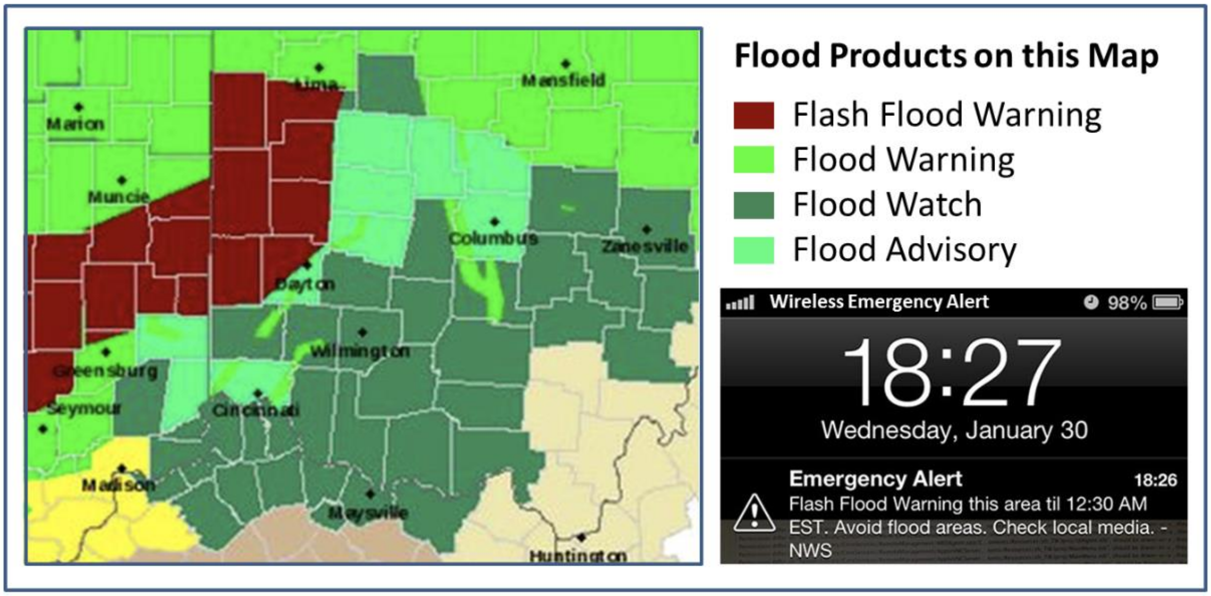


Figure 19. Examples of NWS flood products. Source: https://www.weather.gov/safety/flood-watch-warning

viii. Convective outlook tornado probabilities

The NWS Storm Prediction Center (SPC) issues Day 1, Day 2, and Day 3 Convective Outlooks that depict severe thunderstorm threats across the contiguous United States (NWS 2022b). The categorical forecast specifies the level of the overall severe weather threat as numbers (e.g., 5), descriptive labeling (e.g., high), and colors (e.g., magenta). The probabilistic forecast directly expresses the best estimate of severe convective weather occurring within 40 kilometers (25 miles) of a point. The level of categorical risk in the Day 1–3 Convective Outlooks is derived from probability forecasts of tornadoes, damaging winds, and large hail on Days 1 and 2, the two time periods for which SPC probability forecasts are issued. Convective outlook Day 1 and Day 2 tornado probabilities were obtained in GIS format from IEM (IEM 2022a). An example of the Day 1 convective outlook tornado probabilities ≥ 2% are shown in Figure 20.

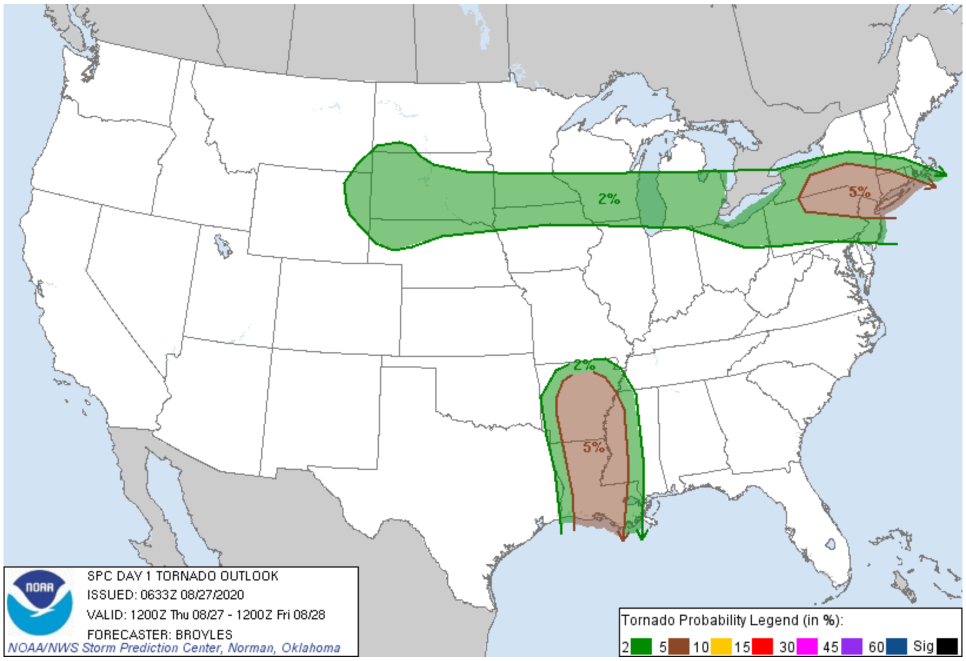


Figure 20. SPC convective outlook tornado probabilities issued at 0600 UTC on August 27, 2020.

ix. Tornado watches and warnings

Tornado watches and warnings are issued by NWS WFOs. A tornado watch suggests that tornadoes are possible in and near the watch area, while a tornado warning indicates that a tornado has been sighted or indicated by weather radar and there is imminent danger to life and property (NWS 2022c). Tornado watches and warnings were obtained in GIS format from IEM (IEM 2022a). The parameters used to filter the data for each type of watch and warning are shown in Table 6.

x. Evacuations

There is no publicly available database (known to us) for evacuation orders. Evacuations are ordered by a variety of entities in the United States, depending on the region. Kruger et al. (2020) documented the evacuation laws in eight southern U.S. states and found that the legal authority for issuing tropical cyclone-related evacuations ranged from local officials to the governor, with that role shared by a variety of local and state officials in several states (Figure 21).

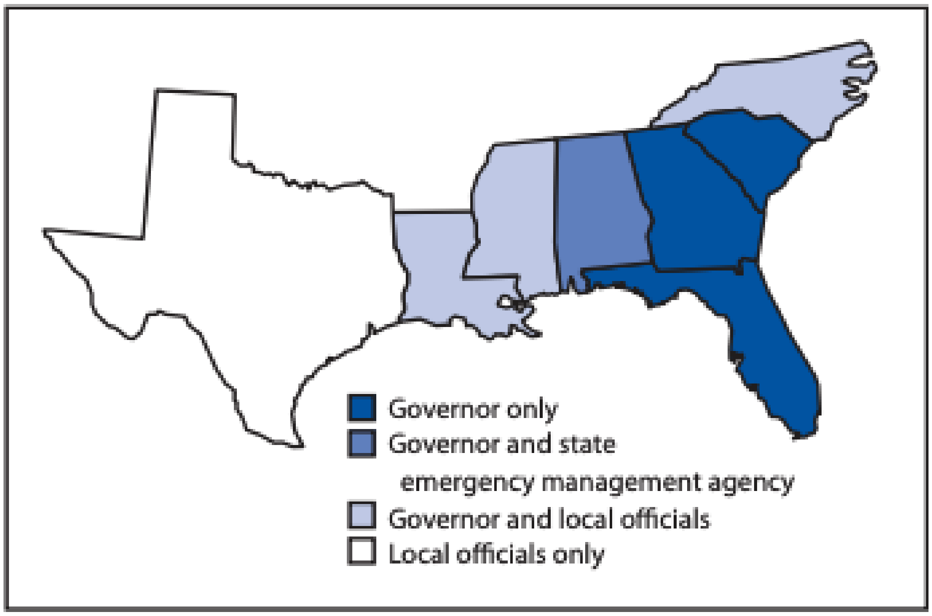


Figure 21. Government agencies granted legal authority to order large-scale evacuation during natural disasters - eight southern U.S. coastal states, December 31, 2018. Source: Kruger et al. 2020.

For the purposes of this study, evacuation data were identified via web searches[6] and compiled in spreadsheets. The majority of this information was found on local news websites (articles and social media) and government websites (official and social media). For the 2020 cases of Laura, which made landfall in southwest Louisiana, and Marco, which passed just south of the mouth of the Mississippi River, several comprehensive lists of evacuations were found posted by local news outlets online, and most of the entries on those lists had links to the official sources of the evacuation orders. Therefore, we have reasonable confidence in our evacuation data for those cases. On the other hand, evacuation orders for Henri in 2021 were much harder to find. Evacuation orders in the New York and New England areas are issued by local municipalities (on the town/city level in most cases), and we were unable to find comprehensive listings by news outlets like we did for Laura and Marco in 2020. This allowed for each municipality to use its own dissemination practices and wording for evacuation guidance (see also Cuite et al. 2017), the latter of which made ascertaining when a voluntary evacuation was ordered particularly difficult. For Henri, “voluntary evacuations” were not always found to be explicitly ordered as such, but rather indicated by the opening of shelters and other resources. For these reasons, although we are reasonably confident that we identified most if not all mandatory evacuation orders, we are less confident about our database of voluntary evacuation orders for Henri.

A summary of all meteorological and evacuation order data is provided in Table 7.

[6] Web searches were conducted using the search term "Hurricane [X] [Year] evacuation" and other combinations of the words therein.

Table 7. Summary of the meteorological datasets collected and used for analysis.

Product name	Issuance times	Forecast period used	Issued by	Data source	Data format
Wind					
Cone of uncertainty	0300, 0900, 1500, 2100 UTC (intermediate 0000, 0600, 1200, 1800 UTC)	0–120 hours	NHC	NHC GIS web	shapefile
Tropical storm (34-kt) and hurricane (64-kt) wind speed probabilities	0000, 0600, 1200, 1800 UTC	0–120 hours (cumulative)	NHC	NHC GIS web	shapefile
Inland tropical storm and hurricane watches	~48 hours before expected landfall	all issued	WFO	IEM	shapefile
Inland tropical storm and hurricane warnings	~36 hours before expected landfall	all issued	WFO	IEM	shapefile
Rainfall					
Excessive Rainfall Outlook (ERO)	Issued 0830 UTC. Day 1 updated 1600 and 0100 UTC. Days 2-3 updated 2030 UTC.	Day 1, Day 2, and Day 3 (and all updates)	WPC	WPC FTP	Grib2 (2020), shapefile
Flood and flash flood watches	varies	all issued	WFO	IEM	shapefile
Flood and flash flood warnings	varies	all issued	WFO	IEM	shapefile
Storm Surge					
Storm surge (SS) watches	~48 hours before expected landfall	all issued	NHC	NHC GIS web	KML
Storm surge (SS) warnings	~36 hours before expected landfall	all issued	NHC	NHC GIS web	KML
Tornado					
Convective Outlook Tornado Probabilities	Day 1, 2, 3, 4–8 issued 1-5 times/day	all issued	SPC	IEM	shapefile
Tornado watches	varies	all issued	WFO	IEM	shapefile
Tornado warnings	varies	all issued	WFO	IEM	shapefile
Other					
Evacuation orders	varies	all issued	state & local	web search	NA

b. Relating meteorological data and survey responses

All respondents were asked to provide their zip code during Wave 1. Spatial boundaries for each zip code were obtained using R (v. 4.0.2) and the U.S. Census Bureau Zip Code Tabulation Area (ZCTA) dataset from the *tigris* package (v. 1.6.1). In addition, response start and end times were recorded for each survey wave. We used this combination of spatial and temporal data to determine the weather risk messages (WRM) relevant to each survey response.

Because we did not know where respondents were located within each zip code, we considered a respondent spatially within a WRM product if any part of their ZCTA intersected with the area. Respondents were matched temporally with the WRM that was valid at the time the survey was started. Figure 22 shows examples of how various response zip codes were considered within (black outline) and not within (grey outline) various WRM areas (shown in red).

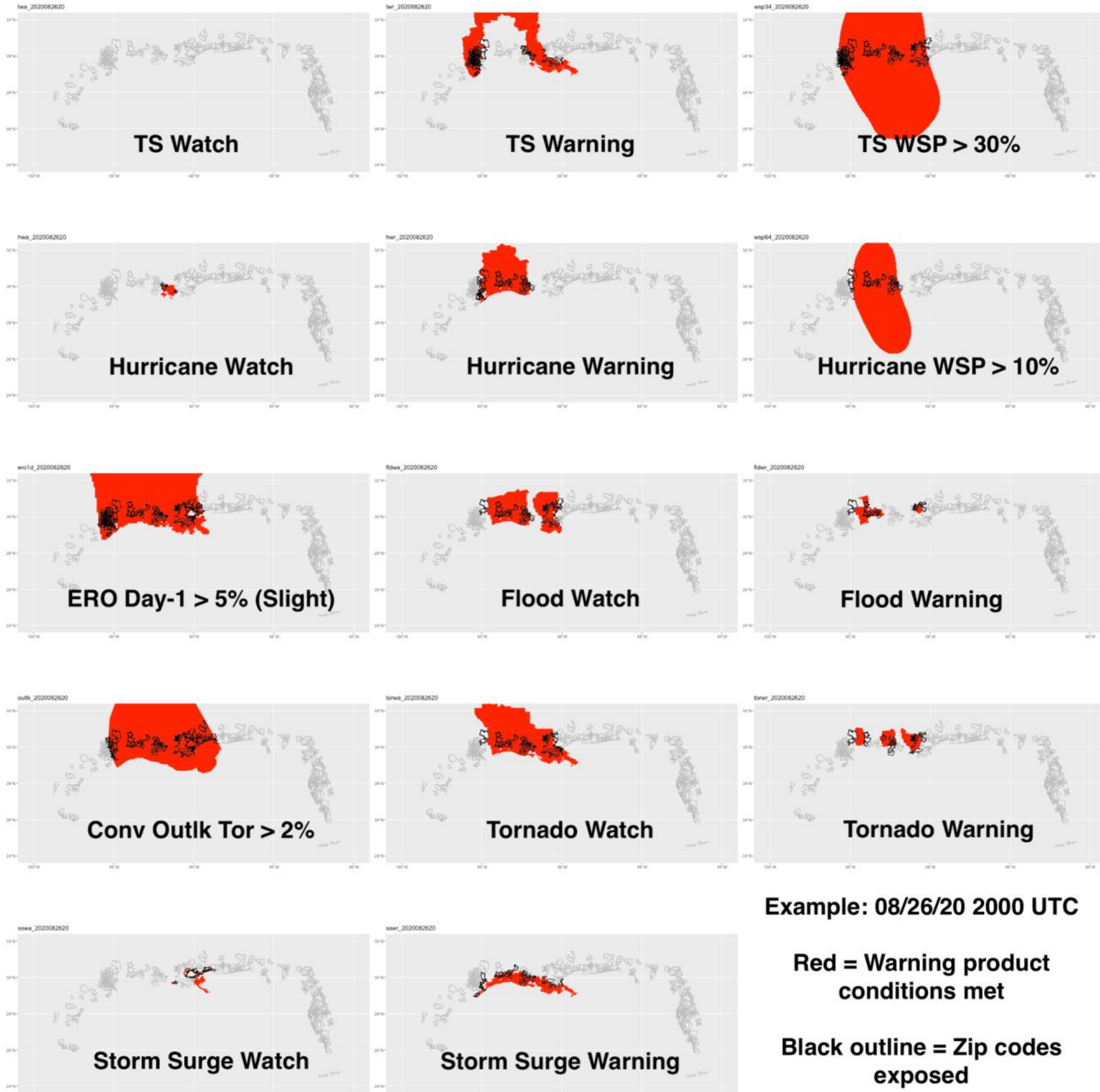


Figure 22. Examples of how various response zip codes were considered within (black outline) and not within (grey outline) various WRMP areas (shown in red) from Hurricane Laura on August 26, 2020, at 2000 UTC. Note that for non-deterministic products (e.g., wind speed probabilities and EROs) a single threshold value is chosen to allow for a binary (within / not within) classification.

c. Defining and characterizing TC exposure versus non-exposure

Because we surveyed over broad geographic areas, both along the coast and inland from it, the respondents in our sample did not have equivalent chances—or necessarily even high chances—of being directly affected by TC hazards, especially as the TC threat evolved and became more refined. To compare survey responses of people in the areas that had a greater versus lesser chance of being directly affected by TC hazards and the risks they pose, we defined and characterized TC exposure versus non-exposure and categorized people into these mutually exclusive groups at each wave.

We defined TC exposure by drawing on the concept of exposure in risk literature, which is the chance of being affected by a hazard (Walpole and Wilson 2020; SRA 2022). Accordingly, we defined TC exposed areas as those that had a higher chance of being affected by TC hazards based on the information available in real time by NWS forecasters and emergency response officials (rather than based on the information available retrospectively, after the TC, based on what areas were actually affected). Specifically, with input from our NOAA collaborators, we categorized people as TC exposed during a given wave if they lived in an area that met one of any of these conditions: (a) was inside the NHC's cone of uncertainty, (b) was under a hurricane or tropical storm watch or warning, (c) was under a storm surge watch or warning, (d) was inside the area where the tropical storm wind speed probability was forecast to be 30% or greater, or (e) was under a mandatory or voluntary evacuation order. These forecast products and evacuation orders represent, in the moment, meteorologists' and emergency response officials' best assessment of the types, magnitude, and locations where TC risks could occur—that is, of potential TC exposure—as well as the best information a member of the public could obtain about their exposure to TC risks leading up to the time they took the survey.

Respondents were matched spatially to these products if any part of the zip code they resided in overlapped by any amount with any of these products. Respondents were matched temporally if any of these products overlapped with their zip code based on a time window of 12 hours before they started the survey through when they completed the survey. We chose this 12-hour window as a feasible and reasonable amount of time that one could become newly or differently exposed to TC risks based on the issuance of these products and have time to acquire this updated risk information. The outcome of this matching was that each respondent was designated "exposed" or "not exposed" for each predictive wave.

Figure 23 provides an example result from the survey that shows how categorization of survey respondents as TC exposed versus not exposed supports more nuanced data analysis of the behaviors of people who reside in areas where the TC risks increased versus decreased as the threat evolved. Specifically, the results show how TC exposed respondents got forecast information on average more frequently from most sources over the prior 24 hours than those who were not exposed, and they tended to get information more frequently at Wave 3 than at Waves 1 or 2.

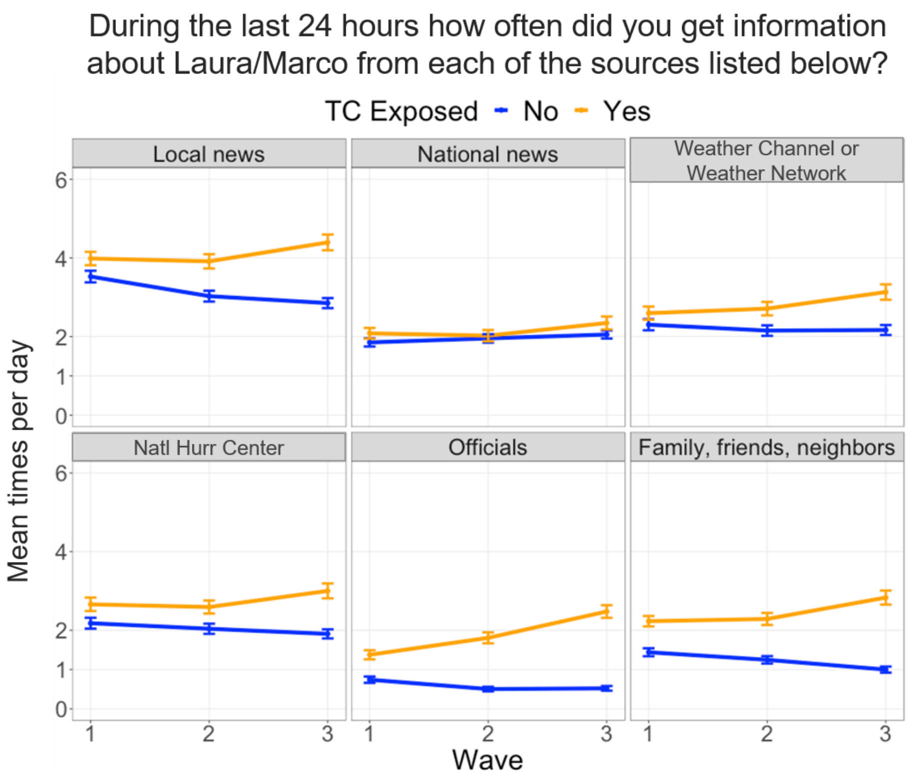


Figure 23. Example survey result showing mean frequency of getting information over the prior 24 hours from different sources across each survey wave for respondents categorized as TC exposed (orange) and not exposed (blue).

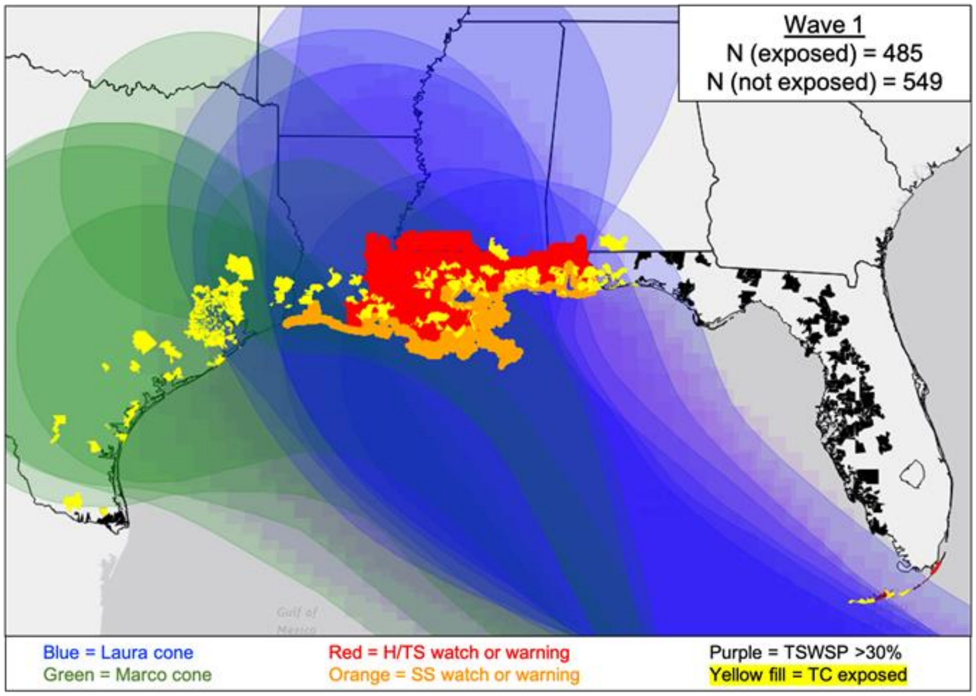
i. TC exposed versus non-exposed areas for Laura and Marco (2020)

Figure 24 is a map-based rendering of all curated forecast products used to categorize respondents as TC exposed or not for each predictive survey wave for TCs Laura and Marco. The figure includes mapping of the cones of uncertainty for Laura (blue) and Marco (green), hurricane or tropical storm watches or warnings (red), and storm surge watches or warnings (orange) that were in effect during the time periods that Wave 1 (Figure 24a), Wave 2 (Figure 24b), and Wave 3 (Figure 24c) were fielded. The zip codes denoted in yellow are those for which the respondent was categorized as TC exposed based on their spatial and temporal mapping, and black zip codes are those for which the respondent was categorized as non-exposed. The number of respondents in each of these areas is denoted in the upper right-hand portion of each figure.

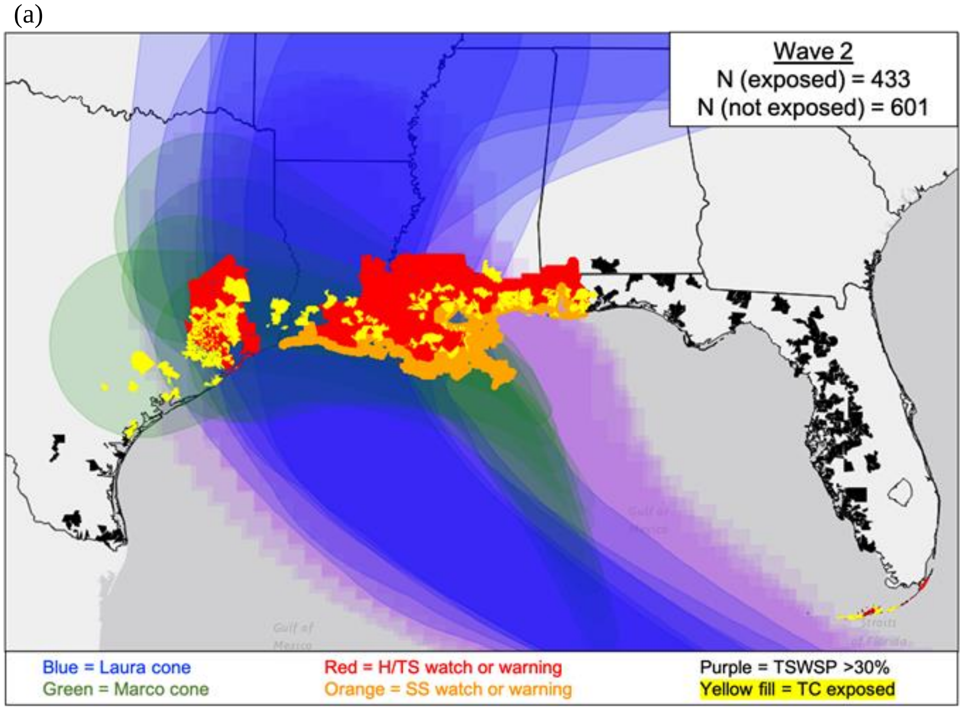
The figure illustrates that a very broad area, from Texas to the Florida panhandle, was TC exposed during Wave 1[7] (Figure 24a). The TC exposed area shrank only slightly on either side during Wave 2 (Figure 24b). The TC exposed area was more refined by Wave 3 (Figure 24c) due to Marco having nearly made landfall (with its center passing just south of the mouth of the Mississippi River) prior to that wave and to Laura being close to landfall. Note that in this situation, our analysis designates fewer respondents exposed in each wave, and respondents who

[7] Those who were not exposed at Wave 1 were exposed when we decided approximately 24 hours prior to field the survey.

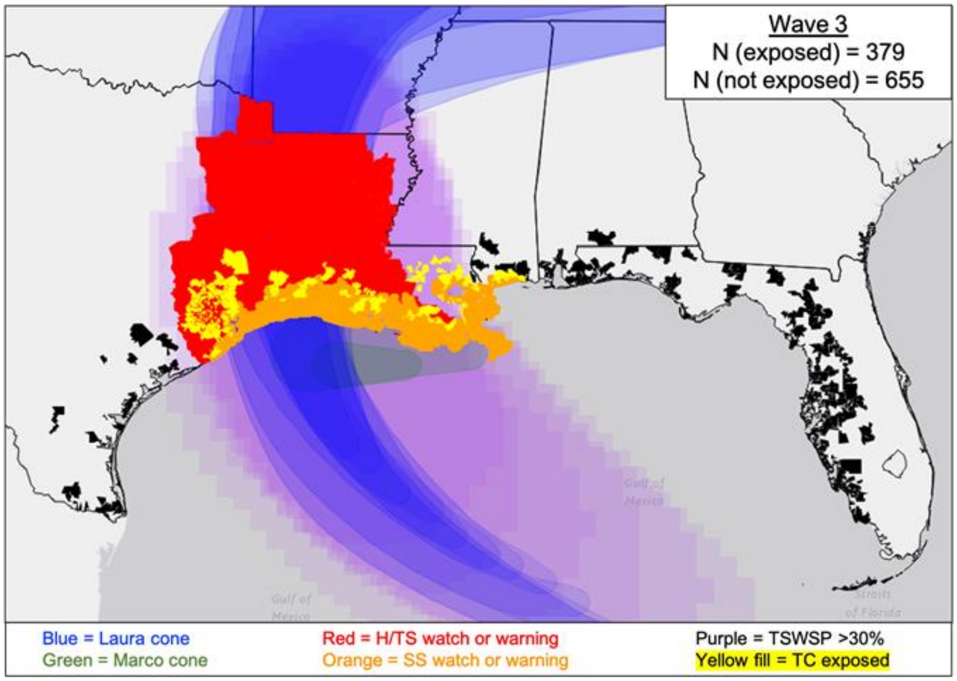
are not exposed in an earlier wave do not become exposed in a later wave. For a TC with less accurate forecasts, however, more complex patterns of exposed versus not exposed could occur.



(a)



(b)



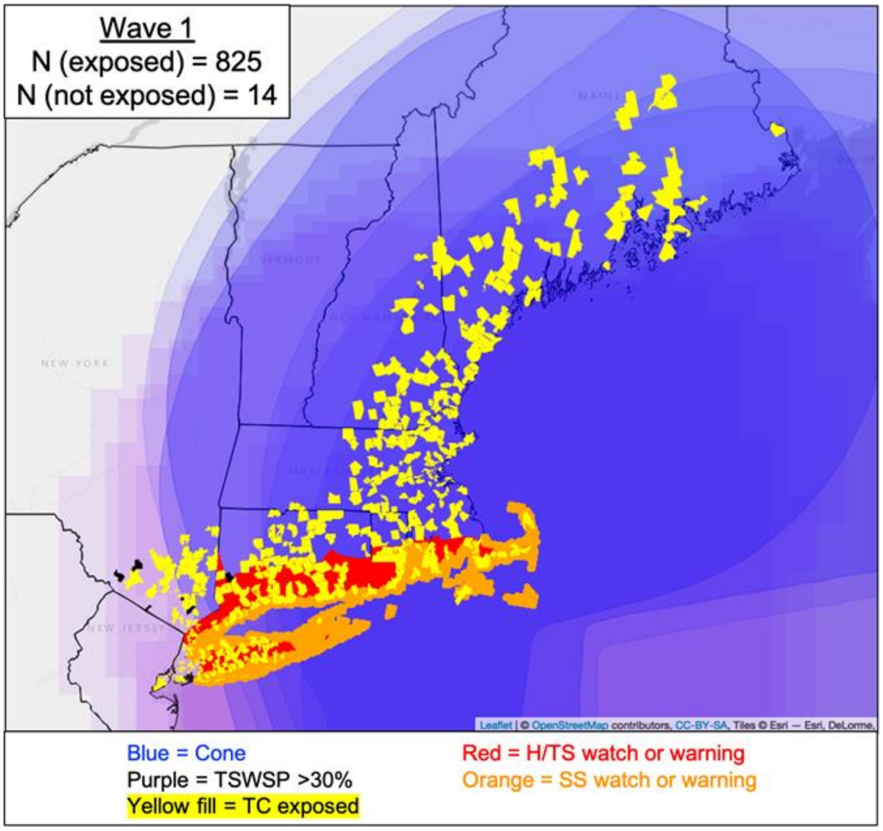
(c)

Figure 24. Map-based rendering of all forecast products used to categorize respondents as TC exposed or not for TCs Laura and Marco for (a) Wave 1, (b) Wave 2, and (c) Wave 3. Mapped are the cones of uncertainty for Laura (blue) and Marco (green), hurricane or tropical storm watches or warnings (red), and storm surge watches or warnings (orange) that were in effect during each wave and corresponding zip codes of respondents who were categorized as TC exposed (yellow) or not (black).

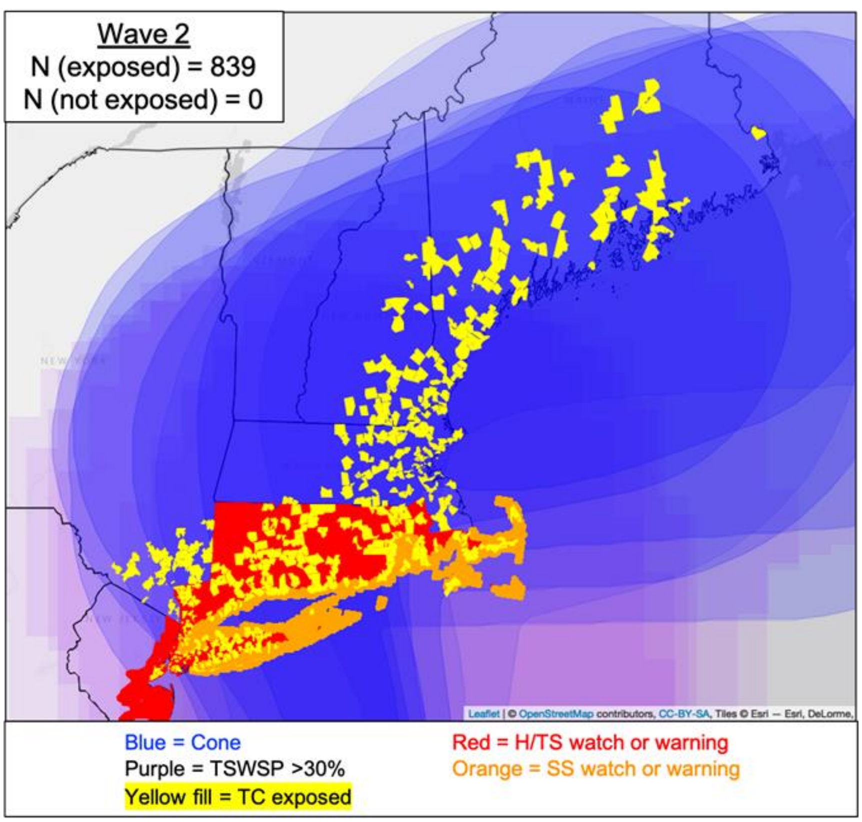
ii. TC exposed versus non-exposed areas for Henri

Figure 25 is a map-based rendering of all curated forecast products used to categorize respondents as TC exposed or not for each predictive survey wave for TC Henri. The figure includes mapping of the cones of uncertainty (blue), hurricane or tropical storm watches or warnings (red), and storm surge watches or warnings (orange) that were in effect during the time periods that Wave 1 (Figure 25a), Wave 2 (Figure 25b), and Wave 3 (Figure 25c) were fielded. The zip codes denoted in yellow are those for which the respondent was categorized as TC exposed based on their spatial and temporal mapping, and black zip codes are those for which the respondent was categorized as non-exposed. The number of respondents in each of these areas is denoted in the upper left-hand portion of each figure.

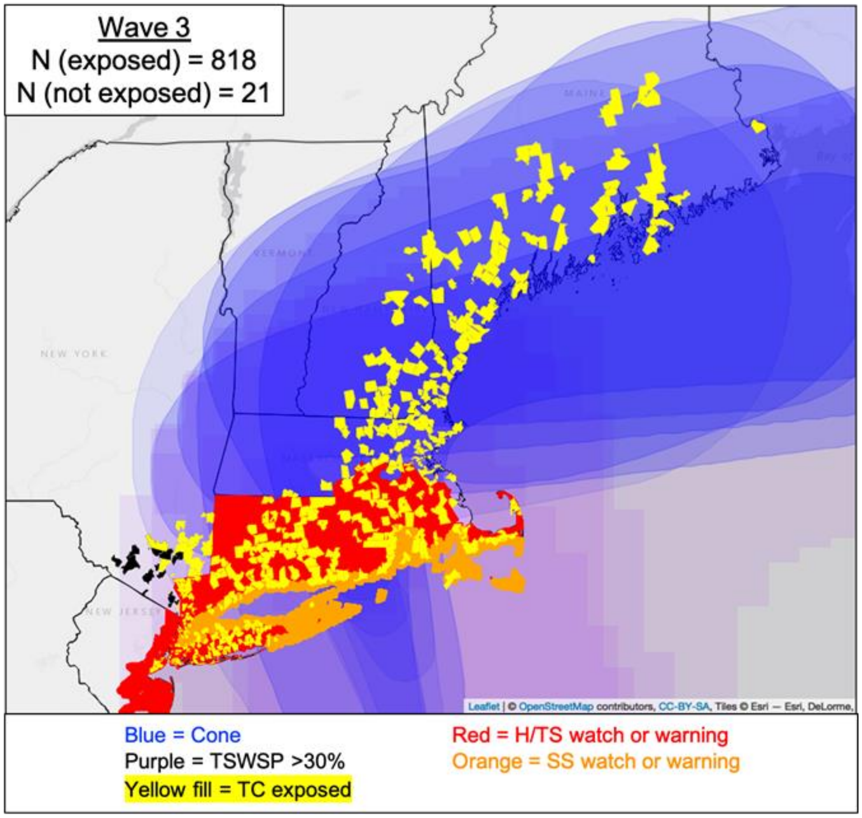
In comparison to the exposure plots for Laura and Marco (Figure 24), the exposure plots for Henri illustrate that respondents from nearly the entire, broad area surveyed were TC exposed across all waves. In fact, no one was categorized as not exposed in Wave 2 (Figure 25b). This illustrates how the evolution of TC exposure, at least based on our classification approach, differs substantially for a TC that tracks more perpendicular to the coast (Laura) versus more parallel to the coast (Henri).


Wave 1
N (exposed) = 825
N (not exposed) = 14
MAINE
VERMONT
NEW YORK
NEW JERSEY
Leaflet | © OpenStreetMap contributors, CC-BY-SA, Tiles © Esri — Esri, DeLorme,
Blue = Cone
Purple = TSWSP >30%
Yellow fill = TC exposed
Red = H/TS watch or warning
Orange = SS watch or warning

(a)


Wave 2
N (exposed) = 839
N (not exposed) = 0
NEW YORK
Leaflet | © OpenStreetMap contributors, CC-BY-SA, Tiles © Esri — Esri, DeLorme,
Blue = Cone
Purple = TSWSP >30%
Yellow fill = TC exposed
Red = H/TS watch or warning
Orange = SS watch or warning

(b)



(c)

Figure 25. Map-based rendering of all forecast products used to categorize respondents as TC exposed or not for TCs Laura and Marco for (a) Wave 1, (b) Wave 2, and (c) Wave 3. Mapped are the cones of uncertainty for Laura (blue) and Marco (green), hurricane or tropical storm watches or warnings (red), and storm surge watches or warnings (orange) that were in effect during each wave and corresponding zip codes of respondents who were categorized as TC exposed (yellow) or not (black).

7. Findings and Recommendations about the Methodology

This report documents our design and implementation of a novel methodology to collect multiple surveys that ask the same questions repeatedly of the same individuals (i.e., a longitudinal panel survey) over multiple days for an active, real-world TC as it evolves. We developed this research approach to systematically collect perishable social science observational data to measure whether, when, and how people get TC risk information, perceive the risks, and respond. We report on the implementation of our longitudinal panel survey approach twice, in two different geographic areas and over two hurricane seasons: first in 2020 for TCs Laura and Marco in the Gulf and second in 2021 for TC Henri in New York and New England. Below, we summarize key findings followed by recommendations to NOAA for future work, all pertaining to the methodology.

a. Findings

FINDING 1. Successfully designing and implementing a novel, rapidly deployed, event-specific longitudinal panel survey during the multi-day predictive phase of a real-world TC required (a) identifying the methodological challenges in detail, (b) assessing the feasibility

of overcoming those challenges, (c) developing a detailed but flexible research approach, and (d) working with "a cross-sector village" of committed research team members and external collaborators who contribute the diverse, needed forms of expertise.

Developing and piloting the methodology for a rapidly deployed, event-driven longitudinal panel survey was, in short, not easy given the multiplicity of challenges and parameters that had to be considered. Thoughtfully identifying, considering, and addressing those parameters to yield a successful outcome took many months and the ideas and expertise of many. Our research team drew on our own disciplinary and interdisciplinary expertise and experience to integrate social science survey research and risk analysis theory with atmospheric science knowledge of TC prediction capabilities, uncertainties, and datasets in order to design and implement this methodology. A critical component of our success was developing a flexible, adaptable approach that included thinking through a variety of potential scenarios and developing strategies for managing associated risks in survey implementation, over many months prior to fielding. In addition, it was essential to have a reputable, accessible, and accommodating (but not overpromising) survey company in YouGov. Furthermore, it was invaluable to have the research support and operational expertise of our NOAA collaborators, both overall and at the critical moments in which we were making decisions about whether or not to field the survey for given TC threats. Much like an atmospheric science field campaign or intensive observation period to collect perishable meteorological observations, it takes extensive planning and "a cross-sector village" of people—from research/academia (NCAR and Stanford University), public sector (NOAA), and the private sector (YouGov)—to collect these perishable social science observations.

FINDING 2. The proof-of-concept methodology yielded convenience samples that are older, more White, more retired, more educated, and well-resourced with adequate access to basic needs. Nevertheless, these samples provide a basis for exploring modifications to the fielding approach and incentive structures to get more diverse or more representative samples.

Given the many challenges with designing and implementing this methodology, we (including YouGov) frankly were unsure whether it would work, that is, whether we could successfully get a sample of people to respond to repeated surveys in such a short amount of time, given both the short fielding periods for each wave and the short intervals in between. Prior to fielding, we did not know what type of sample we would get, and, given available funding, it was only possible to get a convenience sample of people from YouGov's existing survey panel who were willing to respond. The samples we obtained provide a foundation for developing scenarios to modify survey recruitment and/or incentive structure to target and retain samples with sociodemographic characteristics that are more diverse and/or more representative of the population in given areas and for scoping the costs of doing so.

FINDING 3. Curating the meteorological data and evacuation orders for integration and analysis with our survey data (a) is a valuable methodological contribution in its own right, (b) but was time-consuming and challenging to do, even with the meteorological expertise of our team members.

The integration of meteorological data and evacuation orders with the survey data allows for rich and novel analyses. One such analysis is the direct comparison of survey respondents' perceptions about the TC risks with contemporaneous forecast information (e.g., whether a respondent thought they were in a hurricane warning when they responded during a given predictive survey wave versus whether they actually were). Another type of analysis is categorizing our sample into groups of people who were TC exposed versus not exposed, which allows for more refined analyses of whether, when, and how respondents perceive and respond to TC risks. Yet, identifying and accessing all data sources, including publicly available data from NWS, was very time-consuming and challenging—even for researchers with ample experience working with meteorological datasets—given the substantial variability in whether, where, how, and for how long different data are archived. Our data curation efforts involved knowing where to find NHC products and scrape needed forecast parameters (e.g., forecast TC intensity); utilizing an excellent non-NOAA data source (Iowa Environmental Mesonet) to collect many NOAA products; investigating differences between tropical storm and hurricane watches and warnings that are issued by NHC versus by WFOs and determining which to use; leveraging a connection with a NOAA colleague to acquire one of the NOAA datasets; and for evacuation orders, doing web searches to determine who issues orders in different jurisdictions and whether mandatory and voluntary orders were issued. Such inconsistency in data provenance presents challenges, not only for this research project, but for research and other forms of data use more broadly.

FINDING 4. Our planned survey fielding approach was that it would take 6 days from the fielding decision to fully field 3 predictive survey waves—i.e., 24 hours to put the survey in the field, 24-hour fielding periods for each predictive survey wave, and 24-hour intervals between waves. In practice, some aspects took less time, and we explored spending less time in some ways. This suggests there is feasibility to getting a sufficient sample for three predictive survey waves for TCs for which there is less lead time before landfall, but these advantages must be weighed against potential disadvantages.

With the two storms, we were able to observe and to test different survey fielding approaches regarding the lead time of our decision to field, wave length, and wave interval. For Laura and Marco, we obtained a sufficient sample for Wave 1 within 18 hours, and we opted for only an 18-hour interval between Waves 1 and 2 based on the evolving track of Marco in order to field most of Wave 2 before it was forecast to make landfall. It therefore took 12 fewer hours than planned, meaning it took only 5.5 days from fielding decision to fully fielding 3 predictive waves. Accordingly, for Henri, we planned to try fielding for shorter periods with shorter wave intervals, and then we further shortened the interval between Waves 2 and 3, again based on the evolving storm track. Moreover, YouGov worked with us and was able to launch Wave 1 within 12 hours of our fielding decision. In total, it took only 3.5 days from fielding decision to fully fielding 3 predictive waves. Based on these data, we learned that there can be some methodological flexibility in how much time it takes to field a rapidly deployed, event-specific longitudinal panel survey based on the TC situation. Further work is needed for additional storms to more fully characterize the pros, cons, and trade-offs of different approaches—such as the ability to get sufficient samples across three waves of data to examine perceptions and behaviors for TCs that form and threaten landfall with shorter lead times versus whether this privileges responses from certain groups and disadvantages responses from more marginalized populations.

Our research also suggests that it is feasible to conduct a rapidly deployed, cross-sectional survey in situations where having only one snapshot of data would be beneficial.

b. Recommendations

We offer a few key research-guided recommendations for NOAA to modify and expand on the methodology reported here. Our recommendations are supported by many of the Priorities for Weather Research report from the NOAA Science Advisory Board (NOAA 2021), particularly from Recommendations ID-1, ID-1.1, ID-4, ID-4.3, ID-6, ID-6.2, and ID-6.3.

RECOMMENDATION 1. The longitudinal panel survey methodology designed and implemented here that is event-specific and rapidly deployed during the predictive phase of a hazardous weather threat yielded collection of novel, perishable social science observational data for real-world TCs as they evolved. Collection of such data for additional TCs should be prioritized to develop more comprehensive datasets that will facilitate more robust understanding of people's perceptions and behaviors in response to forecast and other risk information provided by NOAA and its partners when TCs threaten. This actionable social science in turn will help identify where improvements are most needed in NWS's forecast product suite content and in dissemination of information across the forecast and response system.

The methodology that our cross-sector team (NCAR, Stanford University, NOAA, and YouGov) developed and implemented has produced unparalleled social science observations that begin to fill a critical knowledge gap regarding whether, when, and how people are getting TC forecast and other risk information, are perceiving the risk, and are responding—all as the TC threat evolves and approaches landfall. These perishable data are essential for NOAA and its partners to understand how the evolving TC meteorological observations and forecast and preparedness information they provide shape people's perceptions and behaviors, for actual, real-world TC threats and where improvements could be made.

Although having such social science observational data is wildly illuminating, given that the baseline was having no such data, having data from only two TC events is limiting. The equivalent would be if the atmospheric science community gathered meteorological observations for only Laura/Marco and Henri—or any two TCs—and deemed this sufficient to understand TC processes. Having data from only two storms limits scientific capabilities to robustly understand what findings do and do not generalize across populations (e.g., demographics, cultures, natural and built infrastructures) and across TC scenarios (e.g., steady versus rapidly changing intensities, shifting tracks, inland flooding versus surge flooding versus wind hazards, TCs that form farther out in the ocean and thus have more days of forecasting and media attention versus ones that form closer to land). Scaling this systematic data collection approach over multiple hurricane seasons would provide a more comprehensive set of data, which would support analytic capabilities to better understand the social mechanisms and the "boundary conditions" that explain the informational, perceptual, and behavioral processes that people engage in in response to dynamic TC forecast information. Exploring these features is critical for understanding both why and when different TC scenarios influence perceptions and responses, which in turn is necessary to develop effective communication and policy interventions to respond nimbly and ably to similar events.

Scaling the rapidly deployed, event-specific longitudinal panel survey methodology is a multi-year effort. We suggest two possible approaches. One suggestion is to invest over the next three years in collecting data for several more TCs to further test and revise the methodology—for instance, to field over shorter periods or to get a more diverse sample (see Recommendation 1a)—and to develop and more fully build out capabilities outlined below in Recommendations 2–5. Another suggestion is to move forward with implementing the methodology we have developed for a minimum of 20 TCs over the next few years, as a form of Intensive Social Science Observation Periods (ISSOPs).[8] Either of these approaches should involve evaluating the survey, including the survey length, measurement quality, and content.

RECOMMENDATION 1a. Different sampling approaches should be developed to acquire more diverse survey samples, including targeted efforts toward (a) more socioeconomic diversity, including more vulnerable populations, and (b) samples that are representative of the population in the geographic areas at risk from different TCs.

Interconnected with the Recommendation 1 need to collect data for additional TCs, there is a need to design and test different approaches for obtaining survey samples that are more socioeconomically diverse—e.g., on age, race, education, and employment status—or that are based on more nuanced, composite indicators of vulnerability. Related efforts should be made to pursue whether it is feasible to obtain longitudinal data from samples that are demographically representative of the population. Such data would allow for richer analyses of individual and socio-cultural differences in informational, perceptual, and behavioral responses, and of barriers that historically marginalized populations encounter when faced with TC risks. Developing different sampling strategies requires considering the feasibility of diverse and/or representative sociodemographic characteristics on existing survey panels, providing different incentive structures (e.g., higher incentives for demographic groups who are less likely to respond), and having a flexible survey fielding approach (e.g., fielding each predictive survey wave for longer to allow people more time to respond).

There are multiple possible approaches to developing different sampling strategies. One approach is to modify our existing sampling strategy with YouGov, by employing targeted recruitment of specific demographic groups and offering higher incentives. Another approach is to work with YouGov (or another survey company) to augment their existing survey panel by additional recruiting or by using complementary methods to obtain a more diverse or more representative sample in the coastal areas of interest for our research. A third approach is to work with NOAA (and potentially private sector companies) to develop a standing panel of participants specifically for conducting this research.

RECOMMENDATION 2. A mechanism could be developed for analyzing in near real time the rapidly deployed, event-specific social science observational survey data to identify critical misperceptions and/or lack of awareness about TC risks and to provide near real-time, actionable input to NWS to guide TC forecast messaging interventions—in other words, to operationalize "incident" TC risk communication alongside operational TC

[8] IOPs or Intensive Observation/Observing Periods are common in the atmospheric science community to collect physical science observations.

forecasting. NOAA should invest in research and development to explore developing this capability.

Meteorological observational data characterize the state of the atmospheric system and can be compared to data collected at prior time steps to detect important changes, all of which inform weather predictions and messaging (e.g., about TC intensities, tracks, hazards, and related changes as the TC threat evolves). The meteorological observational system is extensive and includes regular, ongoing data collection (e.g., from satellites, radar, buoys) as well as supplemental, targeted data collection (e.g., from dropsondes). The social science observational data collected via the methodology discussed here could be thought of and utilized in similar ways. The data characterize the state of the human system, and people's perceptions and behaviors regarding TC risks can be compared with actual meteorological risks as best known at that time. If there are critical mismatches—such as important misperceptions or lack of awareness among some groups—this knowledge could be relayed to NWS in near real time and used to guide decisions in how TC forecast and emergency response information is messaged by NWS and its partners. Such near real-time analysis and feedback to NWS could be done based on data from single, rapidly deployed, event-specific surveys (e.g., cross-sectionally) or from data collected over multiple waves (e.g., longitudinally). Different strategies may need to be developed for different TC scenarios (e.g., rapid intensification, shifting TC tracks). Different strategies also should consider the different sampling approaches, discussed in Recommendation 1a, so that messaging interventions align with the populations surveyed.

Developing the mechanism for this operational, "incident" TC risk communication[9] effort would require building out four main, interrelated components: (a) streamlining the workflows for survey implementation and for access and use of the corresponding meteorological data, including automating as much of the workflows as possible; (b) developing capabilities for rapid data analysis, also automating as much as possible, and capacities for rapid data interpretation; (c) evaluating the utility of the different social science observations resulting from the survey, specifically by working closely with NOAA partners to identify which results are most important and useful for NWS to receive in real time to inform their forecast messaging decisions, and modifying the survey instrument as needed; and (d) establishing communication mechanisms with NWS to provide the "incident" TC risk communication, to guide whether, how, and for whom they might create TC forecast messaging interventions to communicate the risks differently or for different groups.

Importantly, the workforce and infrastructural resources needed to initially explore and then more fully build out this operational, "incident" TC risk communication capability should be co-developed in close partnership with NOAA, as part of the research-to-operations effort.

RECOMMENDATION 3. It should be explored how to expand the longitudinal panel survey methodology designed and implemented here (event-specific, rapidly deployed, during the predictive phase) to other types of hazardous weather threats, beyond TCs. For example, this methodology could be used for weather threats that tend to be longer-fused and spatially broad (e.g., winter storms, atmospheric rivers, heat) and to threats that tend to be shorter-fused and spatially localized (e.g., severe convective storms, fire weather).

[9] This draws on the idea of incident meteorologists or IMETs providing event-based, specific forecasts for user decision-making.

Although we have tested the methodology developed here for TCs, it has potential to develop knowledge about people's perceptions and behaviors in the context of other types of hazardous weather. This research provides an important methodological foundation to build on, but several aspects must be explored to extend it to other hazards, especially those that vary in temporal scale (i.e., lead time) and spatial scale (i.e., areal coverage).

One important consideration is the population size in different geographic areas, especially for more spatially localized threats, such as tornadoes or fire weather, and for threats to more rural areas (e.g., the upper Midwest, Western United States) regardless of the spatial scale of the threat. Relatedly, if an existing panel of a survey company is used to sample from, as we did with YouGov for our proof-of-concept with hurricanes, the panel feasibility numbers for different areas must be considered in conjunction with anticipated response rates and attrition rates. Next, it should be explored how to field at least three survey waves during the predictive phase of a weather hazard over a shorter time frame. This is important for extending this methodology to shorter-fused hazards, and having more flexibility to field multiple waves in less time also would allow the methodology to be used for, say, TCs that form closer to landfall and therefore offer less lead time for surveying before impact.

Relatedly, for the different weather hazards, a plan must be developed about which operational NWS and experimental products could be used to identify the geographic areas at risk at lead times of multiple days (e.g., 144 hours or 6 days) before the hazard begins. It also would be important to develop relationships with personnel at relevant operational NWS offices, analogous to our relationship with NHC, who could be available to provide expert input that is valuable when deciding whether or not to field the longitudinal panel survey for a given threat. Finally, TCs are unique in that they are the only weather hazard that is named by NWS, which allowed us to refer to the TC name in our survey to ensure that all respondents were considering the TC we were interested in studying, especially when multiple TCs existed in the Atlantic at once. It should be explored how to effectively refer to other weather hazards, which are unnamed by NWS, to cue the survey respondent to the threat of interest to reduce measurement error.

RECOMMENDATION 4. NOAA/NWS meteorological data and products as well as associated emergency response orders should be made more easily accessible to a broad range of researchers and other users, with consistent data formats, clear and long archival periods, and standardized units (when possible).

As described in Finding 3, the process of curating meteorological data and evacuation orders was time-consuming and challenging for this project and its team members, and these hurdles no doubt exist for other researchers and users. Efforts are needed to improve such data provenance, particularly in support of open-science initiatives (NOAA 2021).

A few NWS product-specific needs emerged from this project. First, all WPC data, including EROs and quantitative precipitation forecasts (QPF), should be available via public archive for as long as space allows. Second, there should be a single location where WFO-based advisories, watches, and warnings can be queried and downloaded, preferably in GIS format. Although the IEM is an excellent resource, it is a volunteer effort and hence it could be discontinued at any time with little notice. Third, SPC Convective Outlooks in GIS format should be archived in a way that they can be downloaded over a range of dates and times. Currently, they can only be downloaded one forecast time at a time.

As discussed above, finding information about evacuation orders for this project involved extensive searching of news, government, and social media websites. There is a need for an accessible, consistent, and searchable database of basic information on evacuation orders (e.g., start time, location, mandatory versus voluntary). Such a database would enable longer-scale studies of evacuation behavior and facilitate interdisciplinary research that requires the integration of evacuation orders with other types of social and physical science datasets.

Overall, all NOAA/NWS meteorological data (that are not excessively large) should be archived in a single, searchable, publicly available location. Preferably, these data should be available in GIS format to allow for easy integration with non-meteorological data (e.g., social data, health data). Importantly, all datasets should be accompanied by documentation that is clear, concise, complete, and understandable by non-experts. The provision of these data and metadata with dedicated resources to support its development and maintenance is important in its own right but is also essential for facilitating inter- and transdisciplinary research that integrates atmospheric and social sciences.

RECOMMENDATION 5. A dashboard or other web-based platform should be developed to make publicly available the longitudinal social science observational data as well as the corresponding meteorological and evacuation data. These data should be accompanied by detailed metadata about survey development, data quality control, data treatment, and data source.

A dashboard or other platform for data sharing should be developed in support of open science initiatives (NOAA 2021). Making these data and corresponding metadata publicly available encourages more extensive, innovative, and equitable scientific discovery and is in-line with the research and observations supported by NOAA being a public good. The dashboard would be useful for sharing the data already collected in 2020 and 2021, and it would be especially useful if there is support for collecting the longitudinal social science observational data for additional TCs per Recommendation 1. Moreover, if there is support for Recommendation 2, the dashboard could be further developed to support the operational "incident" TC risk communication efforts.

Acknowledgements

This research is funded by NOAA OAR grants NA19OAR0220119 and NA20OAR4590469, and it is supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement 1852977.

We want to express our tremendous gratitude to the team at YouGov who worked so closely, flexibly, and consciously with us to design and implement this survey: Sam Luks, Marissa Shih, Caitlin Collins, and Sara Sermarini. We truly could not have done this research without their expertise and assistance. We also want to express our tremendous gratitude to the team of folks at NOAA who gave us the opportunity to do this research to develop this methodology, given the many challenges and uncertainties surrounding it. Furthermore, the NOAA team provided helpful input about our research design, and they were invaluable in providing expertise and ideas to support our decision-making to field for TCs Laura and Marco in 2020 and TC Henri in 2021. In alphabetical order, thank you Robbie Berg, Mike Brennan, Gina Eosco, Frank Marks, Shirley Murillo, Micki Olsen, Jessica Schauer, Jen Sprague-Hilderbrand, Valerie Were, and Castle Williamsberg. We also thank Jenn Boehnert at NCAR who conducted the GIS analysis to identify the zip codes within different distances of the coast for all of the Gulf and Atlantic states, which was the basis for our survey samples. We thank Emily Laidlaw for her careful proofread and review of the report. And, we thank Mark Klein at NWS WPC who provided us with 2020 ERO data. Last but not least, we thank the thousands of people who took the time to respond to our surveys while at risk of TCs Laura, Marco, and/or Henri. The longitudinal data they provided is unparalleled in the interdisciplinary meteorological and social science community, and the data help us better understand how people think about and respond to real-world TC risks as they evolve in order to provide improved hurricane risk communication to end users.

References

Anderson, J., Kogan, M., Bica, M., Palen, L., Anderson, K., Stowe, K., Morss, R., Demuth, J.L., Lazrus, H., Wilhelmi, O., and Henderson, J., 2016: Far far away in Far Rockaway: Responses to risks and impacts during Hurricane Sandy through first-person social media narratives. *Proceedings of the 13th International Conference on Information Systems for Crisis Response and Management*, Rio de Janeiro, Brazil.

Baker, E.J., 1991: Hurricane evacuation behavior. *International Journal of Mass Emergencies and Disasters*, 9, 287-310

Bandura, A., 1977: Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215, https://doi.org/10.1037/0033-295X.84.2.191

Bostrom, A., Morss, R.E., Lazo, J.K., Demuth, J.L., and Lazrus, H., 2016: A mental models study of hurricane forecast and warning production, communication, and decision making. *Weather, Climate, and Societ*y, 8, 111-129.

Bostrom, A., Morss, R.E., Lazo, J.K., Demuth, J.L., and Lazrus, H., 2018: Eyeing the storm: How residents of coastal Florida see hurricane forecasts and warnings. *International Journal of Disaster Risk Reduction*, 30A, 105-119.

Cialdini, R.B., and Trost, M.R., 1998: Social influence: Social norms, conformity and compliance. In D.T. Gilbert, S.T. Fiske, and G. Lindzey (Eds.), *The handbook of social psychology* (pp. 151-192). McGraw-Hill. https://psycnet.apa.org/record/1998-07091-021

CONVERGE, 2023: Training modules: Collecting and sharing perishable data. https://converge.colorado.edu/resources/training-modules/

Cuite, C.L., Shwom, R.L., Hallman, W.K., Morss, R.E., and Demuth, J.L. 2017: Improving coastal storm evacuation messages. *Weather, Climate, and Society*, 9, 155-170.

Dash, N., and Gladwin, H., 2007: Evacuation decision making and behavioral responses: Individual and household. *Natural Hazards Review*, 8, 69-77.

Demuth, J.L., 2018: Explicating experience: Development of a valid scale of past hazard experience for tornadoes. *Risk Analysis*, 38, 1921-1943.

Demuth, J.L., Morss, R.E., Morrow, B.H., and Lazo, J.K., 2012: Creation and communication of hurricane risk information. *Bulletin of the American Meteorological Society*, 93, 1133-1145.

Demuth, J.L., Morss, R.E., Palen, L., Anderson, K., Anderson, J., Kogan, M., Stowe, K., Bica, M., Lazrus, H., Wilhelmi, O., and Henderson, J. 2018: "sometimes da #beachlife ain't always da wave": Understanding people's evolving hurricane risk communication, risk assessments, and responses using Twitter narratives. *Weather, Climate, and Society*, 10, 537-560.

Demuth, J.L., Morss, R.E., Wong-Parodi, G., Schumacher, A., Walpole, H., Herbert, N., 2023: Longitudinal studies of risk perceptions and behavioral responses for natural hazards. In B. Fisher-Liu and A. Mehta (Eds), *Handbook of Risk, Crisis, and Disaster Communication.*

Gladwin, C.H., Gladwin, H., and Peacock, W.G., 2001: Modeling hurricane evacuation decisions with ethnographic methods. *International Journal of Mass Emergencies and Disasters,* 19**,** 117-143.

Griffin, R.J., Neuwirth, K., Dunwoody, S., and Giese, J.K., 2004: Information insufficiency and risk communication. *Media Psychology,* 6, 23-61.

Harlan, S.L., Sarango, M.J., Mack, E.A., and Stephens, T.A., 2019: A survey-based assessment of perceived flood risk in urban areas of the United States. *Anthropocene*, 28, 100217.

Huang, S.-K., Lindell, M.K., and Prater, C.S., 2016: Who leaves and who stays? A review and statistical meta-analysis of hurricane evacuation studies, *Environment and Behavior*, 48, 991–1029.

IEM (Iowa Environmental Mesonet), 2022a: Archived NWS Watch, Warnings, Advisories. https://mesonet.agron.iastate.edu/request/gis/watchwarn.phtml.

IEM, 2022b: Download SPC Convective Outlooks. https://mesonet.agron.iastate.edu/request/gis/spc_outlooks.phtml.

Johnson, B.B., and Swedlow, B.M., 2019: Cultural theory's contributions to risk analysis: A thematic review with directions and resources for further research. *Risk Analysis*, 41, 429-455.

Kruger, J., et al., 2020: Hurricane Evacuation Laws in Eight Southern U.S. Coastal States — December 2018. *CDC Morbidity and Mortality Weekly Report*, **69**(36), 1233-1237, http://dx.doi.org/10.15585/mmwr.mm6936a1.

Lazo, J.K., Bostrom, A., Morss, R.E., Demuth, J.L., and Lazrus, H., 2015: Factors affecting hurricane evacuation intentions. *Risk Analysis,* 35, 1837-1857.

Lazrus, H., Morrow, B.H., Morss, R.E., and Lazo, J.K., 2012: Vulnerability beyond stereotypes: Context and agency in hurricane risk communication. *Weather, Climate, and Society,* 4, 103-109.

Lindell, M.K., 2012: Response to environmental disasters. *The Oxford Handbook of Environmental and Conservation Psychology,* S. D. Clayton, Ed., Oxford University Press, 391-412.

Meyer, R., Broad, K., Orlove, B., and Petrovic, N. 2013. Dynamic simulation as an approach to understanding hurricane risk response: Insights from Stormview lab. *Risk Analysis*, 33, 1532-1552.

Meyer, R.J., Baker J., Broad K., Czajkowski J., and Orlove, B., 2014: The dynamics of hurricane risk perception: Real-time evidence from the 2012 Atlantic hurricane season. *Bulletin of the American Meteorological Society*, 95, 1389-1404.

Morss, R.E., and Hayden M.H., 2010: Storm surge and "certain death": Interviews with Texas coastal residents following Hurricane Ike. *Weather, Climate, and Society,* 2, 174-189.

Morss, R.E., Cuite C.L., Demuth, J.L., Hallman, W.K., and Shwom R.L., 2018: Is storm surge scary? The influence of hazard, impact, and fear-based messages and individual differences on responses to hurricane risks. *International Journal of Disaster Risk Reduction,* 30A, 44-58.

Morss, R.E., Demuth J.L., Lazo, J.K., Dickinson, K., Lazrus, H., and Morrow, B.H., 2016: Understanding public hurricane evacuation decisions and responses to forecast and warning messages. *Weather and Forecasting*, 31, 395-417.
Morss, R.E., et al., 2017: Hazardous weather prediction and communication in the modern information environment. *Bulletin of the American Meteorological Society*, 98, 2653-2674.
Morss, R.E., Lazrus, H., Bostrom, A., and Demuth, J.L., 2020: The influence of cultural worldviews on people's responses to hurricane risks and threat information. *Journal of Risk Research*, 23, 1620-1649.
NHC (National Hurricane Center), 2022a: NHC Data in GIS Format. https://www.nhc.noaa.gov/gis/.
NHC, 2022b: NHC Tropical Cyclone Graphical Product Descriptions. https://www.nhc.noaa.gov/aboutnhcgraphics.shtml
NHC, 2022c: Glossary of NHC terms. https://www.nhc.noaa.gov/aboutgloss.shtml
NHC, 2022d: Hurricane and Tropical Storm Watch/Warning Breakpoints. https://www.nhc.noaa.gov/breakpoints
NOAA SAB (National Oceanic and Atmospheric Administration Science Advisory Board), 2021: *A Report on Priorities for Weather Research.* https://sab.noaa.gov/wp-content/uploads/2021/12/PWR-Report_Final_12-9-21.pdf
NWS (National Weather Service), 2022a: Flood Warning vs. Watch. https://www.weather.gov/safety/flood-watch-warning
NWS, 2022b: SPC Products. https://www.spc.noaa.gov/misc/about.html
NWS, 2022c: Understand Tornado Alerts. https://www.weather.gov/safety/tornado-ww
Rogers, RW., 1983: Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. Social Psychophysiology: A Sourcebook, J. Cacioppo and R. Petty, Eds., Guilford Press, 153-176.
Schwartz, L.M., Woloshin, S., Black, W.C., and Welch, H.G., 1997: The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, 127, 966-972.
Sjoberg, L., 2000: Factors in risk perception. *Risk Analysis*, 20, 1, 1-11.
Slovic, P., Finucane, M.L., Peters E., and MacGregor, D.G., 2004: Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, 24, 311-322.
Storm Prediction Center, 2022: SPC Products. https://www.spc.noaa.gov/misc/about.html
Tanim, S.H., Wiernik, B.M., Reader, S., and Hu, Y., 2022: Predictors of hurricane evacuation decision: A meta-analysis. *Journal of Environmental Psychology*, 79, 101742.
Taylor, K., Priest S., Fussell Sisco H., Banning S., and Campbell, K., 2009: Reading Hurricane Katrina: Information sources and decision-making in response to a natural disaster. *Social Epistemology*, 23, 361-380.
Trumbo, C., Peek, L., Meyer, M., Marlatt, H., Gruntfest, E., McNoldy, B., and Schubert, W., 2016: A cognitive-affect scale for hurricane risk perception. *Risk Analysis*, 36, 2233-2246.
Walpole, H.D. and Wilson, R.S., 2020: Extending a broadly applicable measure of risk perception: The case for susceptibility. *Journal of Risk Research*, 24, 135-147.
Wong-Parodi, G., and Feygina, I., 2018: Factors influencing (mal) adaptive responses to natural disasters: The case of Hurricane Matthew. *Weather, Climate, and Society*, 10, 747–768.

Wong-Parodi, G., and Garfin, D.R., 2022: Hurricane adaptation behaviors in Texas and Florida: exploring the roles of negative personal experience and subjective attribution to climate change. *Environmental Research Letters*, 17(3), 034033.

Wong-Parodi, G., and Rubin, N.B., 2022: Exploring how climate change subjective attribution, personal experience with extremes, concern, and subjective knowledge relate to pro-environmental attitudes and behavioral intentions in the United States. *Journal of Environmental Psychology*, 79, 101728.

Wong-Parodi, G., Fischhoff, B., and Strauss, B., 2017: Plans and prospects for coastal flooding in four communities affected by Sandy. *Weather, Climate, and Society*, 9(2), 183-200.

Wong-Parodi, G., Fischhoff, B., and Strauss, B., 2018: Effect of risk and protective decision aids on flood preparation in vulnerable communities. *Weather, Climate, and Society*, 10, 401-417.

Wong-Parodi, G., Krishnamurti, T., Gluck, J., and Agarwal, Y., 2019: Encouraging energy conservation at work: A field study testing social norm feedback and awareness of monitoring. *Energy Policy*, 130. https://doi.org/10.1016/j.enpol.2019.03.028

Wu, H-C., Lindell, M.K., and Prater, C.S., 2015a: Process tracing analysis of hurricane information displays. *Risk Analysis,* 35, 2202-2220.

Wu, H-C., Lindell, M.K., and Prater, C.S., 2015b: Strike probability judgments and protective action recommendations in a dynamic hurricane tracking task. *Natural Hazards,* 79, 355-380.

Appendix A. NCAR Longitudinal Hurricane Survey with YouGov: Details of the Survey Parameters, Sample, Programming & Fielding, Costs & Incentive

Survey goal: To collect longitudinal survey data—specifically, a panel study with multiple responses from the same sample of respondents. Data will be collected over just a few days in the context of a real-life hurricane that may (or does) make landfall in the mainland United States.

Survey parameters

- Want to have the survey ready no later than June 1 in order to field this summer when there's a hurricane threat. (This includes initial programming by YouGov, pilot testing by NCAR, and revisions as requested by NCAR based on pilot testing, as noted below.)
    - Want to field by end of September 2020, given the election
- Survey will be web-based, mobile-friendly, and use YouGov's existing panel
- Will field 2 to 3 survey waves during the predictive phase, when the hurricane is approaching
    - Each wave will be fielded for a period of no more than 24 hours.
    - Will have 24-hour intervals between waves
    - Each of the predictive survey waves will ask the same set of questions except for Wave 1, which will have a few extra questions at the end.
    - We are targeting the median survey response time to be 15 min for Wave 1 and 10 min for the rest of the predictive survey waves.
- Will field 1 survey wave, the post-storm wave, approximately 7–14 days after the hurricane
    - This wave can be fielded over a few days.
    - The post-storm wave will have its own set of questions, different from the predictive survey waves.
    - We are targeting the median survey response time for the post-storm wave to be 15 min.
- We want balanced waves, with all respondents responding to all waves for our final sample of at least 700 respondents.
- We also want data for incompletes and for people who drop out across waves, so we can analyze whether there are patterns among these non-respondents.

Survey sample

- Aiming for a final sample of at least 700 respondents
- To get the same person responding to all waves, YouGov will screen for birth year + gender and then eliminate them if these aren't the same (or within a year or two for birth year).
- Sample will be English-speaking, have Internet access, live in areas of the mainland United States at risk from hurricanes, and be part of YouGov's existing panel
- YouGov can provide weighted data if we want—typically weight is based on age, gender, race, education, and possibly income.
- We will survey a broad swath of coastal areas, with areas chosen at the state level except for possibly Florida, which we may divide into the eastern vs. western side of the peninsula.

- YouGov's feasibility numbers—which are the estimated number of people who would respond to the survey—for 50, 75, and 100 miles inland from the coast are provided in Table 1.
- When we field the survey, we will let YouGov know:
    - Which states
    - Whether to pull the sample for 50, 75, or 100 miles inland.
- Because we are using YouGov's existing panel, YouGov will provide the core profile variables.
- YouGov also will provide respondents' locations by zip code and as lat/lon randomly shifted by up to several hundred meters in order to preserve people's anonymity.
- To QC people speeding through the survey, YouGov removes the top 2% of people by time to complete by age group.

Survey programming and fielding

- YouGov will program the predictive and post-storm survey waves.
- YouGov will provide NCAR with a link, and NCAR will pilot test the survey.
- NCAR will provide YouGov with needed revisions to the survey questions or functionality, which YouGov will implement.
- At a minimum, YouGov will need 1 day of lead time after our "go" decision to field the survey in order to upload the final set of zip codes for fielding and program the hurricane name into the repeated text field.
- As possible, YouGov would prefer a longer lead time before our "go" decision, i.e., 2 or 3 days. NCAR will give as much lead time as possible to YouGov, for instance, letting YouGov know that we're considering a certain storm and accordingly eyeing certain states.
- It's not a problem to field on the weekend, as YouGov can work on weekends. We just need to give them a heads-up. Importantly, NCAR must notify everyone on the team (Sam, Marissa, Caitlin).
- Communicating about the longitudinal survey and incentives
    - In order to get people to "click" into the survey,
        - YouGov will not convey at Wave 1 of the survey that it is a multi-wave survey.
        - NCAR also will not convey in its introductory language that the survey is multi-wave.
    - At the beginning of subsequent survey waves, YouGov will communicate that this is a next survey about the hurricane so that people don't think they've already responded.
    - If needed to incentive people, at the beginning of the 3rd wave, YouGov will communicate the final incentive bonus.

Survey costs and incentives

- YouGov will adjust the incentives for later waves (Wave 3 and Wave 4) along the way based on the re-contact rate. Two scenarios are provided in the tables below: one with a good re-contact rate and a high incentive at Wave 4 (Table 2), and one with a high re-contact rate and lower incentive at Wave 4 (Table 3). Both scenarios have a Wave 1 sample size of N=1277.

Table A1. YouGov state feasibility numbers at 50, 75, and 100 miles inland from the coast, by state.

State	Panel feasibility for zip codes within 50 miles of coast	Panel feasibility for zip codes within 75 miles of coast	Panel feasibility for zip codes within 100 miles of coast
AL	130	164	194
CT	544	544	545
DE	194	194	194
DC	132	132	133
FL	4072	4085	4087
GA	152	203	311
LA	411	438	455
ME	281	284	293
MD	825	844	862
MA	883	1007	1023
MS	93	122	140
NH	285	346	349
NJ	1430	1431	1431
NY	1872	1925	2022
NC	292	405	687
PA	1062	1413	1627
RI	168	168	168
SC	308	382	538
TX	1084	1165	1390
VA	1128	1203	1267
WV	10	44	59

Table A2. Scenario with a conservative attrition rate and higher incentive for Wave 4.

Stage	**N**	**Length**	**Rate**	**Incentive**	**Total**	**Attrition rate**
Wave 1	1277	15 min	$13.50	$2.20	$20,049.00	
Wave 2	958	10 min	$10.00	$2.20	$11,688.00	25.0%
Wave 3	795	10 min	$10.00	$2.20	$9,699.00	17.0%
Wave 4	700	15 min	$12.00	$11.00	$16,100.00	11.9%
Setup, programming, hosting, preparation of deliverables	1		$7,200.00		$7,200.00	
Grand total					$64,736.00	

Table A3. Scenario with an optimistic, low attrition rate and lower incentive for Wave 4.

Stage	**N**	**Length**	**Rate**	**Incentive**	**Total**	**Attrition rate**
Wave 1	1277	15 min	$13.50	$2.20	$20,049	
Wave 2	1149	10 min	$10.00	$2.20	$14,018	10.0%
Wave 3	1034	10 min	$10.00	$2.20	$12,615	10.0%
Wave 4	931	15 min	$12.00	$2.20	$13,220	10.0%
Setup, programming, hosting, preparation of deliverables	1		$7,200		$7,200	
Grand total					$67,102	

Appendix B. Predictive and Post-storm Surveys for 2020 Hurricanes Laura and Marco

The published predictive and post-storm survey instruments can be accessed here:

Demuth, J., R. Morss, G. Wong-Parodi. (2023) “2020 Hurricanes Laura and Marco predictive and post-storm survey instruments”, in *Public longitudinal panel surveys collected during and after hazardous weather threats: Hurricanes*. DesignSafe-CI. https://doi.org/10.17603/ds2-j9hc-xy24

Appendix C. Predictive and Post-storm Surveys for 2021 Hurricane Henri

The published predictive and post-storm survey instruments can be accessed here:

Demuth, J., R. Morss, G. Wong-Parodi, A. Schumacher, N. Herbert, H. Walpole. (2023) “2021 Hurricane Henri predictive and post-storm survey instruments”, in *Public longitudinal panel surveys collected during and after hazardous weather threats: Hurricanes.* DesignSafe-CI. https://doi.org/10.17603/ds2-05s8-ah21