

Exploring Community Differences in Tornado Warning Reception, Comprehension, and Response across the United States

Joseph T. Ripberger, Carol L. Silva, Hank C. Jenkins-Smith, Jinan Allan, Makenzie Krocak, Wesley Wehde, and Sean Ernst

ABSTRACT: Effective risk communication in the weather enterprise requires deep knowledge about the communities that enterprise members serve. This includes knowledge of the atmospheric and climate conditions in these communities as well as knowledge about the characteristics of the people living in these communities. Enterprise members often have access to data that facilitate the first type of knowledge, but relatively little social or behavioral data on the populations they serve. This article introduces an effort to overcome these challenges by developing a database of community statistics and an interactive platform that provides dynamic access to the database. Specific emphasis is given to one set of statistics in the community database: estimates of tornado warning reception, comprehension, and response by county warning area in the contiguous United States. Exploration of these estimates indicates significant variation in reception and comprehension across communities. This variation broadly aligns with tornado climatology, but there are noticeable differences within climatologically comparable regions that underline the importance of community-specific information. Verification of the estimates using independent observations from a random sample of communities confirms that the estimates are largely accurate, but there are a few consistent anomalies that prompt questions about why some communities exhibit higher or lower levels of reception, comprehension, and response than models suggest. The article concludes with a discussion of next steps and an invitation to use and contribute to the project as it progresses.

<https://doi.org/10.1175/BAMS-D-19-0064.1>

Corresponding author: Joseph T. Ripberger, jtr@ou.edu

Supplemental material: <https://doi.org/10.1175/BAMS-D-19-0064.2>

In final form 25 March 2020

©2020 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

AFFILIATIONS: Ripberger, Silva, Jenkins-Smith, Allan, and Ernst—Center for Risk and Crisis Management, University of Oklahoma, and National Institute for Risk and Resilience, Norman, Oklahoma; **Krocak**—Center for Risk and Crisis Management, University of Oklahoma, and National Institute for Risk and Resilience, and Cooperative Institute for Mesoscale Meteorological Studies, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma; **Wehde**—East Tennessee State University, Johnson City, Tennessee

Members of the weather enterprise, including National Weather Service (NWS) forecasters, emergency managers, broadcast meteorologists, and private partners, have many roles and responsibilities. These roles and responsibilities range from issuing forecasts and warnings during high impact weather events to outreach and public education campaigns during less turbulent periods. Effective education and risk communication across this range requires deep knowledge of the communities that enterprise members serve. This includes knowledge about atmospheric and climate conditions in communities as well as knowledge about the people in these communities. Enterprise members often have access to a wide variety of data that facilitate the first type of knowledge, but relatively little data on the populations they serve. As a result, it can be difficult to answer basic questions: 1) What risks do the people in a community worry about or neglect? 2) Do they generally receive, understand, and respond to forecasts and warnings? 3) What sources of information do they rely on and trust? Absent reliable answers to these questions, it is challenging to develop public education and risk communication strategies that fit the specific needs of diverse communities. Perhaps more importantly, it is challenging to identify best practices between communities and/or track changes within communities as enterprise members experiment with new education and communication strategies.

Recognizing these challenges, multiple scientific reports note the urgent priority of developing data collection capacities in the social and behavioral sciences (e.g., National Research Council 2010, 2012; National Academies of Sciences, Engineering, and Medicine 2018). Most recently, for example, the U.S. National Academies of Sciences, Engineering, and Medicine (2018) report on “Integrating Social and Behavioral Sciences within the Weather Enterprise” states the following:

Advancing the social and behavioral sciences requires the regular collection and sharing of high-quality data, including ongoing observations that may need to be sustained over periods of months, years, or even longer. This data collection serves many purposes, for instance, to better understand how key factors within a given population or organization vary over time, locations, and across different groups; to help detect gradual trends or abrupt changes in those factors over time or in response to particular events; and to explore possible correlations and causal relationships with other observed variables of interest.

In this article, we introduce an effort to develop a database of community statistics and an interactive platform that provides dynamic access to some of this information. The approach we use leverages population data from the Severe Weather and Society Survey and known subpopulation characteristics from the U.S. Census to estimate statistics of interest for communities across the country. We demonstrate the 1) approach, 2) database, and 3) platform by using them to explore differences in tornado warning, reception, comprehension, and response in counties and county warning areas (CWAs) across the United States. We conclude with a short discussion of future research and development.

Approach: Downscaling population surveys

While more data are certainly necessary, nationally representative surveys that target the U.S. adult population are increasingly common (e.g., Eastern Research Group 2018). The Severe

Weather and Society Survey (Wx Survey) represents one such effort. Run by the University of Oklahoma Center for Risk and Crisis Management, the Wx Survey is a yearly survey of the U.S. public that includes two types of questions: 1) baseline questions that measure core concepts such as risk perceptions; forecast and warning reception, comprehension, and response; knowledge about hazards; and trust in information sources; and 2) one-time questions and experiments that address various topics, such as the impact of uncertainty and probabilistic information on risk judgments and protective action decision making (see Silva et al. 2017, 2018, 2019). Large population surveys such as this provide valuable information about the population as a whole, but rarely address differences across geographic subpopulations.

Somewhat analogous to climate downscaling, where scientists use global climate models to produce local-scale weather and climate predictions, survey researchers are actively developing small area estimation (SAE) techniques that downscale data from large population surveys to subpopulations, such as states, counties, and districts. Currently, there are two primary SAE techniques: disaggregation and multilevel regression and poststratification (MRP). When applying disaggregation, researchers compile as many comparable datasets as possible, and then use responses from survey participants who live in the same geographic area (e.g., county) to calculate a given statistic within that area. While intuitive, disaggregation is data intensive—it requires sufficient sample size in each geographic unit to produce reliable estimates. Most large population surveys do not collect enough observations in each geographic area to produce these estimates; this is especially true in low population areas. In addition, disaggregation techniques typically neglect “nesting” patterns within and across surveys that can bias estimates across geographic areas. For example, most disaggregation techniques ignore the possibility that surveys by different groups using different questions and data collection methodologies are likely to generate different errors that researchers must account for when using the data to make inferences. Failure to do so can result in incorrect estimates and overstatements of confidence in those estimates. As a prime example, some statisticians argue that failure to account for nesting was one of the primary mistakes that some opinion analysts made when (incorrectly) forecasting the results of the 2016 U.S. presidential election (Shirani-Mehr et al. 2018). MRP is less data intensive than disaggregation and it allows researchers to account for nesting. It uses regression analysis to identify demographic and geographic patterns in areas where data are available to produce estimates in areas where data are relatively sparse (Park et al. 2004).

While validation is always necessary, there is an emerging consensus among survey researchers that MRP is a viable alternative to disaggregation when demographic and geographic patterns are evident in the data (Lax and Phillips 2009; Buttice and Highton 2013). As such, researchers from many different fields and agencies are using this technique to estimate a wide variety of community statistics. For example, scientists at the U.S. CDC are using MRP to estimate the prevalence of public health outcomes in census blocks, tracts, districts, and counties across the country (Zhang et al. 2014, 2015; Wang et al. 2018); researchers at Pew Research Center are using it to identify news media consumption habits in U.S. cities (Pew Research Center 2019);¹ and opinion analysts are using it to forecast election outcomes in U.S. states (Wang et al. 2015; Kiewiet de Jonge et al. 2018). MRP is also gaining traction among researchers who study climate change and extreme weather events. In these fields, researchers are using MRP to estimate the geographic distribution of climate change opinions (Howe et al. 2015;² Mildenerger et al. 2017; Bergquist and Warshaw 2019), climate change messaging effects (Warshaw 2018), household disaster preparedness (Howe 2018), and extreme heat risk perceptions (Howe et al. 2019).³

¹ Visit www.journalism.org/interactives/local-news-habits/ to interact with and learn more about these estimates.

² Visit <https://climatecommunication.yale.edu/visualizations-data/ycom-us-2018/> to interact with and learn more about these estimates.

³ Visit <https://climatecommunication.yale.edu/visualizations-data/heatwave-risk-perceptions/> to interact with and learn more about these estimates.

In this ongoing project, we employ data from the Wx Survey in combination with MRP to create a database of community statistics that members of the weather enterprise can use to increase knowledge about the populations they serve. The project is ongoing because data collection for the Wx Survey is ongoing; each wave provides new data that we are using to track baseline measures and develop indicators of new concepts. Here we demonstrate the project approach, database of community statistics, and interactive platform by using them to explore differences in tornado warning reception, comprehension, and response in CWAs across the United States.

Before we begin, we note a few of the principles and practices that guide the project. The first principles are transparency and reproducibility. We publish an open access report that presents an overview of the methodology and survey instrument for each wave of the Wx Survey (Silva et al. 2017, 2018, 2019). In addition, the survey data and code necessary to reproduce the estimates we develop for the database are available in a public repositories.⁴ The next principle is measurement. While some concepts can be measured with a few relatively simple survey questions, others, including many concepts of interest to the weather enterprise, are more complex and therefore require more attention to detail in the construction of the measures. We take care to do this by explicitly assessing and documenting the reliability of the measures we include in the database using a combination of psychometric techniques such as factor analysis and item response theory [for a detailed explanation of this process, see Ripberger et al. (2019)]. The last and perhaps most important principle is consistency. The value of the database stems from an ability to track similarities and differences between communities and within communities over time. This will require that we continue to collect and analyze these data into the future.

Database: Tornado warning reception, comprehension, and response

The most recent waves of the Wx Survey (Wx18 and Wx19) include multiple questions that measure tornado warning reception, comprehension, and response. The questions (examples shown in Table 1) prompt subjective assessments of warning reception, comprehension, and response, and provide an objective test of tornado warning comprehension (see online supplemental materials for a complete list of questions; <https://doi.org/10.1175/BAMS-D-19-0064.2>). We use these questions because a concurrent study (Ripberger et al. 2019) indicates that they provide statistically reliable scales that adequately discriminate between people with low, average, and high reception, comprehension, and response tendencies. The scales measure these concepts with scores from item response

⁴ Survey data are available at <https://dataverse.harvard.edu/dataverse/wxsurvey>. Codes for data analysis are available at <https://github.com/loucr/cm/wxsurvey>.

Table 1. Example survey questions we use to measure tornado warning reception, subjective comprehension, objective comprehension, and responses.

Scale	Example question(s)
Reception	Sometimes people miss tornado WARNINGS because they are doing something that makes it difficult to pay attention to the weather. For example, people often miss tornado warnings when they are sleeping. How confident are you that you would receive tornado warnings in the following situations? If you are...sleeping, in a car, at work, etc.
Subjective comprehension	In general, do you understand the difference between watches and warnings? How would you rate your understanding of tornado watches and warnings?
Objective comprehension	If the National Weather Service issues a tornado warning for your area, how much time do you have before the tornado arrives? If the National Weather Service issues a tornado watch for your area, how much time do you have before the tornado arrives?
Response	For some people the time of day influences tornado warning reception, understanding, and/or responsiveness. If a tornado WARNING were issued for your area tomorrow at [RANDOM TIME], how confident are you that you would take protective action in response to the warning?

theory (IRT) models, which indicate each survey participant’s scale scores by comparing them to other participants. IRT is a common methodology that scientists in education and psychology use to 1) assess the quality of test and survey questions and 2) grade participant’s answers to questions. In this project, we use information from IRT models to assess the reliability of the questions we use to measure concepts like warning reception, comprehension, and response (Ripberger et al. 2019) and to estimate scale scores for each survey respondent on each of these measures. Typically, IRT scores are given as z scores that denote how many standard deviations above or below the mean a person is on a scale. Here, we convert the z scores to percentile scores to facilitate interpretation. The percentiles indicate where a given person scores relative to others.

We use MRP to estimate mean reception, subjective comprehension, objective comprehension, and response percentiles among people who live in counties and CWAs across the U.S. MRP involves three steps—multilevel regression, prediction, then poststratification. In the first step, we estimate the following models:

$$y_i = \beta^0 + \alpha_{j[i]}^{\text{gender}} + \alpha_{k[i]}^{\text{age}} + \alpha_{j[i],k[i]}^{\text{gender*age}} + \alpha_{l[i]}^{\text{race}} + \alpha_{m[i]}^{\text{ethnicity}} + \alpha_{s[i]}^{\text{area}},$$

where

$$\alpha_j^{\text{gender}} \sim N(0, \sigma_{\text{gender}}^2), \quad j = 1 \text{ or } 2,$$

$$\alpha_k^{\text{age}} \sim N(0, \sigma_{\text{age}}^2), \quad k = 1, 2, \text{ or } 3,$$

$$\alpha_{j,k}^{\text{gender*age}} \sim N(0, \sigma_{\text{gender*age}}^2), \quad j = 1 \text{ or } 2 \text{ and } k = 1, 2, \text{ or } 3,$$

$$\alpha_l^{\text{race}} \sim N(0, \sigma_{\text{race}}^2), \quad l = 1, 2, \text{ or } 3,$$

$$\alpha_m^{\text{ethnicity}} \sim N(0, \sigma_{\text{ethnicity}}^2), \quad m = 1 \text{ or } 2,$$

$$\alpha_s^{\text{area}} \sim N(\beta^{\text{climatology}} \times \text{climatology}_s, \sigma_{\text{area}}^2), \quad s = 1, \dots, 115.$$

The models have two levels. Individually, a participant’s percentile score on each scale varies as a function of the participant’s demographic profile (gender, age, a gender–age interaction, race, and ethnicity) and geographic area (CWA).⁵ CWA effects vary in relation to climatology (mean number of tornado event days per year).⁶ We use these variables in the models because a concurrent study indicates that they influence tornado warning reception, comprehension, and response (Ripberger et al. 2019).

The panels in Fig. 1 display the group estimates from these models; the rows in Table 2 display the estimated effects of tornado climatology. They also provide information about the factors that impact tornado warning reception, comprehension, and response. For example, Fig. 1 shows that men and women demonstrate roughly comparable levels of reception, objective comprehension, and response, but men have more confidence in subjective warning comprehension than women. More notably, the estimates indicate relatively significant variation across age and race groups, as well as variation across CWAs. While a complete discussion of each estimate falls outside the scope of this article [see Ripberger et al. (2019) for an extended discussion], it is important to note the amount of variation across CWAs by measure. The models indicate relatively large differences in subjective and objective comprehension, moderate differences in reception, and small differences in tornado warning response across CWAs. The coefficient estimates in Table 2 tell the same story—tornado climatology has a relatively strong effect on tornado warning reception and comprehension, but little effect on warning response. These findings suggest that geography, and the community differences that overlap with geographic boundaries, likely exert more direct influence on

⁵ The CWA is the nesting variable in this equation.

⁶ We use the NOAA Storm Events Database to calculate the mean number of tornado events per year in each CWA. Note that tornado reports in the Storm Events Database are in segments.

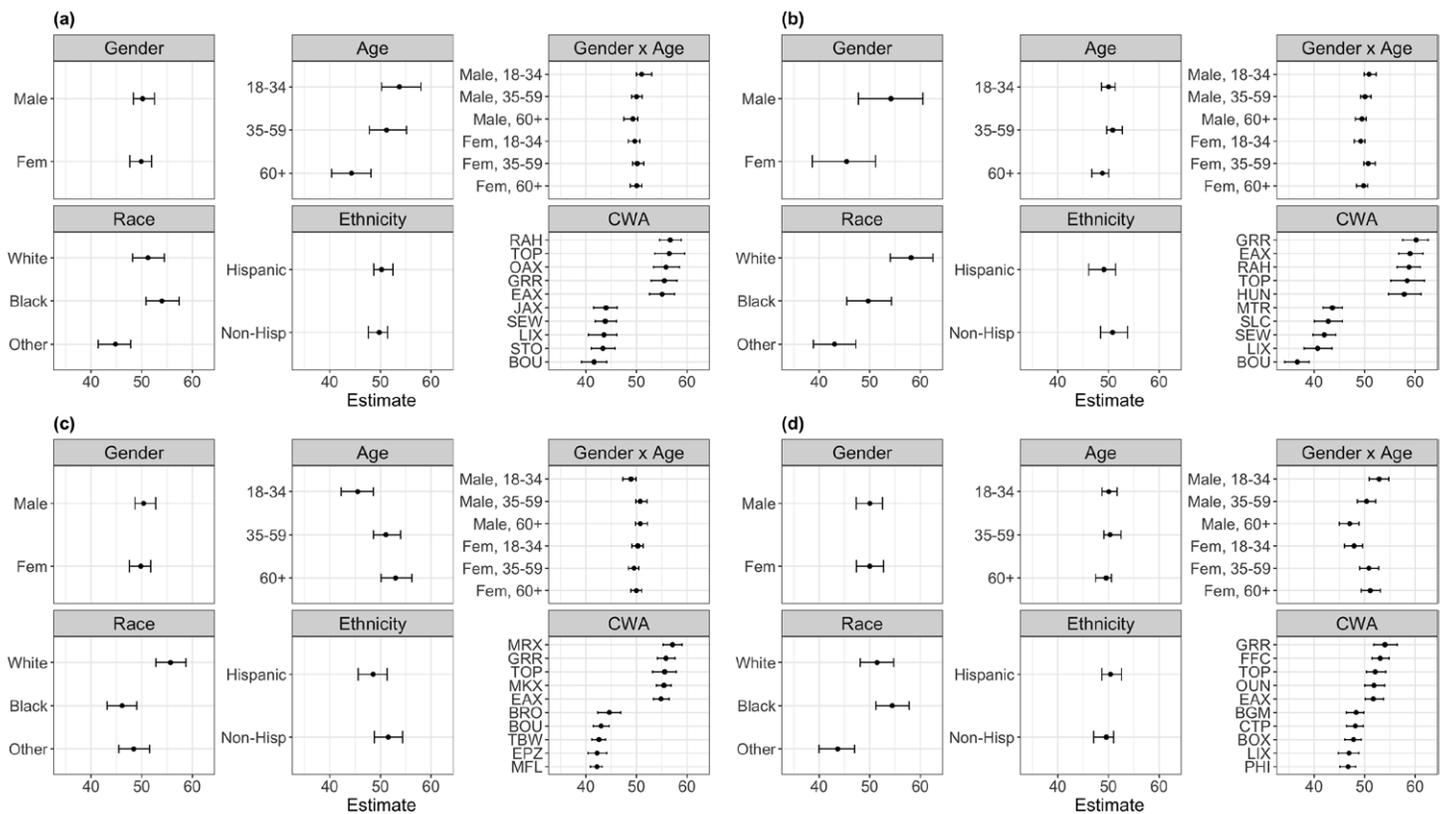


Fig. 1. Group estimates from regression models that predict tornado warning (a) reception, (b) comprehension (subjective), (c) comprehension (objective), and (d) response. The estimates compare group percentile scores; the points indicate median estimates from the models and the intervals indicate 50% uncertainty ranges. The CWA panels show the top and bottom five CWAs for each measure.

Table 2. Summary of the regression models we use to predict tornado warning reception, subjective comprehension, objective comprehension, and response. The models were estimated in a Bayesian framework using the stanarm package in R. The point and uncertainty estimates (in parentheses) were computed from simulations. The point estimates are median values and the uncertainty estimates are median absolute deviation (MAD) values that are analogous to standard errors in frequentist regression frameworks.

	Reception	Subjective comprehension	Objective comprehension	Response
Estimates				
Intercept	-0.06 (0.37)	-0.15 (0.49)	-0.10 (0.32)	-0.04 (0.37)
Tornado climatology	0.12 (0.02)	0.20 (0.02)	0.09 (0.02)	0.03 (0.02)
Error terms				
Gender	0.52	0.76	0.37	0.51
Age	0.46	0.24	0.29	0.25
Gender × age	0.13	0.10	0.10	0.16
Race	0.43	0.50	0.34	0.46
Ethnicity	0.53	0.53	0.42	0.54
Area (CWA)	0.13	0.15	0.11	0.07
<i>Residual</i>	0.95	0.90	0.56	0.97
Obs.	5,996	5,996	5,996	5,996

warning reception and comprehension than on response. Note, however, that these models do not account for the effect of warning reception and comprehension on response. As such, they do not allow for the possibility that geographic differences indirectly influence tornado warning response because they impact reception and comprehension. Because of this, we hesitate to conclude that geography has no impact on tornado warning response.

In addition to basic insight, the estimates shown in Fig. 1 and Table 2 illustrate the parameters we use in step two, the prediction phase of MRP. Here, we use the regression models (the parameters from Table 2) to predict reception, comprehension, and response scale percentiles across demographic groups in each CWA. For example, one demographic group is female, ages 18 to 34, white, non-Hispanic in the Norman, Oklahoma (OUN), CWA. The models predict percentile scores of 62, 63, 55, and 52 on the reception, subjective comprehension, objective comprehension, and response scales, respectively, for this group. These percentile scores indicate that, on average, the women in this group exhibit levels of reception, subjective comprehension, objective comprehension, and response that are greater than or equal to 62%, 53%, 55%, and 52% of people across the country. Because the models provide estimates for two gender groups (male and female), three age groups (18 to 34, 35 to 59, and 60+), three race groups (white, black, other race), and two ethnicity groups (non-Hispanic and Hispanic), we can use them to make 36 such predictions in each CWA across the country.

In step three, the poststratification phase of the MRP analysis, we weight the demographic group predictions in each CWA by population frequency, which we calculate using data from the U.S. Census.⁷ For example, 10.8% of adults in the OUN CWA match the demographic group we describe above—female, ages 18 to 34, white, non-Hispanic. This percentage provides the weight (multiplication term) we use when averaging predictions across demographic groups to calculate aggregate estimates of reception, subjective comprehension, objective comprehension, and response in OUN and other CWAs. More formally, we use the following formula to aggregate scale estimates for each CWA, where r is the demographic group, N is the population frequency, and θ is the prediction:

$$Y_{CWA}^{MRP} = \frac{\sum_{r \in CWA} N_r \theta_r}{\sum_{r \in CWA} N_r}.$$

In combination, these steps—multilevel regression, prediction, and poststratification—allow us to estimate an average person percentile (APP) score for each CWA in the contiguous United States (CONUS) on each measure. These estimates compare the average percentile of all adults who live in a CWA to the distribution of all adults across the country. For example, an APP estimate of 62 indicates that, on average, adults who live in that CWA are above the national average; they score higher than 62% of U.S. adults across the country.

The maps in Fig. 2 display APP estimates of tornado warning reception, subjective comprehension, objective comprehension, and response by CWA. The inset plots in Fig. 2 show the distribution of these estimates across the CWAs. In combination, the figures illustrate multiple findings. Most notably, they indicate significant and systematic (nonrandom) variation in reception and comprehension across the country. CWA APP scores range from 38 to 61 (a span of 23 percentiles) on the reception scale, 32 to 69 (37 percentiles) on the subjective comprehension scale, and 37 to 60 (23 percentiles) on the objective comprehension scale. Response scores, by comparison, exhibit less variation across CWAs; a minimum APP of 45 and maximum of 54 (only 9 percentiles). As we explain above, these findings suggest that warning reception and comprehension are more likely to vary across communities than warning response, but again, these

⁷ County resident population estimates by age, sex, race, and Hispanic origin are available at <https://www2.census.gov/programs-surveys/popest/datasets/2010-2018/counties/asrh/>.

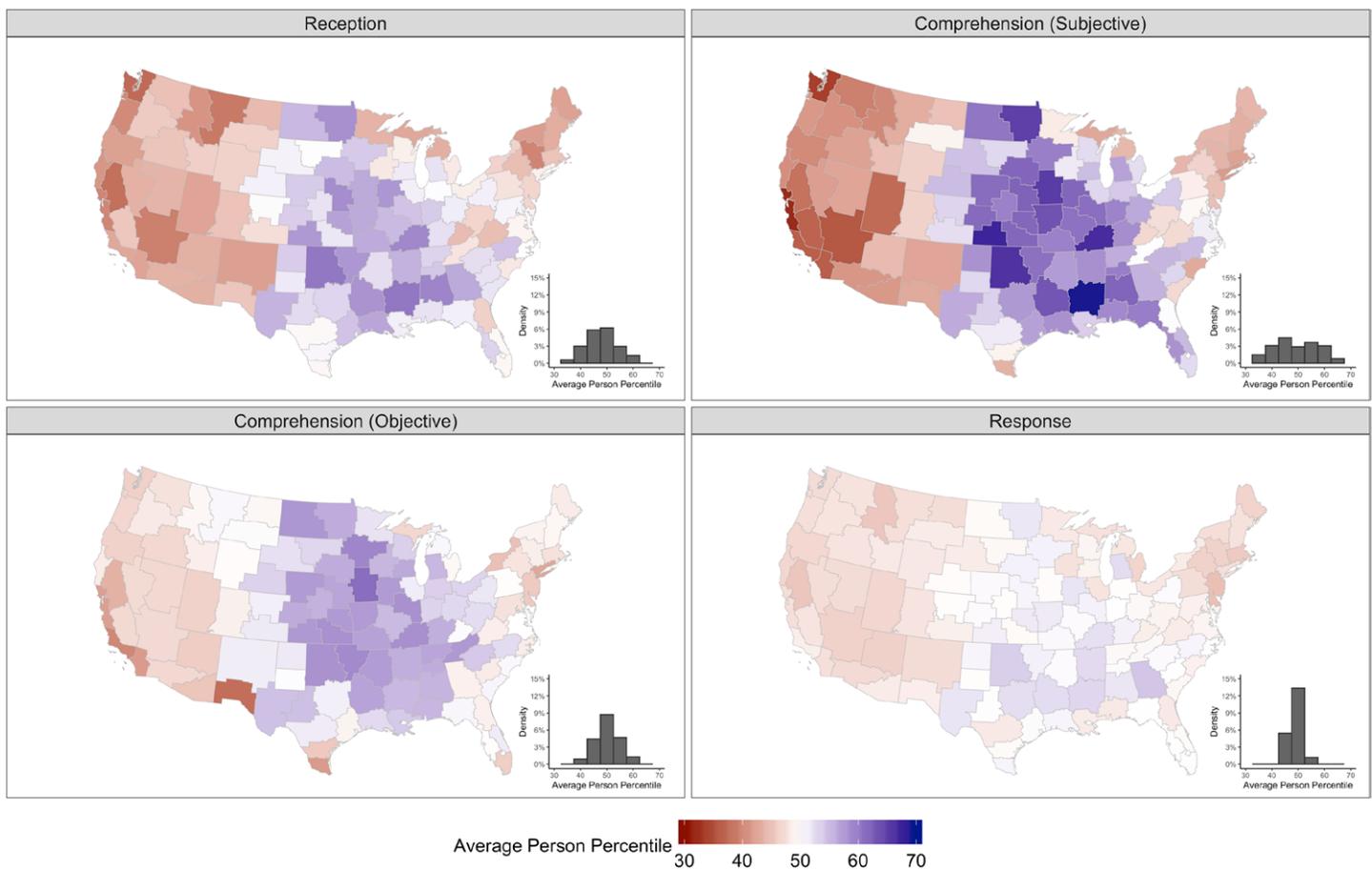


Fig. 2. Average person percentile (APP) estimates of tornado warning (top left) reception, (top right) subjective comprehension, (bottom left) objective comprehension, and (bottom right) response by CWA in the CONUS. The inset plots indicate the frequency distribution of APP estimates across CWAs.

estimates do not address the possibility that reception and comprehension likely impact response in ways that generate systematic differences across communities.

Despite differences in the amount of variation, the maps in Fig. 2a show relatively consistent geographic discrepancies across all the scales, including warning response. On average, the APP estimates indicate that reception, comprehension, and response are lowest in western CWAs, slightly below average in eastern CWAs, and above average in the central portion of the United States. Unsurprisingly, this pattern roughly mimics tornado climatology (e.g., Brooks et al. 2003; Krocak and Brooks 2018), implying that exposure and experience likely prompt adaptation in many communities. In communities that routinely experience tornadoes, people develop strategies, plans, and technologies that enhance confidence in warning reception and acquire the information necessary to interpret warnings when they get them. The same may be true of warning response, but the relationship is more subtle, likely because most people in most communities plan to take protective action if they receive a tornado warning—assuming they receive it and know what it means. On the whole, this adaptation is probably positive and unavoidable; people in communities that experience the most tornadoes are the most likely to receive warnings, know what they mean, and take protective action in response. Nevertheless, tornadoes are possible almost everywhere in the United States and people who live on the coasts can move—both temporarily and permanently—throughout the country. These factors prompt some concern about the low levels of reception and comprehension in some communities, especially those in the west.

In addition to patterns across regions, the maps in Fig. 2a show noteworthy differences within regions that are more difficult to explain with tornado climatology alone. In many

cases, adjacent communities that experience comparable threats exhibit different levels of tornado warning reception, comprehension, and (to some extent) response. For example, there is a 8-percentile-point difference in subjective comprehension estimates between the Norman CWA (APP = 66) and the Fort Worth, Texas, CWA (APP = 58), despite roughly comparable tornado climates. There is a roughly analogous 7-percentile-point difference between the Peachtree City, Georgia, CWA (APP = 49) and the Birmingham, Alabama, CWA (APP = 56) in objective comprehension. Differences like this create important opportunities for research and learning within the weather enterprise. What is regionally unique about the Norman and Birmingham areas that might generate relatively high levels of subjective and objective tornado warning comprehension relative to neighboring communities? Are the warning forecast offices, broadcast meteorologists, emergency managers, and private partners engaging in education and risk communication practices that are especially effective? Are the cultures in these communities especially attentive to and knowledgeable about severe storms? These estimates and the comparisons they facilitate will allow us to begin to address these important questions.

As with all forecast models, the estimates we produce in this project are subject to uncertainty. This necessitates constant verification. Much like forecast verification, we accomplish this by comparing predictions (forecasts) to observations. In this case, the predictions are the APP estimates we produce using MRP models and the observations that come from independent surveys of people in selected CWAs. If the estimates are accurate, they will be consistent with the observations. We began assessing this in 2018 by independently surveying a representative sample of 50 adults in a random sample of 30 CWAs. While a sample size of 50 people in each CWA is not sufficient to draw generalizable conclusions about the populations in each CWA, we assume that 50 observations provide basic information about the communities in the sample. Nevertheless, we protect against outlying observations (extreme values) by partially pooling the mean values we calculate across CWAs by assuming they come from a random distribution. This allows us to produce a mean estimate (observation) for each community that downweights the influence of outlying observations. We believe that this step is important given the relatively small sample size in each CWA, but the results we present below are largely consistent with analyses that do not use pooling.

Figure 3a plots a comparison between independent survey observations in 30 CWAs and the APP estimates we produce using MRP models. As the plots indicate, there is a relatively strong correlation between the independent survey observations and the estimates. This is

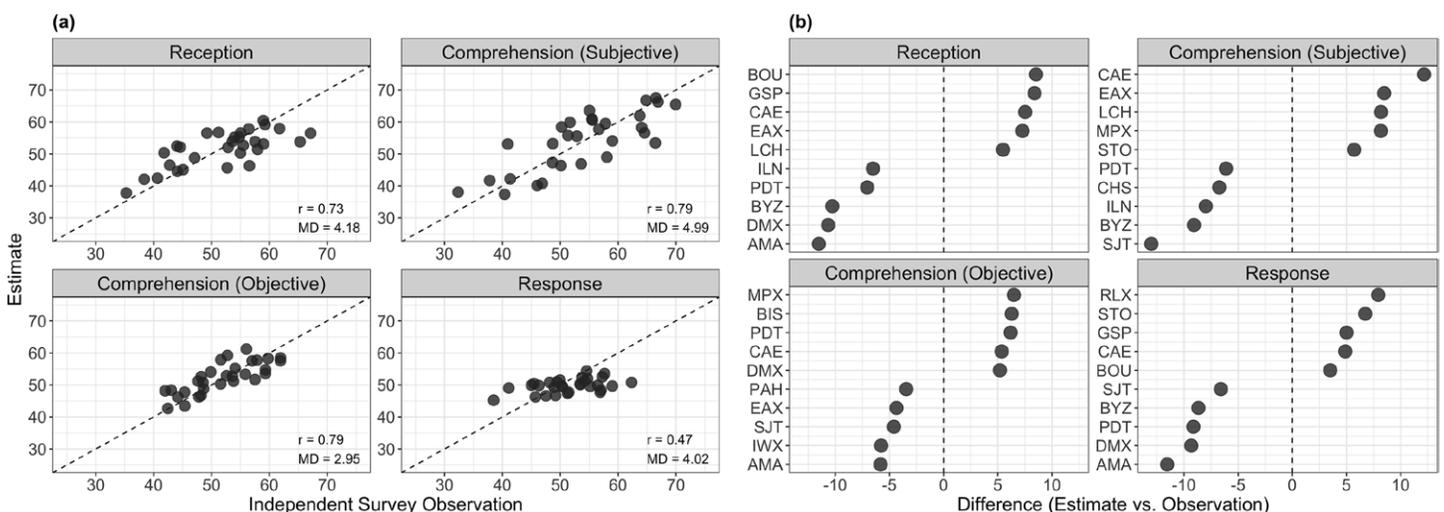


Fig. 3. (a) A comparison between independent survey observations in 30 CWAs and the APP estimates we produce using MRP models. **(b)** The top and bottom five overestimates and underestimates by measure.

especially true of the tornado warning reception, subjective comprehension, and objective comprehension estimates, where the correlation coefficients are 0.73, 0.79, and 0.79, respectively. The response coefficient drops (0.47), but there is still a positive relationship between the estimates and observations. These results suggest that the MRP models are generally able to differentiate between communities that are more and less likely to receive, correctly interpret, and take protective action in response to tornado warnings. In addition to discrimination, the models provide fairly accurate predictions. Relatively low mean absolute differences (MD) between the estimates and observations demonstrate this point. The models predict reception observations within an average of 4 percentiles, subjective comprehension 5 percentiles, objective comprehension 3 percentiles, and response within an average of 4 percentiles.

In forecast terminology, these results indicate that the community estimates we produce have skill, but they are not perfect. The estimates overshoot observations in some CWAs and undershoot them in others. The panels in Fig. 3b plot the top five overestimates and underestimates by measure. Positive values indicate instances where community (MRP) estimates suggest higher levels of reception, comprehension, and response than the observations; negative values indicate the opposite. Interestingly, this analysis reveals a few errors that are consistent across measures. For example, the estimates are consistently higher than observations in the Columbia, South Carolina, CWA (CAE) and consistently lower than observations in the Amarillo, Texas, CWA (AMA). We propose two possible explanations for these estimation errors: systematic bias in the models or anomalous communities. We cannot rule out the possibility of systematic bias, but we can say that there is nothing obvious about the errors we observe that might suggest a bias—they do not relate in systematic ways to sample size, demographic differences, or geographic factors like tornado climatology. We therefore lean toward believing that these are anomalous communities wherein people are either more or less likely to receive tornado warnings, know what they mean, and take protective action than models suggest. In other words, there is something unique about the people in these communities that distinguishes them from communities with comparable demographic and geographic profiles. Perhaps a significant proportion the people in the Columbia CWA are recent transplants who have yet to acquire the type of experiences that strengthen tornado warning reception, comprehension, and response? Or maybe recent experiences in combination with memorable historic events (such as the Amarillo tornado of 1949) stimulate especially high levels of reception and comprehension in communities such as those in the Amarillo CWA [for more on experience, see Demuth (2018)]? This study does not provide answers to these important questions, but we hope that the estimates in the database we are developing will encourage and allow more research on why some communities demonstrate abnormally high or low levels of warning reception, comprehension, and response.

Platform: The Severe Weather and Society Dashboard

As we note throughout the discussion above, we believe that this database of statistics will help members of the weather enterprise answer basic questions about the people in the communities they serve. We also believe that it will provide a resource for administrators and researchers who are working to identify differences and best practices between communities and/or monitor changes within communities as enterprise members experiment with new education and communication strategies. We can only achieve these goals if enterprise members, administrators, and researchers have an opportunity to use and interact with the database.

The Severe Weather and Society Dashboard (WxDash) is meant to provide this opportunity. WxDash (available at <https://crcm.shinyapps.io/wxdash>) is a continuously evolving interactive platform that allows users to explore the characteristics of communities across the country. For instance, it currently provides information on the tornado warning reception comprehension, and response measures we describe above. It also provides information on public trust

in weather information sources, perceptions about the efficacy of protective action, and vulnerability to beliefs about a variety of tornado myths (Klockow et al. 2014; Allan et al. 2017). In addition to this set of composite scales, WxDash provides information on risk perceptions across a variety of hazards [see Allan et al. (2019) for more information], data on tornado warning and extreme weather information sources, and information on how people interpret verbal probability phrases such as “high chance” or “low probability” in severe weather forecasts [Fig. 4d; see Lenhardt et al. (2019) for more information]. In addition to interacting with these data, users are able to download a database of the estimates we produce, the raw survey data we use to calculate them, and the code necessary to reproduce the calculations.

Future: Research and development

Each wave of the Wx Survey provides new data that we use to track baseline measures and develop indicators of new concepts. As the project continues, we expect to move in multiple directions. Most notably, we are working to develop and validate estimates for the database and modules for the platform. These estimates and modules will allow enterprise members to identify and explore significant changes over time that may relate to changing demographics or new education and communication strategies. We are also working to increase the utility of the estimates in the platform by providing them on geographic scales that are more suitable to NWS partners in emergency management (i.e., counties) and broadcast media (i.e., designated market areas). Last, we are designing and validating comparable scales for other hazards. For example, we are building a new set of composite scales that will measure public reception, comprehension, and responsiveness to tropical cyclone and winter weather forecasts and warnings.

As we move in these new directions, we hope that fellow social and behavioral scientists will assist us by using the database to improve the models and address questions of the sort

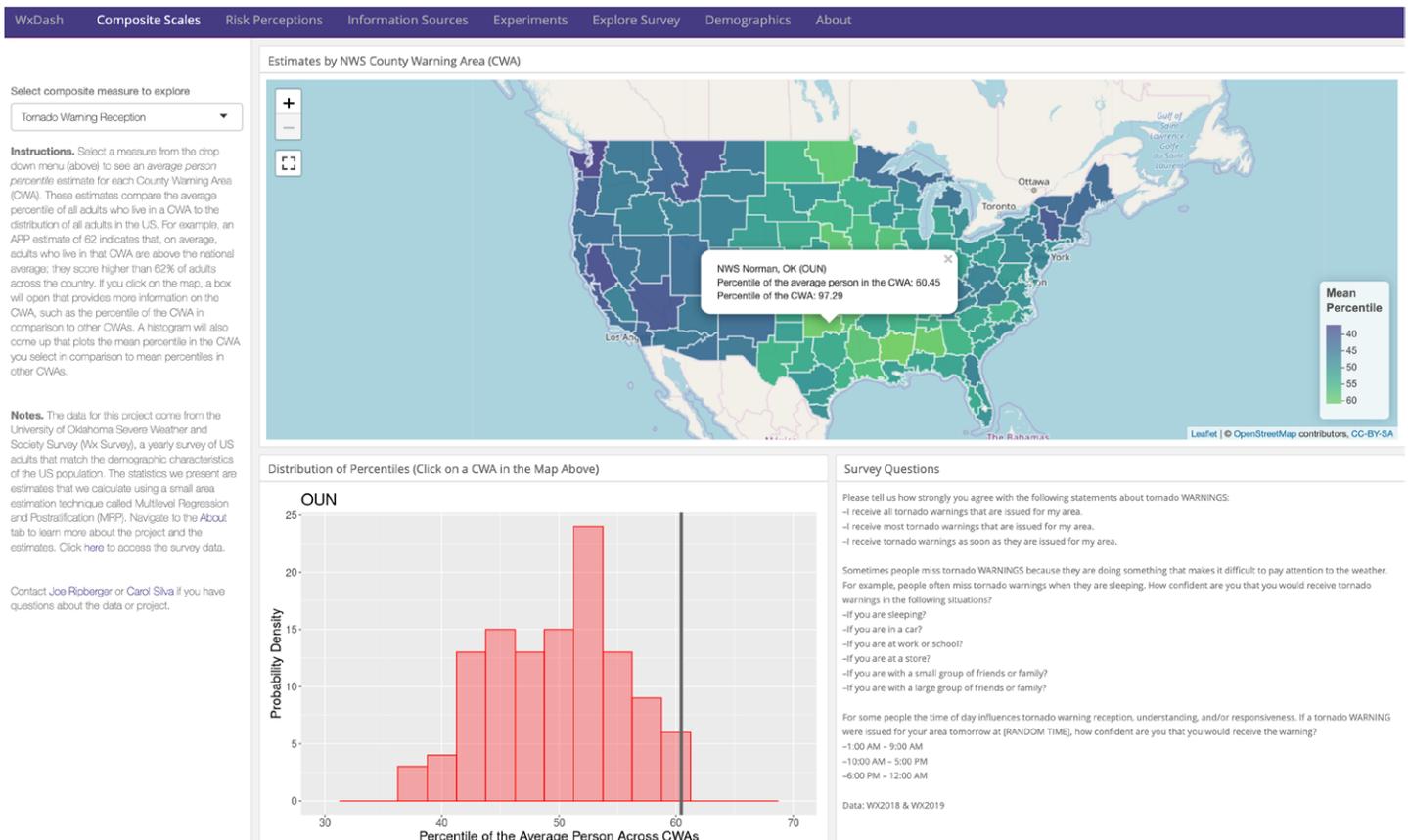


Fig. 4. Example screenshot from WxDash (<https://rcm.shinyapps.io/wxdash>).

we pose in this article. For example, the data show significant differences in tornado warning reception and comprehension between adjacent communities that experience roughly comparable levels of tornado threat. Why is this the case? Are enterprise members in some communities engaging in education and risk communication practices that are especially effective? If so, what can we learn from these practices and can we use them to improve forecast and warning reception and comprehension in different locations? The data also show a variety of anomalies where models consistently suggest lower levels of reception, comprehension, and response than observations indicate (and vice versa). People in the Amarillo CWA (AMA), for example, demonstrate relatively high values on these measures despite modest predictions from the models. The opposite is true of people in the Columbia CWA (CAE), where the models predict higher levels than observations indicate. Why? Is there something unique about the people in these communities? Might migration patterns, recency, or especially memorable events help us explain the patterns we observe in these communities? Maybe the patterns relate to differences in socioeconomic status and vulnerability that the models do not yet include? The database of community statistics we are developing for this project will allow us to address these questions and many others, but we cannot do it alone.

We also hope that NWS forecasters, emergency managers, broadcast meteorologists, and private partners will join this effort by providing feedback on the database and platform. What do you need to know about your communities to improve education and risk communication? What concepts require measurement? Following measurement, how can we best distribute and use the database we are developing to support our collective effort toward a weather-ready nation?

This is a new long-term project, and there will be many opportunities for input and improvement. As with all research of this type, the measures are imperfect and the estimates are uncertain. Nevertheless, we are optimistic it represents an important step toward providing enterprise members with useful information about the people and communities they serve.

Acknowledgments. Data collection for this project was funded by the OU Office of the Vice President for Research. Data analysis was funded by National Oceanic and Atmospheric Administration Project OAR-USWRP-R20, “FACETs Probability of What? Understanding and Conveying Uncertainty through Probabilistic Hazard Services,” and National Oceanic and Atmospheric Administration Project NA18OAR4590376, “Communicating Forecast Uncertainty and Probabilistic Information: Experimenting with Social Observation Data in the Hazardous Weather Testbed.”

References

- Allan, J. N., J. T. Ripberger, V. T. Ybarra, and E. T. Cokely, 2017: Tornado risk literacy: Beliefs, biases, and vulnerability. *13th Bi-Annual Int. Conf. on Naturalistic Decision Making and Uncertainty*, Bath, United Kingdom, University of Bath, 284–290.
- , ——, W. W. Wehde, M. J. Krocak, C. L. Silva, and H. J. Jenkins-Smith, 2019: Geographic distributions of extreme weather risk perceptions in the United States. University of Oklahoma Center for Risk and Crisis Management Rep., 24 pp., <https://github.com/oucr/cm/wxsurvey/blob/master/geographic%20distributions%20of%20extreme%20weather%20risk%20perceptions/manuscript.pdf>.
- Bergquist, P., and C. Warshaw, 2019: Does global warming increase public concern about climate change? *J. Polit.*, **81**, 686–691, <https://doi.org/10.1086/701766>.
- Brooks, H. E., C. A. Doswell III, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640, [https://doi.org/10.1175/1520-0434\(2003\)018<0626:CEOLDT>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0626:CEOLDT>2.0.CO;2).
- Buttice, M. K., and B. Highton, 2013: How does multilevel regression and post-stratification perform with conventional national surveys? *Polit. Anal.*, **21**, 449–467, <https://doi.org/10.1093/pan/mpt017>.
- Demuth, J. L., 2018: Explicating experience: Development of a valid scale of past hazard experience for tornadoes. *Risk Anal.*, **38**, 1921–1943, <https://doi.org/10.1111/risa.12983>.
- Eastern Research Group, 2018: National Weather Service hazard simplification: Public survey. National Weather Service Final Rep., 161 pp., www.weather.gov/media/hazardsimplification/HazSimp%20Public%20Survey%20-%20Final%20Report%20-%2006-01-18.pdf.
- Howe, P. D., 2018: Modeling geographic variation in household disaster preparedness across U.S. states and metropolitan areas. *Prof. Geogr.*, **70**, 491–503, <https://doi.org/10.1080/00330124.2017.1416301>.
- , M. Mildeberger, J. R. Marlon, and A. Leiserowitz, 2015: Geographic variation in opinions on climate change at state and local scales in the USA. *Nat. Climate Change*, **5**, 596–603, <https://doi.org/10.1038/nclimate2583>.
- , J. R. Marlon, X. Wang, and A. Leiserowitz, 2019: Public perceptions of the health risks of extreme heat across US states, counties, and neighborhoods. *Proc. Natl. Acad. Sci. USA*, **116**, 6743–6748, <https://doi.org/10.1073/pnas.1813145116>.
- Kiewiet de Jonge, C. P., G. Langer, and S. Sinozich, 2018: Predicting state presidential election results using national tracking polls and multilevel regression with poststratification (MRP). *Public Opin. Quart.*, **82**, 419–446, <https://doi.org/10.1093/poq/nfy023>.
- Klockow, K. E., R. A. Peppler, and R. A. McPherson, 2014: Tornado folk science in Alabama and Mississippi in the 27 April 2011 tornado outbreak. *GeoJournal*, **79**, 791–804, <https://doi.org/10.1007/s10708-013-9518-6>.
- Krocak, M. J., and H. E. Brooks, 2018: Climatological estimates of hourly tornado probability for the United States. *Wea. Forecasting*, **33**, 59–69, <https://doi.org/10.1175/WAF-D-17-0123.1>.
- Lax, J. R., and J. H. Phillips, 2009: How should we estimate public opinion in the states? *Amer. J. Polit. Sci.*, **53**, 107–121, <https://doi.org/10.1111/j.1540-5907.2008.00360.x>.
- Lenhardt, E. D., R. N. Cross, M. K. Krocak, J. T. Ripberger, S. R. Ernst, C. L. Silva, and H. C. Jenkins-Smith, 2019: How likely is that chance of thunderstorms? A study of how National Weather Service forecast offices use words of estimative probability and what they mean to the public. University of Oklahoma Center for Risk and Crisis Management Rep., 26 pp., <https://github.com/oucr/cm/wxsurvey/blob/master/how%20likely%20is%20that%20chance%20of%20thunderstorms/manuscript.pdf>.
- Mildeberger, M., J. R. Marlon, P. D. Howe, and A. Leiserowitz, 2017: The spatial distribution of Republican and Democratic climate opinions at state and local scales. *Climatic Change*, **145**, 539–548, <https://doi.org/10.1007/s10584-017-2103-0>.
- National Academies of Sciences, Engineering, and Medicine, 2018: *Integrating Social and Behavioral Sciences within the Weather Enterprise*. National Academies Press, 198 pp.
- National Research Council, 2010: *When Weather Matters: Science and Services to Meet Critical Societal Needs*. National Academies Press, 198 pp.
- , 2012: *Weather Services for the Nation: Becoming Second to None*. National Academies Press, 86 pp.
- Park, D. K., A. Gelman, and J. Bafumi, 2004: Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Polit. Anal.*, **12**, 375–385, <https://doi.org/10.1093/pan/mp024>.
- Pew Research Center, 2019: For local news, Americans embrace digital but still want strong community connection: Methodology. Pew Research Center, www.journalism.org/2019/03/26/local-news-methodology/.
- Ripberger, J. T., M. J. Krocak, W. W. Wehde, J. N. Allan, C. Silva, and H. Jenkins-Smith, 2019: Measuring tornado warning reception, comprehension, and response in the United States. *Wea. Climate Soc.*, **11**, 863–880, <https://doi.org/10.1175/WCAS-D-19-0015.1>.
- Shirani-Mehr, H., D. Rothschild, S. Goel, and A. Gelman, 2018: Disentangling bias and variance in election polls. *J. Amer. Stat. Assoc.*, **113**, 607–614.
- Silva, C. L., J. T. Ripberger, H. J. Jenkins-Smith, and M. J. Krocak, 2017: Establishing a baseline: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. University of Oklahoma Center for Risk and Crisis Management Rep., 31 pp., <http://risk.ou.edu/downloads/news/WX17-Reference-Report.pdf>.
- , ——, ——, ——, and W. W. Wehde, 2018: Refining the baseline: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. University of Oklahoma Center for Risk and Crisis Management Rep., 29 pp., <http://risk.ou.edu/downloads/news/WX18-Reference-Report.pdf>.
- , ——, ——, ——, S. Ernst, and A. Bell, 2019: Continuing the baseline: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. University of Oklahoma Center for Risk and Crisis Management Rep., 33 pp., <http://risk.ou.edu/downloads/news/WX19-Reference-Report.pdf>.
- Wang, W., D. Rothschild, S. Goel, and A. Gelman, 2015: Forecasting elections with non-representative polls. *Int. J. Forecasting*, **31**, 980–991, <https://doi.org/10.1016/j.ijforecast.2014.06.001>.
- Wang, Y., J. B. Holt, F. Xu, X. Zhang, D. P. Dooley, H. Lu, and J. B. Croft, 2018: Using 3 health surveys to compare multilevel models for small area estimation for chronic diseases and health behaviors. *Prev. Chronic Dis.*, **15**, E133, <https://doi.org/10.1016/10.5888/pcd15.180313>.
- Warshaw, C., 2018: Spatial variation in messaging effects. *Nat. Climate Change*, **8**, 360–361, <https://doi.org/10.1038/s41558-018-0143-8>.
- Zhang, X., J. B. Holt, H. Lu, A. G. Wheaton, E. S. Ford, K. J. Greenlund, and J. B. Croft, 2014: Multilevel regression and poststratification for small-area estimation of population health outcomes: A case study of chronic obstructive pulmonary disease prevalence using the Behavioral Risk Factor Surveillance System. *Amer. J. Epidemiol.*, **179**, 1025–1033, <https://doi.org/10.1093/aje/kwu018>.
- , ——, S. Yun, H. Lu, K. J. Greenlund, and J. B. Croft, 2015: Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the Behavioral Risk Factor Surveillance System. *Amer. J. Epidemiol.*, **182**, 127–137, <https://doi.org/10.1093/aje/kwv002>.