DRAFT VERSION

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

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

Center for Risk and Crisis Management
University of Oklahoma

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, such as what risks do the people in a community worry about or neglect? Do they generally receive, understand, and respond to forecasts and warnings? 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., NRC 2010 and 2012; NAS 2018). Most recently, for example, the NAS report on "Integrating Social and Behavioral Sciences Within the Weather Enterprise" (2018) states that:

> *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 sub-population characteristics from the US Census to estimate statistics of interest for communities across the country. We demonstrate the (1) approach, (2) database, (3) and platform by using them to explore differences in tornado warning, reception, comprehension, and response in counties and county warning areas (CWAs) across the US. 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 US adult population are increasingly common (e.g., ERG 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 US 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, Silva et al. 2018, and Silva et al. 2019). Large population surveys such as this provide valuable information about the population as a whole, but rarely address differences across geographic sub-populations.

Somewhat analogous to climate downscaling, survey researchers are actively developing small area estimation (SAE) techniques that downscale data from large population surveys to sub-populations, such as states, counties, and districts. Currently, there are two primary SEA 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. MRP is less data intensive. It uses regression analysis to identify demographic and geographic patterns in areas where data are available to produce reliable estimates in areas where data are relatively sparse (Park, Gelman, and Bafumi 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 US 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; Zhang et. al. 2015; Wang et. al. 2018); researchers at Pew Research Center are using it to identify news media consumption habits in US cities (Pew 2019[1]); and opinion analysts are using it to forecast election outcomes in US states (Wang et. al. 2015; Kiewiet de Jonge and Sinozich 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]; Mildenberger 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 https://www.journalism.org/interactives/local-news-habits/ to interact with an 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 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 US.

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, Silva et al. 2018, Silva et al. 2019). In addition, the survey data and code necessary to reproduce the estimates we develop for the database are available in a public repository (https://github.com/oucrcm). The next principle is *measurement*. While some concepts can be measured with a few relatively simple survey questions, others, including many of those 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 (e.g., Ripberger et al. 2019). The last and perhaps most important principle is *consistency*. The value of the database stems from an ability to identify similarities and differences between communities and within communities over time. This requires continuous data collection at routine intervals using constant methodologies and measures across relevant geographic domains.

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, shown in Table 1, prompt subjective assessments of warning reception, comprehension, and response, and provide an objective test of tornado warning comprehension. We use these questions because a concurrent study (Ripberger et al. 2019) indicates that they provide 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 theory (IRT) models that indicate each survey participant's latent scale scores by comparing them to other participants. To facilitate interpretation, we present the scale scores as percentiles.

[Table 1]

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

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

$\alpha^{gender}_{j} \sim N(0,\ \sigma^2_{gender})$, j = 1 or 2

$\alpha^{age}_{k} \sim N(0, \sigma^2_{age})$, k = 1, 2, or 3

$\alpha^{gender*age}_{j,k} \sim N(0, \sigma^2_{gender*age})$, j = 1 or 2 and k = 1, 2, or 3

$\alpha^{race}_{l} \sim N(0, \sigma^2_{race})$, l = 1, 2, or 3

$$\alpha_m^{ethnicity} \sim N\left(0, \sigma_{ethnicity}^2\right), m = 1 \text{ or } 2$$
$$\alpha_s^{area} \sim N\left(\beta^{climatology} * climatology_s, \sigma_{area}^2\right), s = 1, \ldots, 115$$

The models have two levels. Individually, a participant's percentile score on each scale (y_i) varies as a function of the participant's demographic profile (*gender*, *age*, a *gender-age* interaction, *race*, and *ethnicity*) and geographic *area* (CWA). CWA effects vary in relation to *climatology* (mean number of tornado event days per year[4]).

[Fig. 1]

The panels in Fig. 1 display the random intercept 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 influence on warning reception and comprehension than on response.

[Table 2]

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, age 18 to 34, white, non-Hispanic in the NWS Norman, OK (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. 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

[4] 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.

US Census.[5] For example, 10.8% of adults in the OUN CWA match the demographic group we describe above—female, age 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 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 US adults across the country.

[Fig. 2]

The maps in Fig. 2a display APP estimates of tornado warning reception, subjective comprehension, objective comprehension, and response by CWA in the CONUS. Fig. 2b plots the distribution of these estimates across CWAs. In combination, the figures illustrate multiple findings. Most notably, they indicate significant variation in reception and comprehension across the country. APP scores range from 37 to 61 (a span of 24 percentiles) on the reception scale, 32 to 69 (37 percentiles) on the subjective comprehension scale, and 38 to 61 (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.

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 middle portion of the US. Unsurprisingly, this pattern roughly mimics tornado climatology (e.g., Brooks et al. 2003), 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

---

[5] County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin are available here: https://www2.census.gov/programs-surveys/popest/datasets/2010-2018/counties/asrh/.

take protective action in response. Nevertheless, tornadoes are *possible* almost everywhere in the US 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 9 percentile point difference in subjective comprehension estimates between the Norman, OK CWA (APP = 66) and the Fort Worth, TX CWA (APP = 57), despite roughly comparable tornado climates. There is a roughly analogous 7 percentile point difference between the Peachtree City, GA CWA (APP = 49) and the Birmingham, AL 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 tornado warning comprehension? Are the WFOs, 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 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 are confident that 50 observations provide enough information to convey 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 conservative mean estimate (observation) for each community that is more likely to replicate in future surveys.

[Fig. 3]

Fig. 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 especially true of the tornado warning reception, subjective comprehension, and objective comprehension estimates, where the correlation coefficients are 0.68, 0.75, and 0.75, respectively. The response coefficient drops (0.43), 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 (MAD) between the estimates and observations demonstrate this point. The models predict reception observations within an average of 3 percentiles, subjective comprehension within 5 percentiles, and objective comprehension and response within an average of 4 percentiles.

In forecast terminology, these results indicate that the estimates we produce from MRP models 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, SC CWA (CAE) and consistently lower than observations in the San Francisco, CA CWA (MTR). There are many possible culprits; we propose one of two explanations for these 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 towards 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. We hope that the estimates in this database will encourage and allow more research on why some communities demonstrate abnormally high/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.

[Fig. 4]

The Severe Weather and Society Dashboard (WxDash) is meant to provide this opportunity. WxDash, which is 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 (Fig. 4a), 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, Peppler, & McPherson 2014; Allan et al. 2017). In addition to this set of composite scales, WxDash provides information on risk perceptions across a variety of hazards (Fig. 4b; see Allan et al. 2019 for more information), data on tornado warning and extreme weather information sources (Fig. 4c), and information on how people interpret words of estimative 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, 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 area). Lastly, 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 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 San Francisco, CA CWA (MTR), for example, demonstrate relatively high values on these measures despite low predictions from the models. Why? Is there something unique about the people in MTR? Perhaps people are migrating *to* MTR *from* areas that experience more tornadoes and bringing homegrown knowledge with them? Or maybe the differences are related to socioeconomic status and the access to information that status affords? 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 towards a weather-ready nation?

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

References

Allan, J.N., Ripberger, J.T., Ybarra, V.T. and Cokely, E.T., 2017. Tornado Risk Literacy: Beliefs, Biases, and Vulnerability. Naturalistic Decision Making and Uncertainty, pp.284-290.

Allan, J.N., Ripberger, J.T., Wehde W.W., Krocak M.J., Silva C.L., and Jenkins-Smith, H.J., 2019. Geographic Distributions of Extreme Weather Risk Perceptions in the United States.

University of Oklahoma Center for Risk and Crisis Management. Link: https://github.com/oucrcm/wxsurvey/blob/master/geographic%20distributions%20of%20extreme%20weather%20risk%20perceptions/manuscript.pdf.

Bergquist, P. and Warshaw, C., 2019. Does Global Warming Increase Public Concern about Climate Change?. The Journal of Politics, 81(2), pp.686-691.

Buttice, M.K. and Highton, B., 2013. How does multilevel regression and poststratification perform with conventional national surveys?. Political analysis, 21(4).

Eastern Research Group, 2018. National Weather Service Hazard Simplification: Public Survey. Final Report for the National Weather Service. Link: https://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 US States and Metropolitan Areas. The Professional Geographer, 70(3), pp.491-503.

Howe, P.D., Marlon, J.R., Wang, X. and Leiserowitz, A., 2019. Public perceptions of the health risks of extreme heat across US states, counties, and neighborhoods. Proceedings of the National Academy of Sciences, 116(14), pp.6743-6748.

Howe, P.D., Mildenberger, M., Marlon, J.R. and Leiserowitz, A., 2015. Geographic variation in opinions on climate change at state and local scales in the USA. Nature Climate Change, 5(6), p.596.

Kiewiet de Jonge, C.P., Langer, G. and Sinozich, S., 2018. Predicting State Presidential Election Results Using National Tracking Polls and Multilevel Regression with Poststratification (MRP). Public Opinion Quarterly, 82(3), pp.419-446.

Klockow, K.E., Peppler, R.A. and McPherson, R.A., 2014. Tornado folk science in Alabama and Mississippi in the 27 April 2011 tornado outbreak. GeoJournal, 79(6), pp.791-804.

Lax, J.R. and Phillips, J.H., 2009. How should we estimate public opinion in the states?. American Journal of Political Science, 53(1), pp.107-121.

Lenhardt, E.D., Cross, R.N., Krocak, M.K., Ripberger, J.T., Ernst, S.R., Silva, C.L., Jenkins-Smith, H.C., 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. Link: https://github.com/oucrcm/wxsurvey/blob/master/how%20likely%20is%20that%20chance%20of%20thunderstorms/manuscript.pdf.

Mildenberger, M., Marlon, J.R., Howe, P.D. and Leiserowitz, A., 2017. The spatial distribution of Republican and Democratic climate opinions at state and local scales. Climatic change, 145(3-4), pp.539-548.

National Academies of Sciences, Engineering, and Medicine, 2018. Integrating Social and Behavioral Sciences Within the Weather Enterprise. National Academies Press.

National Research Council, 2010. When weather matters: Science and services to meet critical societal needs. National Academies Press.

National Research Council, 2012. Weather Services for the Nation: Becoming Second to None. National Academies Press.

Park, D.K., Gelman, A. and Bafumi, J., 2004. Bayesian multilevel estimation with poststratification: State-level estimates from national polls. Political Analysis, 12(4), pp.375-385.

Pew Research Center, 2019. For local news, Americans embrace digital but still want strong community connection: Methodology. Link: https://www.journalism.org/2019/03/26/local-news-methodology/.

Ripberger, J.T., Krocak, M.J., Wehde, W.W., Allan, J.N., Silva, C. and Jenkins-Smith, H., 2019. Measuring Tornado Warning Reception, Comprehension, and Response in the United States. Weather, Climate, and Society.

Silva, C. L., Ripberger, J.T., Jenkins-Smith H.J., and Krocak M.J., 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. Link: http://risk.ou.edu/downloads/news/WX17-Reference-Report.pdf.

Silva, C. L., Ripberger, J.T., Jenkins-Smith H.J., and Krocak M.J., Wehde, W.W., 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. Link: http://risk.ou.edu/downloads/news/WX18-Reference-Report.pdf.

Silva, C. L., Ripberger, J.T., Jenkins-Smith H.J., and Krocak M.J., Ernst, S., Bell, A., 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. Link: http://risk.ou.edu/downloads/news/WX19-Reference-Report.pdf.

Wang, W., Rothschild, D., Goel, S. and Gelman, A., 2015. Forecasting elections with non-representative polls. International Journal of Forecasting, 31(3), pp.980-991.

Wang, Y., Holt, J.B., Xu, F., Zhang, X., Dooley, D.P., Lu, H. and Croft, J.B., 2018. Using 3 Health Surveys to Compare Multilevel Models for Small Area Estimation for Chronic Diseases and Health Behaviors. Preventing chronic disease, 15.

Warshaw, C., 2018. Spatial variation in messaging effects. Nature Climate Change, 8(5), p.360.

Zhang, X., Holt, J.B., Lu, H., Wheaton, A.G., Ford, E.S., Greenlund, K.J. and Croft, J.B., 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. American journal of epidemiology, 179(8), pp.1025-1033.

Zhang, X., Holt, J.B., Yun, S., Lu, H., Greenlund, K.J. and Croft, J.B., 2015. Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the behavioral risk factor surveillance system. American journal of epidemiology, 182(2), pp.127-137.

DRAFT VERSION

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

Reception	Subjective Comprehension	Objective Comprehension	Response

1. Please tell us how strongly you agree with the following statements about tornado WARNINGS:
i. I receive all tornado warnings that are issued for my area.
ii. I receive most tornado warnings that are issued for my area.
iii. I receive tornado warnings as soon as they are issued for my area.
2. 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?
i. If you are sleeping?
ii. If you are in a car?
iii. If you are at work or school?
iv. If you are at a store?
v. If you are with a small group of friends or family?
vi. If you are with a large group of friends or family?
3. 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 receive the warning?
i. 1:00 AM – 9:00 AM
ii. 10:00 AM – 5:00 PM
iii. 6:00 PM – 12:00 AM
	1. In general, do you understand the difference between watches and warnings?
	2. How would you rate your understanding of tornado watches and warnings?
	3. How would you rate your understanding of severe thunderstorm watches and warnings?
	4. Forecasters, websites, and phone applications often use maps to display tornado watches and warnings. How would you rate your understanding of maps?
	5. Forecasters, websites, and phone applications also use radar images to communicate tornado risk. How would you rate your understanding of radar images?
	6. 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 understand the warning?
	i. 1:00 AM – 9:00 AM
	ii. 10:00 AM – 5:00 PM
	iii. 6:00 PM – 12:00 AM
		1a. This alert is issued when severe thunderstorms and tornadoes are possible in and near the area. It does not mean that they will occur. It only means they are possible. [*50% of participants*]
		• Tornado Watch*
		• Tornado Warning
		• Don't Know
		1b. This alert is used when a tornado is imminent. When this alert is issued, seek safe shelter immediately. [*50% of participants*]
		• Tornado Watch
		• Tornado Warning*
		• Don't Know
		2. If the National Weather Service issues a tornado warning for your area, how much time do you have before the tornado arrives?
		• Less than 1 hour
		  • How many minutes are there between when tornado WARNINGS are issued and when tornadoes arrive? [< 30 minutes]*
		• 1 to 24 hours
		• 1 to 3 days
		• More than 3 days
		3. If the National Weather Service issues a tornado watch for your area, how much time do you have before the tornado arrives?
		• Less than 1 hour
		• 1 to 24 hours*
		  • How many hours are there between when tornado watches are issued and when tornadoes arrive? [1 to 3 hours]*
		• 1 to 3 days
		• More than 3 days
		4. Approximately how large is the area included in an average tornado watch?
		• Around the size of a city
		• Around the size of a county
		• Around the size of multiple counties*
		• Around the size of a state*
		• Around the size of multiple states*
			1. Please tell us how strongly you agree with the following statements about tornado WARNINGS. If you have never received a tornado WARNING, please tell us how you think you will respond if you receive a WARNING in the future:
			• I always take protective action when tornado warnings are issued for my area.
			2. Sometimes people receive tornado WARNINGS but *do not take protective action* because they are busy or doing something that makes it difficult to respond. For example, people often decide not to take protective action in response to tornado warnings when they are sleeping. How confident are you that you would *take protective action in response* to tornado warnings in the following situations?
			i. If you are sleeping?
			ii. If you are in a car?
			iii. If you are at work or school?
			iv. If you are at a store?
			v. If you are with a small group of friends or family?
			vi. If you are with a large group of friends or family?
			3. 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?
			i. 1:00 AM – 9:00 AM
			ii. 10:00 AM – 5:00 PM
			iii. 6:00 PM – 12:00 AM

Notes: * indicates "correct" response to objective comprehension survey questions.

Table 2: Summary of the regression models we use to predict tornado warning reception, subjective comprehension, objective comprehension, and response.

	Reception	Subjective Comprehension	Objective Comprehension	Response
Estimates				
Intercept (β^0)	-0.06 (0.37)	-0.15 (0.49)	-0.10 (0.32)	-0.04 (0.37)
Tornado Climatology (β)	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
Residual	*0.95*	*0.90*	*0.56*	*0.97*
Obs.	5996	5996	5996	5996

Notes: 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.

DRAFT VERSION

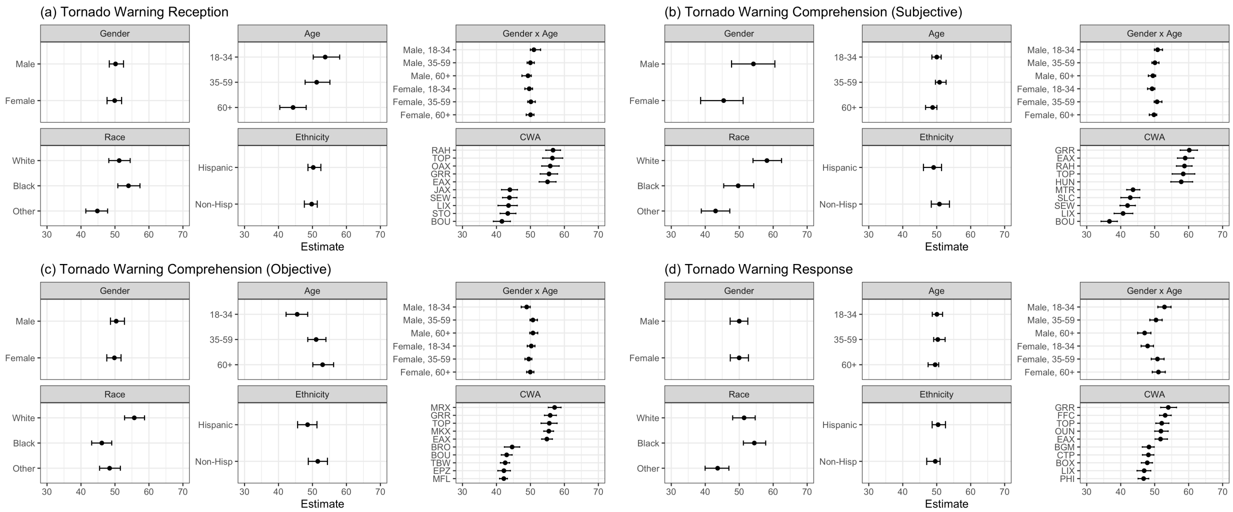

Fig. 1: Random intercept estimates from regression models that predict tornado warning (a) reception; (b) comprehension (subjective); (c) comprehension (objective); and (d) response. The estimates compare group percentile scores; the points indicate median estimates from the models and the intervals indicate 50% uncertainty ranges. The CWA panels show the top and bottom 5 CWAs for each measure.

DRAFT VERSION

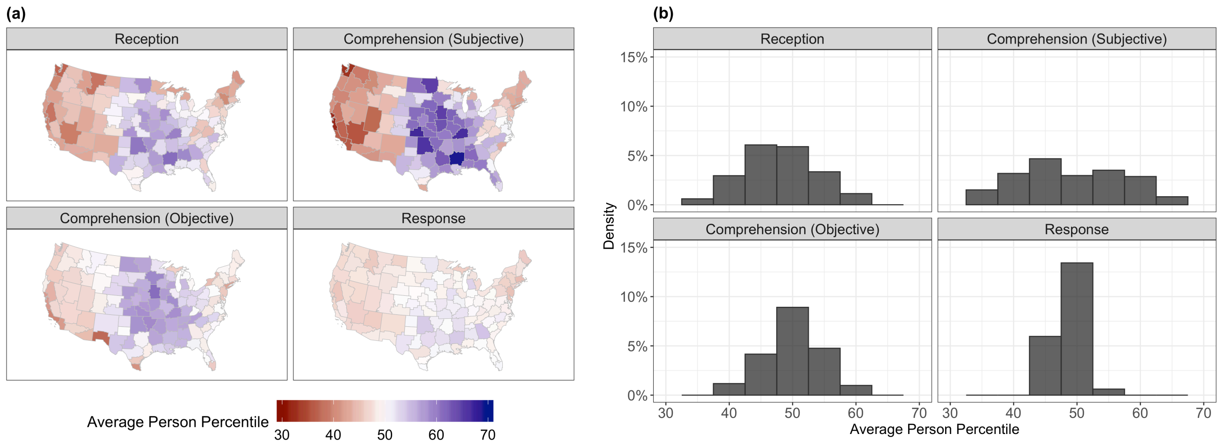

Fig. 2: Average person percentile (APP) estimates of tornado warning reception, subjective comprehension, objective comprehension, and response (a) by CWA in the CONUS; (b) the frequency distribution of APP estimates across the by measure.

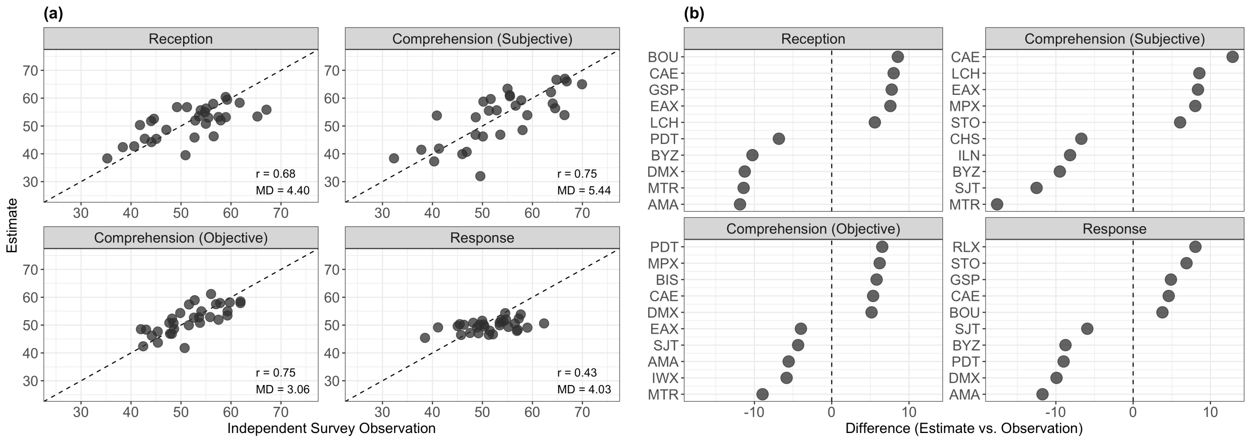

Fig. 3: (a) A comparison between independent survey observations in 30 CWAs and the APP estimates we produce using MRP models; (b) the top and bottom 5 overestimates and underestimates by measure.

DRAFT VERSION

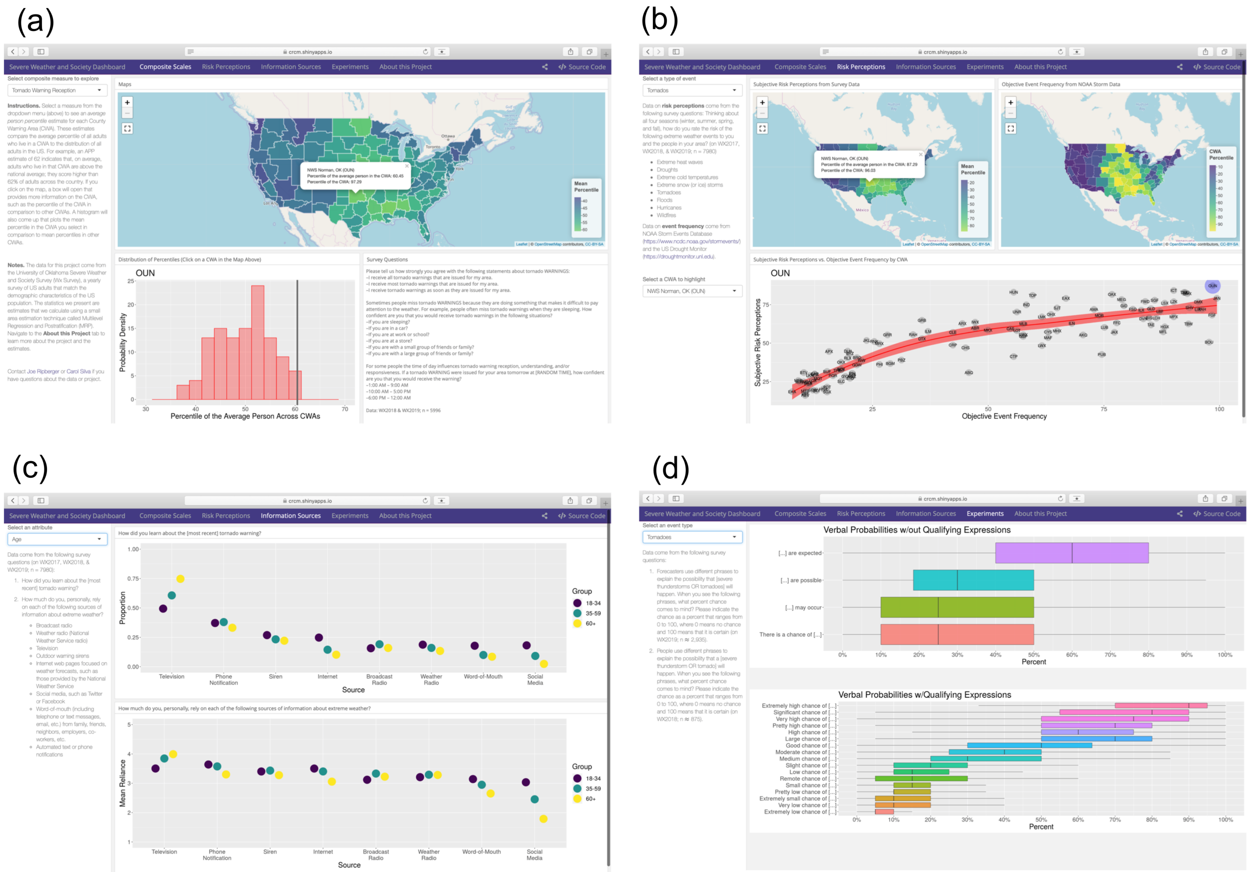

Fig 4: Example screenshots from WxDash that depict (a) tornado warning reception; (b) tornado risk perceptions in comparison to objective frequency; (c) tornado warning and weather information sources; and (d) interpretations of words of estimative probability.