UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

DEFINING A RELATIVELY NORMAL TORNADO DAY: AN EXPLORATION OF THE CLIMATOLOGY AND POTENTIAL USE OF RELATIVE RISK FORECASTS FOR TORNADOES

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

By

SEAN R. ERNST
Norman, Oklahoma
2023

DEFINING A RELATIVELY NORMAL TORNADO DAY: AN EXPLORATION OF THE CLIMATOLOGY AND POTENTIAL USE OF RELATIVE RISK FORECASTS FOR TORNADOES

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Harold Brooks, Chair

Dr. Joseph Ripberger

Dr. Elinor Martin

Dr. Jens Redemann

Dr. Jason Furtado

Dr. Kelvin Droegemeier

Dr. Justin Reedy

Acknowledgements

Ten years ago, nearly to the day, my Dad sent me a news article about how Congress was considering ordering the NWS to extend the length of time that tornado warnings are issued for to an hour in length. The time extension of warnings was being considered as a response to the deadly Central Oklahoma tornadoes of late May 2013, the worst of which, an EF5 tornado that struck the Oklahoma City suburb of Moore on May 20th, had been broadcast live on television stations across the nation. I remember watching the helicopter footage from my grandparent's Massachusetts home that afternoon and feeling deeply conflicted about what I saw, as I tried to reconcile the excitement I felt at watching a live tornado with the heartbreak I felt at seeing the damage it had left behind. In the days after the tornado, as my feelings began to turn towards a desire to find some way to help, Dad forwarded the tornado warning article to me and suggested that I email the scientist interviewed in it, Dr. Kelvin Droegemeier. Dad thought that reaching out to Dr. Droegemeier, who in the article advocated for studying tornado warning communication before making drastic changes to the warning, could open up opportunities for me to begin using my passion for weather and science to make a positive difference. It took a few days to work up the courage to reach out to a Real Life Scientist™ for the first time, but not 30 minutes after I sent an email asking for advice on how to learn more about tornado warnings for a school project, I received a reply from Kelvin wherein he told me he wanted to mentor me on my project and have me visit him at the University of Oklahoma! Suffice to say, the 10 years since have been an absolute whirlwind, and I have a veritable flood of incredibly supportive people I need to take a moment to thank for making my journey to a PhD dissertation possible.

Academically, I have to begin with Dr. Kelvin Droegemeier and Dr. Alicia Knoedler, who recruited me to OU and mentored and supported me through my four years of undergraduate study here. It was a roller coaster ride at times, but I am certain I would never have finished my degree without their guidance and support – and one of the most important hugs of my life from Dr. Knoedler. I am also grateful to Dr. Daphne LaDue for taking me under her wing from freshman to senior year as a student researcher, as I learned so much about the proper rigor needed to produce good social science work from my research with her! My Hollings Scholarship mentors, Dr. Kim Klockow-McClain and Jen Sprague-Hilderbrand also deserve a shout out here for introducing me to many of the NWS scientists and administrators that have helped me accelerate my career since.

My good fortune with mentors continued into graduate school, where I was blessed to have an incredible group of people guide me through my graduate studies here at OU. Dr. Kenzie Krocak deserves greater praise than anything I can put into words on paper – she encouraged me to reach out to Dr. Joseph Ripberger about graduate school opportunities, and has put up with me popping by her office on a daily basis to chat about Legos, volcanoes, or whatever random stories I felt like telling for FIVE YEARS now! I couldn't ask for a better role model, inspiration, mentor, and friend, and I can't wait to collaborate on projects with Kenzie for years to come. Dr. Joseph Ripberger, of course, deserves credit for teaching me much of the social science methodology I know today, as he taught me the ins and outs of coding in R, writing scientific papers, and everything I know about using statistics to analyze social data – I wouldn't be the scientist I am without his guidance, and I'm honored that I got to be the first doctoral student that Joe has seen through the convocation hooding ceremony! And of course, we cannot forget the unforgettable Dr. Harold Brooks, who has always pushed me to do my very best and helped me organize my thesis and dissertation ideas into achievable, meaningful research – while performing the unenviable task of keeping my ego in check with well-placed clever commentary! Harold, I'll try to make sure that I always answer the prompt when I'm asked a question… unless the data tells me that I should do otherwise!

The other members of my Doctoral committee also deserve special recognition here, both for adapting to the schedules of a large group of busy scientists, but also for giving me challenges that pushed me to my limit in the General Exam and dissertation. Drs. Elinor Martin, Jason Furtado, Jens Redemann, Kelvin Droegemeier, and Justin Reedy – I can't thank you enough for being a key part of my academic career, and for the well of support and kindness that you've shown me every step of the way. It means the world to me that I know that I can talk to any of you and receive advice, a kind word, or help solving a tough problem whenever I am in need – you all have my utmost respect and appreciation.

There are also a few teachers I want to mention that encouraged my weather interests before college. Mrs. Henderson and Mr. Dillon in elementary school managed my limitless energy and helped me learn to point it towards creative learning endeavors, while JP Sullivan, Jen Alexander, Dawn Crane, and Alan Laubenstein were kind, supportive high school mentors that taught me the foundational skills I'd need to succeed as a scientist. My Sensei at Luth's Family Karate – Mike and Jen Luth – also taught me important lessons about patience and self-control that I use to this day. I have to thank Dr. Brian Johnson, one of the kindest people I've ever met, for mentoring me in my OU undergraduate years and for bringing my writing skills to the level they are today. I can't thank the many teachers I have had over the decades enough for the love of learning they instilled in me that will serve me for the rest of my life.

I also want to thank the mentors, scientists, and support staff at the School of Meteorology, the Institute for Public Policy Research and Analysis, and the Storm Prediction Center for being such an important part of my everyday life. Danika Hines-Barnett, and before her Christie Upchurch and Shelby Hill, played a crucial role in helping me get past the many administrative hurdles of undergraduate and graduate school, and provided emotional support that made the hardest days feel a bit more manageable. At the SPC, Patrick Marsh and Israel Jirak provided mentorship and insight for much of my graduate career, granting me valuable access to the inner workings of the SPC and Hazardous Weather Testbed that I would never have had otherwise. And of course, Carol Silva, Hank Jenkins-Smith, Jinan Allen, Kuhika Gupta, Nina Carlson, Sam Price, Rhonda Hill, and all the current and former members of IPPRA have welcomed me to the most unbelievably healthy workplace and have always been there with a smile and a hand to help me up when I'm feeling down. They never cease to remind me that I am a hard-working scientist who is a key part of an amazing team – and will continue to be! I look forward to many more lunches with the team to come, and hopefully someday a helicopter for Hank!

Although the amazing people I work with have made my research the least stressful part of my life, I also have to thank the incredible friends I have made for making life outside of work re-energizing and exciting. Back home, I have my yurt buddies Dan Woodworth, Conor McNierney, Sean Nishi, Harrison Erb, and Xal Bilotta to thank for making every trip I take home a blast – and I can't forget Brooke Martino, Jamie DiPalma, and May Brackett for always making time to see me when I visit! Online, my good friends Justin McReynolds, Alvaro, and Karmela kept me sane during the pandemic and continue to be some of the best gaming buddies around. Here in Norman there are simply too many special people to count – at IPPRA, Anna Wanless, Abby Bitterman, and Sam Stormer keep me in line, while in the School of Meteorology my friends Max Ungar, Alec Prosser, Lauren Pounds, Ben Fellman, Francesca Lappin, and Morgan Schneider have all been there through the crazy ups and downs of the past few years. I also have to mention Addison Alford, Rachel Miller, Kenzie Krocak, Matt Flournoy, and Tony Lyza for helping me prepare to tackle my thesis, general exam, and now dissertation, and for all the invitations to spend an afternoon at 405 or BIG with a drink in hand. Finally, I have to thank Tabitha Kloss for being part of the best team in OU Meteorology undergrad history, and my roommates during graduate school, Andy Wilkins and Melanie Schroers, for making home a great place to be.

I'm going on way too long here – this should surprise nobody – but of course I can't wrap up my acknowledgements without thanking my family for nearly three decades of putting up with my nonsense, and for forging me into the person I am today. Mom and Dad gave me the support and love I needed always and pushed me to chase the dreams *I* wanted to catch. They never called me crazy for wanting to be a meteorologist, and instead helped me feed my hunger for knowledge with meteorology textbooks, visits with broadcast meteorologists, and trips to weather exhibits at museums! My family friend Blair Talcott-Orloff also gets special mention here, for introducing me to so many meteorologists in New England and always cheering on my weather passion. I may have (and still do!) butt heads with my sister Jill, but I know few people that will give it to me straight like she does – because of how much she cares about me and wants me to be the best I can be. My Nana and Papa made every Monday magic, as they transported me to distant planets, helped me explore mysterious woods, and taught me that you can learn something from *everyone* you meet. My Meme and Grampy always welcomed me with "hugs and kisses XOXO" on lazy summer days at the lake – Grampy, I miss you, and I know you're always keeping me safe when I find myself in the most dangerous storms. My Aunt Susan shared with me her love of travel and technology, and I can't possibly forget all the year-round fun I've had with the rambunctious Ryders! Aunt Heather and the Fraleys have supported me always if from afar, and I will always look back fondly on our visits to the big playground in Pinehurst where Papa and I would chase Chase and Trace around all afternoon. And of course, my sweet cats Snickers and Peanut, who have welcomed me with a calming purr every time I return home. Every one of my family members worked together to teach me that kindness and love are the two most important things you can offer to others, and I would not be the person I am without those lessons.

Six pages of thanks, and I still cannot fit all the names I wish I could! To those out there who have been a part of my journey and have supported me – thank you. No person is an island, and no success rests on any one person’s shoulders. Each of us stand on the shoulders of giants that themselves stand on giants. We’re in it together, and I can’t wait to see what we’ll be able to do!

Table of Contents

List of Tables

Table of Figures

Abstract

The risks posed by rare and severe weather events are both the most impactful to the public and the most difficult to communicate, due to their low absolute probability of occurrence but high impact at a given point. Recent research on rare risk communication has found that probabilistic information is the most effective way to communicate the risk posed by weather to members of the public, but also that small absolute probabilities can be misinterpreted due to known cognitive biases. One potential solution to communicating rare event likelihoods is a value called relative risk, which this work holistically investigates in the context of tornado risk communication. Relative risk is defined as the ratio of the forecast likelihood of an event to the background likelihood of that event occurring, the quotient of which describes how many times more likely than "normal" an event is for a given forecast. To achieve a broader understanding of how relative risk for tornadoes might impact different aspects of tornado risk communication, a 1950-2021 climatology of tornado relative risks was studied alongside focus group and survey data collection that investigated the reception of relative risk information by broadcast meteorologists and members of the public.

First, the 1950-2021 US observed tornado dataset maintained by the Storm Prediction Center (SPC) was used to calculate relative risks for tornadoes for all events across the 71 year period. Observed tornado reports were used to calculate the climatological likelihood of a tornado within 25 miles of a point for every day of the year, as well as Practically Perfect Hindcasts (PPHs) for tornadoes for every day in the 71-year dataset. Dividing these PPHs by the daily 1950-2021 tornado climatology produced a full series of relative risk values. Analysis of the 71 years of relative risk data revealed that the highest values of relative risk occur across the western and northern regions of the contiguous US where tornadoes can be rare, while the lowest

values of relative risk were observed in the southern US where tornadoes occur year-round. Relative risks values of 5, 20, 50, 100, 250, and 700 times more likely than normal were observed to occur across the 71-year dataset at similar rates to the absolute likelihood values of 2, 5, 10, 15, 30, and 45% used in the SPC probabilistic outlook, and were thus chosen to be used in in mapped presentations of relative risk to potential users. Overall, these results suggest that relative risks likely have the greatest value for communicating tornado events that occur in areas with established tornado climatology when tornadoes occur during off-season times of year or in places that they are more infrequent.

Next, focus group interviews with broadcasters revealed a great deal of suspicion of towards viewers' ability to interpret probability information in any form. Broadcasters in these interviews believed that relative risk information might be seen as overblown, potentially inducing unnecessary panic in their viewers. Relative risk was seen by broadcasters as a useful tool for their personal use in quantifying how unusual a tornado event was, and a few suggested they would present relative risk during tornado for events occurring outside of tornado season or outside of tornado-prone regions.

Finally, members of the public were also surveyed about their perceived level of concern and likelihood of response when shown different levels of relative risk information. Participants on average reported large increases in concern when shown only relative risk information but presenting both absolute and relative risk information led to better differentiation in level of concern across increasing absolute likelihood risk levels. Later experiments added nuance to these findings, and showed that although relative risk information at values as high as 500 times more likely than normal had no effect on participant trust in future forecasts if that forecast were

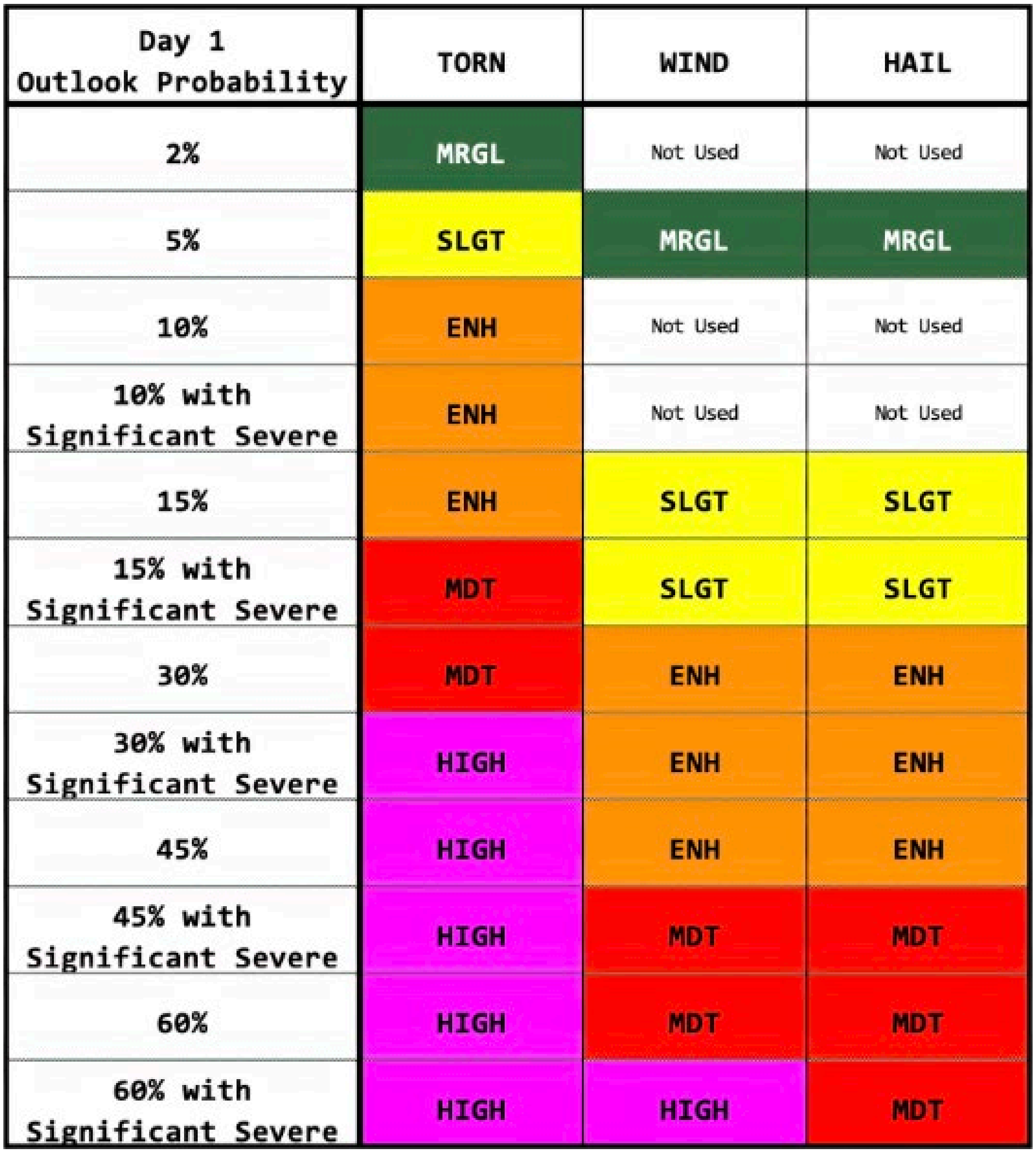
a false alarm, there was also no change to participant concern or likelihood of response across increasing levels of relative risk with a constant absolute risk value.

Through a series of three investigations, this dissertation attempts to develop and test a potential forecast product in a rigorous and methodical way that combines key understandings from both the meteorological and social sciences. Following this process helps ensure that both the potential range of values the product may contain and the range of responses that product recipients will display are well-understood. These rigorous efforts have revealed that although relative risk information could be valuable in specific situations for communicators like broadcast meteorologists, it does not appear to have significant effects on individual risk assessment or decision-making for members of the public. Thus, relative risk does not appear to offer a silver bullet for improving the communication of rare events like tornadoes.

Chapter 1 – Introduction

On June 1st, 2011, I recall spending my morning gym class playing ultimate frisbee on the field outside of my central Massachusetts school. The weather was sultry, with high dewpoints in place after a line of storms had passed through early in the morning. Dark, heavy cumulus clouds filled the sky, their increasing height and greying undersides hinting towards the unstable nature of the air that day. Strong winds aloft sent each new updraft hurrying towards the east, while I had to correct my frisbee tosses to account for a strong southerly wind at the surface. To a meteorologist, all these signs combined would hoist a massive red flag that today was not going to be your average New England severe weather day – and indeed, that afternoon a high-end EF3 tornado would leave a trail of destruction visible from satellite from Springfield, MA in the west to south of Boston in the east. As a precocious young student absolutely fascinated by the weather, I was very much tuned into forecasts that had highlighted the severe weather threat for days leading up to the 1st. But I was surprised to find that my friends and classmates were scoffing at what I considered a highly unusual and significant threat for our area. They believed that tornadoes simply did not happen in New England and were so confident that they offered to bet me that not a single tornado would occur in the state that day. I tried to find a way to highlight how unusual this threat was, how today was not like most severe weather days for us, but I was not able to find a helpful way to do so beyond comparing the setup to one you might expect in tornado alley.

Later, during the data collection process for my Masters thesis, the June 1st 2011 tornado event came up again in focus group conversations with broadcast meteorologists. In that study, I interviewed broadcasters about their use of the Storm Prediction Center's (SPC) convective outlook, a five-tiered scale that communicates severe weather risk based on the likelihood of tornadoes, wind, and hail within 25 miles of a point (see fig 1.1). During these focus groups, a broadcaster from a part of the country that infrequently sees severe weather mentioned that they rarely show the convective outlook on air because the highest levels of risk in the outlook, “Moderate” and “High”, are almost never issued for their area (Ernst 2020). The broadcaster was frustrated that, because the high likelihoods for severe weather within 25 miles of a point that are tied to the highest risk categories (fig 1.1) are unlikely to ever occur in their area, they were in effect limited to communicating risk using only the first four levels of a five-tier scale. Further, other broadcasters in the study noted how “a 5% [tornado likelihood] for [my area is] high but for a typical person it’s low”, even if such a subjectively low likelihood is significant given their location (Ernst 2020). The testimonials of these broadcasters led me to wonder what could be done to better communicate the risk of tornadoes based on the context of when and where those tornadoes are forecast to occur, to somehow convey the idea that a “5% risk of tornadoes for my area is high.”



Day 1 Outlook Probability	TORN	WIND	HAIL
2%	MRGL	Not Used	Not Used
5%	SLGT	MRGL	MRGL
10%	ENH	Not Used	Not Used
10% with Significant Severe	ENH	Not Used	Not Used
15%	ENH	SLGT	SLGT
15% with Significant Severe	MDT	SLGT	SLGT
30%	MDT	ENH	ENH
30% with Significant Severe	HIGH	ENH	ENH
45%	HIGH	ENH	ENH
45% with Significant Severe	HIGH	MDT	MDT
60%	HIGH	MDT	MDT
60% with Significant Severe	HIGH	HIGH	MDT

Figure 1.1: A table that displays the breakdown of the modern day SPC categorical outlook by the probabilities of tornadoes, thunderstorm wind, and hail that are tied to each categorical level. Note that the category names are shortened, such that "MRGL" represents "Marginal," "SLGT" represents "Slight", "ENH" represents "Enhanced Slight" or "Enhanced", "MDT" represents "Moderate", and "HIGH" represents "High" risk. The probabilities on the left side of the table relate to the forecast likelihood of a severe weather event within 25 miles of a point. This graphic is originally from Grams et al. (2014).

My interest in adding regional context to severe weather forecasts eventually led me to consider how climatology information could be used to add context to forecasts of tornadoes. Dr. Harold Brooks, my graduate advisor, pointed me towards the concept of relative risk, a value expressing the ratio of the likelihood of a forecast outcome to the background likelihood of that outcome (Spiegelhalter 2017). Note that the word "risk" with relative risk is used colloquially to describe tornado likelihood, similar to how the word is used in the SPC convective outlook to describe the categorical risk tiers (Grams et al. 2014). While the SPC uses the word "risk" in this colloquial way to describe the likelihood of a severe weather impact, risk scholars have previously defined risk as the product of three factors: the likelihood of a hazard occurring; the exposure of people and property to the hazard; and the vulnerability of those exposed to harm from the hazard (FEMA 2017). As I will be discussing SPC outlook products at length for the remainder of this dissertation, I will use "risk" in a colloquial manner to describe absolute risks (likelihood probability of a hazard) and relative risks (the likelihood probability divided by the climatological probability of a hazard). That aside, relative risk has been proposed as a tool for communicating rare but severe weather events since as early as the 1990s (Murphy 1991). Relative risks have generally only been studied for use in the communication of medical risks, not meteorological ones (Lipkus 2007; Fagerlin et al. 2011; Spiegelhalter 2017; Trevena et al. 2013; Costa-font et al. 2021), although at least one recent study investigated a value similar to relative risk called an "odds ratio" in communicating forecasts of the likelihood of freezing temperatures (LeClerk and Joslyn 2012).

Given the lack of published research into relative risk as a communication tool for tornado forecasts, and my own interest in identifying a better way to communicate tornado hazards that occur at unusual places and times, I decided to embark on a comprehensive study of the climatology, communication, and reception of relative risk forecast information for tornadoes. It is important that any study of relative risk as a potential forecast product involve not only the meteorological foundations of the product but also its communication and reception, given that some prior National Weather Service (NWS) messaging efforts have suffered unintended consequences when social outcomes of forecast messaging are overlooked. Examples include the addition of the two additional risk tiers of “Marginal” and “Enhanced” to the SPC outlook in 2014 leading to confusion and complaints from NWS partners (Ernst et al. 2021) and the increased fear of false alarm effects in the aftermath of the Joplin tornado reducing tornado warning lead time (Brooks and Correia 2018). Fortunately, a bounty of recent studies can provide a template for developing new forecast products with both meteorological and social considerations accounted for in all development steps.

Of those studies that exemplify the integration of meteorological and social science insights throughout the process of forecast product development, three deserve extended mention here. First is Krocak’s (2020) dissertation work developing and testing the Potential Severe Timing (PST) product, which visualizes the expected 4-hour period where severe weather is most likely to occur across a forecast area in the US. In her dissertation, Krocak (2020) identified that the SPC convective outlook lacks timing context that users ranging from emergency managers to members of the public are seeking, as the outlook forecasts cumulative severe weather occurrence across the US for 24-hour periods while severe weather events at a given point on average unfold over a 4-hour period. To provide more nuanced severe weather timing information, Krocak (2020) developed the PST, which highlights the 4-hour period within which severe weather would be most likely to occur during the day. The original PST design was then iteratively refined through live testing in the Hazardous Weather Testbed's Spring Forecasting Experiment. Key to the iterative development process was the use of feedback from not only the forecasters issuing PSTs but also from emergency managers and members of the public that would receive PSTs if they became operational (Krocak 2020). Incorporating feedback from potential users during the development of PSTs allowed Krocak (2020) to both develop a visualization of the PST that was easier for all potential users to understand and identify a strong desire in key user groups for the information provided in the product. By blending practical meteorological research with rigorous social science data collection and analysis, Krocak (2020) was able to develop, test, and present a refined version of an information-gap-filling forecast product that was ready for dissemination to NWS partners and the public.

Also worth mention is the still in-progress research work to refine the NWS Weather Prediction Center's (WPC) Winter Storm Severity Index (WSSI), a product that incorporates forecasts for different winter weather hazards with winter weather vulnerability information to forecast an overall winter storm impact level (Semmens et al. 2022). The WSSI, in its original format, was intended to act as a situational awareness tool for forecasters and to communicate potential impacts from winter weather hazards to NWS partners, media, and members of the public (WPC 2020). However, the index was built in a meteorology-focused development process, and some users were dissatisfied with the flaws of the live version of the WSSI (Kastman et al. 2019). To address these flaws, social scientists at the Nurture Nature Center were contracted to collect surveys of the professional userbase that interacts with the index most frequently and recommend changes (Semmens et al. 2022). Semmens et al. (2022) used a series of focus groups with partners to identify weaknesses in the original design of the WSSI, and thus identified that adding headlines with major hazard descriptions, an improved legend with better descriptions of expected impacts for the different levels of the index, and impact timing information in the form of a rolling forecast could help core partners make more informed decisions to prepare for winter storm threats. The improvements to the WSSI identified by Semmens et al. (2022) are now in the process of being implemented by the WPC, and the Nurture Nature Center team is continuing to interview partners and members of the public to ensure that these changes lead to improved reception of the WSSI by users.

One last example of integrated social research in forecast product development is the ongoing work with the Probabilistic Hazard Information (PHI) tornado warning initiative. This product is being developed with feedback from broadcast meteorologists in the Hazardous Weather Testbed, a group not often studied at length in meteorology research despite their position as the primary source of weather information for most members of the public (Ripberger and Silva 2022; Obermeier et al. 2022). As one of the most well-known products to emerge from the National Severe Storms Lab's Forecasting a Continuum of Environmental Threats (FACETs) initiative, the PHI product forecasts the likelihood of severe weather from individual storm cells and is intended to be the next evolution of the traditional tornado warning (Obermeier et al. 2022). To better understand how broadcasters, as key forecast disseminators, would react to being given PHI products, researchers brought groups of broadcasters to the Testbed to simulate wall-to-wall coverage of tornado events using PHI at a mock-up television station (which they named KPHI-TV). Using broadcasters' feedback during and after these simulations, the KPHI-TV researchers learned that broadcasters preferred to present the PHI products to their viewers over legacy products and expected that they could use PHI to justify their coverage decisions to their management, but that the 2-minute update cycle of the probabilistic PHI plumes could overwhelm broadcasters with rapid changes in spatial coverage and affect their ability to communicate tornado risk (Obermeier et al. 2022). The work done by the KPHI-TV team highlights the importance of testing forecast products during their development phase with key message disseminators as well as with decision-makers and the publoc, while also serving as an example of how to integrate user feedback early in the development process to more effectively operationalize new forecast products.

Following the lead of the rigorous experimentation performed by Krocak (2020), Semmens et al. (2022), and Obermeier et al. (2022), I developed a plan to identify whether relative risk forecasts for tornadoes could improve the communication of rare tornado events to the public such that low absolute risks were no longer underestimated when tornadoes are not expected. The research work to answer this broad question came in the form of three separate studies. First, in Chapter 2 of this dissertation, I investigated the climatology of and variation in tornado relative risk that occurs across the 1950-2021 period, using the work previously done by Brooks et al. (2003) and Krocak and Brooks (2018) as a template for developing a daily tornado climatology and thus relative risks. Doing so allowed me to identify the range of values of relative risk that could be expected to occur across different tornado events, as well as how relative risk values compare in frequency and coverage to absolute risk values for tornadoes. In Chapter 3, I analyzed a series of focus group interviews with broadcast meteorologists, where I asked participants to define how they currently use SPC convective outlook probabilities, whether they thought relative risk forecasts for tornadoes would be of value to them, and how they might present those relative risk forecasts to their viewers. Similar to how Obermeier et al. (2022) learned about the potential value broadcasters could generate using PHI products, broadcaster feedback on relative risk products helped me understand whether relative risk products were something broadcasters could see themselves using, and whether relative risk products had flaws that more experienced communicators of forecast information could more readily identify. Finally, in Chapter 4, I visualized data from a series of survey questions asking the public how they would respond when shown relative risk tornado forecast information. These visualizations allowed me to identify how relative risk information changed these participants' perception of risk from tornadoes as compared to absolute risk products. By comparing the results from these three individual studies in Chapter 5, this dissertation examines the production of relative risk information, the dissemination of that information, and the reception of relative risk information by the public, thus developing a holistic understanding of how relative risk forecast products could impact tornado and possibly other rare weather event risk communication in the future.

Chapter 2 – Development of a 1950-2021 Tornado Relative Risk Database

2.1 Introduction

Convective severe weather, including tornadoes, hail, and straight-line winds, presents a significant challenge for forecasting and communicating weather risk to impacted publics. Due to the localized nature of convective severe weather events, the likelihood of any given point being impacted by a storm, even during a thunderstorm outbreak, is very low. However, the impacts of convective severe weather can be significant and lasting where they do occur, with severe storms dealing billions of dollars in damage and claiming dozens of lives in an average year (NCEI 2023). The relative rarity of severe convective weather events and the serious impacts they can have make forecasting and communicating the risks posed by these storms an essential but difficult task (Murphy 1991). To communicate rare but severe convective events, the National Weather Service's Storm Prediction Center (NWS SPC) has developed the convective outlook product, which presents a categorical forecast of severe weather risk levels based on the likelihood probabilities of local thunderstorm wind, hail, and tornado reports within 25 miles of a point (Grams et al. 2014).

Before the development of the convective outlook, weather forecasters attempted to overcome the forecast and risk communication challenges presented by tornadoes through deterministic, "yes or no" forecasts of tornado potential across large areas. The first tornado forecast of any kind issued in the US was J. P. Finley's experimental tornado predictions, first issued in March 1884 (Finley 1884; Galway 1985). Finley's tornado forecasts communicated whether or not conditions favorable for tornadoes were present across 18 districts that Finley drew across the US east of 105° W longitude, although the skill of these forecasts in identifying tornado events was somewhat questionable (Murphy 1996). No major advancements in tornado forecasting occurred after Finley's forecasts until 1948, when Major E. J. Fawbush and Captain R. C. Miller at Tinker Air Force Base issued a deterministic prediction for tornadoes on March 25th of that year (Maddox and Crisp 1999). Fawbush and Miller's tornado forecast, spurred by their base commander after an unwarned tornado hit the base on the 20th, led to the activation of a severe weather plan that signficantly reduced equipment losses when Tinker was again struck by a tornado that evening.

After that fateful 1948 forecast, tornado forecast science began to rapidly evolve, with the Severe Local Storms Warning Service (the predecessor to the modern SPC) beginning to issue a daily categorical "convective outlook" in 1973 (Hitchens and Brooks 2012). This outlook originally highlighted areas of "Moderate" and "High" risk for severe weather (including tornadoes), with a third category, "Slight", added to describe severe weather risks that were less significant than the Moderate category in 1974 (Hitchens and Brooks 2012). Note that the categorical words in the outlook are a type of Words of Estimative Probability (WEPs), which are words that attempt to convey the likelihood of severe weather occurring across the forecast domain (Lenhardt et al. 2020). Thus, the convective outlook can be considered the point at which day-of tornado forecasts began to shift from a deterministic frame to a probabilistic one (as modern tornado watches and warnings, issued in the hours-to-minutes before a tornado occurs, have remained deterministic in nature to the current day).

The categorical outlook first presented purely probabilistic forecasts of tornado likelihood in 2001 (Edwards and Ostby 2022) after Brooks et al. (1998) tied the three subjective categories in the outlook to the likelihood of severe weather occurring within 25 miles of a point. Designing likelihood probabilities to help define the convective outlook was a challenge, however, as although forecasters make forecast judgements based on their probabilistic true beliefs in the outcome of a weather event (Murphy 1985), they also correctly believe that the small forecast probabilities for rare events can be perceived as low enough to ignore by members of the public (Murphy 1991; Kahneman 2011; Shivers-Williams and Klockow 2020). As the probability of severe weather occurring at a given point, even during an outbreak, is incredibly low, the forecasters at the SPC sought to inflate the forecast probabilities in the outlook so that the new product could avoid known biases regarding the interpretation of small numbers. The simplest way to do so, which the SPC chose to use for the probabilistic outlook, is to expand either the valid time or area covered by the forecast (Chaudhry et al. 2018; Brooks et al. 2003).

The effect of adjusting the temporospatial coverage of a probabilistic forecast can be seen when comparing the maximum forecast probability for tornadoes within 25 miles of a point in the SPC probabilistic outlook (60%) to the maximum forecast probability of 100% (presented as 10 on a 10 point scale) in The Weather Channel's TOR:CON index, where tornadoes are forecast for an area within 50 miles of a point (The Weather Channel 2018). Seeking to achieve this inflationary effect on forecast probabilities such that they would be "large enough to feel" but not so large that average forecast values would approach 100%, the SPC experimented with estimating tornado probabilities for 120, 80, and 40km^2 forecast areas (Shivers-Williams and Klockow 2021). Applying those forecast areas to average "High" risk severe weather coverage revealed that the probabilities of severe weather during an outbreak at the 120 km^2 range were generally larger than 60%, while those for the 40km^2 area could be as low as 6% (Shivers-Williams and Klockow 2021). Thus, the SPC settled on the 80 km^2 forecast area (corresponding to a 25-mile radius) for forecasting severe weather and its average "High" risk likelihood probabilities for severe weather of roughly 25% (Shivers-Williams and Klockow 2021). Using the 25-mile forecast range, the SPC linked absolute likelihood values for tornadoes of 2, 5, 10, and 15% to the "Slight" risk, 30% to the "Moderate" risk, and 45 and 60% (without "hatching", which forecast the likelihood of EF2 or greater tornadoes to be greater than 10%) to the "High" risk in the categorical outlook. The absolute likelihoods for "Slight" were further broken out to incorporate the "Marginal" and "Enhanced" categories seen in Fig. 1.1 in 2014 (Grams et al. 2014).

Although the SPC expanded the radius of probabilistic outlook's forecast to the likelihood of severe weather within 25 miles of a point to address forecasters' concerns about communicating small probability values (Shivers-Williams and Klockow 2020), recent work suggests that even these inflated probability values are seen as too low by users. Shivers-Williams and Klockow (2020) found that members of the public are more likely to take action when shown larger likelihoods of a tornado impact, even if they are told that the forecast likelihood is for a broader area (Shivers-Williams and Klockow 2020). Ernst (2020) added that broadcast meteorologists from regions of the US where severe weather is uncommon rarely share the probabilistic outlook on air, as they feel that warning their viewers of a "5% chance of tornadoes" would fail to communicate the significance of such a tornado risk for their areas. Similar results were found in a study of emergency managers by Klockow-McClain et al. (2020), who found that "Slight" risks and their associated probabilities had different meanings for emergency managers in different parts of the US. These emergency managers suggested that the SPC try to find ways to add more context to the outlook by relating probabilistic outlook forecasts to the local climatology and vulnerability information for convective severe weather (Klockow-McClain et al. 2020).

Even before studies of SPC convective outlook use suggested that users sought more context information about the risks forecast in the product, Murphy (1991) suggested that communication of likelihood probabilities for severe weather hazards could be improved through including information about the background likelihood of those hazards. One way to do so is through dividing forecast likelihoods by climatological likelihoods, which would result in an "X times more likely today than average" estimate that could complement probability forecasts of rare and severe events (Murphy 1991). Spiegelhalter (2017) also theorized that the use of this ratio, which he refers to as "relative risks," could help communicate the risks posed by very rare but severe events like earthquakes and volcanic eruptions. Indeed, some risk communication experts have argued that for earthquakes, risk messaging should highlight the large *relative* increase in the *absolute risk* of a major earthquake after a possible foreshock, even though the absolute likelihood of a major earthquake may still be very low (Woo and Marzocchi 2014). Several studies of probabilistic communication in the weather domain have also found that reference class information, including background likelihoods of occurrence, is key to effective probability communication (Gigerenzer et al. 2005; Strathie et al. 2015; Juanchich et al. 2017; Ripberger et al. 2022).

While risk communicators in a variety of hazard domains have adopted or studied relative risk as a communication tool, no study has yet attempted to investigate the distribution of climatology-adjusted risks – which we will call relative risks, as suggested by prior literature (Murphy 1991; LeClerk and Joslyn 2012; Spiegelhalter 2017) – for the hazards forecast in the SPC convective outlook. Here, I develop a method for calculating the relative risk for tornadoes using the SPC tornado report database and Practically Perfect Hindcasts (PPHs) for past days where tornadoes occurred, as a proxy for absolute risk. This method was then used to calculate relative risk values for every day from 1950-2021, which I then analyzed to answer key questions about the distribution of relative risks for tornadoes in the US. First, I sought to better understand the range of relative risk values for tornadoes, by answering what the highest relative risk values in the 1950-2021 period were, when and where they occurred, and how often the highest relative risk values would resolve. Second, I compared the frequency of occurrence of relative risk values to absolute risk values across the 71-year database, so that I could identify a series of values of relative risks I could delineate in areal plots of relative risk values that compared well to the frequency with which the tornado probabilities used in the probabilistic outlook occur. Finally, I selected a small number of case studies of notable tornado outbreaks to subjectively study, and in doing so gain an understanding of how relative risk might “look” in the forecasts for the most significant tornado events.

2.2 Methods

2.2.1 Data

Relative risk is defined as the ratio between the absolute risk of a tornado on a given day, in this case the PPH for tornadoes, by the climatological average risk of tornadoes on that day, which is calculated from the PPH across each day of the year from 1950-2021. Tornado report data used in this study were retrieved from the SPC severe report database. Local NWS offices collect tornado reports and forward them to the SPC, where they are quality controlled and converted to central standard time for database use (SPC 2022). I used the complete 1950-2021 dataset as made available by the SPC during Fall 2022 to create a tornado climatology, as prior tornado climatology studies have used the complete dataset from 1950 to 2021 for their exploration of tornado data (Schaefer and Schneider 2002; Brooks et al. 2003; Verbout et al. 2006; Brooks et al. 2014; Krocak and Brooks 2018). Although errors in the tornado database, including biases due to population, measurement methodology, and database input errors, have been noted by previous studies (Verbout et al. 2006; Potvin et al. 2019; Edwards et al. 2021), the large amount of data used in this analysis and the Gaussian smoothing applied to the data help to minimize the impact of such errors in the resulting daily climatology.

2.2.2 Climatology Development

The development of the daily tornado climatology for this work closely follows the methods presented by Brooks et al. (2003) and Krocak and Brooks (2018). Note that, because I followed the methodology of these previous studies, I did not perform any sensitivity testing on the effects of different Gaussian smoothing parameters or shorter periods of record (e.g. a 30-year climatology instead of a 71-year climatology) on the final product that I used to calculate relative risks. To begin the calculation of the 71-year tornado climatology, the latitude-longitude pairs of the start point of >(E)F1 tornado tracks in the SPC tornado dataset from 1950-2021 were assigned to an 80km × 80km (~50mi) Lambert conformal grid across the contiguous US, centered around 39.8°N latitude (following Krocak and Brooks 2018). Although coarse, the grid dimensions were chosen based on those used to calculate previous tornado climatologies as well as PPH research efforts that have used an 80km × 80km grid to approximate the 25-mile radius used in the probabilistic outlook likelihood forecast (Brooks et al. 2003; Hitchens and Brooks 2012, 2014; Hitchens et al. 2013; Krocak and Brooks 2018, Gensini et al. 2020). Once tornado track start points were matched to the grid, grid boxes with a tornado report for each day in the dataset were given a value of unity (1, or 100%), and grid boxes without a report were set to zero (see fig. 2.1). I chose to only include EF1 and greater tornadoes due to the significant discrepancies in EF0 tornado reports before and after the introduction of WSR-88D radar systems in the 1970s (Potvin et al. 2019, Edwards et al. 2021).

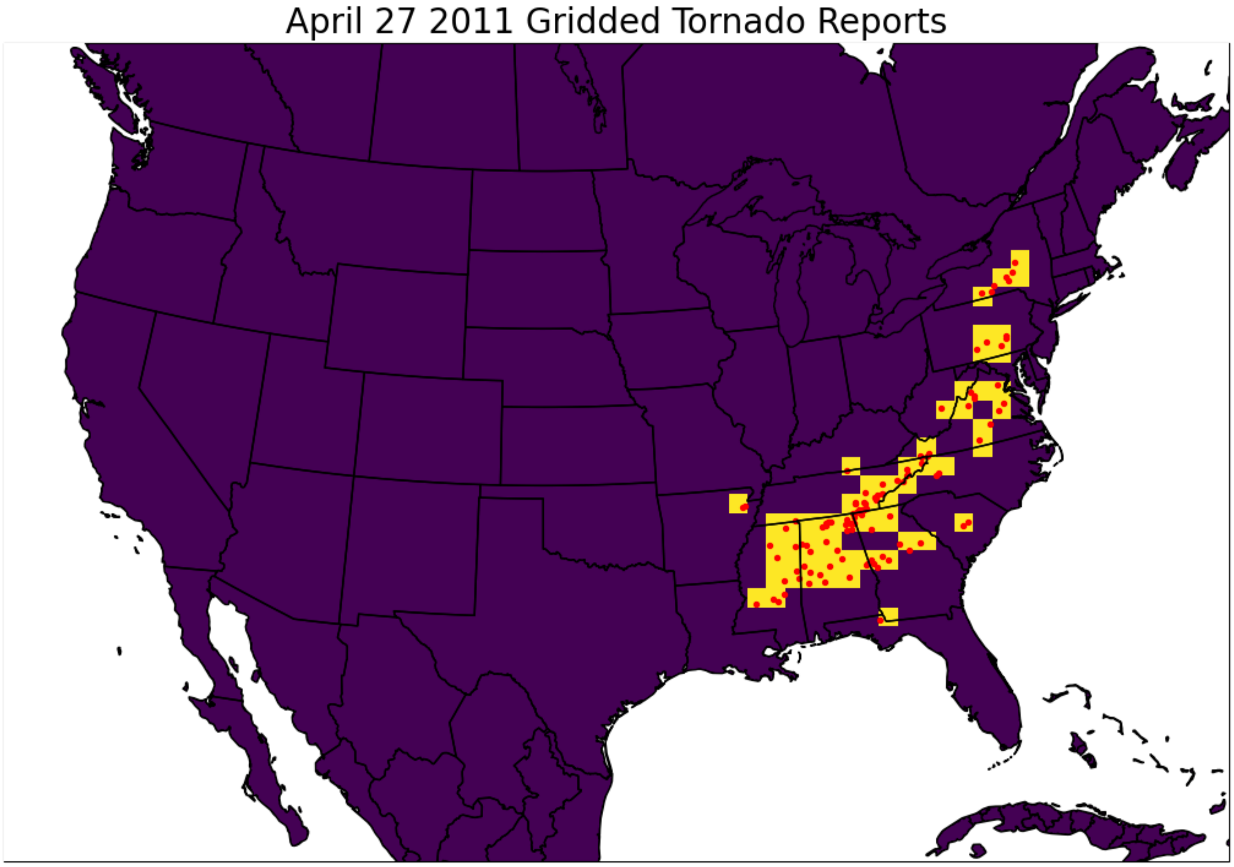


Figure 2.1: Example of gridded tornado reports for the 12z-12z period from April 27th to April 28th, 2011. Red dots represent EF1 or greater tornado path start points, while yellow grid squares represent grid spaces given a value of unity due to the presence of a tornado path start point within their bounds. Purple grid spaces have a value of 0, as no tornado start points were identified within those grid spaces.

Once the tornado report grids were prepared, I applied a two-dimensional Gaussian filter across space and a one-dimensional filter across time to generate more consistent probability fields for the daily tornado climatology (Brooks et al. 2003). As tornadoes are rare events that can vary greatly in occurrence from day to day and year to year, this smoothing ensures that resulting daily tornado likelihoods are not overly influenced by extreme outlier events (such as the April 3rd, 1974 Super Outbreak, among others). The spatial Gaussian filter applied to the data took the form of the equation:

$$P = \sum_{n=1}^{N} \frac{1}{2\pi\sigma^2} e^{-d^2/_{2\sigma^2}}$$

where P is the spatially smoothed probability, N is the total number of grid boxes with a value of unity, d is the distance from the hindcast grid point to the location of the report, and σ is the standard deviation of the Gaussian distribution, also called the smoothing parameter. This study used a smoothing parameter of 120km, or 1.5 grid spaces, as is used in previous tornado climatology development studies (Brooks et al. 2003, Hitchens et al. 2014; Krocak and Brooks 2018). The temporal Gaussian filter took the form of the equation:

$$P = \sum_{n=1}^{N} \frac{1}{2\pi\sigma^2} e^{-t^2/_{2\sigma^2}}$$

where all values are identical to the spatial smoothing equation, save for *t* describing the number of days distant a value is from the day that is being smoothed. The smoothing parameter used for time was 15 days, again based on previous tornado climatology calculations (Brooks et al. 2003, Krocak and Brooks 2018). Note that to calculate the temporal smoothing values in the first and last days of the year, I copied the gridded data for each year twice over to create a “wrapped” year with 366 × 3 total days of gridded data (in other words, repeating each year three times). As each year of data is smoothed individually, “wrapping” the data by extending each year ensures that the boundary conditions at the ends of the dataset do not introduce non-physical values to the smoothed dataset. An example of a single tornado day after smoothing can be seen in Figure 2.2.

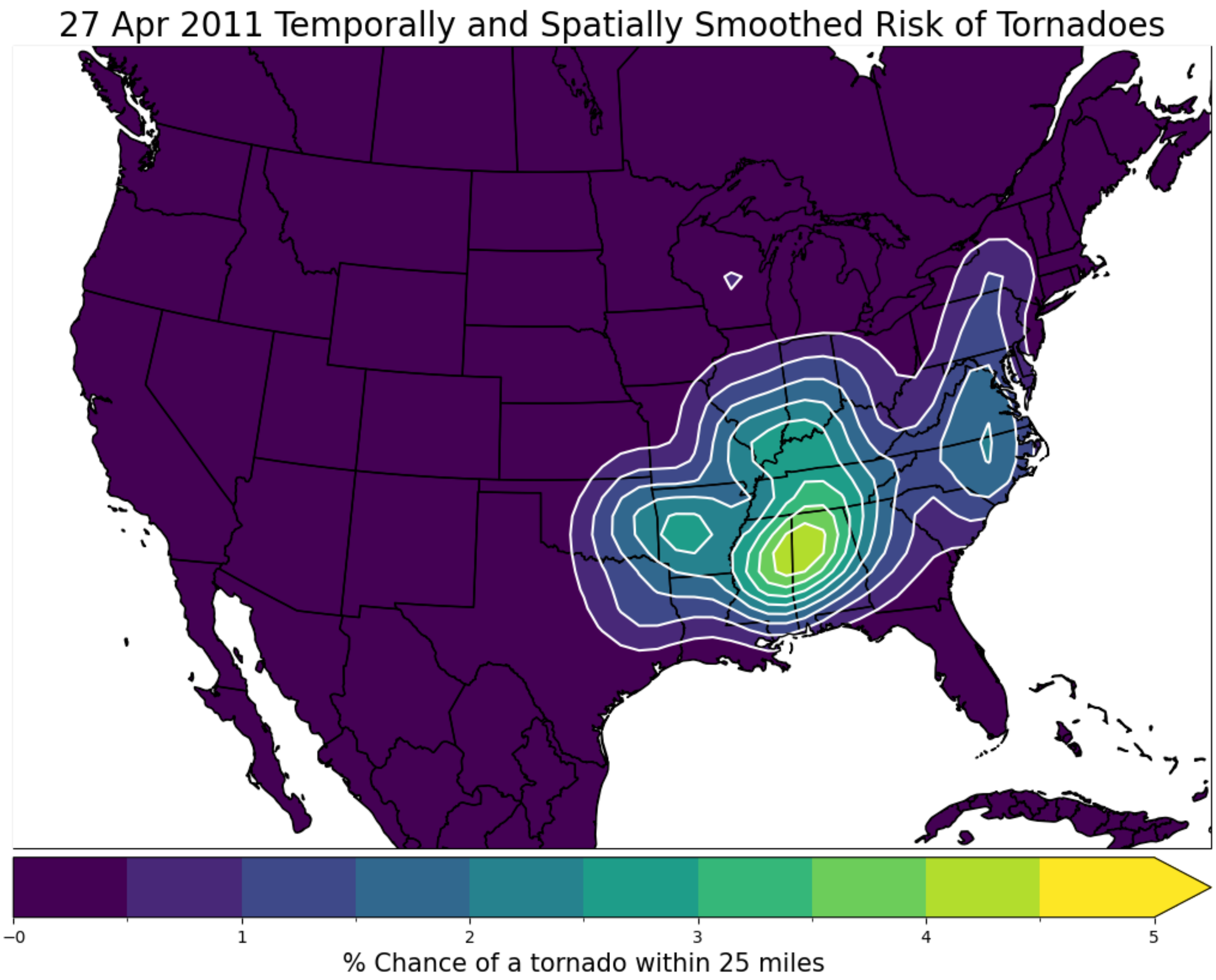


Figure 2.2: Example of gridded tornado reports for the 12z-12z period from April 27th to April 28th, 2011, after spatial and temporal Gaussian smoothing. Note that the maximum probability of a tornado on the legend has been reduced to 4.2% from unity for each grid square with a tornado by the smoothing process.

After smoothing was applied, the middle 366 days of gridded data were removed from each "wrapped" triplet of years, to be used to calculate the average across all smoothed days of data. Note that an additional day was included in each year to represent the leap day of the 29th of February. I then calculated the daily mean of the smoothed values of tornado likelihood across all years of the dataset to complete the daily tornado climatology (for an example of the final product see fig. 2.3).

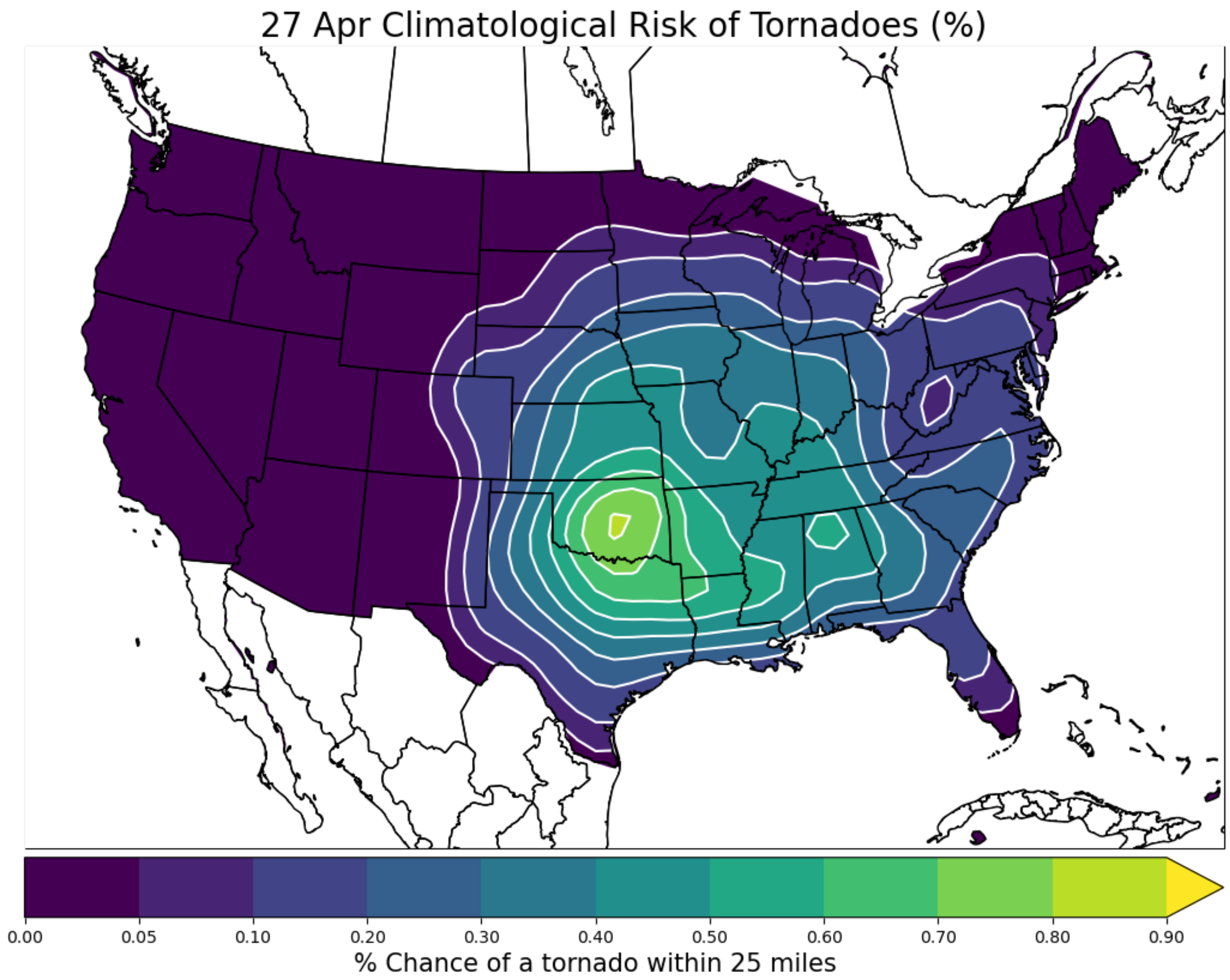


Figure 2.3: Example of the final climatological risk of a tornado for the 12z-12z period from April 27th to April 28th. Note that the values in the legend have changed again and now display tenths to hundredths of a percent likelihood of a tornado within 25 miles of a point on a given April 27th.

2.2.3 Calculation of Daily Relative Risks

The next step after developing the daily tornado climatology was to create SPC outlook probabilities for every day in the SPC tornado report dataset. Here I used PPHs (Practically Perfect Hindcasts) for each day's absolute risk instead of previously issued probabilistic outlook forecasts, because use of PPH allowed me to investigate and compare tornado events from before the SPC probabilistic outlook was first issued 2001 in addition to those after the probabilistic outlook went live. PPHs were first described in the severe weather forecast context by Brooks et al. (1998) as a method for calculating "a forecast that is consistent with that which a forecaster would make given perfect knowledge of the reported events beforehand and the operational constraints associated with the forecasting system" (Hitchens et al. 2013). Using the SPC tornado report data, I can calculate PPH values for every day across the entire 1950-2021 period, allowing us to compare tornado events before and after 2001 without any influence of forecaster biases. Following the previously established process used to calculate PPHs for tornado reports (Hitchens et al. 2013; Gensini et al. 2020), I applied spatial Gaussian smoothing to the previously gridded tornado start point data with a 1.5 (120km) smoothing parameter. The 1.5 smoothing parameter was chosen to create the PPHs so that the output would match the "operational constraints of the system", as the SPC probabilistic outlook is constrained to forecasting tornadoes within 25 miles of a point (Hitchens et al. 2013). As such, a grid square with a tornado start point (and thus a tornado probability value of unity) is smoothed to a maximum value of 0.707 (7.07%) centered at the centerpoint of the grid square where the tornado occurred, decreasing to 0.233 (2.33%) at surrounding grid points. To further mimic the true SPC probabilistic outlook, the PPH probabilities were masked such that only grid point values greater than 2% were recorded, as 2% is the lowest forecast likelihood for tornadoes included in the outlook (see fig. 2.4).

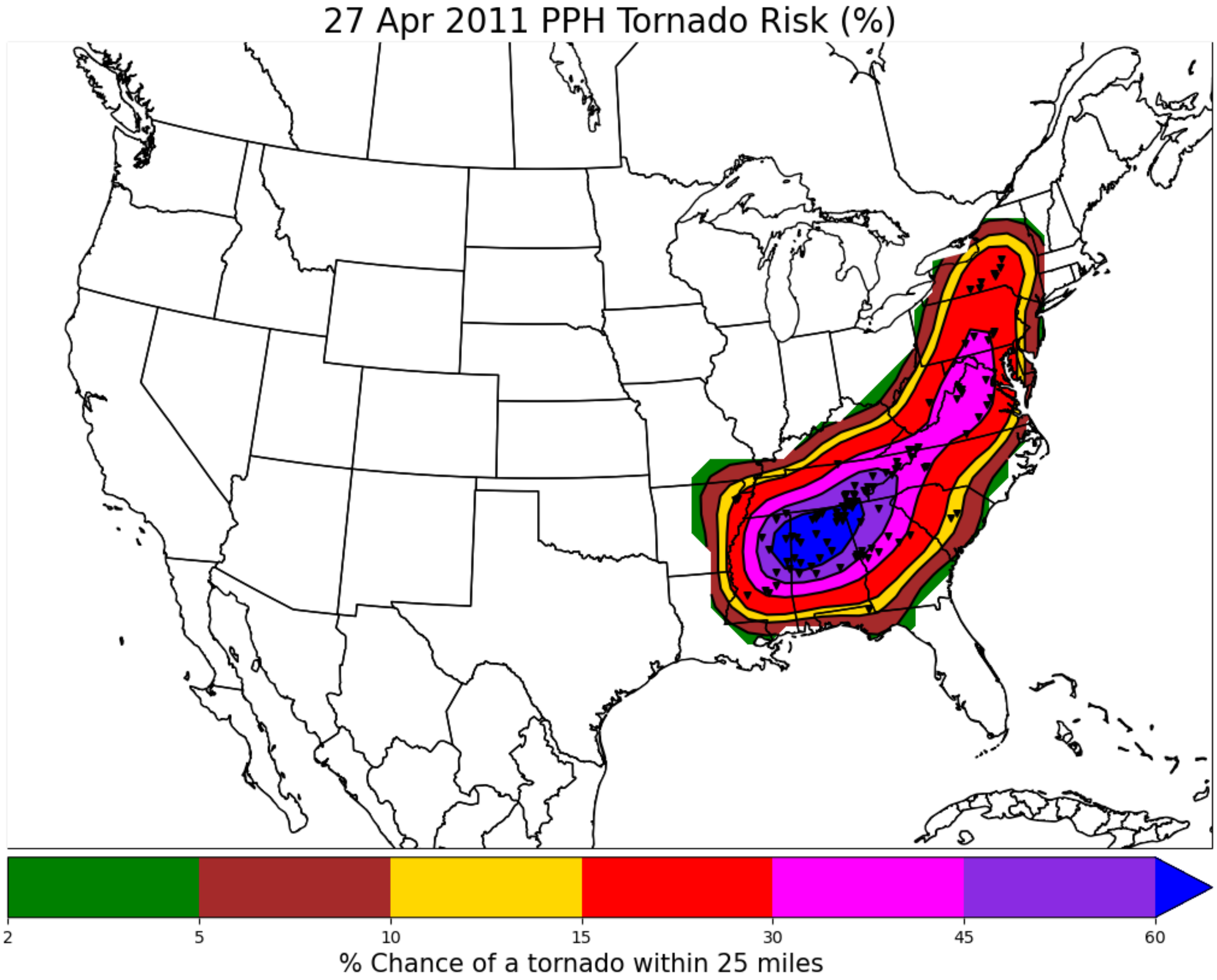


Figure 2.4: Spatially smoothed gridded tornado start point data for the 12z-12z period from April 27th to April 28th. The color table and values in the legend match those in the SPC probabilistic outlook for tornadoes, to highlight how this Practically Perfect Hindcast (PPH) represents a "perfect" version of the SPC probabilistic outlook issued on this day.

Once PPH was calculated for each day from 1950 to the end of 2021, I calculated relative risks for each day in the 71-year dataset using the following formula:

$$Relative\ Risk = \frac{Absolute\ Risk}{Climatological\ Risk}$$

where the PPH for each tornado day was used to supply the *absolute risk*. Once calculated, the *relative risk* values at grid points outside the borders of the contiguous US were masked to remove spurious values that result from the lack of tornado report data outside of the borders of the US. I then created ranked lists of tornado events by peak relative risk value and area above a given relative risk value, as well as plots detailing the maximum relative risk values across the contiguous US, and the distribution of relative risk and absolute risk pairs for all grid points on all days in the dataset.

Before discussing the distribution of relative risks across the US, I want to discuss a few visual examples of the relative risk calculation process and show what relative risk looks like when plotted on a map. To this end I have created three-panel plots containing the PPH absolute risk, climatological risk, and the relative risk for June 1st, 2011 (fig. 2.5) and May 20th, 2013 (fig. 2.6). The June 1st, 2011 event represents the “high likelihood, low climatology” type of event that relative risk is intended to highlight (and is the tornado event that helped inspire this research), where four EF1+ tornadoes across New England resulted in absolute risk maxima in excess of 10% in areas with underlying climatological risks of 0.01 to 0.05% (NCEI 2011, see fig. 2.5). The overlap results in relative risk values across the region in excess of 100 times more likely than normal (henceforth referred to as “x” e.g., 100x) and as high as 250x across Maine. In contrast, the May 20th, 2013, event represents a “high likelihood, high climatology” type of event, where low values of relative risk can be expected even though there is an increased absolute risk of tornadoes. Based on the coverage of tornadoes on that day, a PPH maxima for EF1+ tornadoes of 15% covered parts of Missouri and Oklahoma, including Moore, Oklahoma, where the most recent (as of publication) EF-5 tornado in the US occurred on this day (NWS 2013). The climatological risk of EF1+ tornadoes is much higher in central Oklahoma in May as compared to New England in June, reaching values as high as 0.8%, which on this day leads to a more modest relative risk of 5 to 25x across the region (fig. 2.6).

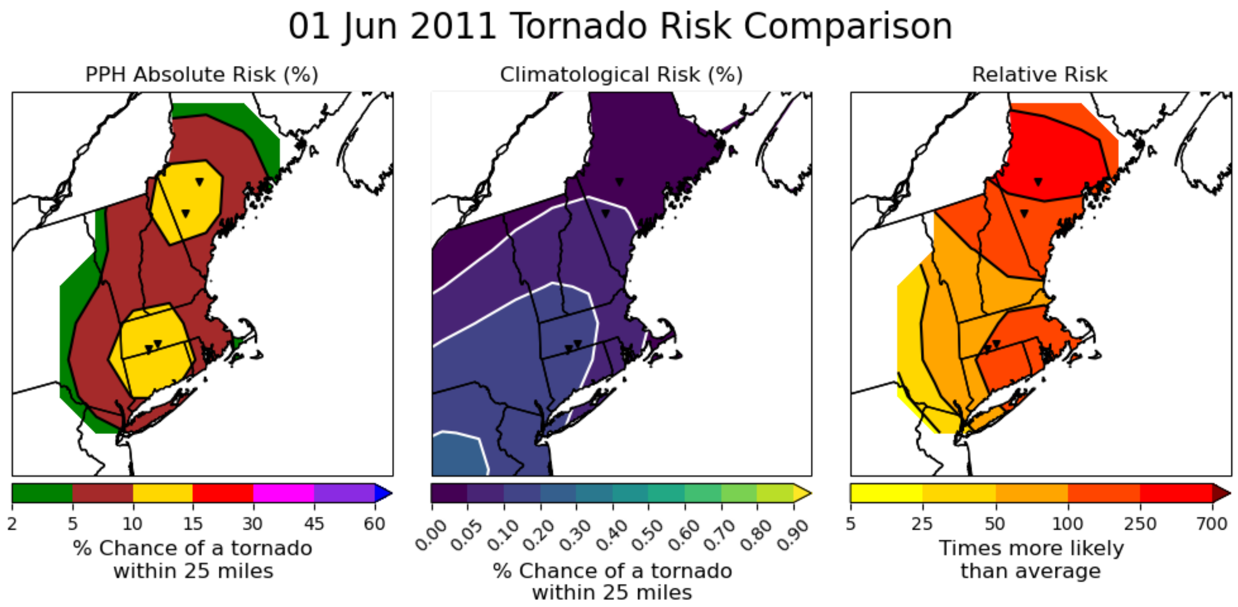


Figure 2.5: Practically Perfect Hindcast (PPH, left), climatological (middle), and relative (right) tornado risk for June 1st, 2011. Black triangles on all three plots represent the start point of individual tornado tracks.

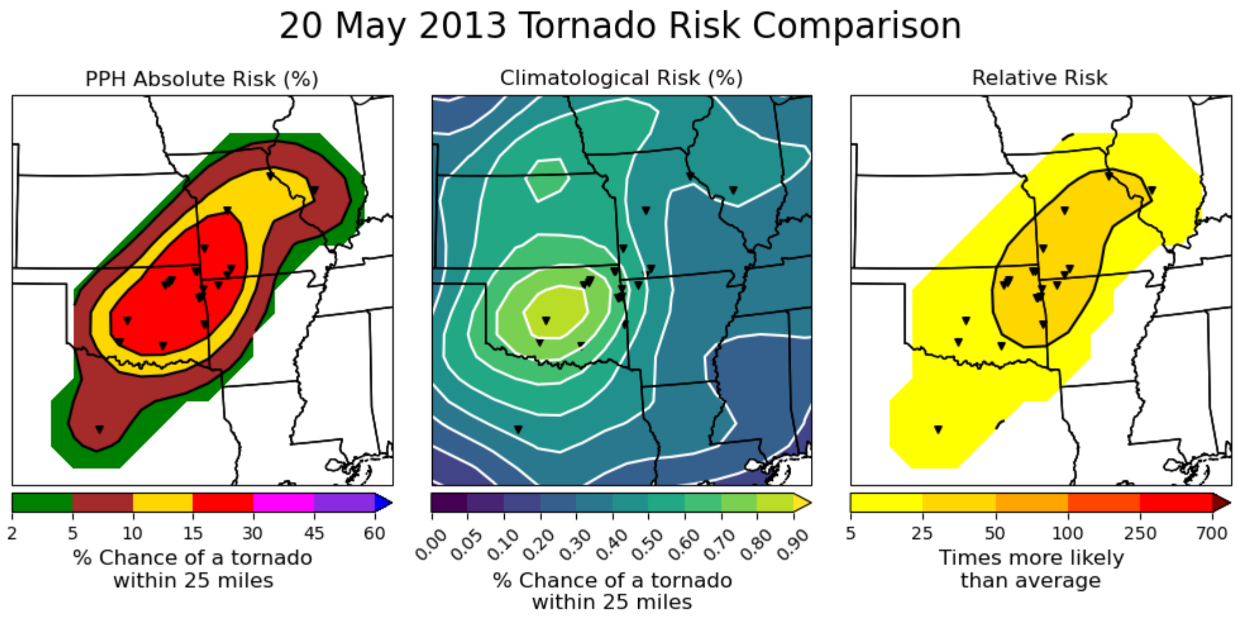


Figure 2.6: Practically Perfect Hindcast (PPH, left), climatological (middle), and relative (right) tornado risk for May 20th, 2013. Black triangles on all three plots represent the start point of individual tornado tracks.

2.3 Results

Although visualizing relative risk for individual tornado events can suggest some patterns in how relative risk responds to changes in absolute and climatological tornado likelihoods, more granular data analysis methods are needed to find stronger evidence for patterns in relative risk across the full 1950-2021 dataset. As a first step in this broader analysis, I used a 2-dimensional histogram to compare the distribution of relative risk values across the entire dataset with the absolute risk values observed at each corresponding grid point and day (see fig. 2.7). Relative risk values in this plot are charted on a logarithmically scaled x-axis due to the wide range of values observed (as low as 2, and as high as 2800), while individual bin counts are also plotted logarithmically due to the large number of data points in the distribution. This plot reveals that relative risk values of about 600 times more likely than normal are rare, save for when absolute risk values are lower than 7%. Further, relative risks are limited by climatological risk values on the left side of the plot – the highest climatological risk that occurs in the dataset is 0.91%, meaning that relative risks will always be a larger value than the absolute risk they are being calculated from (as PPH absolute risk values below 2% are masked) – or to the right of the red line in Figure 2.3, which represents the function $y = x$.

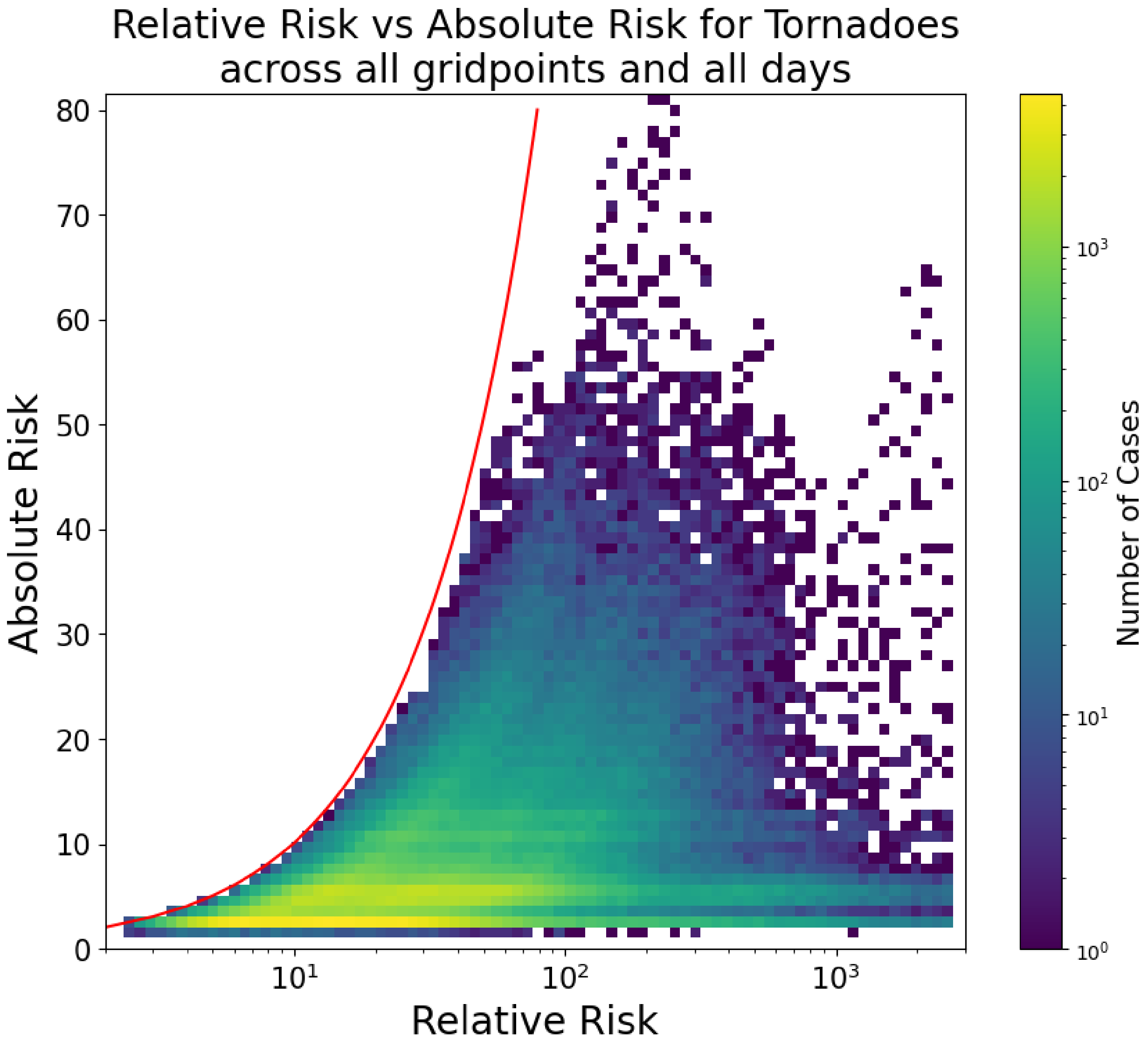


Figure 2.7: 2-dimensional histogram plot of relative risk vs absolute risk pairs for each gridpoint for each day from 1950-2021. The x-axis, which displays relative risk, has been scaled logarithmically, as have the counts for each kernel in the plot (as seen in the colorbar). The red line represents the function $y = x$.

The values of relative risk over 700x identified in Figure 2.3 occur most frequently in the northern and western regions of the country, while regions of the southeast US rarely see relative risk values in excess of 500x (see fig. 2.8). Lower maximum values of relative risk can be found across parts of western Louisiana and far eastern Texas, near the Gulf of Mexico. This distribution is a nearly perfect mirror of the minimum daily climatology for tornadoes across the contiguous US, which has its highest values in southwest Louisiana and east Texas (in other words, tornadoes occur throughout the year along the Gulf Coast regions, see fig. 2.9). Values of minimum daily tornado likelihood in excess of 0.04% extend across the South and lower Midwest regions, but the same regions of the Northern and Western US that see maximum relative risk values in excess of 1000x also see minimum daily tornado likelihoods lower than 0.01%. This suggests that the climatological risk for tornadoes drives the range of relative risk values that occur, and that relative risk values may have less value for users in areas with extreme low climatological risks.

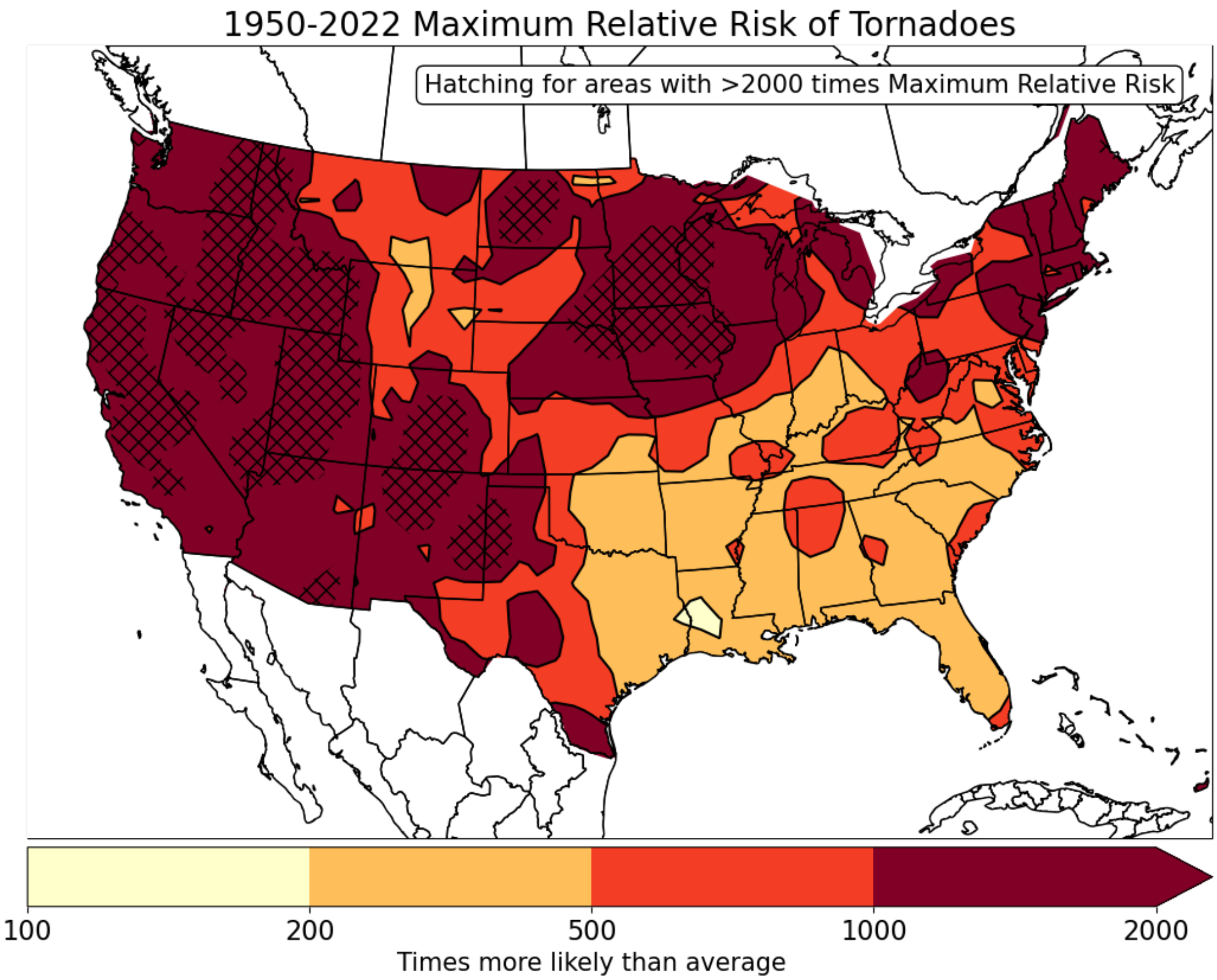


Figure 2.8: The maximum value of relative risk that occurs across the entire SPC tornado dataset from 1950-2021. Note that areas with crosshatching have maximum relative risk values in excess of 2000x.

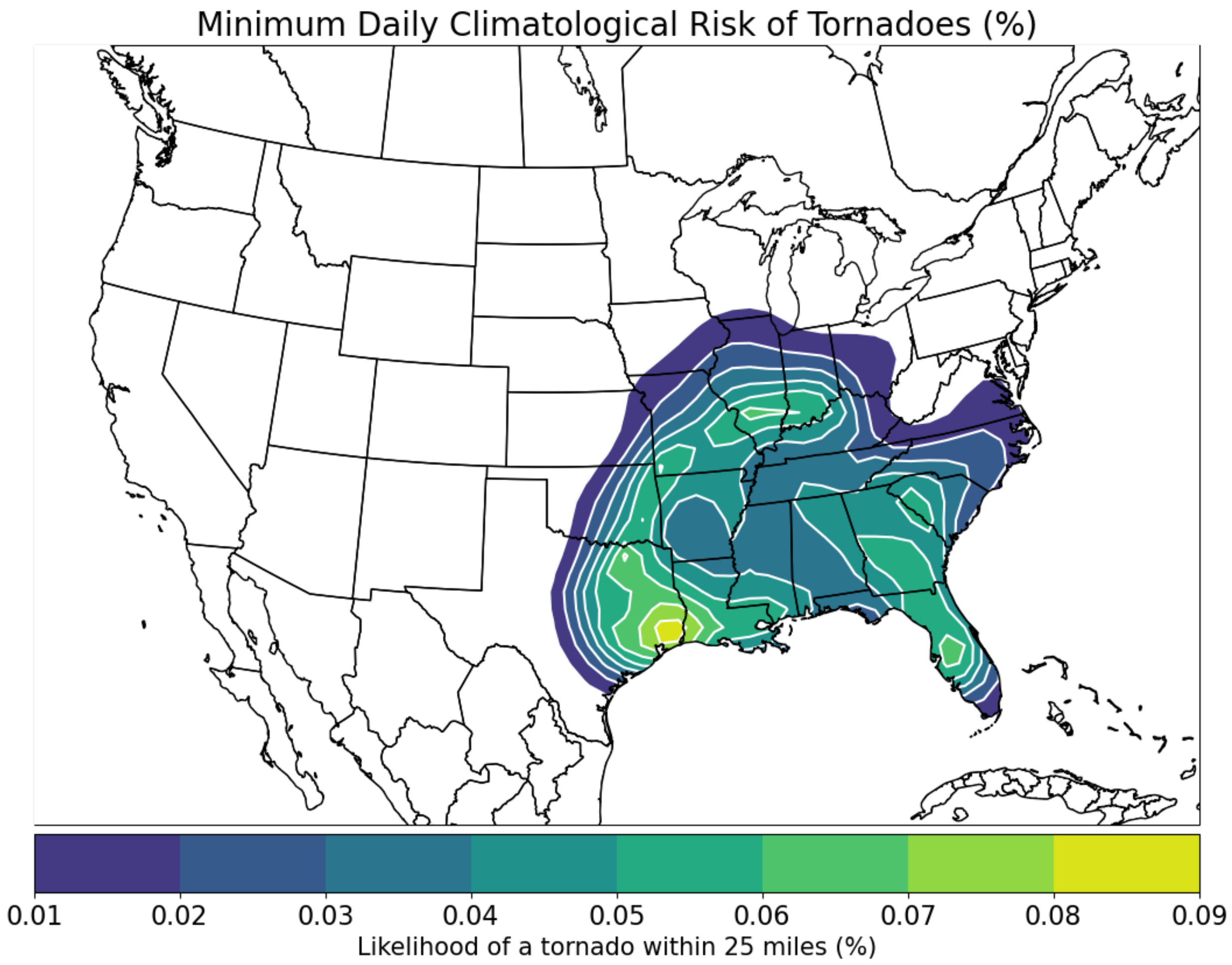


Figure 2.9: Display of the minimum daily likelihood of a tornado within 25 miles of a point across the contiguous US, in % values. Regions within the US with minimum likelihoods below 0.01% are not shaded in this plot.

The impact of minimal climatology events is most clearly seen when viewing a table of the 15 days in the 1950-2021 database with the highest relative risk values. Only one day in the top 15 featured more than 10 tornadoes stronger than (E)F1 (see table 2.1), specifically the tornado outbreak that occurred with a convective line of storms across Iowa on December 15th, 2021. Additionally, most of the top 15 days occur in the fall and winter months, the seasonal minimum for tornado reports across the US. The day with the highest calculated relative risk across the database is December 26th, 1966, where a single tornado rated stronger than (E)F1 occurred in Northern New Mexico, resulting in a relative risk value of 2697x (see fig. 2.10).

Notably, although the Gaussian smoother used to calculate PPH absolute risk smooths grid points with tornado reports to a value of 7.07% chance of a tornado within 25 miles, the extreme relative risk value that occurred on this day was at a grid point away from the tornado report where the absolute risk was smoothed to 2.33% (see table 2.1). Similar extreme relative risk values also occurred on October 17th, 1983, and December 2nd, 1970, where a PPH absolute risk value of 2.33, smoothed from a single tornado, lead to an extreme relative risk value. Tornadoes like these are effectively the entire tornado climatology for where and when that tornado occurred, given that all tornado reports from 1950-2021 are included in the climatology calculation. Such rare tornado events result in a large disparity between a low absolute tornado risk and a very high relative tornado risk, although it is possible that presenting both risk communication formats together may help users interpret their risk better than either format on its own. It also suggests that relative risks may be difficult for individuals to calibrate themselves to, as although absolute risks can only range from 0-100%, relative risks theoretically have no upper limit, as the ratio that defines them can approach infinity as the climatological risk of tornadoes approaches zero.

Table 2.1: Ranking of the 15 days with the highest relative risk at any grid point across the contiguous US. The ranking, date, maximum relative risk grid point value on that date, the Absolute Risk at that grid point, and the national (E)F1+ tornado count on that date are included in the table by column. There are 8922 days with a relative risk greater than zero in total across the 1950-2021 dataset.

RANKING	DATE (M/D/Y)	MAXIMUM RELATIVE RISK VALUE	ABSOLUTE RISK VALUE AT RELATIVE RISK MAXIMUM	NATIONAL 12Z-12Z (E)F1+ TORNADO COUNT
1	12/26/1966	2697.01	2.33	1
2	1/16/2000	2696.91	2.91	1
3	10/17/1983	2691.71	2.33	1

4	10/26/1984	2686.37	7.07	1
5	12/27/1983	2645.01	7.07	2
6	12/15/2021	2617.6	18.78	91
7	9/5/2020	2594.2	3.1	2
8	12/2/1970	2571.16	2.33	1
9	12/27/2019	2541.42	5.66	1
10	2/13/1954	2521.2	4.54	1
11	11/3/1973	2460.66	2.91	1
12	9/24/1986	2459.09	2.33	5
13	11/2/1967	2379.97	7.07	1
14	3/20/1984	2378.52	4.54	4
15	3/8/1960	2373.3	4.54	1

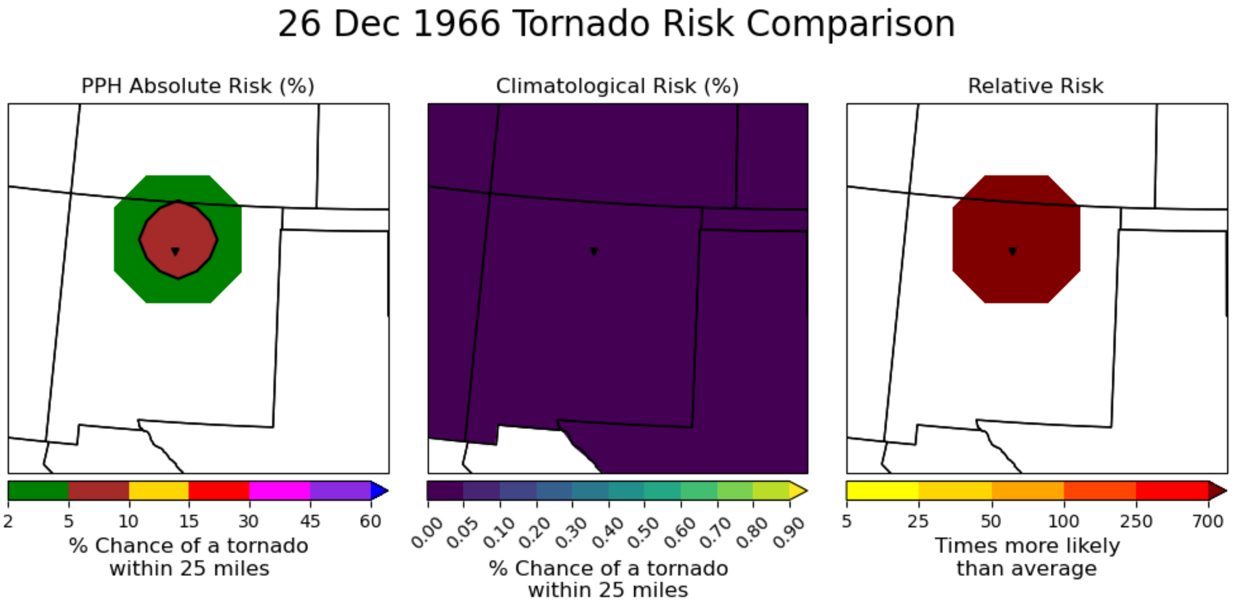


Figure 2.10: Absolute risk (left), climatological risk (left), and relative risk (right) for the 26th of December 1966. Note the lack of contour values in the climatological and relative risk plots, as at no point plotted does the climatological risk have a value over 0.05%, and there are no relative risk values below 700x. The black triangle represents the start point of the tornado recorded on that day.

The impact of regional climatology minima on the list of the highest observed relative risk values across the Contiguous US can be somewhat mitigated by removing grid points from the US West from consideration in the overall distribution of relative risk. Compared to the top 15 relative risk days for the entire Contiguous US, only ranking grid points to the East of 105W (roughly the longitude of the Continental Divide) results in five events with more than 10 EF1 or greater tornadoes (see table 2.2). The maximum value of relative risks in this distribution are also less than the Contiguous US distribution, with December 15th, 2021, moving from the 6th highest to the highest relative risk day, and relative risk values near 1500x for the 15th highest risk vs 2400x for the Contiguous US distribution.

Table 2.2: Ranking of the 15 days with the highest relative risk for grid points east of 105W longitude in the contiguous US. The ranking, date, maximum relative risk grid point value on that date, the Absolute Risk at that grid point, and the national (E)F1+ tornado count on that date are included in the table by column. There are 8217 days with a relative risk greater than zero in total across the 1950-2021 dataset east of 105W, meaning there are 705 days with a relative risk greater than zero west of 105W over the same period.

RANKING	DATE (M/D/Y)	MAXIMUM RELATIVE RISK VALUE	ABSOLUTE RISK VALUE AT RELATIVE RISK MAXIMUM	NATIONAL 12Z-12Z EF1+ TORNADO COUNT
1	12/15/2021	2617.6	18.78	91
2	1/24/1967	2204.05	32.3	27
3	11/1/2000	2203.36	7.08	3
4	11/1/1971	2066.55	8	5
5	1/7/2008	1978.11	4.63	25
6	3/26/2021	1959.1	5.66	1
7	12/1/1970	1941.23	3.52	4
8	1/26/1950	1881.2	5.66	1
9	1/7/1992	1822.98	3.52	6
10	10/18/2007	1774.3	18.34	34
11	10/26/1996	1755.77	14.95	18
12	1/17/1952	1610.08	2.33	1
13	1/14/1992	1543.45	7.44	2
14	10/12/1971	1537.47	7.07	1
15	4/14/1976	1512.53	7.07	6

This change to the distribution of highest end relative risks is clearest when comparing those regionally restricted values to the distribution of absolute and relative risks for the area east of 105W. The distribution of absolute risks across the region east of the Continental Divide has many more examples of risks over 40%, with there being no absolute risks west of the Divide greater than roughly 40% (see fig. 2.11a). The opposite is true for the distribution of relative risks between these two sections of the US, as the distribution is more skewed left for the region east of 105W (see fig 2.11b). The distribution of relative risks across the region west of the Continental Divide has many more examples of risks in excess of 1000x, due to the regional climatology minima there.

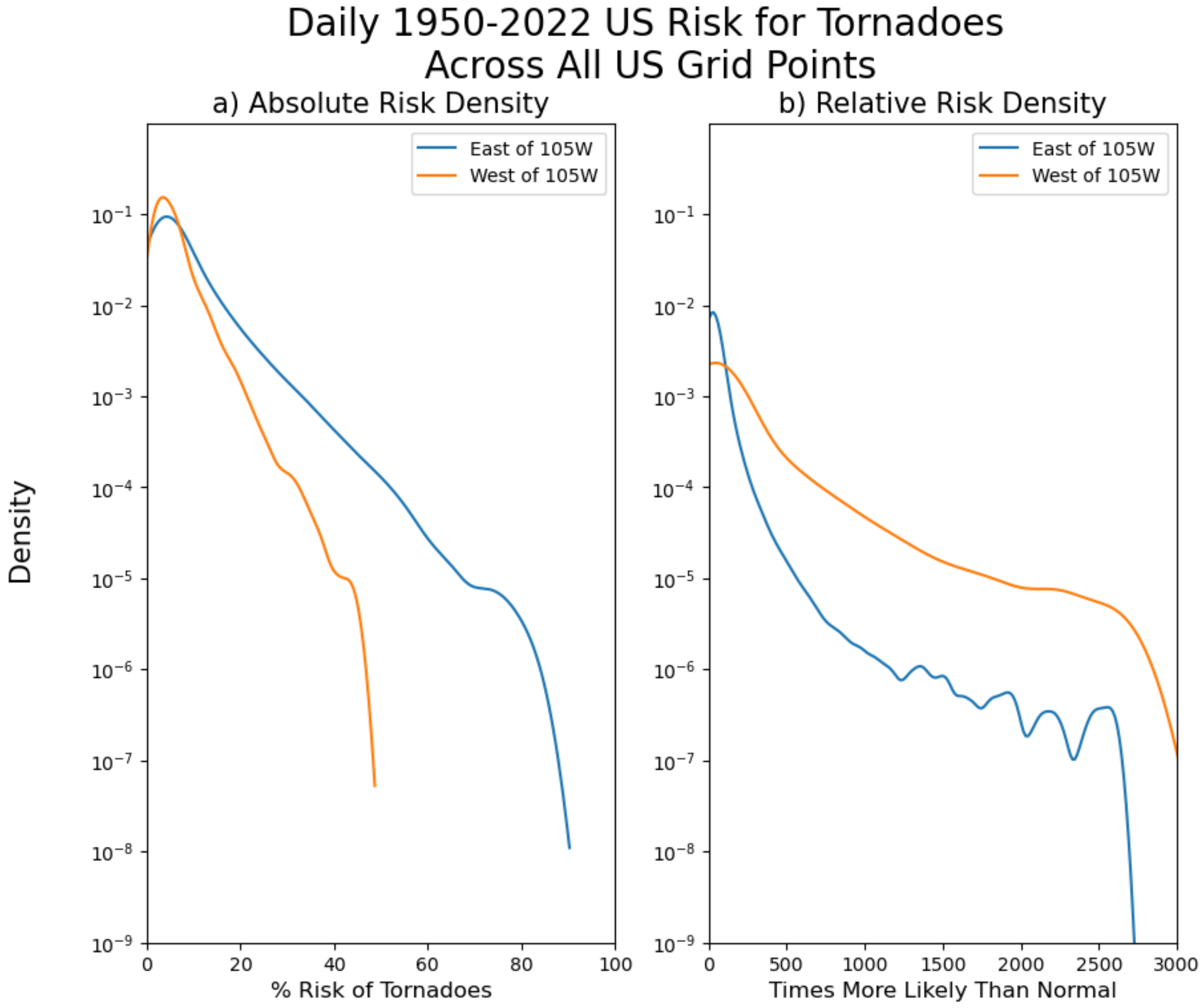


Figure 2.11: Distribution of absolute and relative risk values across all contiguous US grid points in the regions East and West of 105W longitude. The distribution of absolute risk values for each

region are displayed in the left plot (a), while relative risk distributions are shown in the right plot (b). Note that the *y*-axis of both plots is logarithmic.

Although the difference in the risk distributions can be visually described with distribution density plots, the percentiles of each distribution further reveal how removing the grid points west of the Continental Divide reduces the frequency of extreme relative risk values. There is little difference between the percentiles of the distributions for the Contiguous US and the US east of 105W below the 75th percentile of the distribution (see table 2.3). More significant decreases in the value of the relative risk distributions from the Contiguous US to the US east of 105W occur for the 90th and especially the 99th percentile, where the 90th percentile for the US east of 105W distribution is 6.6% lower, and the 99th percentile is 51.1% lower than the Contiguous US 99th percentile value. As observed in the distribution plots, removal of grid points west of the Continental Divide reduces the frequency of extreme high values of relative risk in the dataset.

Table 2.3: Table of percentile values for the Contiguous US and East of 105W relative risk distributions.

PERCENTILE	CONTIGUOUS US RELATIVE RISK DISTRIBUTION (X MORE LIKELY THAN NORMAL)	US EAST OF 105W RELATIVE RISK DISTRIBUTION (X MORE LIKELY THAN NORMAL)
1%	4.21	4.11
10%	7.83	7.65
25%	13.36	13.06
50%	25.85	25.09
75%	52.19	48.75
90%	108.48	90.24
99%	621.97	304.38

Beyond comparing the distributions of relative risk for the Contiguous US and the US east of 105W longitude, percentiles of the risk distributions can also be used to compare the frequency of occurrence of absolute risk values and relative risk values. Table 2.4 displays a list of the absolute risk values used in the SPC probabilistic outlook for tornadoes, with the percentile of the absolute distribution lower than that absolute risk value for both US regions. The value of relative risk for each of the two distributions at that percentile value (e.g., relative risk at the 54.56th percentile in the Contiguous US distribution is equal to 28.95) is also displayed, representing the relative risk value that occurs at the same point in its distribution as the absolute risk value in the SPC forecast. First, note that as grid point values with a less than 2% absolute risk value in the practically perfect hindcast analysis were left as non-existent values, the 2% absolute risk value represents the 0th percentile of the absolute risk distribution. Higher levels of absolute risk are very rare in the distribution, with the 30% absolute risk representing the 99th percentile of the distribution, while the 60% absolute risk value represents the 99.98th percentile of the distribution. These values are nearly identical for absolute risks east of 105W longitude, but the relative risk values associated with those percentiles vary greatly across the two regions. The 60% absolute risk percentile correlates to a relative risk of 2562.59x for the Contiguous US distribution, which decreases to 1924.14 for the distribution east of 105W longitude (see table 2.4). Notable decreases in the value of relative risk matched to absolute risk values across the two distributions also occur at the 15, 30, and 45% absolute risk values.

Table 2.4: Table that displays the percentile of the distribution for the different absolute risk thresholds used in the SPC convective outlook, as well as the relative risk that occurs at that percentile when it is applied to the relative risk distribution. Values for both the Contiguous US and US east of 105W distributions are calculated.

	CONTIGUOUS US DISTRIBUTION		EAST OF 105W (CONTINENTAL DIVIDE)	
ABSOLUTE RISK VALUE	**% of Absolute Risk Distribution**	**Relative Risk Value at that Percentile (X**	**% of Absolute Risk Distribution**	**Relative Risk Value at that Percentile (X**

	lower than that value (%)	more likely than normal)	lower than that value (%)	more likely than normal)
2%	0.0	2.35	0.0	2.35
5%	54.56	28.95	52.74	26.76
10%	84.90	78.79	83.35	65.19
15%	93.36	149.50	92.44	106.98
30%	99.05	644.11	98.88	287.96
45%	99.86	1749.20	99.83	713.22
60%	99.98	2562.59	99.98	1924.14

Beyond ranking and comparing tornado events by their absolute and relative risk values, and whether those events occur east or west of the Continental Divide, the significance of tornado events can be compared in a relative risk framework by comparing the size of the area with a relative risk value in excess of a chosen threshold across events. Table 2.5 displays a list of the 15 event days with the largest area of relative risk values in excess of 100 times greater than normal, the value of which was chosen due to its proximity to the 90th percentile of relative risk values (see table 2.4). Unlike Tables 1 and 2, which reveal that the majority of tornado events with the highest relative risk values have fewer than 10 recorded (E)F1+ tornadoes, all 15 of the events with the largest area of greater than 100x relative risk have (E)F1+ tornado counts greater than 10, and two events (4/27/2011 and 4/3/1974) with over 100 (E)F1+ tornadoes. This list of dates includes many tornado outbreaks that may be familiar to readers that follow tornado history, including the Widespread Outbreak (rank 1), the two Super Outbreaks (rank 5 and 10), the 2021 Mayfield (rank 4) and Midwest tornado outbreaks (rank 11), the 1990 Central US outbreak that included the Hesston F5 tornado (rank 8), and the 2002 Veteran’s Day tornado outbreak (rank 9) among others. This list suggests that significant tornado outbreak events may be best captured by the area of abnormally high tornado risk that they present, more-so than the highest values of relative risk they record.

Table 2.5: Table ranking the top 15 days with the largest number of grid points with a relative risk value greater than 100 times more likely than normal. The rank, date, number of grid points with a relative risk greater than 100x, the area covered by relative risk greater than 100x, and the EF1+ tornado count across the US on that date are all recorded.

RANKING	DATE (M/D/Y)	AREA COVERED BY >100X RISK (10^3 KM^2)	NATIONAL 12Z-12Z (E)F1+ TORNADO COUNT
1	11/22/1992	928.0	43
2	12/23/2015	844.8	24
3	1/10/2020	755.2	36
4	12/10/2021	748.8	51
5	4/27/2011	716.8	115
6	10/26/2010	704.0	26
7	10/17/1971	697.6	21
8	3/13/1990	697.6	49
9	11/10/2002	691.2	51
10	4/3/1974	684.8	133
11	12/15/2021	684.8	91
12	3/6/2017	678.4	38
13	1/17/1958	672.0	28
14	1/29/2013	665.6	33
15	11/15/1988	659.2	36

Given its prominence in the list of tornado days with the largest areal extent of 100x risks and relative recency, December 10th, 2021 deserves greater discussion and analysis through visualizations of the three major risk formats. On that day, an outbreak of tornadic supercells moved across the Midwest US, resulting in a large 30% PPH absolute risk area covering in Tennessee and Kentucky (see fig. 2.12). Climatological risks in this area are seasonally maximized in the spring, but at this time of year are very similar to those in New England in the summer, as seen in figure 2.5. This results in a broad area of 100x relative risk, with nearly the entirety of Kentucky covered by a relative risk of over 250x. Of note, almost every EF1+ tornado that occurred during this event began its track within the 100x area. Off-season events like this one highlight the potential value of relative risk information to weather risk communicators, who may have been able to use it to highlight the unusual nature of this significant tornado outbreak to those in its path.

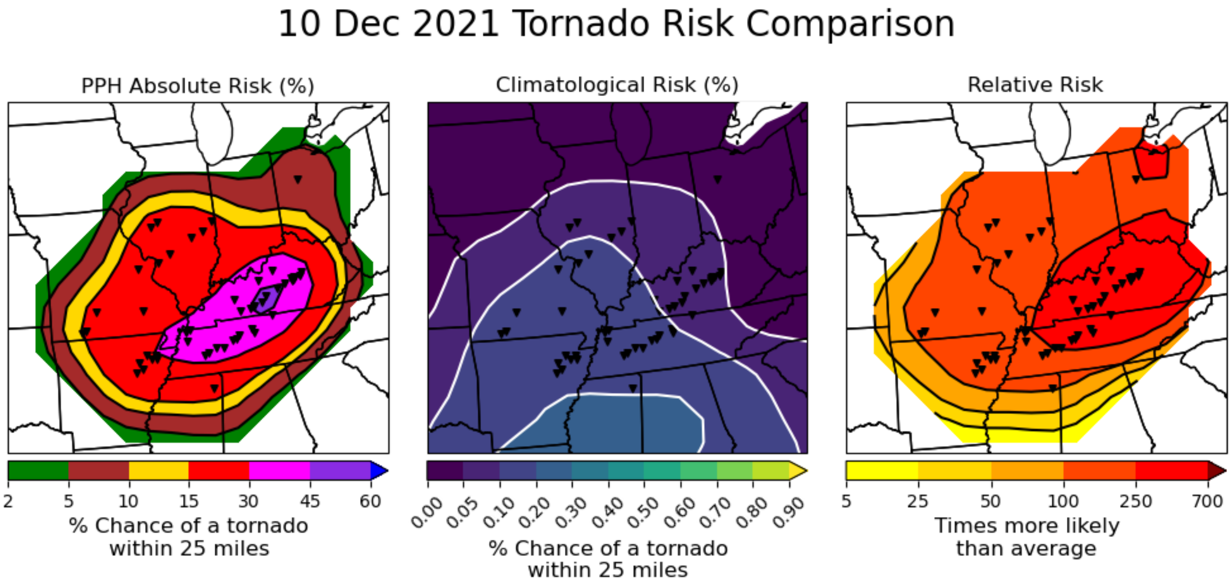


Figure 2.12: Practically Perfect Hindcast (PPH, left), climatological (middle), and relative (right) tornado risk for December 10th, 2021. Black triangles on all three plots represent the start point of individual tornado tracks.

A second notable event in the list of tornado days with the largest 100x relative risk areas is April 27th, 2011, often referred to as the second Super Outbreak. The April 27th event is especially interesting as it highlights how relative risk information is most valuable at the local level that it contextualizes over absolute risks. The greatest coverage of EF1+ tornadoes on this day occurred across northern Alabama, where a rare 60% contour is found in the PPH absolute risk analysis (see fig. 2.13). A 30% PPH absolute risk of tornadoes extended well into Pennsylvania, however, with the 15% contour reaching into upstate New York. As the climatological risk for tornadoes generally decreases with increasingly northern latitudes, especially so at this time of year, the relative risk for tornadoes on this day was highest in areas of New York and the Appalachian Mountains in Virginia and West Virginia (see fig. 2.13). The higher values of relative risk in the northern areas of this outbreak highlight how relative risk can add local context to absolute risks, as both areas of the South accustomed to tornado outbreaks in April and areas of the Northeast more used to late-season snow threats can gain a better understanding of the significance of the tornado event in their area.

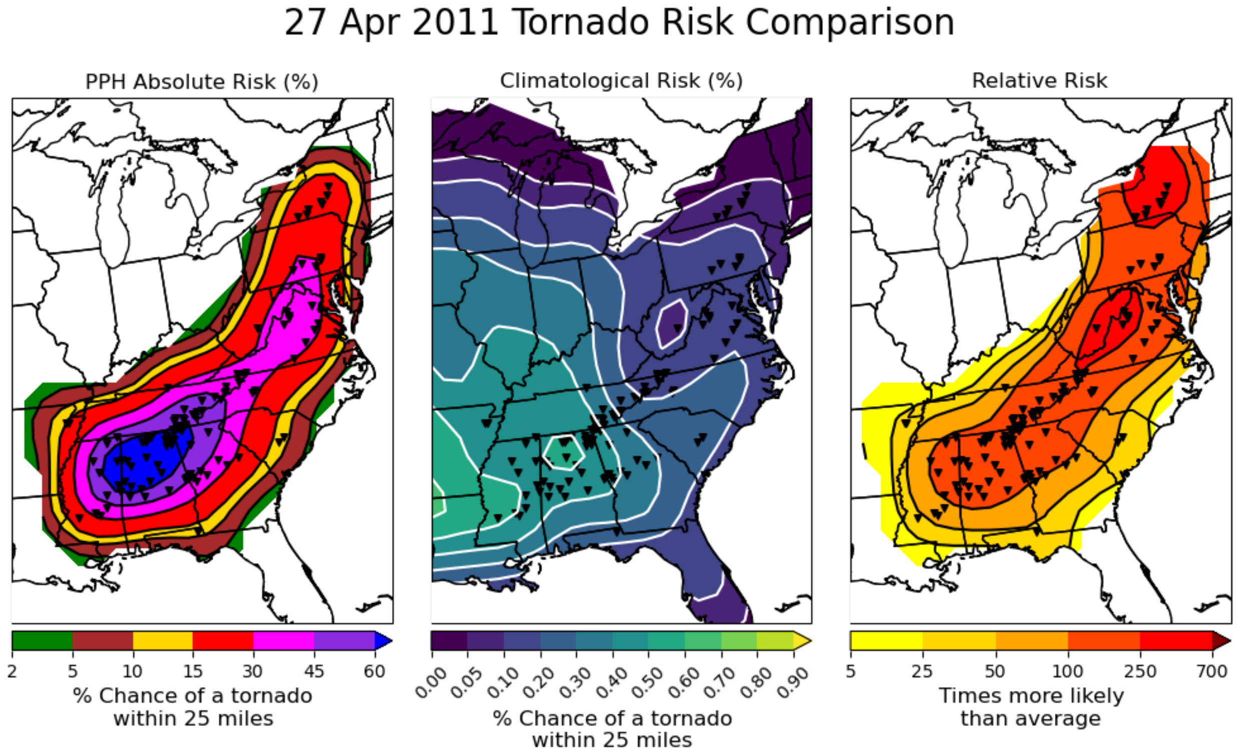


Figure 2.13: Practically Perfect Hindcast (PPH, left), climatological (middle), and relative (right) tornado risk for April 27th, 2011. Black triangles on all three plots represent the start point of individual tornado tracks.

2.4 Discussion and Conclusion

Following the suggestion of Murphy (1991) for communicating rare and severe events, I have developed and explored a dataset of relative risks for (E)F1 and stronger tornadoes across the US from 1950 to 2021. Relative risk values for tornadoes were calculated using publicly accessible tornado report data and peer-reviewed climatology development methods that could be applied to other convective weather hazards, including hail and straight-line wind reports. My results show that the relative risk values for tornado events are heavily dependent on the background tornado climatology at the time and place where that tornado event occurs. The reliance of final relative risk values on climatology values can lead to extreme values of relative risk in areas of the country that see low minimum climatological risks for tornadoes, especially those where the tornado that relative risk is being calculated for is the only tornado at that place and time ever recorded (such as the December 26th, 1966 case). The highest relative risk values in the database occur mostly with single-tornado events in the US west of 105W longitude, which rarely sees tornadoes at any time of year. Removing grid points from west of 105W longitude greatly reduces the frequency of extreme relative risk values in excess of 1000x in the dataset, decreasing the value of the 90th and 99th percentiles of the distribution from 109x and 622x to 90x and 304x, respectively. Alternatively, ranking relative risks by area covered by a risk greater than 100x identifies a list of major tornado outbreaks known to be among the most significant tornado events in US history, suggesting that spatial coverage of relative risks is an important value to consider when determining the significance of a given tornado event.

The objective and subjective analyses of the relative risk database developed in this study raise a number of important questions that should be addressed by future studies. First, the relative risk dataset could benefit from a subjective analysis performed by human coders, who could compare absolute risks, climatological risks, and resulting relative risks to identify different classifications of tornado events based on their risk values and distributions. Based on the subjective analysis performed in this study, I propose that such an investigation consider grouping events by whether they occur in annual and regional peaks or lulls in tornado activity. Additionally, the responses of user groups, including but not limited to emergency managers, broadcast meteorologists, and members of the public, to the different relative risk values and graphics generated for this dataset must be studied before relative risk products are considered for operationalization. Although risk communication theory suggests that adding context to absolute risks for tornadoes would be beneficial for forecast users of all types (e.g. Murphy 1991; Woo and Marzocchi 2004; Spiegelhalter et al. 2017; Klockow-McClain et al. 2020), real world tests are needed to ensure that these theories are realized in practice. Finally, the climatology development and analysis performed in this study should be repeated for straight-line wind and hail hazards as well, to explore the variety of relative risks that could be a part of the complete convective outlook. I will make the baseline code used to perform this analysis available via Github after publication, in the interest of spurring future work on relative risks for other weather hazards.

Future studies may also be able to address some of the limitations of the research performed here. The SPC tornado database has known errors and is restricted by the inherent biases of tornado reporting and the limitations of the damage inspection methods employed to rate tornadoes on the (E)F-scale (Edwards et al. 2021; Potvin et al. 2022). Wind and hail reports in the SPC database are subject to further errors (Edwards, Allen, and Carbin 2018), suggesting future work in this area should be aware that these datasets carry inherent biases that can impact final relative risk values for those hazards. This study also used PPH outlooks to calculate relative risks for past tornado events, to account for the lack of probabilistic outlooks from before 2001. As hindcasts are effectively “perfect” forecasts, future studies of user interpretation of relative risk forecasts may wish to also analyze and compare the distribution of relative risks derived from PPHs to those derived from human forecasts of the likelihood of tornadoes. Further, the use of a 71-year climatology for tornadoes in this study differs from the typical use of a 30-year climatology in other climate studies. As using a complete dataset weather events results in a biased climatology that accounts for all recorded events, and because recent studies have suggested that peak tornado climatology may be shifting eastward from the Great Plains and into the Midwest (Gensini and Brooks 2018), future work should seek to investigate the sensitivity of relative risk values to the application of different 30-year climatologies for tornadoes.

Despite these limitations, I believe relative risk information presents a compelling avenue for future study, as theory suggests that risk communication that includes relative risks may improve recipients' risk comprehension when shown SPC outlook products. Given the efforts of initiatives like the FACETs program to include more probability information in weather risk communication products, it is important to identify ways that severe weather hazard probabilities can be presented that avoid known biases in how individuals process risk information. Although this is an early investigation that only explores the range of relative risk values that are calculated from perfect hindcasts, my findings suggest that relative risk information could add the additional context to rare event forecasts that decision-makers are looking for. However, as described by Murphy (1993), forecast information by itself carries no inherent value, as value is generated by the decisions made by the users that receive the forecast. Relative risk values look useful from my meteorological perspective, but this risk information warrants further quantitative and qualitative study with key user groups to ascertain the value that the product allows them to generate.

Chapter 3 – Investigating Broadcast Meteorologist Use of Tornado Relative Risks

3.1 Introduction

Despite the rapid advance of communication technology following the advent of the smartphone and social media, broadcast meteorologists on television remain a key source of weather information for a majority of the US public as of 2022 (Ripberger and Silva 2022). Local television broadcasters are the highest-regarded and trusted source of general weather information for the greater public and have been for decades (Lazo et al. 2009; Drobot et al. 2014; Ripberger and Silva 2022). Further, studies specifically targeting the dissemination and reception of NWS severe weather warnings have identified that the majority of the public relies on television weather coverage as their source of warning information (Hammer and Schmidlin 2002, Daniels and Loggins 2007; Keul and Holzer 2013; Drost et al. 2016). In accordance with the academically observed value of broadcast weather communication to the public, the NWS recognizes broadcasters as core partners as a key part of the dissemination of NWS products and safety information to the public (NWS 2018). Combined, this means that it is crucial to understand how broadcast meteorologists might reinterpret and present new forecast products like relative risk-based tornado forecasts, as their reinterpretations are what members of the public are most likely to see before a tornado event.

Prior studies of broadcast coverage using tornado forecasts have generally focused on tornado warning dissemination and reception. Nearly 90% of residents surveyed by Hammer and Schmidlin (2002) that were impacted by the May 3rd, 1999 Bridge Creek-Moore tornado reported that the television coverage of the storm was their primary source of warning information. More recently, studies have identified that presenting members of the public with tornado coverage increased their risk perceptions and intent to shelter to protect themselves from the storm (Zhao et al. 2019). The high level of trust that members of the public feel for their local broadcasters (Ripberger and Silva 2022) has also been found to positively impact their likelihood of sheltering from tornadoes (Sherman-Morris 2005). Due to the keystone position that broadcasters hold in the tornado forecast dissemination process, it is especially important to understand how they might receive, and then present, prototype tornado forecast products like relative risk information.

Although relative risk forecasts for tornadoes have never before been presented to broadcasters, some work has been done to understand how broadcast meteorologists repackage the Storm Prediction Center (SPC) convective outlook, which relative risk is derived from, for their viewers. Ernst (2020) identified through interviews with broadcast meteorologists that the convective outlook was a useful tool for making operational decisions and to communicate risk to station management, although broadcasters expressed frustration with the words used in the product. Broadcaster's dislike of the convective outlook words likely plays a role in causing inconsistencies in the product's presentation on television as documented by Williams and Eosco (2021). However, work by Krocak et al. (2022) suggests that the use of numerical levels (e.g. a "level 2 out of 5 risk") may more consistently communicate the threat levels in the SPC outlook to the public. Numerical values like levels may address the concerns identified in Ernst (2020) and Williams and Eosco (2021) and offer broadcasters a clearer and more concise method of presenting the convective outlook product.

Unfortunately, studies of the presentation and use of the SPC convective outlook by broadcast meteorologists have thus far failed to investigate how broadcasters use the probabilistic outlook paired with it. Literature does exist on more general communication of uncertainty and probability by broadcast meteorologists, particularly for the percentages used to communicate precipitation chances. Demuth et al. (2009) noted through focus group interviews with broadcasters that they would present forecast uncertainty to their viewers through word-based verbal or expression-based non-verbal communication methods, but directly expressed doubt in the public's ability to interpret absolute probabilistic information and thought that scales were a better way to present information on TV. In a technical report on the focus groups, the researchers further highlighted that broadcasters felt that the public misunderstood the definition of Probability of Precipitation (PoP) and that including probabilistic uncertainty information in future forecasts could be overwhelming to viewers without a widespread effort to educate the public (Morrow et al. 2008). Efforts by Stewart et al. (2016) supported these findings further, identifying through a survey of professional meteorologists that on average participants believed that only 22% of the public correctly understood the meaning of the PoP. However, 70% of the professionals surveyed defined PoP differently from the NWS definition, suggesting that broader education about the context of probability information for rainfall forecasts might be needed. Overall, broadcasters appear to doubt their viewer's ability to understand probabilistic values, such as the percentages in the probabilistic outlook, and prefer presenting likelihood or probability information through non-numerical methods like scales or words of estimative probability (Lenhardt et al. 2020).

Broadcaster's opinions on what information they should present to their viewers do not always dictate their coverage however, as station management plays a key role in dictating what eventually ends up on air (Obermeier et al. 2022). Station producers often follow the advice of hired consultants when dictating what broadcasts should present during their shows, and consultant feedback may not always be shared with the meteorologist (Henson 2010). As an example of a product that has spread across the nation under the advice of consultants and insistence of station management, many local news outlets have begun using "Code Red" language that highlights impactful weather days on products like the extended range forecast and in graphical cues used during the weather show (e.g. Burnett 2016; Mojica 2019; CBS 4 News 2019). Although some stations have shared definitions of their "Code Red" days, which generally say that “Code Red” is for use during days where impactful severe weather is expected, other reporting suggests that some broadcast meteorologists have little say in when “Code Red” language is used, with the final decision falling to station management (Stelter 2019). The issues surrounding the use of “Code Red” language serve as an example of how coverage decisions can be outside a broadcast meteorologist’s purview and shows that use of a new forecast product may need to first be approved by station management.

Overall, the state of the literature points to three key understandings to address for how broadcasters might use relative risk information to communicate severe convective weather risks. First, broadcasters are an essential part of the weather communication pipeline in the US, and as a trusted source of severe weather information must be a part of the development of any future risk communication product. Second, broadcasters do not all feel that the convective outlook is important to present to viewers and may prefer editing the product or not sharing it at all, a sentiment that may carry over to new convective outlook-adjacent products like relative risk. Third, broadcasters appear to be hesitant to present their viewers with numerical probabilistic values due to the perceived inability of their viewers to understand those values, which may lead them to feel that relative risk information may not be valuable in on-air coverage. Further development of relative risk information as a product thus necessitates broadcast meteorologist input, to ensure that the product can evolve into one that they feel would be of value both to their forecast process and to their viewers.

3.2 Methods

3.2.1 Data Collection

To sample broadcast meteorologists' opinions and perceptions of the value of relative risk information, I first needed to recruit a sample of broadcast meteorologists for a series of focus group interviews. I worked with the Hazardous Weather Testbed (HWT) broadcaster team, which runs the K-Probabilistic Hazard Information-Television (KPHI-TV) project, to recruit my sample of broadcast meteorologists. Over the past seven years, KPHI-TV has built relationships with broadcasters across the country to organize testbed experiments ranging from in-person live presentation of experimental forecast products, to focus group interviews both in-person and online (Ernst 2020; Obermeier et al. 2022). Broadcasters are reached by KPHI-TV through a snowball sampling method (Obermeier et al. 2022), which involves sending emails and letters to potential participants and asking them to forward other potential participants to contact the research team (NSF 2023). For this study, I forwarded an email to the KPHI-TV mailing list asking for participants for hour-long focus group interviews and received 23 responses from broadcasters interested in participating. I then selected 16 of those broadcasters to participate in four focus groups based on their reported availability and location in the US. Of those 16 invitees, a total of 12 were able to participate in my focus group interviews (see fig. 3.1 for a map of broadcaster locations). To protect these broadcasters' identities, I did not conduct any analyses involving their length of tenure at their station or any of their other demographics.

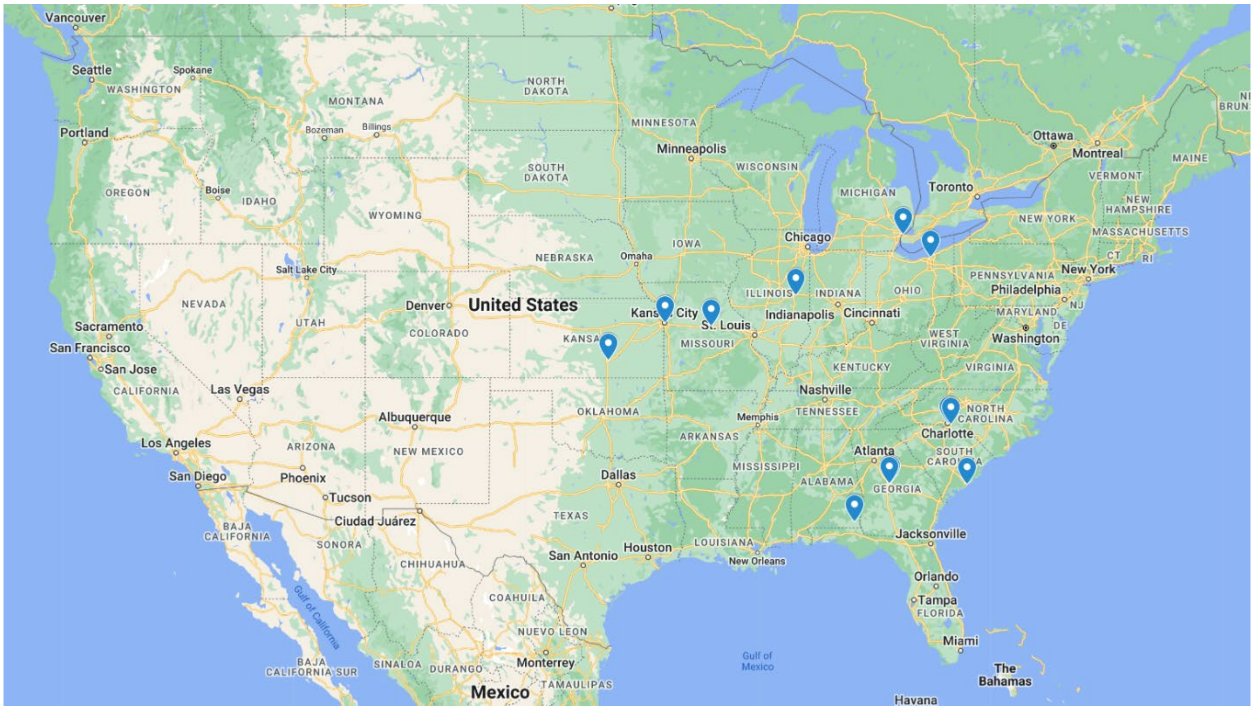
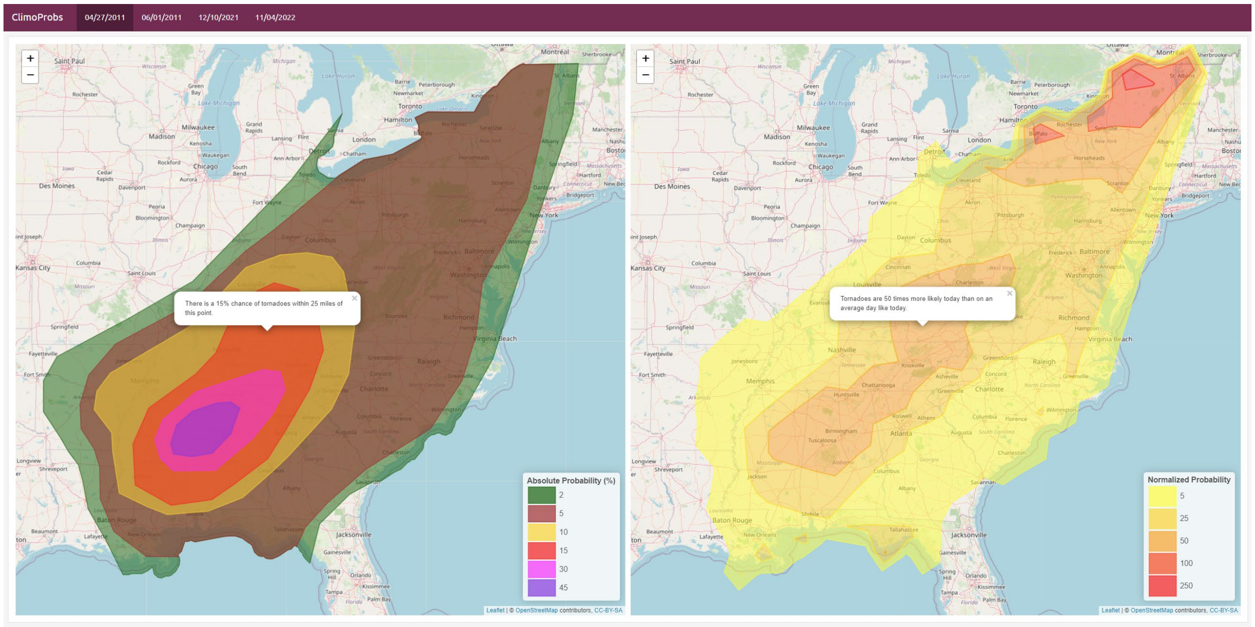


Figure 3.1: Map of where the broadcasters interviewed as part of this study work on air. Note that two broadcasters hailed from the same market in both western North Carolina and central Georgia. All participants hailed from television markets east of the Rocky Mountains, which is discussed later as one limitation of this study.

Focus group interviews were chosen for this analysis due to their proven value in previous studies of broadcast meteorologist communication (Demuth et al. 2009; Ernst 2020; Obermeier et al. 2022) and for their ability to collect data from large numbers of participants in a cost-effective manner. I developed a semi-structured interview guide to conduct my focus groups following guidelines set by established literature (Krueger 2002; Breen 2006; Gill and Baillie 2018). Participants were first read information describing their rights as participants in this research and my expectations for respectful and constructive discussion of the questions asked during the interview (see Appendix A). The moderator then asked participants about their experience with the current SPC probabilistic outlook, and whether or not they presented the probabilities in the outlook to their viewers, to understand participant’s general use of probabilistic information for communicating severe weather risk. After discussion of the current state of affairs, broadcasters were shown a brief PowerPoint introducing them to the concepts that underly relative risk information and how it is calculated for tornado risks (adapted from a presentation given by Ernst et al. 2022). Note that at this time in the development of the product, relative risk information was being referred to as "normalized risk" information. To maintain consistent use of language across this study and its two sibling studies, I have replaced all mentions of "normalized risk" with relative risk from this point on in the paper.

After being shown the PowerPoint explaining relative risk, participants were then given a link to an online web application with a zoomable map of the forecast SPC probabilistic tornado outlook and derived relative risk (see fig. 3.2 for an example) for four tornado events: April 27th, 2011; June 1st, 2011; December 10th, 2021; and November 4th, 2022. These events were chosen as they represented several key scenarios for relative risk presentation identified in previous relative risk studies (see Chapter 2 of this dissertation): a significant in-season, in-region event; an in-season, off-region event; an off-season, in-region event, and a typical in-season, in-region event; respectively.



There is a 15% chance of tornadoes within 25 miles of this point.

Tornadoes are 50 times more likely today than on an average day like today.

Figure 3.2: Example screenshot of the SPC tornado probability forecast and derived relative risk forecast comparison tool that broadcast meteorologists were shown in each focus group. This screenshot shows the tornado probabilities issued in the 1630z outlook on April 27th, 2011, while the relative risk plot calculates relative risk based on those forecast tornado probabilities as compared to tornado climatology for April 27th.

Once given access to the online web application, participants were asked if they thought relative risk information would be useful to them in their forecast process, and whether they would present relative risks on air or on social media for their viewers. The interview guide left room for additional questions if time permitted in this section, including asking participants what they felt the pros and cons of relative risk information were. Participants were also asked if they would prefer to use the map display in the web application to present relative risk information or a numerical display (like the temperature forecast maps used to display the expected temperature for chosen cities and towns). To end the interview, participants were asked to summarize their overall reaction to the relative risk information they were shown during the focus group, and if there were any changes to the product that they felt might be necessary to make to bring relative risk information into an operational format. Interview audio was recorded through the Zoom application that interviews were performed on, and this audio was transcribed for analysis using the Otter.ai online transcription service.

3.2.2 Data Analysis

Once the interviews with broadcast meteorologists were complete and transcribed, I analyzed my interview data using thematic coding analysis. Thematic coding has been described as “a method for identifying, analyzing, and reporting patterns (themes) in data” (Braun and Clarke 2006) that elicits meaningful conclusions from qualitative data through the application of thematic “codes”, or labels, by a human analyst (referred to as a “coder”). These codes are developed either deductively, as a list of topics of interest defined before a coder begins their analysis, or inductively, as a series of labels that emerge organically through a coder's careful reading of the data (Elliott 2018). In this study, I developed a set of deductive codes based on what I wanted to learn from the data and allowed these codes to evolve inductively as the coder read through the dataset. The final list of codes that were applied to the transcribed focus group interviews can be found in Table 3.1. To ensure rigor and the reproducibility of my coding analysis, the coder for this analysis recorded their thoughts and interpretation of the data as analysis was performed in a coding journal (see Appendix B). Future researchers can use a coding journal to understand the process through which the original coder developed their codes and how they applied them to their data (Nowell et al. 2017).

Table 3.1: List of codes applied to transcriptions of the four focus groups performed in this study. Each code, which is a shortened label, is expanded upon with a full name for the code and a description of when that code should be applied to the data.

Code	Full Name	Description
GR8	Things that are great with the SPC outlook	Mentions by broadcasters of things they like about the probabilities in or the design of the current SPC outlook (or positive uses)
BAD	Things that are problematic with the current SPC outlook	Mentions by broadcasters of things they do not like about the probabilities in or the design of the current SPC outlook (not probability understanding related)
USE	Use of Norm/relative Risk	For when broadcasters mention how they currently or would in the future use relative risk information for their own purposes - not including presenting to the public.
SIT	Situation that relative risk would be useful	for when broadcasters discuss WHEN they would use relative risk information or present it to their viewers
SHO	How broadcasters show viewers relative risk	for when broadcasters discuss how they would SHOW relative risk information to their viewers. (or if they wouldn't, and why)

PRO	Pro - positive feedback	Something a broadcaster liked about the normalized risk information
CON	Con - negative feedback	Something a broadcaster didn't like about the normalized risk information
MAP	Mapped normalized/relative risk	For discussion of the visual presentation of relative risk
FAH	False Alarm/Hype Concern	For explicit mentions of being worried that the probabilities lead to hype or false alarms for the public.
PRB	Probability Understanding	For direct discussion of the public's ability to interpret probabilistic information.
LVL	Use of SPC categorical levels	For mentions of using SPC categorical levels instead of or along with probability information.
ADD	Improvement to add to Normalized/relative Risk	When broadcasters suggest ways that normalized/relative risk products could be improved for their use.

Once the transcripts for all four focus group interviews were coded, the coder then extracted each line with a code label and organized these excerpts by code. Once removed from the overall interview transcriptions, the coded excerpts were re-read by the coder to identify common themes across participants and focus groups. Established literature on thematic analysis strongly suggests that analysts interpret their newly-discovered themes and their interactions using visualizations, such as thematic maps (Braun and Clarke 2006; Nowell et al. 2017). Thematic maps for this study were developed around the overarching themes of “Concerns”, “Uses”, and “Presentation” of relative risk information, as these themes represented the deductive concepts that I originally sought to identify in the data (see fig. 3.3). Once sub-themes were related to these three major themes, and supported using quotes from the focus group interviews, the thematic map was compared to the transcribed data, to ensure consistency between the participant's original statements and the organized findings of the study.

3.3 Results

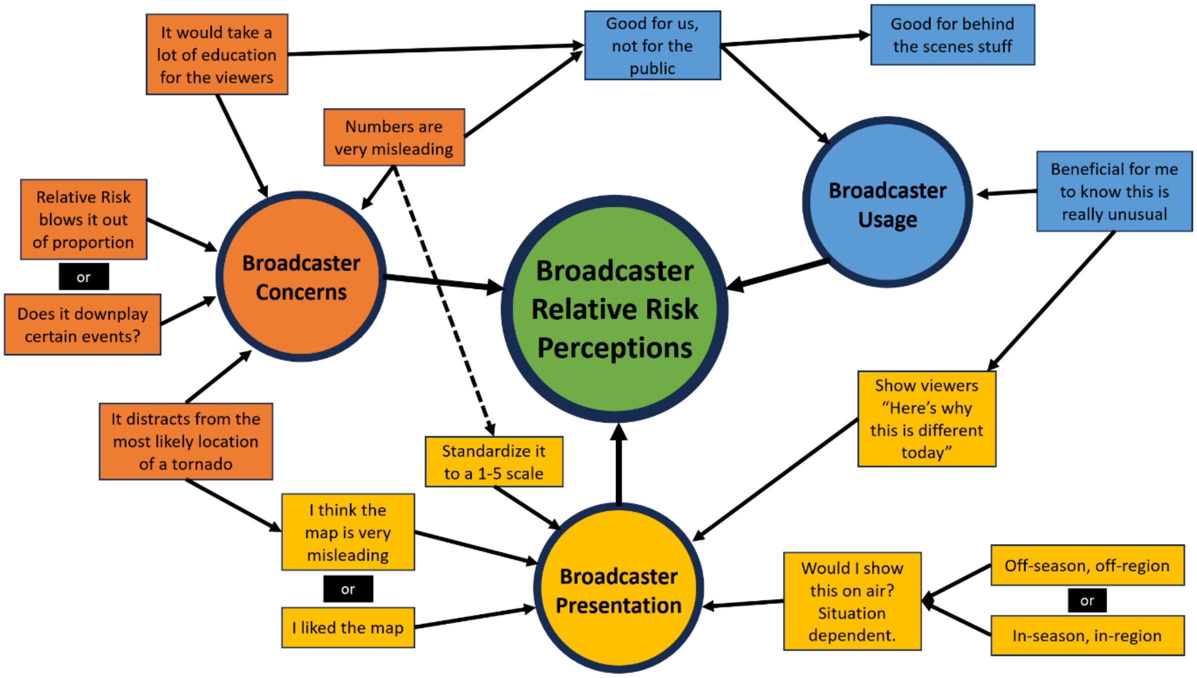


Figure 3.3: Thematic map developed as a product of the thematic coding analysis performed on my transcribed focus group interviews. The central idea of how broadcast meteorologists interpret relative risk information is represented by the central circle labeled "Broadcaster Relative Risk Perceptions", which is broken down into three main theme groupings around "Broadcaster Concerns," "Broadcaster Usage [of Relative Risk]," and "Broadcaster Presentation [of Relative Risk]." The boxes represent themes identified in the data through coding analysis and are presented using quotes from the participants that best summarize each theme. A dashed line is used to connect "Numbers are very misleading" to "Standardize it to a 1-5 scale" to differentiate that line from the line between "Broadcaster Concerns" and "Broadcaster Relative Risk Perceptions".

3.3.1 Broadcaster Concerns about Relative Risk Information

When they were initially asked about their current use of the SPC convective outlook probabilities, broadcasters in every focus group expressed concern that their viewers could not understand probabilistic forecast information. Of the 12 broadcasters interviewed across the four focus groups in this study, 10 directly mentioned that they believed that probabilistic forecasts for severe weather can confuse members of the public, a theme that I named “Numbers are very misleading”. The low absolute probabilities for tornado forecasts were highlighted as a problem by several broadcasters, with some comparing the forecast probabilities to those for rain chances, even though those probabilities are calculated in very different ways (Participants 1, 2, 3, 6, 7, 8, 9, 10, 11, 12). Broadcasters feared that lower probabilities could lead to complacency, as “it’s like 10% chance of rain, I’m not gonna get rained on, so we could actually hurt your overall message” (P10), or that “you tell somebody, there’s a 10% chance of a tornado, they say, okay, 90% it’s not gonna happen” (P1). Participant 6 also added that they were “afraid the viewers will interpret [tornado probabilities] too much like the rain probabilities”. In general, my participants felt that probability information would lead viewers to underestimate their risk from severe weather, and that “location is more important to our viewers” (P11) or that viewers were more interested in having broadcasters “tell me timings, tell me risk, and tell me what I need to watch out for what types of threats” (P7).

The concern these broadcasters felt over the public’s ability to interpret rare risk probabilities was further reflected in their initial reactions to relative risk information, under the theme “Relative risk blows it out of proportion”. Broadcasters feared that higher values of relative risk would be seen as overly threatening, and that viewers would panic if they were shown some of the relative risk values that occurred in the four example events (Participants 1, 2, 3, 4, 6, 9, 10, 11, 12). Participant 2 was concerned that “people [in my area] would completely come apart if I said something like that on air” after Participant 1 suggested that “the person who’s watching for the first time and has never seen the scale, they’re gonna go 250, the hell, where did that come from? They’re gonna think they’ll be in the twister movie that afternoon!” Broadcasters in other focus groups also felt that relative risk information could be misleading to their viewers, “because people would expect maybe more of the damaging storms up in New York [on April 27th 2011] but that’s not what we’re trying to say” (P6) and that “you worry about the Chicken Little factor… if there’s a 100% chance more than normal of having a tornado and you go on air and say that, that’s a hard thing to walk back” (P12). Some participants even suggested they preferred absolute probabilities to the relative risk information, as “a lot of people, even if they don’t understand percentages, they understand the 0 to 100 thing… the times normal, I think that’s just too much for them” (P3).

A few other participants, however, felt that the relative risk information might instead downplay the risk posed by high-end tornado events, asking “Does it downplay certain events?” Participant 4 noted that, when comparing the absolute and relative risk forecasts for April 27th, 2011, “for those areas that are at the 45 hatched absolute, just detailing it with [100 times greater than normal] relative risk is extremely misleading for what happens down there.” Other participants applied this concern more broadly, including Participant 8, who suggested that “in April, May, and June… that’s our severe weather season, we’d expect [tornadoes] anyway. And looking at the relative risk if it downplays particular events you don’t want to downplay.” Finally, Participant 10 noted that they were surprised the relative risk wasn’t higher in the areas most impacted by tornadoes in the forecast for the December 10th, 2021 tornado outbreak, and felt that “there’s going to be tornadoes, more than likely… so [relative risk] wouldn’t hit the way that you would want it to hit in your overall messaging scheme.” Although only these three broadcast meteorologists shared their concern that relative risk information might downplay some tornado events, the contrast between this downplaying concern and others’ fears of relative risk information overinflating tornado risks makes it worth mention.

Broadcaster's fears that relative risk information could overhype or downplay tornado risk further led them to express concern that relative risk maxima on maps could "distract from the most likely location of a tornado". After viewing the four relative risk tornado forecasts created for this experiment, seven participants mentioned that they thought the offsets between absolute and relative risk maxima in several events could confuse viewers, particularly the offset seen on April 27th, 2011 (Participants 1, 3, 4, 6, 9, 11, 12). Participant 12 explained the problem with offset maxima eloquently, as they feared that "seeing red on the relative risk is concerning… and then you realize that's an increased risk more than normal. But the main threat is to the southwest," and that "I could see [the general public] thinking that red is the most likely place to see a tornado, and that's not it." Other participants agreed, noting that compared to more tornado prone areas "a 5% risk in upstate New York is much more significant than normal. But from a mathematical standpoint, your chances are still the same. It's 5%, that's grounded in reason" (P3). Overall, these broadcasters felt that "[the map] misleads because people I feel like would expect more of the damaging storms up in New York, but that's not what we're trying to say if we're looking nationally" (P6).

Given their fears about the public's ability to interpret relative risk information, it is unsurprising that the broadcasters interviewed for this study also expressed concerns about how presenting relative risk "would take a lot of education for the viewers". Six broadcasters (Participants 4, 6, 9, 10, 11, 12) mentioned they'd need to perform on-the-air education describing the use of relative risk information before presenting it, as well as concerns that they would not have time in their already tightly scheduled shows to include a discussion of how to interpret relative risks. Indeed, in broadcast television "we're already up against a time crunch" (P4) and "explaining something like [relative risk] on air will sometimes take more time than I'm allotted" (P6). The main concept that broadcasters felt they needed to educate their viewers on was the climatological context of the forecast, as Participant 4 explained that "I'd have to do a little weather education on what the normal threat for tornadoes is [in my area], what's our normal severe weather season, and then relate that to [relative risk]." Participant 10 summarized the problem succinctly, as "its hard reinventing the wheel and going against people's routines, people's familiarity [with a pre-existing forecast product]."

3.3.2 Broadcaster Usage of Relative Risk Information

Across the four focus groups, the concerns that broadcasters felt towards viewer's ability to understand and use relative risk information led many to suggest that relative risk would be "good for us, not for the public". This sentiment was expressed by seven participants (Participants 1, 2, 3, 4, 5, 10, 12), who felt that the relative risk forecast provided them with valuable information, but that it was overwhelming for the public. For their own use, some broadcasters mentioned that "if I'm able to look and see that today is a 100 times higher day… it's a 10% day, its January, trying to do all that in my head, this takes all of that away" (P1) and that "this would be worthwhile for the meteorologists to know… especially those that are newer, they can know this is how unusual this is, or other dates that people can relate to" (P3). Others suggested that relative risk could become a part of their forecast process, as "it would help me define better on a day where there's tornadoes… if I knew that was showing 100 times more likely, I may be showing reds instead of oranges [on my own impact graphics]" (P5). Finally, broadcasters also thought that relative risk information might be more valuable to emergency managers, "because they're the ones that have to make calls ahead of time" (P3) and that "this is more beneficial information, though, for emergency managers, for partners that are planning for this scenario" (P12). In general, although these seven broadcasters believed "I would use a relative risk in my forecasting" (P12) and that relative risk was "definitely something I'm going to add into the arsenal to use" (P5), they also felt that the product was not "super applicable from an audience standpoint" (P3) and that "relative risk would be good for behind the scenes stuff" (P2).

The belief that relative risk information was best used by experts naturally led broadcasters to see the product as "good for behind the scenes stuff" as mentioned by Participant 2. Participant 5 also highlighted that "I'm using it from a behind the scenes perspective," but that they did see relative risk information as useful to them, particularly to make more informed decisions regarding staffing and planning ahead of a tornado event. Participant 3 mentioned the value of relative risk information for knowing whether they need "normal staffing, or do we put people on standby?" Other participants (P2, P12) also mentioned that relative risks would be helpful for planning purposes. Relative risk information was seen as useful as part of the forecast development process as well, including Participant 5's idea of highlighting a higher level of risk on their personal impact graphics for high relative risk days, while Participant 7 suggested that "the more information that you can give meteorologists, the more information that gives me… to keep people safe".

Beyond scheduling and planning for covering tornado threats, four broadcasters (Participants 1, 3, 11, 12) mentioned that relative risk information for tornadoes was "beneficial for me to know this is really unusual". Participant 3 noted that "this would be worthwhile for the meteorologists, especially those that are newer, they can go 'oh, this is how unusual this is'" and thus prepare more appropriately for their coverage. Even for more experienced broadcasters, "in the offseason, it can be really beneficial to have context of this is really unusual. According to climatology it's 50, 100, 200 times more likely, whatever" (P12). Participant 1 also felt that knowing the relative risk would be useful to “look and see that it’s a 100 times higher day… where I don’t have a lot of time in the morning, I can internalize a lot from the statement, as opposed to having to click through everything… from a workflow standpoint, extremely helpful there.” Overall, these broadcasters saw value to relative risk forecasts for tornadoes for informing them and other experts about the context and rarity of a tornado event, though they preferred to keep the product for their eyes only for fear of misinterpretation.

3.3.3 Broadcaster Presentation of Relative Risk Information

Although the broadcasters I interviewed generally expressed concerns about how their viewers might interpret relative risk products, when pressed about presenting relative risk information, 10 out of the 12 participants answered “would I show it on air? Situation dependent”. Seven broadcasters (Participants 4, 6, 8, 9, 10, 11, 12) felt that relative risk would be most useful to present for a tornado event that was occurring during the “off-season or off-region” tornado events, while three (Participants 4, 7, 9) believed that relative risk information was best shown during high-end “in-season or in-region” events. The group that leaned towards more unusual events felt that relative risk was “helpful in those extreme events to get people to understand this isn’t like last time, heed these warnings” (P6). Relative risks were seen as particularly useful in winter, as “if it’s a high risk in December, I think there’s some significant utility to that particular product” (P8) and that “if it’s a December 10th event, maybe I am more likely to talk about how uncommon it is” (P12). These broadcasters were also adamant that “if it’s spring or fall, when we typically would see a peak in severe weather, it doesn’t add much” (P11), or that “during the busy season on the plains, I don’t think we’d even bother with it” (P8). In contrast, other broadcasters that felt relative risk information would be most useful in “tornado alley, Dixie alley, those areas, I think that’d be the most beneficial for it" (P7). Participant 4 also added that “you’re going to have a higher than normal [relative risk] just because you’re out of season [on December 10th 2021]. So I don’t know if I would have leaned on it so much, then.” Combined, broadcaster’s thoughts on when to present relative risk suggest that they generally see uses for relative risk in supplementing absolute risk information during particularly rare events. There are cases, however, where broadcasters’ concerns about higher values of relative risk “overhyping” a tornado event may lead some to avoid presenting relative risk except during times and places where tornadoes are expected.

Despite their uncertainty in when presenting viewers with relative risk information would be most appropriate, and whether relative risk would be of value to viewers at all, five broadcasters across the four focus groups (Participants 1, 4, 8, 9, 10) came to the conclusion that relative risk forecasts would be useful to “show viewers ‘here’s why this is different today’”. These broadcasters saw relative risk as a useful addition to any discussion of unusual tornado events, as “if it was completely abnormal, a higher risk event way out of season, and we’re in the red, I might show it just as a supplemental” (P4). Other broadcasters mentioned using the “times more likely than normal” or “x times higher” language of relative risk in their forecasts before ever being shown prototype relative risk forecasts in the interviews. Participant 1 mentioned that they currently used the SPC outlook probabilities “to create the statement of ‘the risk today is x times higher’, or ‘the risk in this location is x times higher than in that location’ because I’m able to show where the risk is higher.” Other broadcasters mentioned similar messaging tactics, as Participant 8 described how “you look at 2%, and you think 2%, whatever. But that’s two times as much as a normal day, 5% is five times as much, if we’re talking damaging winds 15 times as much as just a normal day.” Participant 10 went even further, as on rare tornado events they would “do the math and compare that probability to the climatological risk for the day to determine how many times more likely a tornado is to occur within this area today versus any other one of these days in the past." After being shown the prototype relative risk products, Participant 10 further added that "we don't use [our own relative risk] map often, just on higher impact days, I think the last time we did was in 2019. And we had an EF4 tornado that day. So it really helped separate that day… here's why this is different today, and why you need to pay attention more." Although the five meteorologists that shared this idea were not in the majority of participants in this study, their pre-existing uses of relative risk-like information suggests that there may be a role for relative risk in highlighting the unusual nature of some severe weather events to viewers.

The participants in this study were also asked to describe whether they would present relative risk to their viewers in the form of a map (see fig. 3.2), or as number values at selected locations, similar to the temperature or snowfall forecast maps that are commonly used in broadcast weather shows. As with what situation broadcasters felt relative risk information added the most value, there was some disagreement among those interviewed here about which presentation format to use, with eight broadcasters (Participants 1, 2, 4, 5, 6, 8, 9, 10) saying "I liked the map" and three (Participants 6, 11, 12) that felt "I think the map is very misleading". Those that preferred the map cited the convenience of a visual representation and concerns about how discrete numbers are currently shown on maps, as "I definitely would not put cities on there because [people ask]… 'why didn't you call out my city?'" (P2). Another broadcaster mentioned that they "like the dual pane look, because then I get to see the back and forth to it. If I just see one, I want to know what the comparison is" (P4), which highlights both the utility of the map and how broadcasters see relative risk as a supplement to absolute risk information. Generally, the broadcasters that preferred the map worried that showing exact numbers could give viewers a sense of precision in the forecast that does not exist, as “people will remember those numbers and take them verbatim a lot like they do with snow” (P10). In contrast, the broadcasters that preferred a numerical presentation of relative risk information thought that the colorful but broad map could be confusing, and that “locally, [numbers] might be a more helpful tool than showing the big picture map” (P6). Broadcasters opposed to the map may have been more generally opposed to showing relative risk information, however, as Participant 12 also mentioned that “I would personally like to see the map, but I don’t think I would show the map, I think I would just talk about it.”

Finally, several broadcasters sought to address both their concerns and ideas about how they might use relative risk information through a suggestion for how to better present relative risk to their viewers – which came in the form of asking the researchers to “standardize it to a 1-5 scale”. Of the five broadcasters that mentioned that they would rather present tornado risk on a scale (Participants 1, 4, 5, 8, 11), two directly mentioned The Weather Channel’s TOR:CON product (Participants 8 and 11), a 1-10 level scale that suggests the likelihood of tornadoes within 50 miles of a point (The Weather Channel 2018). Participant 1 highlighted that they would prefer a format “kind of like how we have the percentage to level translation, you should have a percentage to level translation for [relative risk].” The interest in converting relative risk to a scale, for some of these broadcasters, was related to their concern that “sometimes simplified is the route to go… I don’t know if by adding more [numbers] we’re necessarily doing our consumers better” (P8). Others were resistant to changing the way they currently present tornado risks, as “we’ve been doing it long enough that we’ve got them trained on this low, medium, high and 1-10 kind of stuff that deviating from that would just put us back to square one” (P5). Regardless, these broadcasters felt that communicating relative risk through a categorical format would be easier for their viewers to interpret and would increase the appeal of relative risk as an on-air tool for them.

3.4 Discussion

When shown relative risk forecast information, broadcast meteorologists initially highlighted several key concerns they had with how members of the public would interpret relative risks. Despite the body of risk communication research that has found that people are not only capable of interpreting probabilistic information, but more likely to protect themselves appropriately from a hazard when provided with probabilistic information (Ripberger et al. 2022a), the broadcasters I interviewed for this study lacked confidence in their viewer's ability to interpret probabilistic forecast information. Other studies have previously identified a lack of broadcaster confidence in their viewers' ability to interpret numerical probabilistic information (Morrow et al. 2008; Demuth et al. 2009), which suggests that broadcaster's opinions on communicating probabilities have not changed since the 2010s. The hesitance that the broadcasters in this study felt towards sharing numerical probability information appears to extend to relative risk information as well, as several broadcasters mentioned concerns that large values of relative risk (the 250 times more likely than normal value in particular) would lead to panic amongst their viewers. Finally, management-imposed restrictions on broadcasters' ability to present weather information, particularly the length of time they have available to present information during their weather show, were also mentioned by broadcasters as an obstacle for them to effectively share relative risk forecasts with their viewers.

Despite their concerns about whether their audience would be able to interpret relative risk information, the broadcasters in this study felt that relative risk forecasts could be a valuable tool for their own behind-the-scenes use. Several broadcasters thought that relative risk information could help them decide whether to increase staffing to cover a given severe weather event, and that relative risk information could help them maintain awareness of the rarity of a given tornado event. Some suggested that they could pass information about the rarity of a tornado event to their viewers in a repackaged format. It is likely that broadcasters would reinterpret the uncertainty information in relative risk in these cases similar to the ways broadcasters use WEPs to discuss the likelihood probabilities of other weather events, as identified in prior studies (Morrow et al. 2008; Demuth et al. 2009; Lenhardt et al. 2020).

Some of the broadcasters in my focus groups were more open to the idea of presenting relative risk information to their viewers without repackaging, with several even mentioning that they currently calculate and share values similar to relative risks to their viewers. Tornado events that occurred during the off-season, or outside the typically defined “tornado alley”, were highlighted by a majority of participants as the events where they would be most likely to present their viewers with relative risk information, although a small minority thought that in-season events in tornado-prone areas would benefit more from relative risk information than uncommon events. A similarly sized majority of broadcasters believed that relative risk information was best presented in a contoured map format, even though a number of broadcasters previously expressed concern about how separation between the locations of maximum tornado likelihood and relative risk in some tornado events could confuse viewers as to what areas would be at greatest risk from tornadoes. That said, there were still a handful of participants that felt that sharing number values of relative risk in a manner similar to how forecast temperatures are displayed on television would be better than using a contoured map. Finally, a minority of broadcasters suggested that they might be more likely to present relative risk information if it was adapted into a categorical scale, like how the probabilistic outlook informs the levels of the

SPC categorical outlook. These broadcasters' desire for a categorical scale again reflects Demuth et al.'s (2009) finding that broadcast meteorologists prefer to translate uncertainty information into a more verbal or descriptive format, versus presenting numerical probabilities to viewers.

3.5 Conclusions

Overall, my findings suggest that while there would be an adjustment period if relative risk information was made available to broadcast meteorologists and the public, there are signs that relative risk would be valuable as a planning and communication tool for broadcasters seeking to highlight unusual tornado events for their audience. Although broadcaster's concerns about the ability of the public to interpret probabilistic information are worth consideration and deeper study, there is evidence in the broader risk communication literature to suggest that viewers would be able to understand probabilistic forecast information like relative risk if the numerical probabilities are properly explained. Further, the broadcasters I interviewed in this study broadly agreed that relative risk information could help them better understand the climatological context of a tornado event, and that they could use relative risk maps to share that information with their viewers when off-season or off-region tornado events occur. In summary, the broadcasters interviewed here consider relative risk to be a situationally useful supplemental product to the SPC probabilistic outlook that can help highlight tornado events that occur during off-season months or outside of regions considered to be the traditional "tornado alley".

Despite the evidence suggesting that further development of relative risk forecast information would be of value to broadcasters, there are a few limitations in this study that should be considered as well. First, this sample was limited to 12 total broadcast meteorologists due to time and resource constraints, and I was unable to recruit any broadcasters from the Northeastern or Western US (see fig. 3.1). Given that those regions are prone to extremely high values of relative risk due to the rarity of tornadoes in there, future work should seek to invite broadcasters from these regions to share their insights on relative risk forecasts for locally major events that occurred in their coverage areas. Focus groups also suffer from limitations inherent to their design, as the group discussion format of the interviews can sometimes lead the voices and opinions of more outgoing participants to steer the conversation. Future studies could seek to build on the findings of this study by performing individual interviews with broadcasters, avoiding any potential bias due to such steering. Finally, analysis for this study was performed by a single thematic coder, where typically at least one additional coder is utilized to test how consistently codes are applied to the dataset across different readers. The rigor of this study could be improved by introducing a second coding analyst to the dataset and using comparative statistics like Cronbach's Alpha to measure how consistently the codes developed in this study are applied.

Future studies could also improve on this work by broadening the sample of broadcasters shown relative risk information and by studying the use of relative risk for different hazards. One option would be to perform a survey of broadcast meteorologists across the United States, where the findings of this study could inform a battery of questions about potential relative risk concerns and use. It may also be worth exploring the interest expressed by some participants of this study to link relative risk to the pre-existing five-level convective outlook scale. Versions of this scale could be tested with both numerical values and words (possibly including words that reference the unusual nature of high relative risk events, like the words "seasonable", "uncommon", "rare", "extreme", or "unprecedented") that could communicate the nature of a tornado event in comparison to climatology. Finally, focus groups like the ones performed for this study cannot capture how broadcasters might utilize relative risk information in a realistic coverage scenario. Broadcasters' potential use of relative risk information may be better understood if study participants were presented with relative risk information before covering a live or simulated severe weather event, as in experiments with KPHI-TV.

Despite the limitations of this study and the future work needed to expand upon its early findings, the broadcast meteorologists interviewed for this study identified that they could use relative risk at least personally to better understand the context of unusual tornado events, and in some cases thought they could use relative risk to communicate that context to members of the public. Further, my focus group interviews have highlighted potential challenges for presenting relative risk information, such as misleading gaps between absolute and relative risk maxima in the map format of the product, or some broadcasters' preference for relative risk values being related to a categorical product, that future development of relative risk can prioritize investigation of to maximize the usefulness of relative risk products for broadcast meteorologists. Overall, relative risk forecast information for tornadoes can help broadcast meteorologists more quickly understand the context of tornado risks and the significance of those tornado risks to their local area, and broadcasters see relative risk as a valuable supplement to the current suite of SPC convective outlook products when unusual tornado events occur.

Chapter 4 – Understanding Public Reception of Relative Risk Information

4.1 Introduction

Extreme and rare weather events, including tornadoes, pose a variety of challenges to communicators seeking to warn the public about potential impacts from storms. Although a variety of solutions have been leveraged to create risk messaging products for weather hazards, ranging from winter storms (LeClerk and Joslyn 2012; Semmens et al. 2022) to hurricanes (Demuth et al. 2012), a group of recent studies have found that the Storm Prediction Center (SPC) convective outlook, which forecasts convective hazards including tornadoes, can be difficult for non-expert members of the public to interpret (Ernst et al. 2021; Krocak et al. 2022; Bitterman et al. 2023). Tornado hazards in particular can be difficult to communicate due to the localized but intense nature of tornado damage, which results in an "all-or-nothing" outcome for those in the path. While prior literature in weather risk communication suggests that probabilistic forecasts communicate hazards to the public more effectively (Joslyn and LeClerk 2011; Ripberger et al. 2022a), Murphy (1991) suggests that rare and severe events like tornadoes be communicated with additional contextual information about the frequency with which the event in question occurs.

4.1.1 Rare Event Communication Challenges

The ability of individuals to interpret probabilistic information is not perfect, and studies have identified several key moderating factors that can influence a person's interpretation of probabilistic risk forecasts. Numeracy, or the ability to correctly interpret probabilities and other mathematical concepts, has been closely linked to risk perception, as studies have shown that individuals that view probabilistic estimates of risk perceive higher levels of risk if they are less numerate (Dieckmann et al. 2009; Trevena et al. 2013). Literature reviews of risk communication generally recommend that communicators account for variation in numeracy across populations at risk due to the relationship between numeracy and probability comprehension (Spiegelhalter 2017; Ripberger et al. 2022a). Further, numerical presentations of risk can be impacted by anchoring effects, described by Kahneman (2011) as when people consider a value for an unknown quantity before estimating that quantity – say, the level of risk that a probabilistic forecast is suggesting. If multiple numerical estimates of probability are presented in a risk message, receivers are most likely to "anchor" to the number they see first and neglect numbers presented later in the message (Visschers et al. 2009). Numeracy level and anchoring effects must be taken into account when studying any new presentation of risk information, especially if that risk information is to be paired with other forms of numerical risk representation.

Although anchoring effects and individual numeracy can impact an individual's ability to interpret all types of probabilistic risk information, unique interpretation challenges also exist specifically for rare events and the low forecast probabilities that capture their likelihood. In general, people tend to either overweight the odds of a rare event occurring or ignore the slim possibility of a rare event occurring altogether (Kahneman 2011). The context of the rare event can explain this split in behavior, as visceral, attention-grabbing rare events (terrorist attacks, nuclear disasters) are generally overweighted compared to more familiar, unassuming rare events (like skin cancer, Kahneman 2011). Further, the presentation of risk information can lead to rare events being overweighted, as Yamagishi (1997) identified that study participants saw a cancer that "kills 1,286 out of 10,000" (a 12.86% chance) as riskier than one that "kills 24.14 out of 100" (a 24.14% chance), a phenomenon now described as "denominator neglect" (Kahneman 2011). The likelihood of a rare event can also be underweighted or ignored when the event has not impacted an individual directly or been discussed at length on social media or in the news, due to the well-documented heuristic of recency bias (Hertwig et al. 2003; Weber 2006; Kahneman 2011). This suggests that people that are less familiar with tornadoes and their impacts may underweight their risks from tornadoes while those who have seen visceral examples of tornado impacts or are shown risk messages that present larger numbers may overweight their risk.

Risk perception of rare events is also impacted by the high rate of rare event forecast false alarms, which are due to rare events generally being always forecast to occur more often than they are actually observed. False alarms occur at a much higher rate than forecast "hits" for both earthquake forecasts (Woo and Marzocchi 2014) and for the tornadoes forecast by the SPC convective outlook (Hitchens and Brooks 2012). This impacts both economic and personal safety decisions made with regards to rare event forecasts. Economically, it may not be "rational" to spend money mitigating the impacts of rare events or suffering the opportunity costs of sheltering actions when a forecast of a rare event more commonly results in a false alarm, as Simmons and Sutter (2013) found with regards to building residential tornado shelters across tornado prone regions of the US. From an individual safety perspective, communicators are often greatly concerned about the "false alarm effect" that can lead to public complacency and loss of trust in risk communicators (Breznitz 1984). Studies of the false alarm effect with regards to tornado forecasts have found evidence of a reduction in protective action behavior and trust in communicators in locales with high false alarm rates (Trainor et al. 2015; Ripberger et al. 2015a), but that the size of the effect may not be large or even statistically significant in many cases (Lim et al. 2019). Further, public definitions of a "false alarm" tornado can vary based on their personal definition of a "hit" versus a "miss" (Trainor et al. 2015), and that missed events that occur without any forecast lead to a much more significant loss in trust than false alarms (Ripberger et al. 2015a). Overall, these findings suggest that false alarms impact public perception and interpretation of rare event forecasts, but that this effect may be an unavoidable and to some extent acceptable part of forecasting inherently uncommon events like tornadoes.

4.1.2 Rare Event Communication Solutions

The challenges presented by the low absolute probabilities and frequent false alarms associated with rare event forecasts are significant, but studies of risk communication across disciplines have suggested a variety of solutions that may help meteorologists communicate rare events risks more effectively. Murphy (1991) suggested that dividing the likelihood of occurrence of a rare event by the climatological likelihood of occurrence of that event, to highlight that even though "the occurrence of the event on this occasion is less likely than its nonoccurrence, but it is considerably more likely "today" than it is climatologically". This climatology-based approach to contextualizing forecast rare event likelihood has also been suggested for earthquake forecasting, as Woo and Marzocchi (2014) note using the March 2011 Tōhoku earthquake in Japan. After a 7.2 magnitude earthquake occurred in the area on March 9th, the weekly forecast odds of a greater than 8.5 magnitude earthquake increased from their background probability of 0.0012% to 0.12% - an absolute probability that is still quite small, but 100 times larger than the background probability (Woo and Marzocchi 2014). The authors suggest that using this relative comparison of the forecast likelihood of an earthquake to its background likelihood may have helped communicate risk before the 2009 L'Aquila earthquake in Italy in particular. After the deadly quake, several scientists were jailed for correctly presenting the absolute risk of an earthquake as low, which members of the public and government regarded as an all-clear message and thus were unprepared when the earthquake did strike the region (Pappas 2012; Woo and Marzocchi 2014; Spiegelhalter 2017).

Studies of health risk communication have also found that communicating small changes in absolute risk by communicating the difference between background and predicted risk results in more protective actions by message recipients. Most recently, with regards to COVID-19 vaccine side effect risk communication, Costa-Font et al. (2021) suggested that presenting rare vaccine side effects, like heart conditions, alongside the base rate of such heart conditions could help reduce public concern about such rare side effects and result in increased vaccine uptake. Nuance to this suggestion is noted by Lipkus (2007), who found across health literature that relative risks that suggest a risk is X times higher than another could lead to overestimation in perceived risk and thus could raise ethical questions about persuasion. More recent literature reviews of relative risk information in the medical sciences also highlight the occurrence of risk overestimation with relative risk presentations and suggest that relative risk formats can be very persuasive for encouraging recipient responses (Fagerlin et al. 2011; Trevena et al. 2013). Other studies of tornado warning response have found a similar impact on concern and protective action from consequence-based messaging (e.g., the use of words like “light” or “devastating” to communicate tornado impacts), wherein more severe language leads to an increase in response to risk information, but also to a decrease in safer “shelter in place” behavior and an increase in dangerous evacuation behaviors (Ripberger et al. 2015b). To more ethically present risks in all contexts, Lipkus (2007) suggests that relative risk information should only be presented alongside absolute risk information, to help message recipients gain a more accurate understanding of the risk posed to them.

Studies of meteorological risk communication have begun to suggest that the benefits of relative risk communication as found in the medical community may have value in communicating weather risks. Interviews with NWS partners, both in the broadcast (Ernst 2020) and emergency management (Klockow-McClain et al. 2020) have identified that high-level weather information users are seeking greater local context in forecasts of weather risk. Broadcasters have identified that the reliance of the SPC outlook on absolute likelihoods of severe weather prevents areas less prone to severe weather from ever receiving higher end outlooks, like Moderate or High, and that they have difficulty describing why an Enhanced or Slight outlook category for their area suggests unusually high likelihoods of severe weather for their area (Ernst 2020). Emergency managers have also requested more information about the frequency with which the different categorical outlook levels are issued in their areas, due to the rarity of higher end outlooks in many parts of the country at different times of year (Klockow et al. 2020). Despite this work with partners, however, few studies have investigated the interest in risk context information within the public.

One of the few studies that has attempted to understand how alternative risk information formats are received by the public is LeClerc and Joslyn (2012), which used odds ratios to present the likelihood of freezing conditions that participants were asked to decide whether to salt roads for. Unlike relative risks, which are the ratio of likelihood of an event occurring vs the background likelihood of that event, odds ratios are the ratio of the odds of an event occurring vs the background odds of the event occurring. Compared to participants shown deterministic forecasts and absolute probabilities of freezing temperatures, those shown odds ratio forecasts were more likely to make cautious decisions with regards to road salting; the same result described by medical studies investigating relative risks (Lipkus 2007; Fagerlin et al. 2011; LeClerk and Joslyn 2012; Trevena et al. 2013). Although road salting decisions made using absolute probability forecasts were found to result in the highest expected value, participants rarely salted roads when shown that the absolute likelihood of freezing temperatures was between 17 and 31%. Overall, LeClerk and Joslyn (2012) conclude that odds ratio information encourages more precautionary action, particularly when absolute likelihoods are low, but that false-alarm effects of odds ratio forecast presentations needed future study.

4.1.3 Hypotheses and experimental goals

Considering prior studies of rare event risk communication and alternatives to deterministic and absolute probability forecasts alongside my own research goals, I designed a set of three experiments to test public reactions to relative risk forecasts for tornadoes based on the absolute probabilities presented in the SPC convective outlook (see Table 4.1 for a full list of hypotheses). First, I sought to identify how perceived concern and likelihood of response varied when members of the public were shown tornado forecasts with absolute risk, relative risk, or both risk formats. Based on findings in both health literature and meteorology studies (Lipkus 2007; Fagerlin et al. 2011; LeClerk and Joslyn 2012; Trevena et al. 2013), I hypothesized that relative risk information will greatly increase participant concern and likelihood of response to tornado hazards, but that combining absolute and relative risk information would only result in a small increase in both measured values. I also hypothesized that less numerate participants would rate their concern more similarly to more numerate participants when shown relative risk information, as the context provided by relative risk information would lower their comparatively heightened risk perceptions (Dieckmann et al. 2009; Trevena et al. 2013). Finally, I hypothesized that participants that can objectively discern tornado watches from warnings (used as a proxy for familiarity with tornado hazards) would have higher concern when shown relative risk information, as prior work has found that familiarity with rare events to be related to an overestimation of personal risk from rare events (Hertwig et al. 2003; Weber 2006; Kahneman 2011).

To expand upon my first experiment, I next sought to identify whether an anchoring effect existed when presenting relative or absolute risk information first, as well as how increasingly large values of relative risk impacted perceived concern, likelihood of response, and participant's perceived reasonableness of the risk information. As anchoring literature suggests that individuals base their interpretations on the first number in sequence they perceive (Visschers et al. 2009; Kahneman 2011), I hypothesized that the larger relative risk values would result in higher levels of concern and likelihood of response when presented before absolute risk values, although I also hypothesized that reasonableness would not change across changes in information presentation order. Further, given the large values of relative risk found when values for the contiguous US were calculated, I wanted to identify how participants' concern, likelihood of response, and perceived reasonableness of the forecast prompt changed when relative risk increased to values as high as 100, 200, or even 500 times more likely than normal. I hypothesized that individuals' concern and likelihood of response would increase less with increasing levels of relative risk if absolute risk was not changed, and that larger relative risk forecasts would be seen as increasingly unreasonable.

Finally, for my third experiment, I wanted to understand how relative risk forecasts for tornadoes that lead to no observed nearby tornadoes – a false alarm event – would impact trust in the forecast product in the future. As false alarm effect literature suggests that the negative effects of missed forecasts for hazardous weather are small (Trainor et al. 2015; Ripberger et al. 2015a; Lim et al. 2019), I hypothesized that most participants would not lose trust in absolute and relative risk forecasts after a false alarm occurred, but that the number of participants that lost trust in the forecast would increase with larger forecast values of relative risk.

Table 4.1: Complete list of hypotheses, broken up by experiment number, for this study.

Hypothesis Number	**Hypothesis Text**
Experiment One	

H1	When presented with **relative risk** information, participants will report **much greater concern and likelihood of response** than those shown only absolute probabilities.
H2	When presented with **combined absolute and relative risk information**, participants will report **increased concern and likelihood of response** than those shown absolute probability information alone, but **less so than for relative risk** information alone.
H3	**Less numerate** participants will **not report higher levels of concern** than more numerate participants when shown **relative risk** information.
H4	Participants able to **objectively discern the difference between a tornado watch and a tornado warning** will have **higher levels of concern** when shown **absolute or relative risk** information than those that cannot.
Experiment Two	
H5	When **relative risk** information is presented to participants **before absolute risk** information, they will report a **higher level of concern and likelihood of response** than if absolute risk information was presented first. **Perceived reasonableness will not change** across changes in presentation order.
H6	When participants are shown **increasingly large values of relative risk**, their **concern and likelihood of response will increase less** between increases in risk value, and their **perceived reasonableness** of the forecast will **decrease**.
Experiment Three	
H7	When participants are told that a tornado forecast prompt containing **absolute and relative risk** forecast information **results in a false alarm** for their area, participants will **not report a significant loss of trust** in the forecast, although **loss of trust will be higher** for **higher relative risk** values than lower ones.

4.2 Methods

To investigate my set of hypotheses, this study uses data from the 2021 and 2022 versions of the Severe Weather and Society Survey (referred to as WX21 and WX22 henceforth), an annual survey of US adults (age 18+) on topics related to their reception, response, and awareness of severe weather forecast information (Krocak et al. 2021; Bitterman et al. 2022). The survey was developed by the University of Oklahoma Institute for Public Policy Research and Analysis (OU IPPRA) and administered by the research and marketing company Quatrics through a dynamic sampling process. As part of the dynamic sampling process, e-mail invitations to the WX21 and WX22 surveys were sent first to a group of panelists that matched the demographic makeup of the US population, with additional invitations sent to panelists from groups that were underrepresented in the sample. This resulted in a demographically representative sample of US Adults that matches the most recent US Census data (see Table 4.2) with 1550 and 1409 respondents for the WX21 and WX22 surveys, respectively. Reference reports summarizing the results of every version of the Severe Weather and Society Survey, as well as de-identified data, can be accessed online at https://dataverse.harvard.edu/dataverse/wxsurvey.

Table 4.2: Demographics of the US public, as determined by US Census estimates, as compared to the participants for the WX21 and WX22 surveys.

	2010-2018 Census Estimate (%)	WX21 Participants (%)	2020-2021 Census Estimate (%)	WX22 Participants (%)
Gender				
Male	51.3	51.3	50.9	51.8
Female	48.7	48.7	49.1	48.2
Age				

18 to 24	12.0	12.0	11.7	11.4
25 to 34	18.0	18.0	17.6	19.3
35 to 44	16.3	16.3	16.8	17.8
45 to 55	16.4	16.4	15.8	16.3
55 to 64	16.7	16.7	16.6	14.9
65 and up	20.6	20.6	21.6	20.3
Ethnicity				
Hispanic	16.3	16.3	83.0	83.0
Non-Hispanic	83.7	83.7	17.0	17.0
Race				
White	77.9	77.9	77.4	78.3
Black or African American	13.0	13.0	13.1	14.6
Asian	5.9	5.9	6.1	3.3
Other Race	3.2	3.2	3.4	3.8
NWS Region				
Eastern	31.6	31.7	31.7	32.2
Southern	27.1	27.1	27.3	27.8
Central	20.7	20.7	20.5	21.7
Western	20.6	20.5	20.5	18.3

Although the WX21 and WX22 surveys both contain a broad suite of questions, this study will only focus on a subset of those questions related to my research questions. In both surveys, I asked participants to answer a series of questions designed to objectively measure their numeracy through the Berlin Numeracy Test (BNT; Cokely et al. 2012). The BNT uses a series of multiple-choice questions to score participant numeracy on a scale from 1-7, starting with a series of four multiple-choice questions that measure base numeracy. The Severe Weather and Society Survey uses the BNT-S version of this test, which adds three additional questions proposed by Schwartz et al. (1997) that increase the sensitivity of the BNT for participants with less experience or education with probabilities and probability math (see Table 4.3). The BNT-S is also an adaptive test, presenting participants with different questions based on their ability to correctly answer preceding questions in the test, reducing the time and effort needed for participants to complete the test without compromising it's scoring as compared to the full BNT (Cokely et al. 2012). Multiple published papers have been able to use the numeracy data from the Severe Weather and Society Survey to draw insightful conclusions about weather messaging interpretation (Ernst et al. 2021; Krocak et al. 2022; Ripberger et al. 2022b). For simplicity in this study, I group participants into a "low" numeracy group, which encompasses numeracy score values of 1-3, and a "high" numeracy group that includes participants with scores that range from 4-7.

Table 4.3: List of questions included in the BNT-S adaptive numeracy test, along with the percentage of participants in the WX2021 and WX2022 surveys that correctly answered each question.

Numeracy Question Prompt	**WX2021 % Correct**	**WX2022 % Correct**
Imagine that we flip a fair coin 1000 times. What is your best guess about how many times the coin would come up heads in 1000 flips? *(Verbatim, Answer = 500.)*	53.35%	57.35%
In the BIG BUCKS LOTTERY, the chance of winning a $10 prize is 1%. What is your best guess about how many people would win a $10 prize if 1000 people each buy a single ticket to BIG BUCKS? *(Verbatim, Answer = 10.)*	46.77%	44.57%
In ACMEPUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car? *(Verbatim, Answer = 0.1)*	16.90%	16.18%
Out of 1000 people in a small town 500 are members of a choir. Out of these 500 members in a choir 100 are men. Out of the 500 inhabitants that are not in a choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability as a percent. *(Verbatim, Answer = 25.)*	11.55%	7.17%
Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)? *(Verbatim, Answer = 30.)*	27.74%	11.00%
Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the	14.77%	2.20%

other numbers. On average, out of 70 throws how many times would the die show the number 6? *(Verbatim, Answer = 20.)*		
In a forest, 20% of the mushrooms are red, 50% are brown, and 30% are white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red? Please indicate the probability as a percent. *(Verbatim, Answer = 50.)*	7.10%	0.78%

The second variable I used to differentiate participants in this study was through a set of questions that were designed to measure whether participants could objectively differentiate between tornado watches and tornado warnings. I randomly assigned half of my participants a prompt that described a tornado watch, while the other half was shown a prompt describing a tornado warning (see table 4.4). Participants were asked to decide, to the best of their knowledge, whether the assigned prompt described a tornado watch, tornado warning, or if they did not know which product the prompt described. I could then compare participant answers across my research questions by whether they were able to correctly define the prompt as a tornado watch or warning, which I call objective tornado watch/warning understanding. Finally, I also tested other demographic variables, including education level, racial identity, and gender, across all three experiments, but did not find significant or novel patterns of response that differed from the overall population and thus do not discuss breakdowns by these variables in this paper.

Table 4.4: Table of prompts shown to participants to measure objective tornado watch/warning understanding, along with the percentage of participants in the WX2021 and WX2022 surveys that correctly answered each question.

Objective Tornado Watch/Warning Question Prompt	WX2021 % Correct	WX2022 % Correct
The next few questions focus on severe thunderstorms and tornadoes. They may be relatively rare in your area, but severe thunderstorms and tornadoes can happen in every state.		

To the best of your knowledge, is the following alert considered a tornado watch or a warning?		
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.	66.12%	66.05%
This alert is used when a tornado is imminent. When this alert is issued, seek safe shelter immediately.	78.76%	79.54%

For my first experiment, in the WX21 survey, participants were randomly assigned one of six conditions and then asked to rank their level of concern and likelihood of changing their plans for the day from 0-100 (see table 4.5). The conditions varied by risk information type (absolute risk, relative risk, or both formats) and by the likelihood value (2% and 20 times greater than normal risk or 5% and 50 times greater than normal risk). I chose the relative risk values of 20 times greater than normal (henceforth shortened to 20x) for the 2% and 50x for the 5% tornado risk levels using the lowest contoured level of the SPC's weekly tornado probability graphic, which is 0.10% (SPC 2022). Participants were told they received this forecast at 8:00 AM on a Saturday to both retain consistency with other SPC forecast experiments in the Severe Weather and Society Survey (Krocak et al. 2021; Bitterman et al. 2023) as well as to have participants imagine their behavior on a day of the week where they were likely to have plans that could be changed. I also chose to ask participants to rank their concern and likelihood of response from 0-100 to frame their responses in a format similar to the probabilities they were being asked to interpret.

Table 4.5: List of prompts and questions shown to participants in the first relative risk experiment during the WX21 survey (Krocak et al. 2021).

Experiment One (WX21)	
Question Group	Question Wording
Forecast Risk Prompt	Now, imagine that it is a *different* Saturday morning at 8:00 AM and you see this forecast:

	Random Assignment: - There is a 2 percent chance of tornadoes today. - There is a 5 percent chance of tornadoes today. - Tornadoes are 20 times more likely today than on an average day like today. - Tornadoes are 50 times more likely today than on an average day like today. - There is a 2 percent chance of tornadoes today; that means that tornadoes are 20 times more likely today than on an average day like today. - There is a 5 percent chance of tornadoes today; that means that tornadoes are 50 times more likely today than on an average day like today.
Level of Concern	On a scale from 1 to 100, where 0 means _not at all concerned_ and 100 means _extremely concerned_, how concerned would you be if you were to get this forecast? (*Verbatim answer, type number between 0 and 100*).
Likelihood of Response	On a scale from 1 to 100, where 0 means _not at all likely_ and 100 means _extremely likely_, how likely is it that you would change your plans for the day if you were to get this forecast? (*Verbatim answer, type number between 0 and 100*).

Our second experiment on public interpretation of relative risk forecast information was conducted in the WX22 survey, and sought to understand how public concern, likelihood of response, and perceived reasonableness changed when participants were shown relative risk forecasts with increasingly large values (see table 4.6). Participants were prompted with a combined relative risk and absolute risk forecast for their area, using the same 8:00 AM on a Saturday timing that was used in experiment one for consistency. Although the absolute risk was kept at 15%, which translates to a Moderate risk for severe weather in the SPC convective outlook, I varied the relative risk shown to participants across five values – 20x, 50x, 100x, 200x, and 500x. I determined the scale for these relative risk values by estimating the relative risk for a group of tornado events, including April 7th, 2006; October 6th, 2010; June 1st, 2011; and May 26th, September 1st, and December 10th, 2021 (see table 4.7), to identify a range of relative risk values that could be reasonably expected to occur across a variety of tornado events. Concern and likelihood of response were again measured from 0-100, to ensure comparability with experiment one, and I added a question about the perceived reasonableness of the forecast to understand whether participants thought higher values of relative risk were justified given the 15% absolute risk of tornadoes (see table 4.6).

Table 4.6: List of prompts and questions used in the second experiment, presented to participants in the WX22 survey (Bitterman et al. 2022).

Experiment Two (WX22)	
Forecast Risk Prompt	Forecasters often use a combination of phrases, scales, probabilities, and graphics to describe the risk of severe thunderstorms and tornadoes in an area. We want to know how you interpret these forecasts. To begin, imagine that it is a Saturday morning at 8:00 AM and you see this forecast: *Random assignment (two different orderings, five different relative risk values)*: - Tornadoes are (20/50/100/200/500) times more likely today than on an average day like today; this means that there is a 15 percent chance of tornadoes in your area today. - There is a 15 percent chance of tornadoes in your area today; this means that tornadoes are (20/50/100/200/500) times more likely today than on an average day like today.
Level of Concern	On a scale from 1 to 100, where 0 means *not at all concerned* and 100 means *extremely concerned*, how concerned would you be if you were to get this forecast? (*Verbatim answer, type number between 0 and 100*).
Likelihood of Response	On a scale from 1 to 100, where 0 means *not at all likely* and 100 means *extremely likely*, how likely is it that you would change your plans for the day if you were to get this forecast? (*Verbatim answer, type number between 0 and 100*).
Reasonableness of Forecast	On a scale from 1 to 100, where 0 means *not at all reasonable* and 100 means *extremely reasonable*, how reasonable does this forecast seem to you? (*Verbatim answer, type number between 0 and 100*).

Table 4.7: Table displaying the events used to define the levels of relative risk that would be presented to participants in experiment two. The location of the absolute risk and climatological

risk (recovered from the SPC's weekly tornado climatology maps, SPC 2023) for each event is provided, as well as the relative risk calculated from the two risk values.

Event Date	Location	Absolute Risk	Climatological Probability (SPC map)	Relative Risk
April 7th, 2006	Northern Alabama	60%	0.6%	100x
October 6th, 2010	Central Arizona	5%	<0.1%	>500x
June 1st, 2011	Western Massachusetts	5%	0.1%	50x
May 26th, 2021	Northern Kansas	15%	1.2%	12.5x
September 1st, 2021	New Jersey	10%	0.1%	100x
December 10th, 2021	Southwestern Kentucky	15%	0.1%	150x

For my third experiment, which was performed immediately after experiment two in the WX22 survey, I asked participants to imagine that the forecast for tornadoes they saw previously was a false alarm, in that no tornadoes occurred in their area after that forecast was made. In the prompt for the one question in this experiment, I highlighted that forecasters can make mistakes in their forecasts, then suggested that the tornado forecast that participants were shown for experiment two resulted in lightning and rain but no tornadoes. I then asked participants to answer on a 5-point Likert scale, from strongly reduce to strongly increase, how much this experience would influence their trust in future forecasts (see table 4.8).

Table 4.8: List of the prompts and questions used to perform experiment three, which was presented to participants in the WX22 survey (Bitterman et al. 2022).

Experiment Three (WX22)	
False Alarm Effect	Forecasters do the best they can, but sometimes they *underestimate* or *overestimate* the probability that a storm system will cause a tornado. We are interested in how these miscalculations might influence your trust in future forecasts. Again, imagine that it is a Saturday morning at 8:00 AM and you see this forecast: (*show same forecast from Forecast Risk Prompt in experiment two*). As the day

	progresses, imagine that you see some rain and lightning, but the storm does *not* produce a tornado within 25 miles of your residence. How would this influence your trust in future forecasts? 1.) It would *strongly reduce* my trust. 2.) It would *reduce* my trust. 3.) It would have *no effect* on my trust. 4.) It would *increase* my trust. 5.) It would *strongly increase* my trust.

4.3 Results

4.3.1 Comparing Absolute and Relative Risk Responses

In my first experiment with relative risk information, I sought to compare how members of the public respond when shown absolute risk forecasts (2% and 5% risk of tornadoes) to when they were shown relative risk forecasts (20 and 50 times greater than normal risk of tornadoes). My first hypothesis (H1) was supported by the data, as when participants of the WX21 survey were shown relative risk information alone, they reported significantly higher levels of concern and likelihoods of response as compared to participants that were shown only absolute probabilities (see fig. 4.1). Participants shown only relative risks were over 30 points more concerned and likely to take action than those shown absolute probabilities across both risk level groups (see Table 4.9), a value that amounts to nearly a third of the possible variation across the 0-100 scale provided to participants to rate their answers.

Figure 4.1 also displays the level of concern and likelihood of response reported by participants that were shown combined absolute and relative risk information in their forecast prompt. Reported values of concern and likelihood of response from this group were significantly larger than those for the absolute risk only group, but meaningfully lower than the relative risk only group, for both levels of risk (see fig. 4.2), supporting my second hypothesis

(H2). Participants prompted with the combined risk format reported average concern values of 27.8 and 37.6 for the "*2% and 20x*" and "*5% and 50x*" risk level groups, respectively (and average likelihood of response of 29.6 and 39.1, see table 4.9). But beyond this hypothesis, and unlike the absolute or relative risk only groups, the participants shown the combined risk format reported an increase of about 10 points in both concern and likelihood of response from the lower to the higher risk prompts (see fig. 4.2). The differences in the two measures from the lower to the higher risk prompts was not significant for either the absolute only or relative only groups (see table 4.9), which suggests that the combined format also led participants to perceive a greater difference in what the 2% and 20x risk prompt and the 5% and 50x risk prompt meant for their forecast risk from tornadoes.

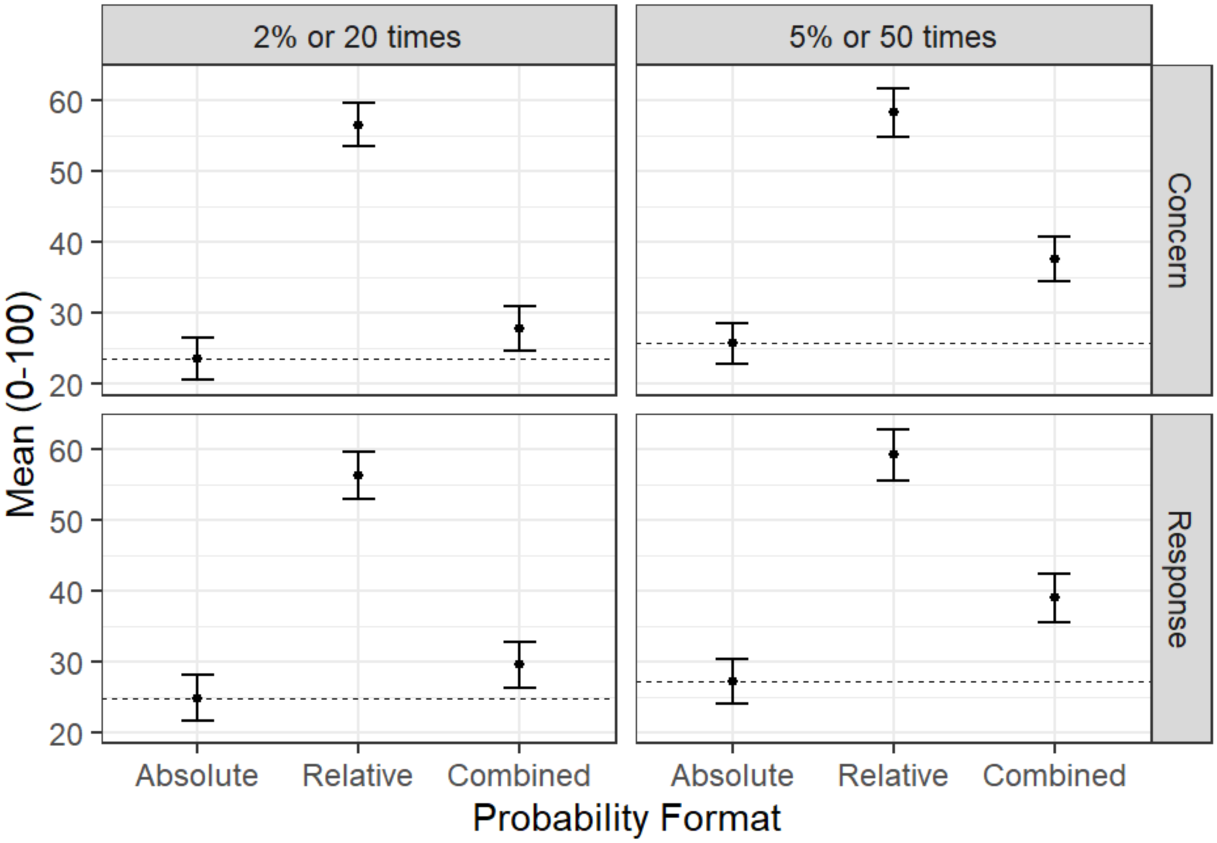


Figure 4.1: Display of the average participant concern and likelihood of response when shown an absolute risk, relative risk, or both formats combined, for the 2% or 20 times greater than normal

and 5% or 50 times greater than normal risk thresholds. Error bars represent the 90% confidence interval around the central tendency. (As an example for how to read the plot, participants in the "Absolute" probability format in the "5 or 50 times" risk group were told that there was a 5% chance of a tornado within 25 miles of their location).

Table 4. 9: Average participant concern and likelihood of response when shown absolute risk, relative risk, or both risk formats, across the two levels of risk shown for experiment one. Values displayed in parenthesis present the standard error for each distribution of responses.

	2% and 20x Risk Level			**5% and 50x Risk Level**		
	Absolute	**Relative**	**Combined**	**Absolute**	**Relative**	**Combined**
Concern	23.5 (1.8)	56.5 (1.8)	27.8 (1.8)	25.7 (1.7)	58.2 (2.0)	37.6 (1.9)
Likelihood of Response	24.8 (1.9)	56.3 (2.0)	29.5 (1.9)	27.2 (1.9)	59.2 (2.1)	39.0 (2.0)

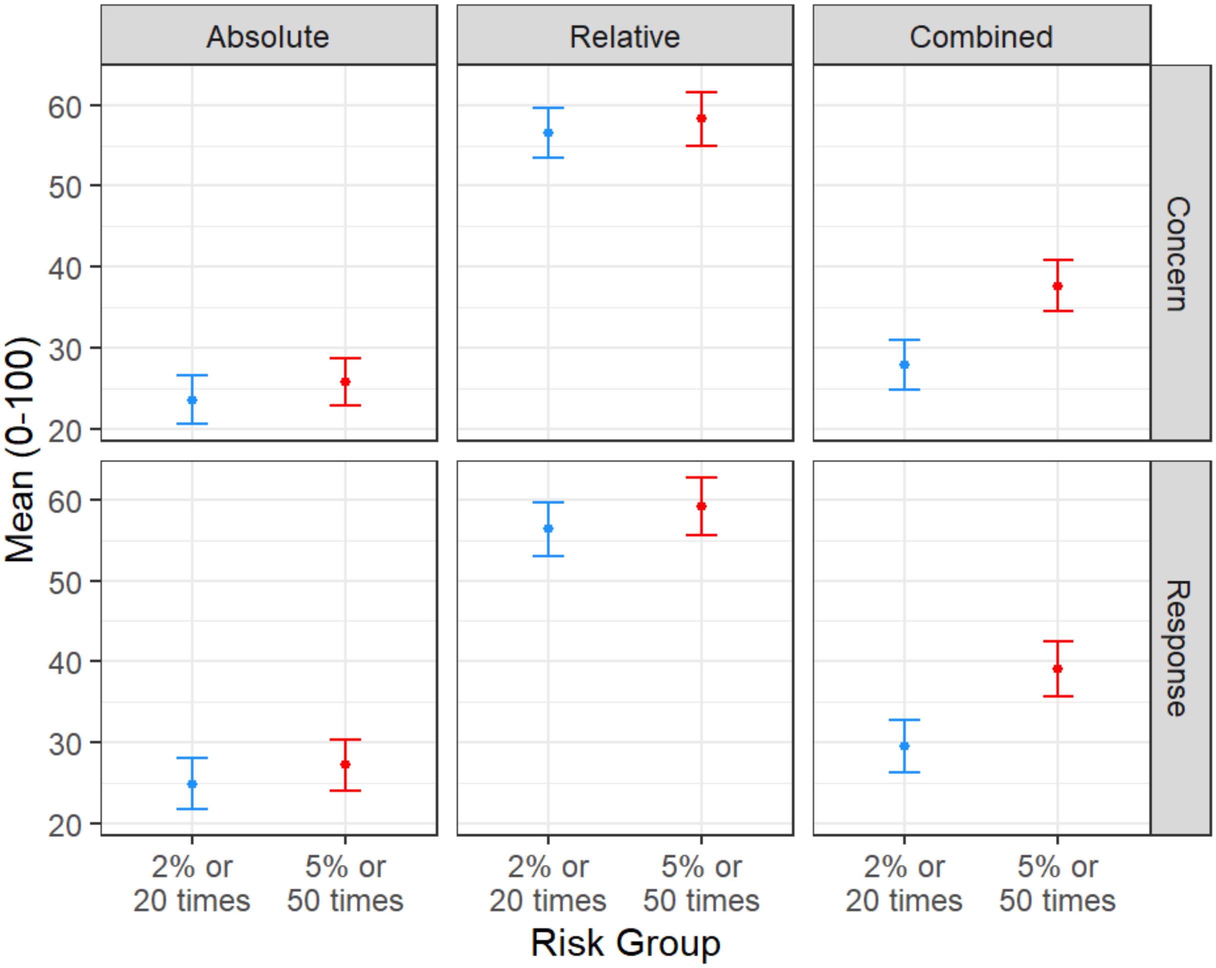


Figure 4.2: Comparison of the average participant concern and likelihood of response across the three information prompt groups, broken out by the risk level presented to participants.

Combining relative and absolute risk information may also improve comprehension of tornado risk information for individuals of different levels of numeracy and tornado watch/warning understanding. As I anticipated in my third hypothesis (H3), more numerate participants were significantly less concerned than less numerate participants when shown only absolute risk, while no significant difference was observed for participants shown only relative risk information (see fig. 4.3). However, less numerate participants were significantly more concerned than more numerate participants when participants were shown the combined risk prompt, which does not support H3. Beyond my hypothesis, more numerate participants displayed a significant increase in concern between the 2% and 5% absolute risk levels while less numerate individuals did not, suggesting the observed lack of increase in concern with increasing absolute risk in the overall survey is driven by low numeracy participants. Although neither group saw much change in concern across the 20x and 50x relative risk prompts, participants from both numeracy levels showed an increase in concern across the combined risk prompts, with a nearly 10-point increase in average concern for less numerate participants and a 5-point increase for more numerate ones.

Unlike this third hypothesis, which was partly supported by the data, my fourth hypothesis (H4) was completely rejected. Participants that incorrectly identified the difference between tornado watches and warnings had higher concern at both risk levels when shown absolute risk information as well as at the 5% or 50x level when shown combined risk information (see fig. 4.4). Additionally, participants in the incorrect watch/warning understanding group reported levels of concern that were not significantly different from correct participants when they were shown relative risk information alone. Although H4 was rejected by these findings, I also identified that incorrect watch/warning participants, similarly to low numeracy participants, displayed a significant increase in concern when moving from the lower to higher risk levels only if they were shown the combined risk information format (see fig. 4.4). Participants that correctly identified tornado watch and warning definitions saw a significant increase in concern across both the only absolute and combined risk information formats across the two levels of risk, similar to the more numerate participants seen in Figure 4.3.

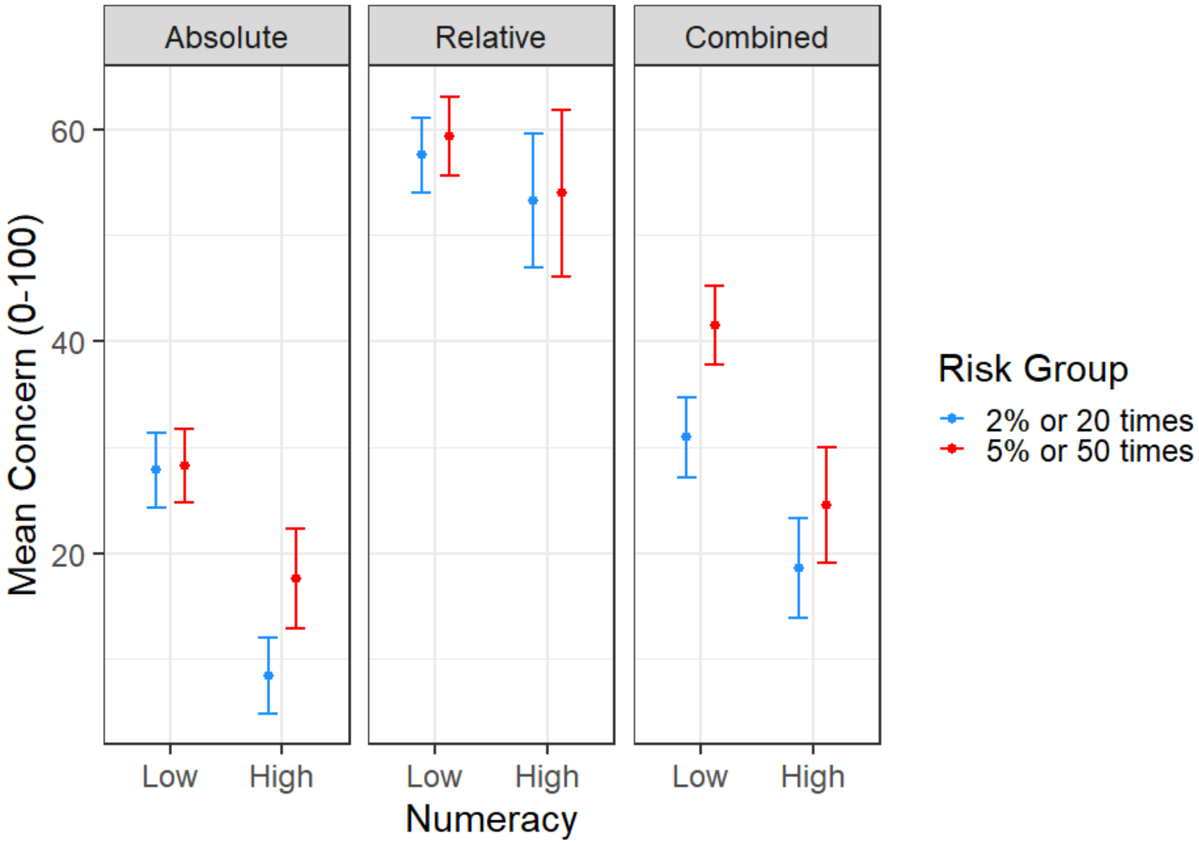


Figure 4.3: Average concern of low and high numeracy participants (numeracy score values 1-3 and 4-7, respectively) across groups shown absolute, relative, and both types of risk information combined at the 2% or 20x or 5% or 50x risk level. Data is broken down by type of risk information shown, numeracy group, and risk level group, with risk levels broken down by color. Error bars represent the 90% confidence interval around the central tendency.

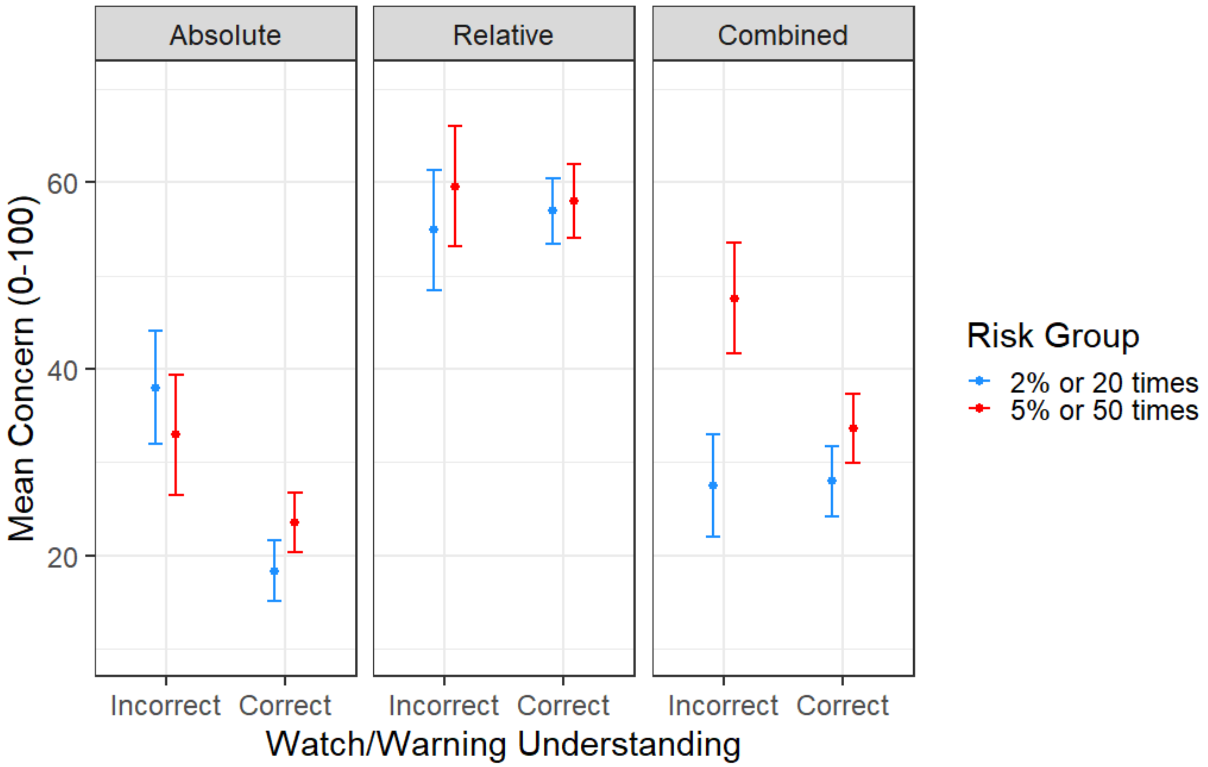


Figure 4.4: Average concern reported by participants that correctly/incorrectly identified tornado watch/warning definitions across groups shown absolute, relative, and both types combined risk information at the 2% or 20x or 5% or 50x risk level. Data is broken down by type of risk information shown, objective watch/warning understanding, and risk level group, with risk levels broken down by color. Error bars represent the 90% confidence interval around the central tendency.

4.3.2 Comparing increasing levels of relative risk

After investigating the impact of different absolute and relative risk forecast prompts on participant concern and likelihood of response, I sought to identify whether an anchoring effect was present in combined presentations of absolute and relative risk, as well as how larger values of relative risk impact perceptions of concern, reasonableness, and likelihood of response. First, my fifth hypothesis (H5) was supported by my data, as participants that were shown relative risk information first had a significantly higher average level of concern and likelihood of response than those shown absolute probabilities first (see fig. 4.5). Further, there was not a significant difference in perceived reasonableness when the order of presentation of absolute or relative risk was changed.

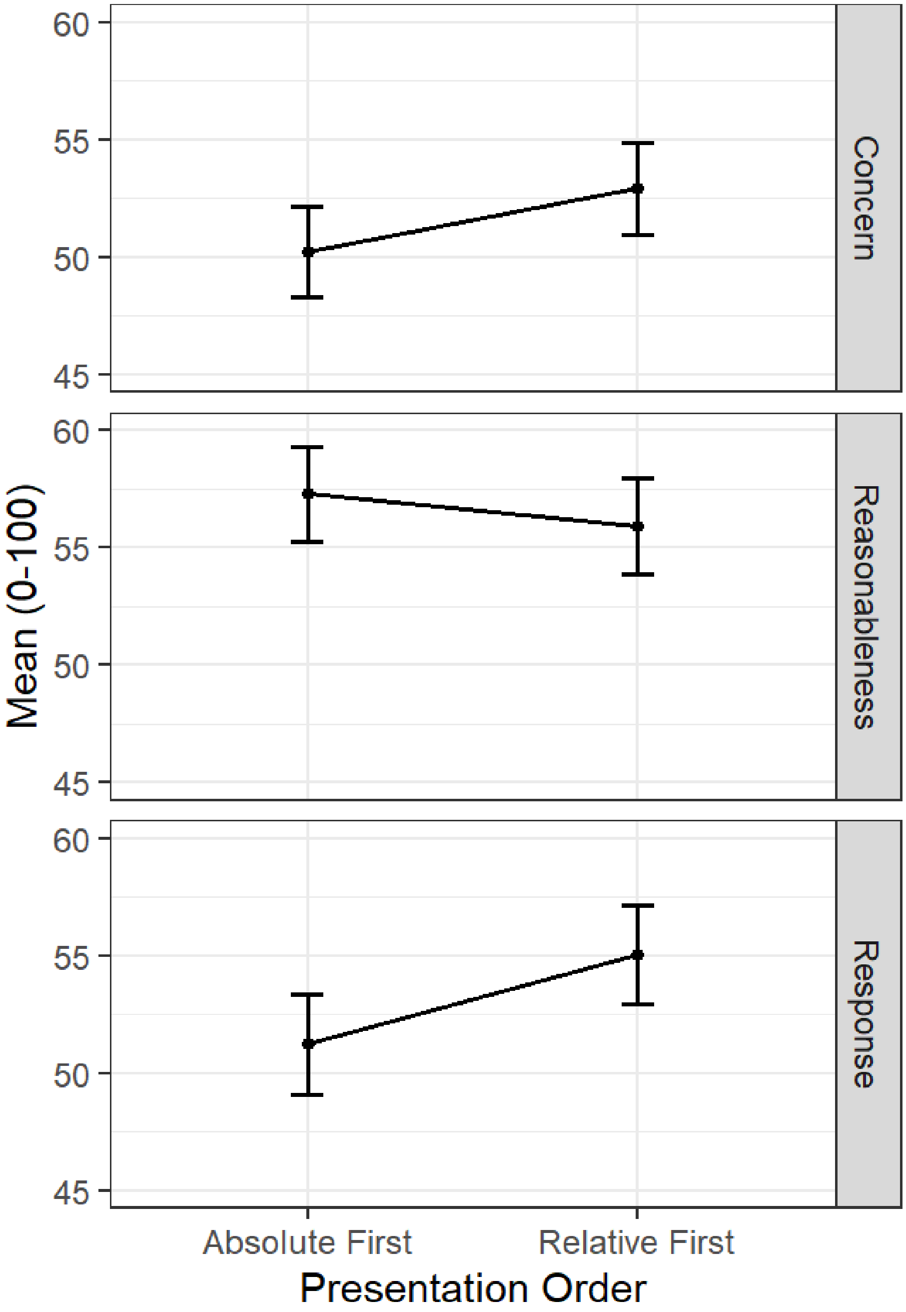


Figure 4.5: Average participant concern, perceived reasonableness, and likelihood of response across participant groups shown absolute risk information first in their forecast prompt versus

those shown relative risk information first. Error bars represent the 90% confidence interval around the central tendency.

After identifying support for H5, I compared my participant responses across different levels of relative risk to investigate the validity of my sixth hypothesis (H6). As the relative risk presented to participants increased from 20x, to 50x, 100x, 200x, and finally 500x, average level of concern remained in the low 50s, increasing slightly but not significantly for participants shown a relative risk of 500x (see fig 4.6). Participants' average likelihood of response decreased a small, but again not significant, amount from the 20x level to the 100x relative risk level, before increasing across the 100x and 500X risk groups. Combined, these results fail to support H6, and suggest that higher values of relative risk presented at a given absolute probability do not have significant impacts on people's perceived concern or likelihood of response. Further, H6's prediction of decreasing perceived reasonableness of the forecast across increasing values of relative risk was not supported, as reasonableness did not significantly change across the levels of relative risk presented here (see fig 4.6). Overall, these findings suggest that participants did not see large values of relative risk, including risk values as high as 500x, as any more concerning or unreasonable to expect or in a forecast of tornado risk than a relative risk of 20x. I also found similar patterns in the measures across participants of different numeracy levels and understanding of tornado watches and warnings, and thus do not present demographic breakdowns for this experiment.

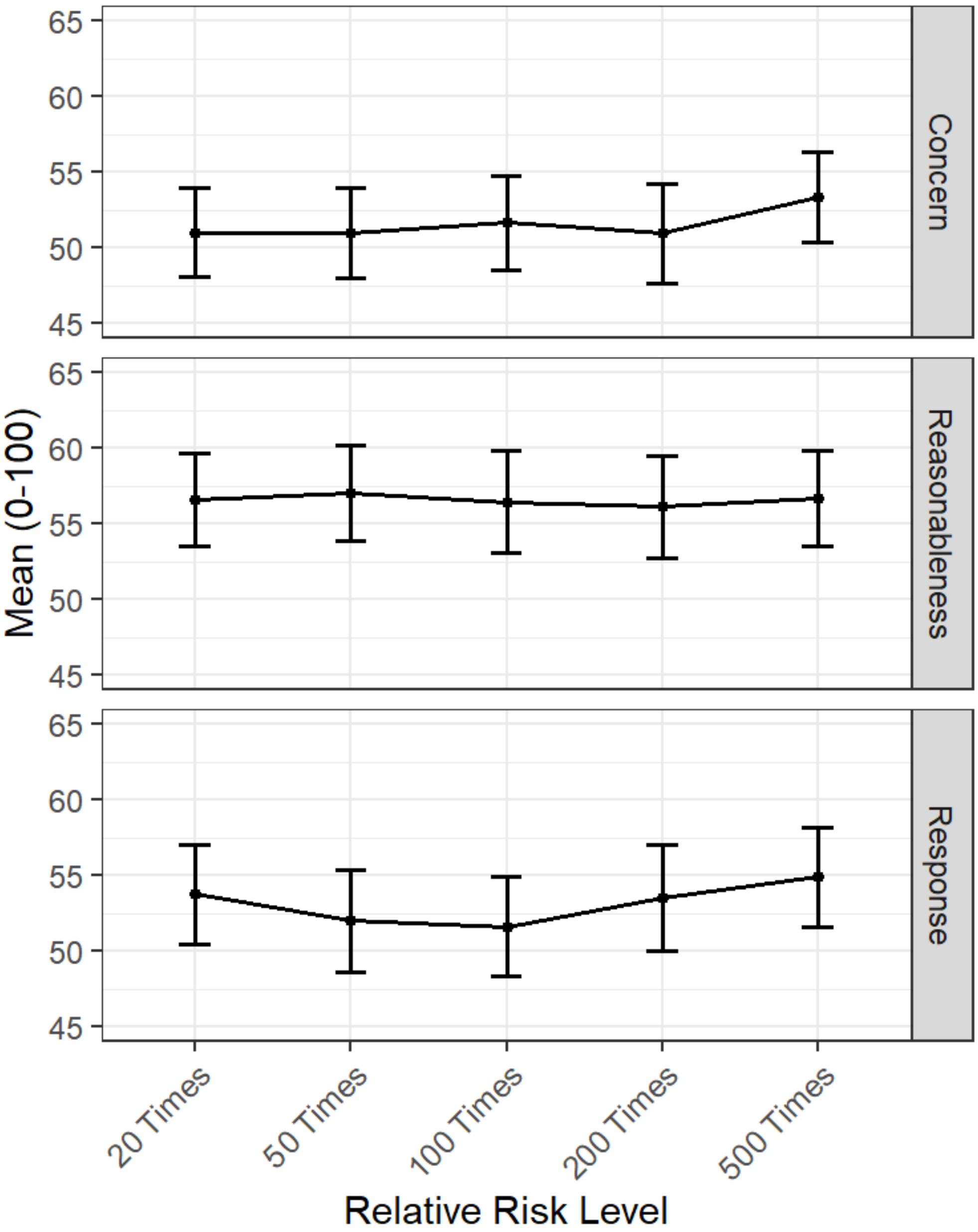


Figure 4.6: Average participant concern, perceived reasonableness, and likelihood of response across the value of relative risk shown to participants. Error bars represent the 90% confidence interval around the central tendency.

4.3.4 Comparing loss of trust across increasing levels of relative risk

Our third experiment sought to investigate my seventh hypothesis (H7), which predicted that participants would not change their trust in the combined absolute and relative risk forecast they were given if that forecast did not result in a tornado near them. I found that the first part of H7 was supported, as over 60% of participants at each relative risk level reported no change in their trust if their forecast prompt resulted in a false alarm (see fig. 4.7). The participants that suggested they would reduce or increase their trust in the forecast after a false alarm were also well balanced, resulting in average participant responses very close to the "No Effect" level of the Likert scale presented to participants (see table 4.10). Thus, the second part of H7 was not supported, as I did not observe an increase in loss of trust with increasing values of relative risk.

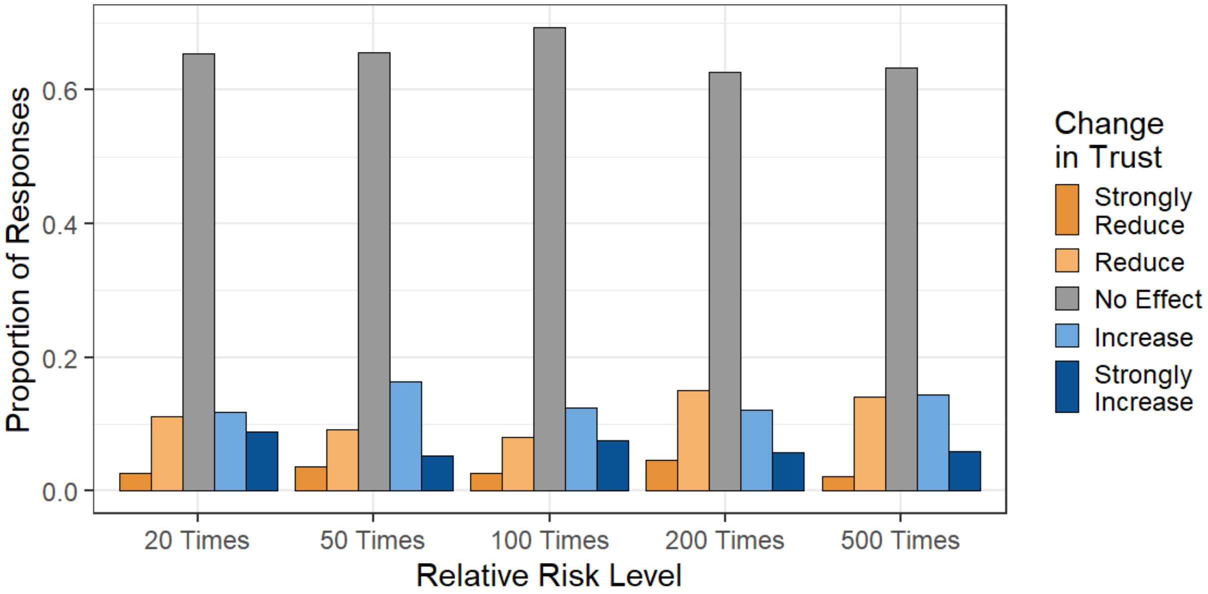


Figure 4.7: Proportion of participants that felt that their trust in a combined absolute risk and relative risk forecast was reduced, increased, or unchanged when that forecast resulted in a false alarm.

Table 4.10: Average responses to the Likert scale question asking participants about their loss of trust after a false alarm (values range from 1 – strongly reduce to 5 – strongly increase). Question

text and scale translations can be reviewed in Table 4.8. Standard error and total number of respondents that were shown each level of relative risk are also reported.

Relative Risk Level	Average Response (1-5)	Standard Error	Participant Count
20 Times More Likely	3.13	0.05	304
50 Times More Likely	3.10	0.05	282
100 Times More Likely	3.14	0.05	264
200 Times More Likely	2.99	0.05	265
500 Times More Likely	3.08	0.05	284

4.4 Discussion

Through a series of three experiments, I explored how members of the public react to relative risk information when asked to imagine a hypothetical scenario where they were given probabilistic forecasts for tornadoes. The results of these survey questions suggest that relative risk information leads to important and measurable changes to individual risk perception and likelihood of response, but that these changes may not be as operationally meaningful as initially theorized. As I hypothesized, survey participants displayed greatly increased concern and likelihood of response when shown relative risk forecasts, aligning with findings in the health and meteorology communication literature that relative risks and other similar risk presentations that incorporate baseline frequency information can more frequently persuade people to take protective actions (Lipkus 2007; Fagerlin et al. 2011; Trevena et al. 2013). Reported participant concern and likelihood of response levels above 55 out of 100 may be evidence of an overestimation of tornado risk, however, given that the 20x and 50x relative risk values were defined by dividing the two lowest absolute probability values in the SPC outlook (2% and 5%) by the lowest value contoured in the SPC climatology graphics (0.1%). Individuals' overestimation of risk when using relative risk forecasts is part of the significant ethical concern surrounding the use of relative risks as highlighted by Spiegelhalter (2017) and others, and suggests that presenting relative risks for tornadoes alone is not an ethically sound way to communicate tornado risk. Additionally, overinflated levels of concern and likelihood of response with regards to a tornado threat may, as suggested by LeClerk and Joslyn (2012), lead to excessive personal costs due to sheltering actions being taken more often. However, combining relative and absolute risk information leads participants to report a level of concern that is higher than what was reported for absolute risk information alone, but appears more subjectively reasonable given the lower absolute likelihood levels present in the first experiment. The combined information format also led participants to report greater separation in level of concern and likelihood of response across the two risk levels (2% and 20x vs 5% and 50x) which suggests that participants interpreted higher and lower levels of personal risk from increasing absolute likelihood forecasts of tornadoes only when relative risk was included in the forecast.

Numeracy and objective ability to discern tornado watches from tornado warnings were also related to changes in participants' level of concern and likelihood of response after they were shown absolute and relative risk information. My prediction that relative risk forecasts would lead less numerate participants report a level of concern that was closer to that of more numerate participants was supported by the data, but only when participants were shown relative risk information alone. When relative and absolute risk information were presented to participants together, less numerate participants rated their level of concern as consistently higher than their more numerate counterparts. However, less numerate participants showed greater increases in their level of concern and likelihood of response between the two levels of absolute likelihood when shown the combined information format than when shown only absolute or relative risk information, which leads me to believe that less numerate members of the public interpret absolute likelihood increases as increases in their personal risk when relative risks are presented with absolute likelihoods. I also identified a similar phenomenon in results from participants that could not correctly define tornado watches or warnings, although my hypothesis that participants who could correctly define watches and warnings would be more concerned about tornadoes was the opposite of the observed results. This may be due to the small impact footprint of tornadoes, as the majority of people that receive tornado warnings (even those with confirmed tornadoes) are never impacted by a tornado, leading to an underestimation of tornado risk due to recency bias as theorized by studies of rare event risk interpretation (Hertwig et al. 2003; Weber 2006; Kahneman 2011).

In my second experiment, I observed evidence for an anchoring effect as described by Visschers et al. (2009) and Kahneman (2011), where participants shown relative risk values before absolute probabilities reported higher levels of concern and likelihood of response without any significant change to their perceived reasonableness of the forecast. This finding supported my hypothesis that an anchoring effect would be present due to the larger values of relative risk (20x, 50x, 100x, 200x, 500x) that I presented to participants when compared to the absolute probability of 15% that was shown to all groups. The observed anchoring effect was strongest across participants’ reported likelihood of response, suggesting that the order that combined absolute and relative risks are presented has a meaningful impact on protective action decisions.

Considering the ethical impacts of a combined risk information anchoring effect, I believe that absolute risk information, as the most unbiased estimate of the potential for tornadoes, should be presented before relative risk information when both are combined. I did not find significant differences in the anchoring effect across demographics, which aligns with Kahneman's (2011) assertion that anchoring effects are due to the inherent quirks of human brain function, and thus universally experienced.

In addition to investigating the presence of an anchoring effect in my second experiment, the second survey questions sought to disprove my hypothesis that increasingly large values of relative risk would see smaller increases in concern and likelihood of response from participants and a decrease in perceived reasonableness of the forecast. Unexpectedly, the data show that there is no significant change in participants' level of concern, likelihood of response, or perceived reasonableness when they are shown relative risk values ranging from 20x to as high as 500x. This result may mean that, when shown absolute and relative risk information together, individuals may hedge their level of concern and likelihood of response based on the absolute likelihood that is shown (which in this case was 15% in all prompts). Although this result suggests that relative risk information had little to no impact on participant's risk judgements, it is possible that participants were more focused on the higher absolute likelihood presented in this experiment than in the first survey, regardless of how much more likely than normal a tornado was forecast to be. It is also worth noting that the consistent level of perceived reasonableness across increasing levels of relative risk suggests that there is not a "cost" to presenting extreme values of relative risk alongside absolute risk information in the form of public distrust of forecasts.

Finally, I hypothesized that loss of trust in forecasts for tornadoes that included relative risk information would be minimal if participants were told that those forecasts ended up being a false alarm, although I also expected a greater loss of trust when forecasts had larger values of relative risk. In the results from the third experiment, I identified that there was little impact on trust in the forecast when participants were told that the forecast they had been shown was a false alarm, but also that there was no significant change in trust in the forecast even with increasing levels of relative risk. The great majority of participants reported that a false alarm would have no effect on their trust in forecasts of tornadoes that presented absolute and relative risk information, counter to the fears presented by LeClerk and Joslyn (2012) about the perception of false alarms after odds ratio forecasts. Although this was an experimental scenario, which incurred no felt costs for a false alarm forecast, this result does lend further credence to the findings of Lim et al. (2019) that the size of the false alarm effect is not large, and that any false alarm effect may be more dependent on an individual's definition of a false alarm for a tornado forecast (as suggested by Trainor et al. 2015).

4.5 Conclusion

When I presented members of the public with forecasts for tornadoes with absolute probability and relative risk information, their responses showed that relative risks on their own could lead people to overestimate their tornado risk. However, combining both information formats led participants to display an increased level of concern when tornado likelihood increased, as compared to participants only shown one format or the other. The results of the second experiment found that the order with which both risk formats were presented suffered from an anchoring effect, where presenting larger relative risk values before smaller absolute probabilities could lead to higher levels of concern and likelihood of response. The presence of this anchoring effect suggests that the most ethical and honest method to present absolute and relative risk information would be to put absolute risk first in the presentation order. In addition, no changes in participants' concern or likelihood of response were noted across increasingly large values of relative risk with an unchanging absolute risk, suggesting that relative risks do not strongly influence individuals' perception of their personal tornado risk. Finally, I found that false alarms after forecasts for tornadoes using absolute and relative risk information did not result in a significant loss of trust in forecasts. Combined, these results suggest that there may be some cases where relative risk information adds value to absolute probability forecasts for tornadoes, particularly when message recipients need to discern between small values of absolute risk. However, I could not find any relationship between relative risk forecasts and perceived concern or overall decision-making intent, particularly for higher values of absolute risk, suggesting that relative risk has at best limited potential for improving tornado risk communication.

In acknowledging these findings, it is also important to highlight that this study has a number of key limitations. First, I did not collect qualitative data on survey respondents' interpretation of relative risk information, as I sought to minimize the length of time it would take participants to complete the already large WxSurvey. Qualitative data of this nature could help explain why participants ranked their level of concern and likelihood of response the way that they decided to, and potentially highlight common thought processes that could better explain participant behaviors. Data collection techniques like focus groups may also reveal greater nuance to how members of the public might interpret relative risk information, answering questions such as what individuals interpret "an average day like today" in the relative risk prompt to mean. Future work should also seek to better understand the range of protective actions that participants might consider taking after being shown relative risk information, similar to the analysis performed in Ripberger et al. (2015b). Such an analysis could highlight how relative risk values relate to changes in what actions people favor, like evacuating before tornadoes instead of sheltering in place, that were not measured in this study. Additionally, I did not test a wide variety of combinations of absolute probabilities and relative risk values, as I used a between-groups design to study differences in concern and likelihood of response, and would have thus resulted in small, non-representative groups of participants for comparing across each condition had I added more combinations. Future studies could use a within-subjects design in future surveys to present a wider variety of absolute and relative risk value combinations for larger numbers of participants to interpret.

Another limitation of the public surveys performed here is the lack of mapped relative risk information provided to participants, as text-based forecasts were prioritized in this study. Future work should investigate whether broadcasters' concerns about how members of the public might interpret relative risk, especially when offsets in absolute and relative maxima occur, are justified by measured public responses. This study also did not account for the factors identified by the first study in this dissertation that could lead to variations in relative risk, such as time of year and geographical location, in the design of forecast prompts presented to members of the public. The level of concern, likelihood of response, and perceived reasonableness that participants report when shown relative risk information regarding peak tornado season in Oklahoma as compared to deep winter in Minnesota may vary significantly, and future work should seek to identify what variations in these measures may occur when participants are given greater local context.

Future efforts to study relative risk forecast products should seek to use a variety of research methods to more deeply explore the risk judgements that people make when evaluating relative risk information. Interviews, either individually or in focus groups, with members of the public may help expose some of these critical judgements, and future work should seek to present interviewees with a variety of combinations of absolute and relative risk to better understand the variation in responses that can occur with different risk value pairs. Additionally, future work should seek to better understand public responses to a variety of impactful weather hazards, including not just the hail and thunderstorm winds also forecast by the SPC, but also winter storms, hurricanes and tropical storms, and both short- and long-term flooding events. It is possible that the visceral and localized nature of tornado hazards leads to inflated levels of concern and likelihoods of response when relative risk information is included in forecasts, and that relative risk information may lead to different changes in concern and response for different weather hazards. Finally, studies of relative risk interpretations should also be expanded to more weather-savvy populations, including the broadcast meteorologists and emergency managers that are core partners for the NWS, to determine how they might use relative risk information to perform their duties as part of the weather enterprise.

Overall, I believe that Murphy (1991) was right to suggest that relative risk information be tested with forecasts of rare and severe events, including tornadoes, but that these tests have revealed that relative risk is likely of only marginal value in presenting tornado risk information to the public. While relative risks, when combined with absolute likelihood information, appear to communicate context for changes in low levels of absolute tornado risk, relative risks did not have any measurable impact on individual concern or likelihood of response at a constant level of absolute risk, even with relative risk values as high as 500 times more likely than normal. The local context provided by relative risks is, on paper, a useful tool for describing the day-to-day and place-to-place significance of hazardous but rare weather events, but it may not be as important to communicate to individuals as information about other aspects of tornado events, including event timing or impact severity.

Chapter 5 – Discussion

For years, forecast product development has centered around the needs and wants of the forecasters producing them. In the last two decades, however, a new generation of scientists have worked to refine the forecast product development process so that it also incorporates user needs and understandings. Multiple efforts across the NWS, from the FACETs initiative (NOAA NSSL 2023a) to the development of the WPC WSSI (Semmens et al. 2022) and the VORTEX-Southeast project (NOAA NSSL 2023b), have started fostering opportunities for interdisciplinary science that helps us learn more about the weather and how people interact with it. Combined, these efforts point to a paradigm shift in how new forecast information is developed, generated, and disseminated to users, ranging from emergency managers to broadcast meteorologists and the public. My investigation into relative risk as a potential information supplement in the SPC convective outlook would not have been possible without the increased support and acceptance of integrated social science and meteorology research these previous efforts have fostered. Indeed, my results in this dissertation build on the decade-plus of efforts made by dozens of other researchers seeking to improve extreme weather risk communication.

Communicating rare and severe events, like the June 1st, 2011, tornadoes that I recall so viscerally, is the great challenge of meteorology today. Although our ability to forecast rare events has improved with time, these forecasts only ever generate value if users are able to interpret and act upon them (Murphy 1993). Recent work has suggested that probabilistic information, rather than deterministic, “yes/no” or “warning/no warning” products, allows non-expert forecast users to make more well-informed protective action decisions and thus generate more value from forecasts (Ripberger et al. 2022a). However, low absolute probabilities for rare events can lead individuals to dismiss those risks, and thus increase their vulnerability to negative consequences if a rare and severe event like a tornado comes to pass (Murphy 1991; Kahneman 2011). Although there are multiple ways to address the low likelihoods of rare events in risk communication, including expanding the time and space that a forecast is valid for, risk communicators have suggested for decades (Murphy 1991; Spiegelhalter 2017) that relative risk information could add valuable context to risks with small absolute likelihoods.

As a well-documented, highly impactful, and extremely rare (at the local level) weather risk, tornadoes present a ripe opportunity for testing the value of forecast relative risk information. Furthermore, broadcast meteorologists, emergency managers, and members of the public have been seeking greater context around tornado forecasts, like those on June 1st, 2011, that relative risk information could potentially provide (LeClerc and Joslyn 2012; Klockow-

McClain et al. 2020; Ernst 2020). To understand whether relative risk forecasts for tornadoes could add the context to forecasts that users are looking for, I first used tornado reports from 1950-2021 to calculate relative risks for every tornado event in that 71-year period. Creating this dataset of relative risks allowed me to see how relative risks for tornadoes change across different tornado events, regions of the country, and time. Analysis of the complete dataset revealed that relative risk values for tornadoes vary greatly across the contiguous US. High values of relative risk, some in excess of 2000 times more likely than normal, were found to occur across the Western and Northern US, where tornadoes are highly infrequent at some or all times of the year. In general, off-season and off-region tornado events resulted in the highest values of relative risk, while much lower values were identified to occur during in-season and in-region events. This analysis also suggested that, because of the decrease in tornado likelihood with increasing latitude in the US, events like April 27th, 2011, could see an offset between the locations of maximum relative and absolute risk values. Finally, comparing the percentile rankings of the absolute likelihood values used in the SPC probabilistic outlook with the percentile rankings of relative risks across the 71-year dataset identified that the 5, 25, 50, 100, 250, and 700 times more likely than normal levels could be roughly compared to the 2, 5, 10, 15, 30, and 45% levels of absolute likelihood for tornadoes. Based on this finding, I divided my maps presenting relative risks across those six levels of relative risk, so that spatial absolute and relative risk distributions during tornado events could be more easily compared.

Once I had data that gave me a better understanding of the distribution of relative risks for tornado events, I sought to identify how broadcast meteorologists would interpret relative risk information and whether they felt that relative risks would be useful to share with their viewers. Across a set of focus group interviews, I found that broadcasters expressed doubts about the ability of their viewers to interpret probability information in general, and that while they felt that absolute risk information could undersell tornado risks, broadcasters believed that relative risk information might come across as overblown if shown on air. Although the assumption that people cannot interpret probability information does not agree with risk communication literature (Ripberger et al. 2022a), broadcasters did suggest that relative risk products could help them personally to better understand and quantify forecast absolute likelihoods for tornadoes during unusual events. Some broadcasters felt that relative risk information may be useful when presented to viewers in map form for rare events that were out of season or outside of the traditional tornado alley, and a few broadcasters even mentioned presenting information similar to relative risks before ever being shown the product. This suggests that broadcasters may be open to using relative risk information as a context aid for events where absolute likelihoods may downplay tornado risks, although they may need to see more evidence of the public's ability to interpret probabilities before using relative risk or absolute likelihood information more often in their shows.

Finally, I presented members of the public with relative risk information through a series of survey questions, to better understand how they would react to the large relative risk values identified in the climatological investigation of relative risk that could occur with extreme events. My initial findings in part supported broadcasters' beliefs that relative risk information could lead to excessive public concern, as on their own relative risks led to very high reported concern and likelihood of response among participants. When combined with absolute risk information, relative risks increased participant concern more modestly, but also led participants to increase their concern more for a 5% and 50x risk scenario as compared to a 2% and 20x scenario. I also observed that participants of varying numeracy levels displayed this larger change in concern across risk levels, suggesting that combining the two risk formats could improve public perception of small changes in tornado likelihood for populations previously identified to have struggled with interpreting SPC products (Ernst et al. 2021).

My second set of survey questions, however, identified that reported levels of concern and likelihood of response changed little for participants when shown increasing relative risk values with a consistent absolute likelihood. Although participants did not perceive increasingly large values of relative risk as less reasonable to expect in a forecast, these results suggest that the value of relative risk for changing how individuals interpret tornado threats may be extremely limited. Overall, the results of my public surveys suggest that relative risk is at best a potential supplement to absolute likelihood information, specifically to help people discern between low levels of tornado likelihood more effectively. I would hesitate to suggest that relative risk be implemented in the probabilistic outlook given these public interpretation results however, especially without studying how individuals change their chosen response actions given different levels of relative risk.

Combined, these three studies display the development path of a potential forecast information product, from establishing its meteorological distribution to identifying its effectiveness through focus groups with and surveys of potential forecast recipients. The results suggest that relative risk is absolutely not a replacement for absolute risk information – as prominent risk communication researchers have warned (Spiegelhalter 2020) – and that relative risks may only add limited value to personal risk assessments made by members of the public at a high time cost in broadcasters' tornado risk coverage. Although this is a personally disappointing result, given my hopes that relative risk could aid communicators like broadcast meteorologists in warning for tornado events like the June 1st, 2011 tornado event I had difficulty explaining to friends, this result is still an important finding that helps us understand the limits of tornado risk communication. The reviews of risk communication literature and tests of relative risk with user audiences performed here have revealed the limitations of this messaging format, but understanding these limitations allows us to focus our efforts on communication techniques and designs with greater promise. Spending money operationalizing ideas with limited positive impact or measurable negative impact is anathema to the construction of a more ethical, equitable, and effective weather risk communication system. Potential risk communication innovations must always be tested with a high level of scientific rigor before a decision can be made on which new forecast product ideas are best suited for operationalization. It is my hope that this dissertation can add to the growing list of literature that helps exemplify a multidisciplinary approach to how we produce and share important weather risk information, and that what I have learned here about relative risk forecasts for tornadoes can help better define our path towards a better rare risk communication system.

References

Beitscher, M., D. Snyder, S. Poulos, A. Schoettmer, J. Lee, G. Goggins, C. Adcock, M. Pfahler, R. Commodari, and A. Wagner, 2023: Forecasting and communicating an historic outbreak. National Weather Service, accessed 13 Feb 2023, https://storymaps.arcgis.com/collections/8243cae7290947b1ab389c56ccb98445?item=1

Bitterman, A., M. Krocak, J. Ripberger, C. Silva, H. Jenkins-Smith, S. Ernst, and S. Stormer, 2022: Maintaining the series: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. Reference Rep., 34pp.

Bitterman, A., M. J. Krocak, J. T. Ripberger, S. Ernst, J. E. Trujillo-Falcón, A. Gaviria-Pabón, C. Silva, and H. Jenkins-Smith, 2023: Assessing public interpretations of original and linguist-suggested SPC risk categories in Spanish. *Wea. Forecasting*, Early Release. https://doi.org/10.1175/WAF-D-22-0110.1

Braun, V. and V. Clarke, 2006: Using thematic analysis in psychology. *Qual. Res. Psychol.*, **3(2)**, 77-101. https://doi.org/10.1191/1478088706qp063oa

Breen, R. L., 2006: A practical guide to focus-group research. *J. Geog. Higher Edu.*, **30(3)**, 463-475, DOI: 10.1080/03098260600927575

Breznitz, S., 1984: *Cry Wolf: The psychology of false alarms*. Lawrence Erlbaum Associates, 280pp.

Brooks, H. E., C. A. Doswell III, and M. P. McKay, 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%3C0626:CEOLDT%3E2.0.CO;2

Brooks, H. E., G. W. Carbin, and P. T. Marsh, 2014: Increased variability of tornado occurrence in the United States. *Science,* **346**, 349-352, https://doi.org/10.1126/science.1257460

Brooks, H. E., and J. Correia Jr., 2018: Long-term performance metrics for National Weather Service tornado warnings. *Wea. Forecasting*, **33**, 1501-1511, https://doi.org/10.1175/WAF-D-18-0120.1

Brooks, H. E., M. Kay, and J. A. Hart, 1998: Objective limits on forecasting skill of rare events. *19th Conferences Severe Local Storms*, American Meteorological Society, Minneapolis, MN, 552-555. https://www.nssl.noaa.gov/users/brooks/public_html/papers/prague2k1.pdf

Burnett, A. J., 2016: 'Impact weather days' alert viewers to potential severe weather. WCVB-TV, accessed 31 October 2022, https://www.wcvb.com/article/impact-weather-days-alert-viewers-to-potential-severe-weather/8235816

CBS 4 News, 2019: What is a 'Code Red' day? Accessed 31 October 2022. https://www.valleycentral.com/news/local-news/what-is-a-code-red-day/

Chaudhry, S. J., M. Hand, and H. Kunreuther, 2018: Extending the time horizon: Elevating concern for rare events by communicating losses over a longer period of time. *Working paper #2018-06*, 66pp. Philadelphia, PA: Risk Management and Decision Processes Center. Accessed April 27 2023, https://riskcenter.wharton.upenn.edu/wp-content/uploads/2018/08/WP201805_Extending-Time-Horizon_Chaudhry-etal.pdf b

Cokely, E. T., M. Galesic, E. Schulz, S. Ghazal, and R. Garcia-Retamero, 2012: Measuring risk literacy: The Berlin Numeracy Test. *Judgem. Decis. Mak.,* **7(1),** 25-47, DOI: http://dx.doi.org/10.1017/S1930297500001819

Costa-Font, J., M. Asaria, and E. Mossialos, 2021: 'Erring on the side of rare events'? A behavioural explanations for COVID-19 vaccine regulatory misalignment. *J. Glob. Health.*, **11**, DOI: https://doi.org/10.7189%2Fjogh.11.03080

Daniels, G. L. and G. M. Loggins, 2007: Conceptualizing continuous coverage: A strategic model for wall-to-wall local television weather broadcasts. *J. Appl. Commun.*, **35(1)**, 48-66, DOI: https://doi.org/10.1080/00909880601065680

Demuth, J. L., B. H. Morrow, and J. K. Lazo, 2009: Weather forecast uncertainty information: An exploratory study with broadcast meteorologists. *Bull. Amer. Meteor. Soc.*, **90(11)**, 1133-1146. DOI: 10.1175/2009BAMS2787.1

Demuth, J. L., R. E. Morss, B. H. Morrow, and J. K. Lazo, 2012: Creation and communication of hurricane risk information. Bull. Amer. Meteor. Soc., 93(8), 1133-1145. DOI: https://doi.org/10.1175/BAMS-D-11-00150.1

Demuth, J. L., R. E. Morss, I. Jankov, T. I. Alcott, C. R. Alexander, D. Nietfeld, T. L. Jensen, D. R. Novak, and S. G. Benjamin, 2020: Recommendations for developing useful and usable convection-allowing model ensemble information for NWS forecasters. *Wea. Forecasting*, **35(4)**, 1381-1406. DOI: https://doi.org/10.1175/WAF-D-19-0108.1

Dieckmann, N. F., P. Slovic, and E. M. Peters, 2009: The use of narrative evidence and explicit likelihood by decisionmakers varying in numeracy. *Risk Anal.*, **29**, 1473-1488, DOI: https://doi.org/10.1111/j.1539-6924.2009.01279.x

Drobot, S., A. R. S. Anderson, C. Burghardt, and P. Pisano, 2014: U.S. public preferences for weather and road condition information. *Bull. Amer. Meteor. Soc.,* **95(6)**, 849-859 DOI: https://doi.org/10.1175/BAMS-D-12-00112.1

Drost, R., M. Casteel, J. Libarkin, S. Thomas, and M. Meister, 2016: Severe weather warning communication: Factors impacting audience attention and retention of information during tornado warnings. *Wea. Clim. Soc.*, **8(4)**, 361-372. DOI: https://doi.org/10.1175/WCAS-D-15-0035.1

Edwards, R. and F. Ostby, 2022: Time line of SELS and SPC. Storm Prediction Center, accessed Mar 13 2023, https://www.spc.noaa.gov/history/timeline.html

Edwards, R., H. E. Brooks, and H. Cohn, 2021: Changes in tornado climatology accompanying the Enhanced Fujita Scale. *J. Appl. Meteor. Climatol.,* **60**, 1465-1482, https://doi.org/10.1175/JAMC-D-21-0058.1

Elliott, V., 2018: Thinking about the coding process in qualitative data analysis. *The Qualitative Report*, **23(11)**, 2850-2861, https://doi.org/10.46743/2160-3715/2018.3560

Ernst, S., 2020: Colorful language: Investigating the interpretation of the Storm Prediction Center's convective outlook by broadcast meteorologists and the US public. Thesis, Dept. of Atmospheric and Geographic Sciences, University of Oklahoma, 115pp.

Ernst, S., J. Ripberger, M. J. Krocak, H. Jenkins-Smith, and C. Silva, 2021: Colorful language: Investigating public interpretation of the Storm Prediction Center convective outlook. *Wea. Forecasting*, **36**, 1785-1797, DOI: https://doi.org/10.1175/WAF-D-21-0001.1

Ernst, S., M. J. Krocak, H. E. Brooks, and J. Ripberger, 2022: Adding context to risk: Climatology of relative probabilities in the SPC convective outlook. *30th Conf. on Severe Local Storms*, Santa Fe, NM, Amer. Meteor. Soc., 1.4, https://ams.confex.com/ams/30SLS/meetingapp.cgi/Paper/407056

Fagerlin, A., B. J. Zikmund-Fisher, and P. A. Ubel, 2011: Helping patients decide: Ten steps to better risk communication. *J. Natl. Cancer Inst.*, **103**, 1436-1443, DOI: https://doi.org/10.1093/jnci/djr318

FEMA, 2017: IS-393.A Lesson 3: Assessing risks. Accessed 21 June 2023, https://training.fema.gov/EMIWeb/IS/IS393a/IS393.A-Lesson3.pdf

Finley, J. P., 1884: Tornado predictions. *Amer. Meteor. J.*, **1**, 84-88.

Galway, J. G., 1985: J. P. Finley: The first severe storms forecaster. *Bull. Amer. Meteor. Soc.*, **66**, 1389-1395, DOI: https://www.spc.noaa.gov/publications/galway/finley1.pdf

Gensini, V. A., A. M. Haberlie, and P. T. Marsh, 2020: Practically perfect hindcasts of severe convective storms. *Bull. Amer. Meteor. Soc.,* **101**, 1259-1278, DOI: https://doi.org/10.1175/BAMS-D-19-0321.1

Gensini, V. A., and H. E. Brooks, 2018: Spatial trends in United States tornado frequency. *npj Clim. Atmos.,* **1**, DOI: https://doi.org/10.1038/s41612-018-0048-2

Gigerenzer, G., R. Hertwig, E. V. Den Broek, B. Fasolo, and K. V. Katsikopoulos, 2005: "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? Risk Anal., 25, 623-629, DOI: https://doi.org/10.1111/j.1539-6924.2005.00608.x

Gill, P. and J. Baillie, 2018: Interviews and focus groups in qualitative research: an update for the digital age. Br. Dent. J., 225, 668-672, DOI: https://doi.org/10.1038/sj.bdj.2018.815

Grams, J., B. Bunting, and S. Weiss, 2014: SPC convective outlooks. Storm Prediction Center, April 9th, 2019, DOI: https://www.spc.noaa.gov/misc/SPC_probotlk_info.html

Hammer, B. and T. W. Schmidlin, 2002: Response to warnings during the 3 May 2022 Oklahoma City tornado: Reasons and relative injury rates. *Wea. Forecasting*, **17(3)**, 577-581. DOI: https://doi.org/10.1175/1520-0434(2002)017%3C0577:RTWDTM%3E2.0.CO;2

Hertwig, R., G. Barron, E. U. Weber, and I. Erev, 2003: Decisions from experience and the effect of rare events in risky choice. *Psychol. Sci.,* **15**, 534-539, DOI: https://doi.org/10.1111/j.0956-7976.2004.00715.x

Hitchens, N. M. and H. E. Brooks, 2012: Evaluation of the Storm Prediction Center's day 1 convective outlooks. *Wea. Forecasting,* **27**, 1580 – 1585, DOI: https://doi.org/10.1175/WAF-D-12-00061.1

Hitchens, N. M. and H. E. Brooks, 2014: Evaluation of the Storm Prediction Center's convective outlooks from day 3 through 1. *Wea. Forecasting,* **29**, 1134 – 1142, DOI: https://doi.org/10.1175/WAF-D-13-00132.1

Hitchens, N. M., H. E. Brooks, and M. P. Kay, 2013: Objective limits on forecasting skill of rare events. *Wea. Forecasting,* **28**, 525 – 534, DOI: https://doi.org/10.1175/WAF-D-12-00113.1

Joslyn, S. L., and J. E. LeClerk, 2011: Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *J. Exp. Psychol.*, **18**, 126-140, DOI: https://psycnet.apa.org/doi/10.1037/a0025185

Juanchich, M. and M. Sirota, 2018: Not as gloomy as we thought: Reassessing how the public understands probability of precipitation forecasts. *J. Cogn. Psychol.*, **31**, 116-129, DOI: https://doi.org/10.1080/20445911.2018.1553884

Kahneman, D., 2011: *Thinking fast and slow*. Farrar, Straus, and Giroux, NY, 499pp.

Kastman, J., and J. A. Nelson, 2020: "Updates to the Winter Storm Severity 664 Index for 2019/20. *30th Conf. on Weather Analysis and Forecasting/26th Conf. on Numerical Weather Prediction*, Boston, MA, Amer. Meteor. Soc., J51.4, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/368357

Keul, A. G. and A.M. Holzer, 2013: The relevance and legibility of radio/TV weather reports to the Austrian public. *Atmos. Res.*, **122**, 32-42. DOI: https://doi.org/10.1016/j.atmosres.2012.10.023

Klockow-McClain, K. E., K. Berry, C. A. Shivers-Williams, M. J. Krocak, K. A. Wilson, J. J. James, G. J. Stumpf, Z. Stanford, A. MacDonald, J. Trujillo, and A. E. Gerard, 2020: Putting multiple probabilistic products before end-users: The 2019 HWT emergency manager experiments. *15th Symposium on Societal Applications: Policy, Research, and Practice*, Boston, MA, Amer. Meteor. Soc., 17.6, https://ams.confex.com/ams/2020Annual/webprogram/Paper369752.html

Knupp, K. R., T. A. Murphy, T. A. Coleman, R. A. Wade, S. A. Mullins, C. J. Schultz, E. V. Schultz, L. Carey, A. Sherrer, E. W. McCaul Jr., B. Carcione, S. Latimer, A. Kula, K. Laws, P. T. Marsh, and K. Klockow, 2014: Meteorological overview of the devastating 27 April 2011 tornado outbreak. *Bull. Amer. Meteor. Soc.*, **95,** 1041-1062. DOI: https://doi.org/10.1175/BAMS-D-11-00229.1

Krocak, M. J., and H. E. Brooks, 2018: Climatological estimates of hourly tornado probability for the United States. *Wea. Forecasting,* **33**, 59-69, DOI: https://doi.org/10.1175/WAF-D-17-0123.1

Krocak, M. J., J. T. Ripberger, S. Ernst, C. L. Silva, H. C. Jenkins-Smith, 2022: Exploring the differences in SPC convective outlook interpretation using categorical and numeric information. *Wea. Forecast.*, **37**, 303-311, DOI: https://doi.org/10.1175/WAF-D-21-0123.1

Krocak, M., J. Ripberger, C. Silva, H. Jenkins-Smith, A. Gaviria-Pabón, A. Forney, and A. Bitterman, 2021: Continuing the series: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. Reference Rep., 33pp.

Krueger, R. A., 2002: Designing and conducting focus group interviews. Accessed 9 April 2020, https://www.eiu.edu/ihec/Krueger-FocusGroupInterviews.pdf

Lazo, J. K., R. E. Morss, and J. L. Demuth, 2009: 300 billion served: Sources, perceptions, uses, and values of weather forecasts. *Bull. Amer. Meteor. Soc.,* **90(6)**, 785-798. DOI: 10.1175/2008BAMS2604.1

LeClerk, J. E., and S. Joslyn, 2012: Odds ratio forecasts increase precautionary action for extreme weather events. *Wea. Clim. Soc.*, **4**, 263-270, DOI: https://doi.org/10.1175/WCAS-D-12-00013.1

Lenhardt, E. D., R. N. Cross, M. J. Krocak, J. T. Ripberger, S. R. Ernst, C. L. Silva, and H. C. Jenkins-Smith, 2020: 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. *J. Operational Meteor.*, **8**, 64-78, DOI: https://doi.org/10.15191/nwajom.2020.0805

Lim, J. R., B. F. Liu, and M. Egnoto, 2019: Cry wolf effect? Evaluating the impact of false alarms on public responses to tornado alerts in the Southeastern United States. *Wea. Clim. Soc.*, **11**, 549-563, DOI: https://doi.org/10.1175/WCAS-D-18-0080.1

Lipkus, I. M., 2007: Numeric, verbal, and visual formats of conveying health risks: Suggested best practices and future recommendations. *Med. Decis. Making*, **27**, 696-713, DOI: https://doi.org/10.1177/0272989x07307271v

Maddox, R. A. and C. A. Crisp, 1999: The Tinker AFB Tornadoes of March 1948. *Wea. Forecasting*, **14**, 492-499, DOI: https://doi.org/10.1175/1520-0434(1999)014%3c0492:TTATOM%3e2.0.CO;2

Mojica, A., 2019: Code red explained. FOX 17 Nashville, accessed 31 October 2022. https://fox17.com/weather/tracking-code-red-days/code-red-explained

Morrow, B. H., J. L. Demuth, and J. K. Lazo, 2008: Communicating weather forecast uncertainty: An exploratory study with broadcast meteorologists. *36th conf. on Broadcast Meteorology*, Denver, CO, Amer. Meteor. Soc., https://www.researchgate.net/publication/280611657_Communicating_weather_forecast_uncertainty_An_exploratory_study_with_broadcast_meteorologists

Murphy, A. H., 1985: Probabilistic weather forecasting. *Probability, Statistics, and Decision Making in the Atmospheric Sciences*. A. H. Murphy and R. W. Katz, Eds., Westview Press, 337-377.

Murphy, A. H., 1991: Probabilities, odds, and forecasts of rare events. *Wea. Forecasting*, **6**, 302-307, DOI: https://doi.org/10.1175/1520-0434(1991)006%3C0302:POAFOR%3E2.0.CO;2

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281-293, DOI: https://doi.org/10.1175/1520-0434(1993)008%3C0281:WIAGFA%3E2.0.CO;2

Murphy, A. H., 1996: The Finley affair: A signal event in the history of forecast verification. *Wea. Forecasting*, **11**, 3-20, DOI: https://doi.org/10.1175/1520-0434(1996)011%3C0003:TFAASE%3E2.0.CO;2

National Weather Service, 2013: The tornado outbreak of May 20th, 2013. NOAA, accessed 13 Feb 2023, https://www.weather.gov/oun/events-20130520

NOAA National Centers for Environmental Information (NCEI), 2011: Monthly Tornadoes Report for June 2011. Accessed 13 Feb 2023, https://www.ncei.noaa.gov/access/monitoring/monthly-report/tornadoes/201106

NOAA National Centers for Environmental Information (NCEI), 2023: U.S. Billion-Dollar Weather and Climate Disasters. https://www.ncei.noaa.gov/access/billions/, DOI: https://www.doi.org/10.25921/stkw-7w73

NOAA National Severe Storms Laboratory, 2023a: FACETs. Accessed 16 June 2023, https://www.nssl.noaa.gov/projects/facets/

NOAA National Severe Storms Laboratory, 2023b: VORTEX Southeast/VORTEX USA. Accessed 16 June 2023, https://www.nssl.noaa.gov/projects/vortexse/

Nowell, L. S., J. M. Norris, D. E. White, and N. J. Moules, 2017: Thematic analysis: Striving to meet the trustworthiness criteria. *Int. J. Qual. Methods*, **16**, 1-13. DOI: 10.1177/1609406917733847

NSF, 2023: Frequently asked questions and vignettes: How should research involving "snowball samples" be handled from a human subjects perspective? Accessed 12 June 2023, https://www.nsf.gov/bfa/dias/policy/hsfaqs.jsp#snow

Obermeier, H. B., K. L. Berry, K. E. Klockow-McClain, A. Campbell, C. Carithers, A. Gerard, and J. E. Trujillo-Falcón, 2022: The creation of a research television studio to test probabilistic hazard information with broadcast meteorologists in NOAA's Hazardous Weather Testbed. *Wea. Clim. Soc.*, **14(3)**, 949-963, DOI: https://doi.org/10.1175/WCAS-D-21-0171.1

Pappas, S., 2012: Italian scientists sentenced to 6 years for earthquake statements. Accessed 27 April 2023, https://www.scientificamerican.com/article/italian-scientists-get/#:~:text=Six%20Italian%20scientists%20and%20a,which%20alarmed%20earth%20scientists%20worldwide.

Potvin, C. K., C. Broyles, P. S. Skinner, H. E. Brooks, and E. Rassmussen, 2019: A bayseian hierarchical modeling framework for correcting reporting bias in the U.S. tornado database. *Wea. Forecasting,* **34**, 15-30, DOI: https://doi.org/10.1175/WAF-D-18-0137.1

Ripberger, J. T., C. L Silva, H. C. Jenkins-Smith, D. E. Carlson, M. James, and K. G. Henson, 2015a: False alarms and missed events: The impact and origins of perceived inaccuracy in tornado warning systems. *Risk Anal.*, **35**, 44-56, DOI: 10.1111/risa.12262

Ripberger, J., A. Bell, A. Fox, A. Forney, W. Livingston, C. Gaddie, C. Silva, and J. Jenkins-Smith, 2022a: Communicating probability information in weather forecasts: Findings and recommendations from a living systematic review of the research literature. *Wea. Climate Soc.*, **14**, 481-498, DOI: https://doi.org/10.1175/WCAS-D-21-0034.1

Ripberger, J., and C. Silva, 2021: WxDashExp. Accessed 18 April 2022, https://crcm.shinyapps.io/WxDashExp/#section-information-sources

Ripberger, J. T., C. Silva, H. C. Jenkins-Smith, and M. James, 2015b: The influence of consequence-based messaging on public responses to tornado warnings. *Bull. Amer. Meteor. Soc.*, **96**, 577-590, DOI: https://doi.org/10.1175/BAMS-D-13-00213.1

Ripberger, J., J. Allan, K. Gupta, A. Bitterman, C. Silva, H. Jenkins-Smith, and E. Cokely, 2022b: There's a chance of what? Assessing numeracy skills of forecasters, emergency managers, and publics to improve TC product uncertainty communication, IDSS, and training. 25pp., https://repository.library.noaa.gov/view/noaa/47959/noaa_47959_DS1.pdf

Schaefer, J. T., and R. Edwards, 1999: The SPC tornado/severe thunderstorm database. Preprints, *11th Conf. on Applied Climatology,* Dallas, TX, Amer. Meteor. Soc., 215-220.

Schwartz, L. M., S. Woloshin, W. C. Black, and H. G. Welch, 1997: The role of numeracy in understanding the benefit of screening mammography. *Ann. Intern. Med.*, **127**, 966-972, DOI: https://doi.org/10.7326/0003-4819-127-11-199712010-00003

Semmens, K., R. H. Carr, B. Montz, and K. Maxfield, 2022: Communicating the complex impacts of severe winter weather through the winter storm severity index. *10th Symposium on Building a Weather-Ready Nation*, Houston, TX, Amer. Meteor. Soc., 5.4, https://ams.confex.com/ams/102ANNUAL/meetingapp.cgi/Paper/387982

Sherman-Morris, Kathleen, 2005: Tornadoes, television, and trust – A closer look at the influence of the local weathercaster during severe weather. *Glob. Environ. Change Part B: Env. Haz.*, **6(4)**, 201-210. DOI: https://doi.org/10.1016/j.hazards.2006.10.002

Shivers-Williams, C. A., and K. E. Klockow-McClain, 2021: Geographic scale and probabilistic forecasts: A trade-off for protective decisions? *Nat. Hazards*, **105**, 2283-2306, DOI: https://doi.org/10.1007/s11069-020-04400-2

SPC, 2023: All tornado probabilities. Accessed 11 April 2023, https://www.spc.noaa.gov/new/SVRclimo/climo.php?parm=allTorn

Spiegelhalter, D., 2017: Risk and uncertainty communication. Annu. Rev. Stat. Appl., 4, 31-60, DOI: https://doi.org/10.1146/annurev-statistics-010814-020148

Spiegelhalter, D., 2020: Relative vs absolute risks: Why relative risks are misleading, and how to communicate absolute risks. The Winton Centre, YouTube, Accessed 16 June 2023. https://www.youtube.com/watch?v=ELNt0XZYsAQ&ab_channel=TheWintonCentre

Stelter, B., 2019: Weatherman who defied ‘Code Red’ alerts is out of a job. CNN, Accessed 23 June 2022, https://www.cnn.com/2019/06/13/media/joe-crain-sinclair-code-red/index.html

Storm Prediction Center, 2022: 1950-2021_actual_tornadoes, accessed 22 August 2022, https://www.spc.noaa.gov/wcm/#data

Strathie, A., G. Netto, G. H. Walker, and G. Pender, 2015: How presentation format affects the interpretation of probabilistic flood risk information. *J. Flood Risk Manag.*, **10**, 87-96, DOI: https://doi.org/10.1111/jfr3.12152

The Weather Channel, 2018: TOR:CON: Tornado condition index. Accessed 12 April 2023, https://youtu.be/NKupn1Fu39Q

Trainor, J. E., D. Nagele, B. Philips, and B. Scott, 2015: Tornadoes, social science, and the false alarm effect. *Wea. Clim. Soc.*, **7**, 333-352, DOI: https://doi.org/10.1175/WCAS-D-14-00052.1

Trevena, L. J., B. J. Zikmund-Fisher, A. Edwards, W. Gaissmaier, M. Galesic, P. KJ Han, J. King, M. L. Lawson, S. K. Linder, I. Lipkus, E. Ozanne, E. Peters, D. Timmermans, and S. Woloshin, 2013: Presenting quantitative information about decision outcomes: A risk communication primer for patient decision aid developers. *BMC Medical Inform. Decis. Mak.*, **13**, DOI: https://doi.org/10.1186/1472-6947-13-S2-S7

Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U.S. tornado database: 1954-2003. *Wea. Forecasting,* **21**, 86-93, DOI: https://doi.org/10.1175/WAF910.1

Visschers, V. H. M., R. M. Meertens, W. W. F. Passchier, and N. N. K. de Vries, 2009: Probability information in risk communication: A review of the research literature. *Risk Anal.*, **29**, 267-287, DOI: https://doi.org/10.1111/j.1539-6924.2008.01137.x

Weber, E. U., 2006: Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). *Climactic Change*, 77, 103-120, DOI: http://dx.doi.org/10.1007/s10584-006-9060-3

Williams, C. A. and G. M. Eosco, 2021: Is a consistent message achievable?: Defining “message consistency” for weather enterprise researchers and practitioners. *Bull. Amer. Meteor. Soc.,* **102(2)**, E279-E295, DOI: https://doi.org/10.1175/BAMS-D-18-0250.1

WPC, 2020: Winter Storm Severity Index (WSSI) Product Description Document (PDD). 8 pp, https://nws.weather.gov/products/PDD/PDD_OplWSSI_2020-2021.pdf

Woo, G. and W. Marzocchi, 2014: Operational earthquake forecasting and decision-making. *Early Warning for Geological Disasters,* F. Wenzel and J. Zschau, E.d., Springer Berlin, Heidelberg, 353-367. DOI: https://doi.org/10.1007/978-3-642-12233-0_18

Yamagishi, K., 1997: When a 12.86% mortality is more dangerous than 24.14%: Implications for risk communication. *Appl. Cog. Psychol.*, **11**, 495-506, DOI: https://doi.org/10.1002/(SICI)1099-0720(199712)11:6%3C495::AID-ACP481%3E3.0.CO;2-J

Zhao, M., H. Rosoff, and R. S. John, 2019. Media disaster reporting, effects on public risk perception and response to escalating tornado warnings: A natural experiment. *Risk Anal.*, **39(3)**, 535-552. DOI: https://doi.org/10.1111/risa.13205

Appendix A: Interview Guide used to organize Focus Group Interviews

Broadcaster interview guide for Normalized Probability products

1) ***Introduce self and focus group rules.***
    a. Remind participants this is being recorded, but that names will not be attached to statements here and that data will be de-identified for presentation and publication.
    b. Remind participants that because this is human subjects research, they can post their own involvement on social media but cannot disclose another's participation. An anonymized group photo will be taken and shared through the @NOAA_HWT Twitter account that everyone is free to retweet. Those who do not wish to be in the photo can turn their cameras off for it.
    c. You may tweet or present images of experimental information and products as long as you highlight the experimental, non-operational nature of these designs. These are not future products – they are a communication experiment!

   d. Thank them for being here!

2) ***Opening – General SPC outlook use, familiarity***
   - **How familiar are you with the individual hazard (tornado, wind, hail) probabilities that define the convective outlook?**
     - **Have you used these to communicate severe weather risk on air or on social media before?**
   - What are the pros and cons of the probabilistic outlook product for you, as it currently stands?
3) ***Midsection – introduce Normalized Risk and goals of product – present slideshow***
   - *Present link to Normalized Risk viewer and run through use of the viewer.*
4) ***Present example outlooks and begin conversation on value of Normalized Risk***
   - Introduce events briefly (with descriptors)
     - November 4th, 2022 – Typical in-season high-probability event (though fall vs spring)
     - June 1st, 2011 – Off-region in-season event
     - April 27th, 2011 – Major in-season in-region event – but offset NR area
     - Dec 10th, 2021 – Major off-season in-region event
   - **Would you use Normalized Risk information in your forecast process for severe weather events? Would you use or present Normalized Risk information on air or on social media?**
     - **If your local competitors shared Normalized Risk information with their viewers, would that change your stance?**
   - What advantages or disadvantages might Normalized Risk information offer in these scenarios?
5) ***Summarize thoughts and conclude.***
   - Do you think that the Normalized Risk information would help or hinder your communication of severe weather hazards to your viewers relative to the current SPC outlook information alone?
   - **If you were to show Normalized Risk information to your viewers, would you prefer to present mapped Normalized Risk visuals, or report values of Normalized Risk for specific locations? Would it depend on the event?**

  - Is there another method of presenting severe weather probabilities that would be more useful to you than the current absolute probabilities or Normalized Risk?

What are your overall thoughts on the Normalized Risk product? Summarize your reaction to the products you've seen today. Where would you like to see development of this experimental information go?

Appendix B: Coding Journal

Codebook/Journal – Broadcaster Relative Risk interviews

5/1/2023

- Edited first focus group to make sure transcription solid – need to do 3 more.
- Initial thoughts
  - First interview was the most unfriendly to the product, if I recall correctly – consensus of "for me, but not thee" use of product, distrust of public interpretation of large numbers in normalized/relative risk product
  - Later ones friendlier to product. Also recall from others that some forecasters mentioned presenting risk contextually already – add a code for this.
  - Probably need a code for concern about the appearance of the product on the map as well.
  - Code for preferring map vs value of risk at a point – generally think map won, not fans of reporting values by location b/c of the "what about my town" effect
  - Definitely need code for concern about public ability to interpret probability information/false alarm effect.
  - For codes to learn more about opinions on the product – need PRO, CON overarching codes?

5/3/2023

- Completed all four transcription reviews.
- Thoughts continued
  - Overall consensus seems to be suspicious of product presentation to viewers. Lots of variation between preferring to present the map or just values - or not at all.
  - Need to add codes addressing concerns about hype-ing forecasts
  - Noted a lot of mentions of "this is more useful in off-season or when unusual than peak season" - situationally specific code?

- Track participant familiarity by participant - can overall gauge familiarity in plots/figs
    - Track map vs number preference also
- I think some of the specific code ideas I had above - I'll remember those when breaking down pros and cons. But keep false alarm/hype as a code.

Initial codes:

Code	Full Name	Description
~~FAM~~	~~Familiarity with SPC~~	~~Direct mention of how familiar the participant is with CURRENT SPC probability product.~~
GR8	Things that are great with the SPC outlook	Mentions by broadcasters of things they like about the probabilities in or the design of the current SPC outlook (or positive uses)
BAD	Things that are problematic with the current SPC outlook	Mentions by broadcasters of things they do not like about the probabilities in or the design of the current SPC outlook (not probability understanding related)
USE	Use of Norm/relative Risk	~~Excerpt where broadcaster describes a situation that they would use normalized risk information. OR already uses it or something similar.~~ for when broadcasters mention how they currently or would in the future use relative risk information for their own purposes - not including presenting to the public.
SIT	Situation that relative risk would be useful	for when broadcasters discuss WHEN they would use relative risk information or present it to their viewers

SHO	How broadcasters show viewers relative risk	for when broadcasters discuss how they would SHOW relative risk information to their viewers. (or if they wouldn't, and why)
PRO	Pro - positive feedback	Something a broadcaster liked about the normalized risk information
CON	Con - negative feedback	Something a broadcaster didn't like about the normalized risk information
MAP	Mapped normalized/relative risk	~~For mentions of preferring to use a map vs numbers to show normalized risk~~ For discussion of the visual presentation of relative risk
FAH	False Alarm/Hype Concern	For explicit mentions of being worried that the probabilities lead to hype or false alarms for the public.
PRB	Probability Understanding	For direct discussion of the public's ability to interpret probabilistic information.
LVL	Use of SPC categorical levels	For mentions of using SPC categorical levels instead of or along with probability information.
ADD	Improvement to add to Normalized/relative Risk	When broadcasters suggest ways that normalized/relative risk products could be improved for their use.

5/8/2023
New code - LVL, for when broadcasters mention turning the percentages to levels (categorical or numerical)
Note that FAM is SPECIFICALLY for probability product familiarity.
Do we need a code for concern about assuming risk is low based on visual of normalized risk not matching absolute risk visuals?
Maybe a code for development of the normalized risk product? New ideas basically.
Finished Group 1 coding. May need to define more codes and go again later.

5/9/2023
Mention of management possibly nixing use of normalized risk - need to be sure to track that
New code - ADD - for when broadcasters suggest changes to normalized risk to better suit them
Group 2 done - feel like I coded less pro/con and used “USE” more on this one, will need to think about that for Group 1 revisit. More MAP as well. Didn’t fill out participant statistics as religiously, may need to revisit that as well.
Kinda also not a fan of the FAM code as described. May need to edit.
Group 3, applied PRB to users wanting impact information - since they don’t want probabilities in the met’s mind.
May need to break USE out - by whether they currently use a relative risk or if they are saying HOW they’d use it.
~~Used FAH on a mention of underselling event severity in group 3 -~~ scratch that, changed to CON
Started applying USE when situational use of the product is discussed
Broadening application of MAP to discussion of visual presentation of product
Finished Group 3, trying to have smaller highlights for codes. General message seems to be “situationally good, I like it when something is out of the ordinary”. Less “I want it but not to show it to the public” responses than I thought. Need to go through again for statistics I was thinking about.

5/10/2023
FAM really does not work and needs to be revised
Applied USE to mentions of answering questions about products - tough to code, that
Finished group 4 - way more hype discussion there than the others, very concerned about numbers being too large. Seems to be a theme overall, as well as doubt in public ability to understand probability.

5/15/2023
Back to group 1 recode
Revised FAM - original goal to capture familiarity with the old scale. Too vague to capture ideas. Instead, broken into:
GR8 - Things that broadcasters appreciate with the current SPC outlook
BAD - Things that broadcasters think the current outlook could improve on
Broke out USE code - original code good but applied too broadly. Now have:
USE - for when broadcasters mention how they currently or would in the future use relative risk information for their own purposes.
SIT - for when broadcasters discuss WHEN they would use relative risk information or present it to their viewers
SHO - for when broadcasters discuss how they would SHOW relative risk information to their viewers.
Revised MAP to reflect coding actually done in analysis.

5/16/2023
Starting to think PRO/CON may be eliminated by new use/sho/sit codes. Will run through other groups and remove as necessary.
Decided not to remove, general good/bad reviews have a space

5/18/2023
Used "SHO" when broadcaster mentioned adding local context to presentation of SPC outlook as it stands - basically the same idea as RR.

5/19/2023
Completed 2nd pass, collating codes in individual documents.
Pulling quotes thoughts

- Overall very strong suspicion of public ability to interpret severe weather probabilities.
- Fondness for numbered levels, even converting RR into numbered levels.
- Concern about education on how to interpret and management?
- Amount of time it takes to explain a big concern
- General sense it is best in off season events

5/22/2023
Begin organizing themes, building support for general ideas
First - suspicion of ability of the public to understand probability - specifically, underestimating low probabilities and overestimating relative risks
Noting a general sense of "public wants to know impact/what will happen"? LVL, FAH, PRB…
By theme -
GR8 - use of current outlook to define what hazard dominates, getting SPC insight on storm potential
BAD - Forecast errors (missed events and false alarms), issues with the outlook words
LVL - Use of numbers to describe risk level (1-5 vs 1-10)
^^^ feel like data from the above three less relevant to this study
Next look thru use/sit/sho and pro/con/add. I think both will have interesting ideas, hone in on when relative risk shown and major likes/dislikes

PRO - adds context to how unusual a situation is (honestly its more "I like it" than anything else…)
CON - highlights the "wrong area"/misleading, numbers are too high vs low absolute probability/outcomes, need for more time/education on air, downplays peak season?

5/23/2023
Continued code breakdowns
ADD - Convert RR into a numbered scale/back into the SPC categorical outlook

USE - "I would use but not for public"/most helpful for meteorologists, help define how unusual an event is, a few already use RR-like products
SIT - When relative and absolute line up, when the event is out of season (and not during season)
SHO - Would show values/maps on social more often, would show for unusual events, would show to the "expert public"

MAP - split between "like map, not a fan of numbers" and "don't like map, doesn't match"

Began filling in themes with data and building thematic map

Tomorrow:
CON -numbers are too high vs low absolute probability/outcomes -> feed into hype/#s overblown concern theme
CON - highlights the “wrong area”/misleading -> capture this in data
CON - need for more time/education on air -> capture this in data
ADD - Convert RR into a numbered scale/back into the SPC categorical outlook -> capture in data and add data from LVL - Use of numbers to describe risk level (1-5 vs 1-10) potentially
USE - “I would use but not for public”/most helpful for meteorologists -> Identify themes in *how* it would be used
Do something with map use?

5/24/2023
Added “CON” quotes to hype/overblown theme, also found 3 broadcasters concerned about RR underselling risk
Added “Concern about risk offset” theme from CON (i.e. wrong area data, funny I did that without realizing I told myself to do it! Good theme.)
Added education theme - a bit tenuous, as some quotes less supportive, but 6 separate broadcaster mentions
Added broadcasters want RR in level form - 5 broadcasters so not as strong as expected

5/25/2023
Defined themes across usage of RR, and added a theme about map use
Begin thematic mapping