



# 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

EMILY D. LENHARDT, RACHAEL N. CROSS, MAKENZIE J. KROCAK, JOSEPH T. RIPBERGER,  
SEAN R. ERNST, CAROL L. SILVA, and HANK C. JENKINS-SMITH  
*University of Oklahoma Center for Risk and Crisis Management, Norman, OK*

(Manuscript received 29 August 2019; review completed 30 March 2020)

## ABSTRACT

One of the challenges when communicating forecast information to the public is properly contextualizing uncertainty. No forecast is ever certain, as no meteorological phenomenon is guaranteed to occur. As such, the uncertainty in forecast information should be communicated in a way that makes sense to end users. Previous studies of the communication of probabilistic information suggest that, although the general public are more apt to communicate uncertainty with words of estimative probability (WEPs), they prefer to receive uncertainty information numerically. Other work has suggested that a combination of numbers and WEPs is the best method for communicating probability, but fewer research studies have assessed the communication and interpretation of probabilistic meteorological information. In this study, we code 8900 tweets from the National Weather Service Weather Forecast Offices (WFOs) and analyze them to find how forecasters communicate probabilistic forecast information to the public via Twitter. This analysis reveals that WFO messaging is dominated by WEPs, with few numerical descriptions of probability. These WEPs are generally vague, unqualified notions of probability that may impede the public's ability to interpret the information that forecasters are trying to communicate. Based on this analysis, two publicly fielded surveys also are analyzed in order to understand how participants tend to interpret the most common qualified and unqualified WEPs that WFOs used on Twitter. Though participants generally interpret qualified WEPs more consistently than unqualified WEPs, both categories featured a wide range of interpretations that suggest both types of WEPs are vaguely defined for the general public.

## 1. Introduction

Following the recommendations by the National Research Council (2006), the National Institute of Standards and Technology (Kuligowski 2013), and academic experts in weather risk communication (e.g., AMS Council 2008; Morss et. al 2008; Joslyn and Savelli 2010), the National Weather Service (NWS) is exploring and expanding the use of probabilistic information to convey weather forecast uncertainty. For example, the National Severe Storms Laboratory (NSSL) is exploring the development of a new forecast and warning paradigm that supplements current watch, warning, and advisory products with constantly evolving, geographically specific, probabilistic

information (Rothfus et al. 2018). However, questions about how to best communicate uncertainty remain, and improper implementation of new messaging could defeat the purpose of these changes by leading to greater confusion for users. This study contributes to this area of research by (1) systematically documenting the words and phrases that NWS forecast offices currently use to communicate probabilistic information in severe thunderstorm and tornado forecasts; and (2) examining how members of the public interpret these words and phrases. The purpose of this study is not necessarily to determine or judge how successful the NWS is in communicating severe weather probability. Rather, it is intended to provide the NWS with a dataset of their most commonly used words of estimative probability

(WEPs) and how the public tends to interpret them. We begin by briefly summarizing previous research. In the subsequent sections, we introduce the data and methods we use for the analysis and highlight the study results. We then conclude by discussing the implications of the study and suggesting directions for future research.

## 2. Summary of previous research

This study focuses on the NWS' use of WEPs, which are a common method of communicating probabilistic information to the general public. An example of the use of a WEP is "severe thunderstorms have a low chance of occurring this evening," where "low chance" is a WEP used to communicate the likelihood of severe thunderstorm occurrence. Unfortunately, these words can be vague in nature and users can have difficulty interpreting their meaning. Despite this confusion, studies show that people prefer to express probabilistic information using words and to receive such information numerically (Fischer and Jungermann 1996, Morss et al. 2008, MacLeod and Pietravalle 2017, Willems et al. 2019, Wintle et al. 2019). Despite past work that indicates people inherently understand words better than numbers (Wallsten et al. 1986a), recent findings suggest that numbers communicate probabilities more effectively (Willems et al. 2019, Wintle et al. 2019) and potentially carry less ambiguity (Fischer and Jungermann 1996, Friedman and Zeckhauser 2014). In fact, as early as the 1960s, researchers noted the difficulty in deriving precise meaning from WEPs, arguing that expressing probability using words is a poor way to convey confidence and should be used sparingly (Kent 1964, Barclay et al. 1977, Beyth-Marom 1982).

Although numbers are potentially more useful in communicating probability, there are different numerical ways people can express probability. For example, ranges of probabilities are a popular method used to express uncertainty about the occurrence of an event, though there is some disagreement in the literature about the effectiveness of ranges in communicating probabilistic information. Friedman and Zeckhauser (2014), for example, argue that there is little logical difference between ranges and single point estimates because most readers condense ranges into single numbers when making decisions. They therefore recommend point estimates as they communicate less ambiguity. Other studies, including Wallsten et al. (1986a) and Budescu et al. (2014), assert that point estimates might be too precise to express an uncertain

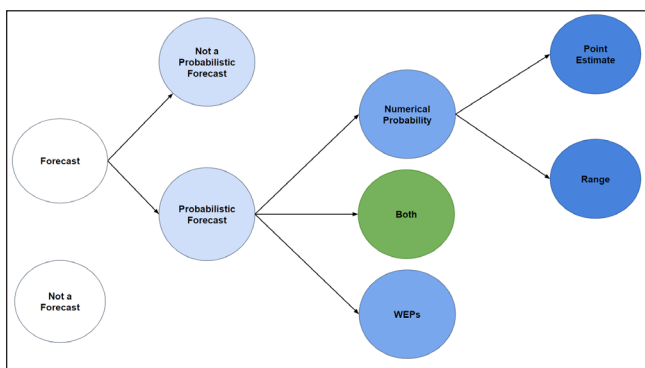
opinion, suggesting that uncertainty intervals are preferable. It also is noted by Morss et al. (2008) and Joslyn and Savelli (2010) that the public anticipates uncertainty when given a point estimate. Overall, studies appear to suggest that WEPs combined with numerical values most effectively convey the intentions of the communicator (Budescu et al. 2009, Budescu et al. 2012, Budescu et al. 2014, Wintle et al. 2019).

Although combining words and numerical estimates in messaging can lead to public interpretations that are more consistent with the expert understanding of the situation, there are other factors that can influence an individual's interpretation of probabilistic information. These factors include the interaction between base rate (an event's frequency of occurrence) and expected outcome severity (the perceived severity of the future event). Previous work has shown that when base rate is kept consistent and the severity of an expected outcome changes, people tend to associate higher probabilities with higher severity outcomes (Harris and Corner 2011). This is called severity bias. For example, when a narrative described a chance of global sea levels rising, participants considering an island barely above sea level interpreted the probability of sea rise as higher than those considering an island elevated far above sea level. In the severe weather domain, severity bias has been studied and shown to result in the same outcome—when given a more severe situation, people tend to perceive a higher probability of occurrence (Losee et al. 2017).

In addition to severity bias, another consideration of probabilistic communication deals with the presence of appropriate context. For example, Fischer and Jungermann (1996) studied the impacts of removing contextual information from drug labels, meaning that participants were asked to provide a numerical estimate for words such as "rarely," "occasionally," and "frequently" in terms of side effects of a particular drug. However, they were not given information about side effect severity or how often the side effect occurs. They found that participants in the study assigned a higher probability to outcomes when the researchers removed context from the label (Fischer and Jungermann 1996). Additionally, researchers found the consideration of base rate was inversely related to frequency, meaning base rates played a larger role when an event was thought to occur less often (Fischer and Jungermann 1996). One other consideration when evaluating the effectiveness of probabilistic communication is the difference in results between WEPs and numbers. Previous research has found that numerical estimates of probability also

can lead people to draw more deliberate, logic-based conclusions, but WEPs can cause people to think more intuitively and express more uncertainty than when numbers are used to communicate such information (Windschitl and Wells 1996). However, regardless of the method of communicating probability, there are many cognitive biases that influence the human decision-making process.

In summary, previous work outlines three main methods of communicating probabilistic information: numerical estimates, WEPs, or a combination thereof. Recent studies suggest that a combination of WEPs and numbers is the best method of communicating probability, though recommendations for the specific method of expressing the numeric aspect of this combination still varies across studies. Figure 1 outlines a schematic of these various methods of expressing probability as suggested by past research and applies them to a meteorological context. We focus on a weather forecast context because, although some studies have addressed how uncertainty in forecast information is communicated (e.g., Morss et al. 2008), less research has attempted to evaluate how the public interprets WEPs in a meteorological context (for an exception see Wallsten et al. 1986b). Therefore, although Fig. 1 focuses on the previously mentioned methods of expressing probability, it also includes a distinction



**Figure 1.** This schematic was developed from past work on the communication of probabilistic information and applied to a meteorological context. It shows the three main methods of expressing probability as suggested by past research in addition to differentiating between whether or not something is a forecast and whether or not a forecast is probabilistic in nature. The green circle represents the method of expressing probability that research suggests is best for communicating probabilistic information. *Click image for an external version; this applies to all figures and hereafter.*

between messages that do and do not communicate forecast information in order to focus on probability information inherent in forecasts.

### 3. Data and methods

There are three main datasets incorporated in this study—two public surveys and 8900 NWS tweets. Although the main focus of this paper is the investigation of probabilistic communication in the NWS Weather Forecast Office (WFO) severe weather tweets, we write the data and methods section in chronological order to best portray the series of events that motivated certain decisions made along the way. As described below, the 2018 survey was administered first in June 2018, and the analysis of NWS WFO tweets was later based on an overview of our initial results. Finally, after coding the database of tweets, we developed and implemented a new WEP question for the 2019 survey.

#### a. 2018 survey

Public interpretations of WEPs that forecasters commonly use were assessed using data from the Severe Weather and Society Survey (WX), a yearly survey of United States adults administered by the Center for Risk and Crisis Management (CRCM) at the University of Oklahoma (see Silva et al. 2018 for more information on WX surveys). WX18 was fielded in June 2018 as an online questionnaire completed by 3000 adults (age 18+) who match the demographic characteristics of the United States population. The sample of participants was provided by Qualtrics, a company that maintains a diverse panel of internet users in the United States who agree to complete online surveys. Qualtrics uses a quota system to produce representative samples, a practice used by many other internet sampling companies.

Data from WX18 allow for examination of how members of the public interpret a variety of qualified WEPs used in the context of severe weather. Interpretations were elicited using the following question:

*People use different phrases to explain the possibility that a [severe thunderstorm/tornado] will happen. When you see the following phrases, what percent chance comes to mind? Please indicate the chance as a percent that ranges from 0 to 100, where 0 means no chance and 100 means that it is certain.*

- *Very low [chance/probability] of a [severe thunderstorm/tornado]*
- *Extremely low [chance/probability] of a [severe thunderstorm/tornado]*
- *Pretty low [chance/probability] of a [severe thunderstorm/tornado]*
- *Small [chance/probability] of a [severe thunderstorm/tornado]*
- *Low [chance/probability] of a [severe thunderstorm/tornado]*
- *Slight [chance/probability] of a [severe thunderstorm/tornado]*
- *Moderate [chance/probability] of a [severe thunderstorm/tornado]*
- *Good [chance/probability] of a [severe thunderstorm/tornado]*
- *Significant [chance/probability] of a [severe thunderstorm/tornado]*

Each respondent was given an opportunity to evaluate a random sample of five phrases. The words “chance” and “probability” were randomized in the experiment to identify possible differences between the two. Although people associated the word “chance” with slightly lower percentages than the word “probability,” the variation was not statistically significant (see Appendix A). The event also was randomized to identify the possibility that event severity might influence perceptions of probability. For both qualified and unqualified WEPs, people associated nominally lower probabilities with severe thunderstorms than with tornadoes, but the differences were negligible (see Appendix B). Although this result does contradict the severity bias noted by Harris and Corner (2011), there is also a base rate issue to consider that may play a role in equalizing the perceived severity of severe thunderstorms and tornadoes. This contradiction speaks to the difficulty in understanding human interpretation of probability due to the many influential cognitive processes and biases at play. Regardless, for the purpose of this paper, responses were not filtered by the displayed event.

#### *b. NWS probabilistic communication*

An initial glance at the results of WX18 showed wide variations in how survey-takers interpreted the

given WEPs in a severe weather context. Therefore, to investigate the practical implications of these results, Twitter messages (“tweets”) from National Weather Service WFO accounts were used to examine how forecasters communicate probabilistic information about severe thunderstorms and tornadoes. With no prior knowledge of the specific language present in the database of messages described below, we hypothesized that we would primarily find the words tested in WX18 (“very low,” “extremely low,” “pretty low,” “small,” “low,” “slight,” “medium,” “remote,” “moderate,” “good,” “large,” “significant,” “high,” “pretty high,” “very high,” “extremely high”).

We obtained the tweets from the CRCM Severe Weather and Social Media data collection program. This data collection program connects with Twitter’s streaming application program interface to automatically and continuously collect and archive tweets that include keywords and phrases related to extreme weather. In addition to the text of each tweet, the program archives metadata provided by Twitter [Twitter Developer (a) and (b) 2020] about the tweet itself and the user who created it, such as time the tweet was sent, number of followers from the tweeting account, number of retweets the original tweet received, etc. (see Ripberger et al. 2014 for more information on the project). This analysis focuses on tweets from WFO accounts in 2018 that (1) included one or more of the following keywords/phrases: severe weather, hail, tornado, severe thunderstorm but (2) did not include one or more of the following words: outlook, warning, watch, advisory. These latter tweets were excluded because they typically relay information about a product (e.g., “Tornado Warning including Miltonvale KS”) rather than a forecast. Given this study’s focus on the communication of probabilistic forecasts, we excluded tweets such as these from data collection for the sake of getting relevant data, rather than simply “cut and paste” notifications about newly issued products. For consistency, the sample was limited to WFO accounts that posted more than 100 tweets that met these criteria. There were 89 WFO accounts that did so. This analysis focused on a random sample of 100 tweets from each of these 89 accounts (n=8900) throughout 2018.

Following dataset construction, we initially coded the tweets based on the schematic provided in Fig. 1, where tweets were coded according to whether they included forecast information, were probabilistic or deterministic, and if they included WEPs or numerically expressed probabilities. Any instance of probabilistic

communication resulted in the tweet being coded as probabilistic, even if it also contained a deterministic phrase. Again, at this part of our investigation we expected tweet WEPs to be like those tested in WX18. Examples of tweets that fell into each category are provided in Table 1. However, rather than finding examples of WEPs such as those tested in WX18, our initial analysis found that a majority of our coded WEPs were less specific and seemingly more vague. Overall, there were very few instances of the WX18 WEPs in the Twitter dataset and an overwhelming number of the less specific ones. Therefore, we did not feel that it made sense to code all WEPs together, especially when one goal was to be able to relate the Twitter results to the initial WX18 results. Furthermore, we speculated that the WEPs from WX18 would more accurately communicate probability than these less specific and more vague terms found. In order to account for this, a new schematic was developed (Fig. 2).

Using the new schematic, tweets with WEPs were broken down by whether the WEPs were qualified or unqualified. The two categories were differentiated by whether the WEP had a qualifier attached to the phrase (see examples in Table 2). If a qualifier was present, then the WEP was determined to be “qualified” and, if a qualifier was absent, then the WEP was determined to be “unqualified.” A similar approach to distinguishing between different types of WEPs was shown in an older study by Kent (1964) in which he mentions a difference between WEPs with and without a modifier. He describes qualified WEPs as those that describe an important matter that does not have a specific order of likelihood. It is not guaranteed to occur but is not impossible. Furthermore, he goes on to say that such a matter should be described using the word “possible” or using a similar WEP without a modifier (Kent 1964). This same sort of reasoning also describes why we decided to make this distinction.

The qualified WEPs are based on those tested in WX18, whereas the unqualified WEPs consist of the vaguer phrases that prompted this separation. Phrases with qualifiers expressing a change in a probability (but without explicitly noting the original probability) were added to the unqualified WEP category because they are relative and not easily quantifiable. Examples of such phrases include “increased chance,” “decreased potential,” and “diminishing threat.” Similarly, WEPs such as “greatest potential,” “highest threat,” and “lesser chance” were categorized as unqualified. Although these phrases do contain a word that acts like a qualifier,

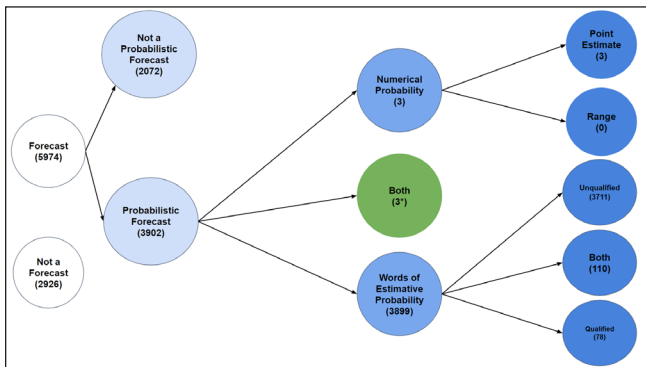
they only make sense in relation to an initial condition or in an external context that is not necessarily given or known by the reader.

The words “risk” and “threat” were included in the WEP category if the context in which they were used seemed to imply a probability of occurrence rather than the severity of the event. When the word “risk” was explicitly referencing the Storm Prediction Center’s (SPC’s) Convective Outlook, the word was not added to the WEP category. This is because the use of phrases like “moderate risk” in the SPC’s Convective Outlooks corresponds to certain numerical thresholds and definitions rather than a probability of occurrence that NWS forecasters assign to an event. Additionally, when the word “threat” was used to describe the types of threats associated with an event, it was not added to the WEP category as it was not used to describe a probability. Finally, some WEPs described probabilities of spatial coverage and time, rather than the likelihood of an event’s occurrence. These were included in the dataset as unqualified WEPs because they were vague and, if interpreted one way, could imply a probability of occurrence. An example tweet (NWS La Crosse, 5 June 2018) with one of these vague phrases, “occasional chance,” is: “Forecast through Saturday looks just June should - mild with occasional storm chances. Keep an eye on Wednesday as severe storms are possible with a risk for damaging winds and possibly large hail too.”

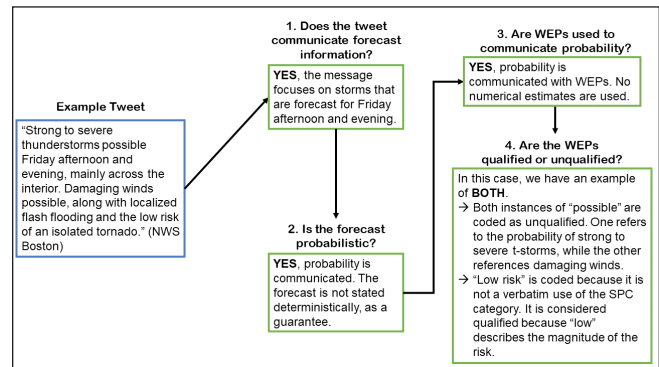
The above criteria, in addition to the schematic in Fig. 2, were used as the guidelines for coding the tweets. All instances of WEPs in a single tweet were coded. Therefore, most tweets were coded multiple times so that all WEPs were accounted for in the final dataset. This method is explained further in Fig. 3, a schematic that gives an example of how we coded each tweet and handled the presence of multiple WEPs in the same message. Any tweets with questionable interpretations were discussed between the two coders before the final classification in order to minimize discrepancies in the data. As such, although messages from the NWS about current weather observations can be interpreted differently by different people (e.g., some might consider a probabilistic forecast to be deterministic based on wording), the guidelines each coder followed ensured the data were coded consistently. In general, discrepancies in interpretations were approached by considering how it likely would be understood by the general public, rather than from a more meteorological and analytical point of view. The minimal subjectivity that remains within the coding scheme should not

**Table 1.** Tweets were categorized based on a specific set of criteria representative of the most common ways of expressing uncertainty that were mentioned in the literature. This table shows an example of a tweet from each of the first three categories. The “Forecast” and “Not a Forecast” categories were not literature-based; rather they were added later so that the focus of this study would remain on communication of forecast information specifically.

Forecast	Not a Forecast
Heavy Rain, Large Hail And Damaging Winds Possible Today And Tonight. #sdwx #mnwx (NWS Aberdeen)	The image below shows the number of hail reports equal to or greater than 3 inches that occurred in El Paso county, CO from 1950 to 2017 (68 yrs). In 2018, we have had at least 2 days were hail >= 3”, with one of these events occurring in the middle of the night. #cowx (NWS Pueblo)
Probabilistic Forecast	Deterministic Forecast
Scattered thunderstorms will continue this evening. While widespread severe weather is not expected, an isolated strong to severe storm is possible. #iawx (NWS Des Moines)	Some scattered thunderstorms are ongoing across the northwest portion of our area tonight. Expect heavy downpours and occasional small hail with these storms. (NWS Wilmington)
Words of Estimative Probability	Numeric Probability
Expect above normal warmth & rain & storm chances today ahead of a developing cold front & strong storm system. Expect the best chances of storms & severe weather by late afternoon/early evening in the Delta, but some conditional threat is possible by late afternoon elsewhere. (NWS Jackson MS)	815am: Thunderstorm activity flaring up with Caribbean disturbance. 40% chance of development within the next 5 days as it moves into the eastern or central Gulf. Heavy rain remains a threat for South FL late this weekend into Memorial Day weekend #flwx <a href="https://t.co/JfcD43cTBy">https://t.co/JfcD43cTBy</a> (NWS Miami)



**Figure 2.** All tweets were coded based on this schematic, and the numbers in parentheses represent the number of tweets from each category. For example, if a tweet relayed forecast information but in a non-probabilistic way, then that tweet would be included in the “Not a Probabilistic Forecast” category and would not be coded beyond that. This schematic was built from Fig. 1, with variations of WEPs added. \*The value here was added to the WEP category in order to focus more on the various types of WEPs used by the NWS. The number shown for the WEP category includes this value.



**Figure 3.** The above schematic gives an example of how an individual tweet was coded. It goes through each step of the coding process, including how the presence of multiple WEPs was handled.

analysis are shown in Table 3. Inter-coder agreement ranged from 91% to 96% (Cohen’s kappa from 0.86 to 0.90), which indicates very minimal subjectivity in the coding process.

c. 2019 survey

WX19 was fielded in June/July 2019 using the same procedure and methodology as WX18 (Silva et al. 2018) and after the database of tweets had been analyzed. We added a question to this survey with the goal of examining how members of the public interpret a variety of unqualified WEPs that WFOs used on Twitter, as the WX18 results did not provide

discredit the methods employed but rather speaks to the inherent subjectivity that lies in forecast interpretations by anyone receiving the information.

To statistically examine the subjectivity of the coding methodology, the two lead researchers separately coded 130 tweets that were randomly selected by a third researcher from the original dataset. The results of this

**Table 2.** List of accepted qualifiers used in tweets that were coded as being probabilistic and qualified.

Low	Slight	Small
Good	Limited	Isolated
Low-end	Some	Marginal
Minimal	Very	Very low
Very small	Big	Bit
High-end	High	Highly conditional
Little	Minimum	Moderate
Not much	Outside	Pretty good
Pretty	Pretty low	Significant
Slim	Some conditional	Very good
Very limited	Very minimal	

**Table 3.** After coding the 8900 tweets used in the dataset, coders were given a random set of 130 tweets out of the dataset to recode in order to test intercoder reliability. High percentage values of agreement indicate that the subjectivity inherent in the coding scheme is minimal.

	Forecast versus Not Forecast	Probabilistic versus Not Probabilistic	WEP versus Numeric	Classification of Qualified versus Unqualified
Agreements (#)	125	118	122	121
Disagreements (#)	5	12	8	9
Accuracy (%)	96	91	94	93
Cohen's kappa	0.90	0.86	0.88	0.86

this information. This question also was added in order to test our hypothesis that the unqualified WEPs would be less effective in communicating severe weather information. Interpretations were elicited using the following survey experiment:

*Forecasters use different phrases to explain the possibility that a [severe thunderstorm/tornado] will happen. When you see the following phrases, what percent chance comes to mind? Please indicate the chance as a percent that ranges from 0 to 100, where 0 means no chance and 100 means that it is certain.*

- *There is a chance of [severe thunderstorms/tornadoes] this afternoon and evening*
- *[Severe thunderstorms/tornadoes] are possible this afternoon and evening*
- *[Severe thunderstorms/tornadoes] may occur this afternoon and evening*
- *[Severe thunderstorms/tornadoes] are expected this afternoon and evening*

Each respondent was given an opportunity to evaluate all phrases. As with WX18, the event was randomized to identify the possibility that event severity may influence perceptions of probability; no significant differences were identified (see Appendix A).

## 4. Results

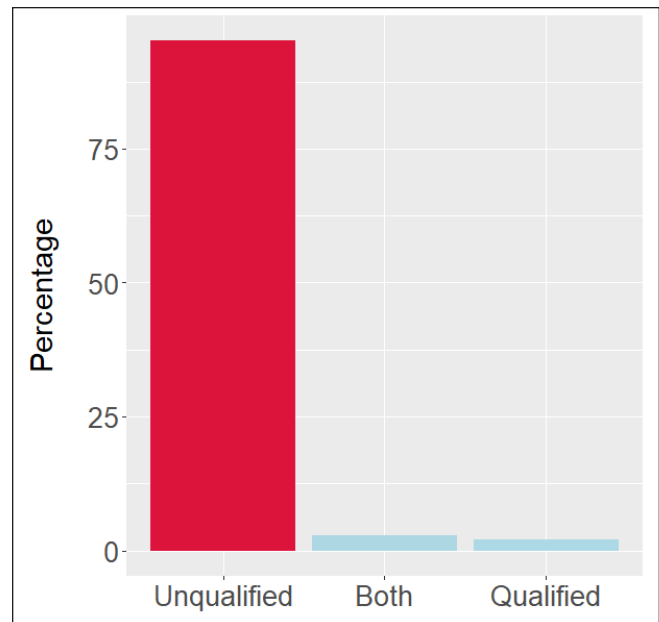
### a. NWS probabilistic communication

Within our database of 8900 severe weather-related tweets collected from 89 WFOs, 67.1% of the tweets relayed forecast information to the public. Of these tweets, 65.3% communicated probabilistic forecast information, with the rest being deterministic forecasts. This resulted in 3902 out of 8900 tweets that expressed probability either numerically, with words, or using a combination of both. Of these probabilistic tweets, 99.9% used some kind of WEP, and only 0.07% contained exclusively numeric expressions of probability. Despite recent literature such as Budescu et al. (2014) and Wintle et al. (2019) suggesting that probability may best be communicated with a combination of numbers and words, only 0.08% of the

tweets used both numbers and words; these tweets were placed into the WEP category due to this paper's focus on WEPs (i.e., the 0.08% of tweets using numbers and words were included in the 99.9% of tweets that used some kind of WEP).

When communicating a severe weather forecast on Twitter, NWS offices primarily used unqualified WEPs to do so. In fact, unqualified phrases were used about 48 times more often than qualified words (Fig. 4). In the rare case that the NWS used a qualified phrase, it was more common for them to combine it with another occurrence of an unqualified word in the same message. For example, “Thunderstorms possible today with a low risk for damaging winds and hail” was tweeted by NWS Quad Cities on 1 Dec 2018. In this tweet “possible” is coded as an unqualified WEP, whereas “low risk” is coded as a qualified WEP. The top ten most commonly used phrases from each category, the percentage of times the phrase was used, and an example tweet containing each phrase, are shown in Tables 4 and 5. For example, out of all the times the NWS used a qualified WEP, 14.4% of the time the tweet included the phrase “low threat.” Although there was not as much variance in the frequency of the qualified WEPs (Table 4), there is a large difference between the most used unqualified WEP and the rest of the phrases. The most commonly used unqualified WEP was “possible,” which was used 39.4% of the time an unqualified WEP was used (Table 5). The next highest unqualified WEP was “expected,” which was used 8.81% of the time (Table 5). Given that unqualified words were used more frequently by the NWS than qualified WEPs, “possible” was the most commonly used WEP overall by the NWS. There was also a larger variety of unqualified WEPs used, with 61 different qualified words and phrases and 79 different unqualified words. Finally, the percentage of times that the top ten qualified and unqualified WEPs were used with respect to all occurrences of WEPs are shown in Table 6.

Another notable result is the overwhelming use of low-magnitude qualified WEPs. Table 4 shows the top ten most-used qualified WEPs, and it is clear that only “good chance” is very likely to be interpreted above 50%. This could be because, in this specific context of thunderstorms and tornadoes, occurrence is inherently low. However, it also could stem from the fact that forecasters receive less probabilistic information about this type of weather than they would about, for example, probability of precipitation and may tend to hedge towards lower magnitude descriptors to avoid false



**Figure 4.** Percentage of tweets that contained unqualified and qualified WEPs along with tweets that contained both. The percentages are of all tweets coded as containing WEPs and not out of the total number of tweets.

alarm chances. Although not the focus of this study, in the future it would be interesting to compare this result to a set of tweets focused on winter, tropical, or heavy rain events.

#### *b. 2018 survey results—qualified WEPs*

In the 2018 Severe Weather and Society Survey, we asked participants to numerically interpret qualified WEPs by providing their assessment of the implied probabilities for each one. As seen in Fig. 5a, the central tendency of the distribution increases from smaller probability WEPs such as “very low” and “small” to higher probability WEPs such as “significant” and “good.”<sup>1</sup> To quantify this tendency, we calculated the mean values, the lower and upper bounds of the interquartile range (IQR), and standard deviations for all analyzed WEPs from the WX18 survey data (Table 7). One result is that the presence of the qualifier “very” did not significantly change the mean interpretations for either “small” or “low.” Perhaps more telling of

<sup>1</sup> In addition to the words seen in Fig. 5a, “extremely low,” “remote,” “medium,” “high,” “large,” “pretty high,” “very high,” and “extremely high” also were assessed in the WX18 survey. We did not include them in our results because they were never used in any NWS tweets in our dataset.



**Table 4.** For each of the top ten qualified WEPs observed in the database of severe weather tweets, an example is given for how that phrase is used by WFOs. The specific WFO for which the example is from is shown in parentheses. The percent that each word was used out of every occurrence of a qualified WEP is shown on the right. Tweets with an asterisk (\*) contain multiple WEPs that were coded, as there was not an available example with only the example WEP. The additional WEPs present in those tweets that are not being used as examples here are underlined>, and the one being used as an example is in bold.

Qualified WEPs	Example	Percent (%)
Low Threat	A few isolated thunderstorms moving into the I-69 corridor this evening. Low threat for severe wind, hail. #miwx (NWS Detroit)	14.4
Low Chance	Warm and somewhat humid weather is in store for the holiday weekend. Low chances for severe storms Friday. #iawx #ilwx #mowx (NWS Quad Cities)	12.3
Slight Chance	Cooler temperatures for the area after during the week. Slight chance for some afternoon thunderstorm development Wednesday afternoon. (NWS Aberdeen)	11.2
Low Risk	@joshgalka88 Low risk for severe weather at this time. (NWS Detroit)	5.88
Small Chance	Pleasant conditions this weekend. There is a small chance for an afternoon thunderstorm on Saturday. #swiwx (NWS Milwaukee)	5.35
Good Chance	@DPA_Insight Mountain areas will have a good chance at some convection and thunderstorm activity this afternoon. Not the coast, though... (NWS San Diego)	4.28
Limited Threat*	Another round of storms is possible this afternoon for eastern #OKwx and western #ARwx, possibly severe. Main threats are damaging winds with limited threat for tornadoes. Stay weather aware today. (NWS Tulsa)	3.21
Low Confidence*	@IowaStormChasr Yes, severe weather is likely to occur over the next 2 weeks. It's June in Iowa, which is the peak severe weather month for this state. That being said, low confidence with any significant severe weather event atm. (NWS Des Moines)	3.21
Isolated Chance*	RT @NWSNorthPlatte: Thunderstorm chances continue Friday and Saturday with an isolated chance of severe weather both days. #NEwx (NWS Cheyenne)	2.14
Low Potential	Low potential for severe storms this afternoon. 60 mph winds and 1 inch hail are the main threat. #iawx #ilwx #mowx (NWS Quad Cities)	2.14

the variance in interpretations, standard deviations from the mean range from 14.7% for “small” to 26.9% for “significant” (Table 7), and the average standard deviation for this set of qualified WEPs is 17.9%. The next highest values of standard deviation after “small” are for “very low,” “very small,” “slight,” and “low” (Table 7). On the other hand, higher standard deviation values are calculated for “moderate,” “good,” and “significant” (Table 7), which could indicate that it is more difficult for participants to interpret these higher probability WEPs than the lower ones.

The standard deviation values for each qualified WEP imply that people’s interpretation of words that indicate a low probability tend to be within the same relatively restricted range. Another way to illustrate this is with the lower and upper bounds of the IQR of the data (Table 7). The WEPs “very low” and “very small” both range from 5% to 20%, with 5% being the 25th quantile and 20% being the 75th quantile. “Small”

and “low” range from 10% to 25%. For the higher probability words, “moderate” ranges from 25% to 50%, “significant” from 60% to 90%, and “good” has bounds of 30% to 60%. Overall, we see that the lower probability WEPs have IQRs of about 15%, whereas these higher probability WEPs have IQRs of about 30%.

#### c. 2019 survey results—unqualified WEPs

In the 2019 Severe Weather and Society Survey, participants were asked to numerically interpret unqualified WEPs. Some of the most frequent unqualified WEPs from the Twitter results (Table 5) were evaluated in this portion of the survey. For “chance,” “may,” and “possible,” the distributions were very similar with participants associating values near 50%, 25%, and 10% most frequently (Fig. 5b). Conversely, the distribution for “expected” had much more variance

**Table 5.** For each of the top ten unqualified WEPs observed in the database of severe weather tweets, an example is given for how that phrase is used by WFOs. The specific WFO for which the example is from is shown in parentheses. The percent that each word was used out of every occurrence of an unqualified WEP is shown on the right. Each tweet contains only one WEP that was coded, as shown in bold.

Qualified WEPs	Example	Percent (%)
Possible	Strong to severe storms are possible this afternoon through 9 pm. The main threat will be damaging wind gusts of 40-60 mph...with heavy rain possible. Keep an eye on the weather today. #alwx (NWS Birmingham)	39.4
Expected	Strong to severe thunderstorms are expected late this afternoon through this evening. Stay weather aware and be ready to seek shelter if necessary. (NWS DC/Baltimore)	8.81
May	Storm SE of Yuma, CO capable of producing dime size hail and wind gusts of 40 MPH. Storm may continue to strengthen as it moves northeast. #cowx (NWS Goodland)	8.60
Could	Scattered showers and storms could bring a brief downpour, gusty winds or even pea sized hail. #MNWX #NDWX (NWS Grand Forks)	7.67
Chance	The week will start with a chilly wet Monday morning lasting into the afternoon near the Lake. The pick day of the week will be Tuesday. The end of the week features a warming trend with daily shower and thunderstorm chances (NWS Milwaukee)	5.64
Potential	The potential for severe weather returns next week across south central NE and north central KS. If you haven't practiced your severe weather drills and procedures yet this spring, now is a great time to do so. Stay tuned for updates as we head into early next week #NEwx #KSwx (NWS Hastings)	4.53
Likely	Severe Storm near Newberry headed for UF Campus in Gainesville. Strong winds and hail are likely. #jaxwx #flwx (NWS Jacksonville)	4.00
Threat	Tornado threat continues today across central #NCwx as TD #Florence continues to migrate to the west/northwest. Click the following link to download our latest #Florence briefing: <a href="https://t.co/PCyRIBYFOU">https://t.co/PCyRIBYFOU</a> (NWS Raleigh)	3.11
Not Expected	Here is the latest update concerning the storm system that will move through the state tonight and Sunday. Precipitation is already moving through eastern Oklahoma and headed our way. Severe weather is not expected. #arwx (NWS Little Rock)	2.59
Can't Be Ruled Out	@timothy_tessmer Hail is not a significant concern. However, like yesterday, pea or penny-sized hail cannot be ruled out. (NWS Northern Indiana)	2.48

and peaked near 80% (Fig. 5b). These results likely reflect the fact that these terms are relatively vague, and thus participants are anchoring to common probabilities like 25% and 50%.

Like the WX18 data, we calculated the mean values, the lower and upper bounds of the IQR, and standard deviations for all analyzed WEPs from the WX19 data (Table 8). The unqualified WEP with the lowest mean value was “chance” with an average of 30.9% (Table 8), and “expected” had the highest average of 59.4%. As also seen in Table 8, standard deviations for the unqualified words tested in WX19 range from 21.3% for “chance” to 27.8% for “expected.” The average standard deviation was 23.3%. This value is

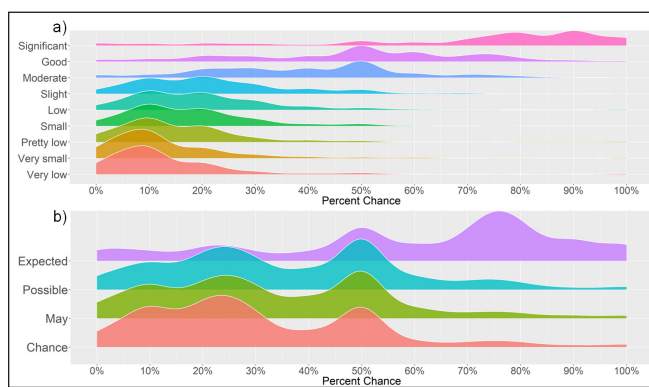
higher than the average standard deviation for qualified WEPs (17.9%), suggesting that unqualified WEPs, although extremely common in NWS tweets, likely communicate less information to the public and are less consistently understood. Supporting this claim, “may” and “possible” have a 25th quantile of 20% and a 75th quantile of 50%. “Chance” ranges from 15% to 50% and “expected” goes from 45% to 80% (Table 8).

## 5. Conclusions and discussion

Previous studies within non-meteorological contexts have suggested that the communication of probabilistic information often differs in terms of how

**Table 6.** Percent use of the top ten qualified and unqualified WEPs relative to the total 3902 tweets that use WEPs.

Qualified WEPs	Percent (%)	Unqualified WEPs	Percent (%)
Low Threat	0.494	Possible	38.1
Low Chance	0.421	Expected	8.51
Slight Chance	0.384	May	8.30
Low Risk	0.201	Could	7.41
Small Chance	0.183	Chance	5.45
Good Chance	0.146	Potential	4.37
Limited Threat	0.110	Likely	3.86
Low Confidence	0.110	Threat	3.00
Isolated Chance	0.073	Not Expected	2.51
Low Potential	0.073	Can't Be Ruled Out	2.40



**Figure 5.** Density curves show the distribution of responses when survey respondents were asked to assign a percentage to various a) qualified and b) unqualified WEPs.

the communicator wants to portray probability and how the audience would prefer to receive such information. Generally, people tend to communicate probabilistic information using words, but want to receive the same type of information numerically (Fischer and Jungermann 1996, MacLeod and Pietravalle 2017, Willems et al. 2019, Wintle et al. 2019). Other work suggests that a combination of words and numbers may be ideal for increasing understanding of probability (Budescu et al. 2009, Budescu et al. 2012, Budescu et al. 2014, Wintle et al. 2019). The main intent of this study was to understand which methods NWS WFOs use to communicate probability in severe thunderstorm and tornado forecasts and to combine this information with how members of the public understand the types of communication that are used.

Initially, we expected that the NWS would communicate probabilities using qualified WEPs or numbers, with few instances where both were used

simultaneously. While coding the tweets, it became apparent that the NWS primarily communicated probabilistic information using WEPs. These WEPs were either unqualified or qualified, with the NWS primarily using unqualified WEPs. Again, previous studies have shown that communicating probability with words tends to be less effective than with numbers (Willems et al. 2019, Wintle et al. 2019). Therefore, these results raise questions about how effective the NWS WFO severe weather-related tweets (those included in the dataset) are at communicating probability with any type of WEP. However, the efficacy of qualified versus unqualified WEPs also depends on the motivation and goals of the WFOs. If the goal is to simply inform the public of the existence of incoming severe weather, then using unqualified WEPs may achieve that goal. On the other hand, if the WFOs want each member of the audience to have the same understanding of the exact probability of severe weather, then the overwhelming use of unqualified WEPs that result in a wide range of interpretations may not be reaching that goal.

Because the motivation of WFO messaging was not a part of this study, the overall efficacy or success of WEP use cannot be conclusively defined. In addition to motivation behind WFO messaging, information about how NWS forecasters perceive various WEPs also would help to strengthen these conclusions. Investigating this topic was beyond the scope of this study, as the goal of this effort was not necessarily to assess how effective the use of WEPs in Twitter messages is on behalf of the NWS accounts. Rather, we wanted to combine these datasets in order to provide information to the NWS for reference when improving probabilistic communication in the future.

When looking at how the public interprets both qualified and unqualified WEPs within a meteorological

**Table 7.** Out of all WEPs included in WX18, those that also were found in the severe weather tweets were analyzed. Respondents were asked to give each word a numerical value. The mean percentage for each WEP is given, as well as standard deviation from the mean and the 25th and 75th quantiles.

Qualified WEPs	Mean (%)	Standard Deviation (%)	25th Quantile (%)	75th Quantile (%)
Very Low	14.0	14.8	5	20
Very Small	14.4	15.2	5	20
Pretty Low	17.2	16.3	10	20
Small	18.7	14.7	10	25
Low	20.7	15.9	10	25
Slight	22.7	15.8	10	30
Moderate	39.6	19.5	25	50
Good	48.8	22.2	30	60
Significant	70.7	26.9	60	90
AVERAGE		17.9		

**Table 8.** The top four unqualified WEPs found in the severe weather tweets were included in WX19. Respondents were asked to give each word a numerical value. The mean percentage for each WEP is given, as well as standard deviation from the mean and the 25th and 75th quantiles.

Unqualified WEPs	Mean (%)	Standard Deviation (%)	25th Quantile (%)	75th Quantile (%)
Chance	30.9	21.3	15	50
May	33.9	22.0	20	50
Possible	36.2	22.1	20	50
Expected	59.4	27.8	45	80
AVERAGE		23.3		

context, the qualified WEPs that represented lower probabilities had the lowest standard deviations. These lower deviations from the mean may indicate that participants interpreted the WEPs similarly and that there is less confusion over what terms such as “very low” and “small” mean as compared to words such as “good” and “expected.” This could further imply that people are better at interpreting words that represent lower probabilities (which they see more often in weather forecasts) as compared to their less used, higher probability counterparts, but more research into the interpretation of higher-end WEPs is needed.

Overall, variability in perceived risk was higher for numerical interpretations of unqualified words than for the qualified examples tested in WX18. The only qualified WEPs that had a higher standard deviation than some of the unqualified words were “good” and “significant.” When comparing the average of the standard deviations for the qualified and unqualified WEPs, the unqualified WEPs had a higher mean

deviation. This implies that participants are less certain about what unqualified WEPs mean and indicates that the terms tested in WX19 are more vague than those tested in WX18. For example, if the NWS sends out a message saying, “storms are expected,” people could interpret that to mean anything between a 45% to 80% chance of storms occurring. This implies that in general, using unqualified WEPs to communicate severe weather forecasts is ineffective if the goal is to provide consistent and easy-to-interpret information about forecast uncertainty that people can use to make protective decisions. In fact, based on the results of this study and past research, we suggest the NWS implement a combination of words and numbers when communicating probabilistic forecast information. This could be done in a way similar to Kent (1964) who suggested creating a set list of WEPs and assigning numerical values to those words. In practice, the values would be given alongside the WEPs, though the method of expressing the quantitative values (for example, as a

point estimate or a range) is subject to debate.

However, we recognize that the use of numbers to convey probability, especially in the case of severe weather and tornadoes, might not always be a viable option. Based on the study results, we recommend increasing the use of qualifiers when WEPs are used. (For example, instead of saying “storms are possible today,” a forecaster could say “the chance of storms is low.”) As seen from the survey comparison, people tend to interpret qualified WEPs more consistently than their unqualified counterparts, giving people a better understanding of forecast information. Other issues that should be addressed include compounding uses of unqualified WEPs by using multiple in the same message, as well as the words “risk” and “threat” being treated differently depending on the context. Although these were not the focus of this study, also improving these issues is necessary for the understanding of NWS messages to be more consistent among all audiences.

Although these findings provide important results about how WFOs communicate and how members of the public interpret probabilistic information, we note that this study is relatively narrow in its focus on severe thunderstorms and tornadoes. It is quite possible that WFOs communicate probabilistic information in different ways for different hazards. For example, we hypothesize that forecasters will use more precise WEPs (and numerical expressions of probability) in domains where they receive relatively specific probabilistic guidance (such as the probability of precipitation) and less precise WEPs in domains where they receive less specific guidance. Likewise, there may be differences in communication for relatively small-scale, discrete events (like thunderstorms and tornadoes) and large-scale events like winter storms and hurricanes. Given these possibilities, more research is necessary before we can generalize the results of this study to other forecast contexts.

In addition, we note that this study focuses exclusively on WFO text messages on Twitter, one of many communication outlets that WFOs use during severe weather events. It is likely that forecasters express probabilities in different ways on different mediums. For example, they might use text on Twitter to note basic uncertainties and graphics to provide more specific probabilistic information to people who have more time to engage with the information. Likewise, they sometimes supplement relatively short tweets with more in-depth Facebook messages and/or YouTube videos that use more precise expressions of probability

to convey forecast information. Again, more research is necessary before we can generalize the results of this study to other methods of communication.

Finally, we note that the survey results utilized in this study focus on consistency of interpretation across segments of the United States adult population. We do not examine consistency with forecaster intentions, nor do we examine the implications of inconsistent interpretation. We therefore urge future research on what forecasters mean when they use different WEPs and, perhaps more importantly, what happens when people misinterpret these meanings. Do people draw incorrect inferences about risk and therefore make poor decisions when they misjudge the probability of an event because they misinterpret a vague statement like “severe storms are possible tomorrow”? This is an open question that requires more attention as researchers continue to consider how to best communicate probabilistic information in weather forecasts.

*Acknowledgments:* The data for this project were collected with support from the University of Oklahoma. The analysis was conducted with support from NOAA’s Office of Weather and Air Quality through the United States Weather Research Program. We also would like to thank our reviewers for their valued feedback on this paper.

**APPENDIX A**

WX18 Question Variations

Qualified WEPs	Chance Mean (%)	Chance Standard Deviation (%)	Probability Mean (%)	Probability Standard Deviation (%)	Mean Difference (%)	Standard Deviation Difference (%)
Very low	13.6	14.5	14.4	15.1	-0.87	-0.60
Very small	13.7	14.2	15.0	16.1	-1.26	-1.94
Pretty low	16.8	15.9	17.6	16.7	-0.81	-0.80
Small	17.7	14.4	19.7	15.0	-2.05	-0.65
Low	19.9	15.1	21.5	16.6	-1.57	-1.54
Slight	22.0	16.4	23.4	15.2	-1.34	-1.13
Moderate	39.4	20.0	39.9	18.9	-0.48	1.11
Good	47.8	21.5	49.8	22.9	-2.04	-1.36
Significant	70.6	25.8	71.0	28.1	-0.40	-2.24

**APPENDIX B**

WX19 Question Variations

Qualified WEPs	T-Storm Mean (%)	T-Storm Standard Deviation (%)	Tornado Mean (%)	Tornado Standard Deviation (%)	Mean Difference (%)	Standard Deviation Difference (%)
Very low	14.1	13.9	14.0	15.8	0.09	-1.88
Very small	14.3	12.9	14.4	17.1	-0.12	-4.16
Pretty low	17.2	15.4	17.1	17.2	0.16	-1.86
Small	19.8	14.6	17.6	14.7	2.18	-0.11
Low	21.4	14.7	20.1	16.8	1.22	-2.09
Slight	23.2	16.2	22.1	15.4	1.11	0.80
Moderate	40.8	18.8	38.4	20.2	2.43	-1.41
Good	49.2	21.1	48.3	23.1	0.84	-1.92
Significant	72.0	26.2	69.6	27.5	2.44	-1.22

Qualified WEPs	T-Storm Mean (%)	T-Storm Standard Deviation (%)	Tornado Mean (%)	Tornado Standard Deviation (%)	Mean Difference (%)	Standard Deviation Difference (%)
Chance	33.3	21.5	28.5	20.9	4.83	0.51
May	37.4	21.8	30.4	21.7	7.06	0.08
Possible	38.7	21.8	33.7	22.2	4.99	-0.36
Expected	62.0	26.2	56.8	29.1	5.16	-2.86

## REFERENCES

- AMS Council, 2008: Enhancing weather information with probability forecasts. An information statement of the American Meteorological Society. *Bull. Amer. Meteor. Soc.*, **89**, 1049–1053. [Crossref](#).
- Barclay, S., R. V. Brown, C. W. Kelly III, C. R. Peterson, L. D. Phillips, and J. Selvidge, 1977: *Handbook for Decision Analysis*. Decisions and Designs, Inc., 66–68 pp.
- Beyth-Marom, R., 1982: How probable is probable? A numerical translation of verbal probability expressions. *J. Forecast.*, **1**, 257–269. [Crossref](#).
- Budescu, D., S. Broomell, and H.-H. Por, 2009: Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychol. Sci.*, **20**, 299–308. [Crossref](#).
- \_\_\_\_\_, H.-H. Por, S. B. Broomell, 2012: Effective communication of uncertainty in the IPCC reports. *Clim. Change*, **113**, 181–200. [Crossref](#).
- \_\_\_\_\_, H.-H. Por, S. B. Broomell, and M. Smithson, 2014: The interpretation of IPCC probabilistic statements around the world. *Nature Clim. Change*, **4**, 508–512. [Crossref](#).
- Fischer, K., and H. Jungermann, 1996: Rarely occurring headaches and rarely occurring blindness: Is rarely = rarely? The meaning of verbal frequentistic labels in specific medical contexts. *J. Behav. Decis. Mak.*, **9**, 153–172. [Crossref](#).
- Friedman, J. A., and R. Zeckhauser, 2014: Handling and mishandling estimative probability: Likelihood, confidence, and the search for Bin Laden. *Intell. Natl. Secur.*, **30**, 77–99. [Crossref](#).
- Harris, A. J. L., and A. Corner, 2011: Communicating environmental risks: Clarifying the severity effect in interpretations of verbal probability expressions. *J. Exp. Psychol. Learn. Mem. Cogn.*, **37**, 1571–1578. [Crossref](#).
- Joslyn, S., and S. Savelli, 2010: Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteorol. Appl.*, **17**, 180–195. [Crossref](#).
- Kent, S., 1964: Words of estimative probability. *Stud. Intel.*, **8**, 49–65. [Available online at [www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol8no4/html/v08i4a06p\\_0001.htm](http://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol8no4/html/v08i4a06p_0001.htm).]
- Kuligowski, E. D., F. T. Lombardo, L. T. Phan, M. L. Levitan, and D. P. Jorgensen, 2014: Final Report, National Institute of Standards and Technology (NIST) Technical Investigation of the May 22, 2011. NIST, U. S. Dept. of Commerce. [Crossref](#).
- Losee, J. E., K. Z. Naufel, L. Locker Jr., and G. D. Webster, 2017: Weather warning uncertainty: High severity influences judgment bias. *Wea. Climate Soc.*, **9**, 441–454. [Crossref](#).
- MacLeod, A., and S. Pietravalle, 2017: Communicating risk: Variability of interpreting qualitative terms. *EPPO Bulletin*, **47**, 57–68. [Crossref](#).
- Morss, R. E., J. L. Demuth, and J. K. Lazo, 2008: Communicating uncertainty in weather forecasts: A survey of the U. S. public. *Wea. Forecasting*, **23**, 974–991. [Crossref](#).
- National Research Council. (2006): *Completing the forecast: Characterizing and communicating uncertainty for better decisions using weather and climate forecasts*. The National Academies Press. [Crossref](#).
- Rothfus, L. P., R. Schneider, D. Novak, K. Klockow-McClain, A. E. Gerard, C. Karstens, G. J. Stumpf, and T. M. Smith, 2018: FACETs: A proposed next-generation paradigm for high-impact weather forecasting. *Bull. Amer. Meteor. Soc.*, **99**, 2025–2043. [Crossref](#).
- Silva, C., J. Ripberger, H. Jenkins-Smith, M. Krocak, and W. Wehde, 2018: Refining the baseline: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. University of Oklahoma Center for Risk and Crisis Management. [Available online at [risk.ou.edu/downloads/news/WX18-Reference-Report.pdf](http://risk.ou.edu/downloads/news/WX18-Reference-Report.pdf).]
- Twitter Developer (a), cited 2020: Tweet object accessed 24 Jan 2020 at [developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object](https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object).
- Twitter Developer (b) cited 2020: User object accessed 24 Jan 2020 at [developer.twitter.com/en/docs/tweets/data-dictionary/overview/user-object](https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/user-object).
- Wallsten, T. S., D. V. Budescu, A. Rapoport, R. Zwick, B. Forsyth, 1986a: Measuring the vague meanings of probability terms. *J. Exp. Psychol. Gen.*, **115**, 348–365. [Available online at [faculty.ucr.edu/~ramiz/papers/Vague\\_meanings\\_of\\_probability\\_terms.pdf](http://faculty.ucr.edu/~ramiz/papers/Vague_meanings_of_probability_terms.pdf).]
- \_\_\_\_\_, S. Fillenbaum, and J. A. Cox, 1986b: Base rate effects on the interpretations of probability and frequency expressions. *J. Mem. Lang.*, **25**, 571–587. [Crossref](#).
- Willems, S. J. W., C. J. Albers, and I. Smeets, 2020: Variability in the interpretation of Dutch probability phrases - a risk for miscommunication. *J. Sci. Comm.*, **19**. [Crossref](#).
- Windschitl, P. D., and G. L. Wells, 1996: Measuring psychological uncertainty: Verbal versus numeric methods. *J. Exp. Psycho. Appl.*, **2**, 343–364. [Crossref](#).
- Wintle, B. C., H. Fraser, B. C. Wills, A. E. Nicholson, and F. Fidler, 2019: Verbal probabilities: *Very likely* to be *somewhat* more confusing than numbers. *PLOS ONE*, **14**: e0213522. [Crossref](#).