



Communicating Probability Information in Hurricane Forecasts: Assessing Statements that Forecasters Use on Social Media and Implications for Public Assessments of Reliability

ZOEY ROSEN

Department of Journalism and Media Communication, Colorado State University, Fort Collins, CO

MAKENZIE J. KROCAK

University of Oklahoma National Institute for Risk and Resilience, Norman, Oklahoma

JOSEPH T. RIPBERGER

University of Oklahoma National Institute for Risk and Resilience, Norman, Oklahoma

RACHAEL CROSS

School of Meteorology, University of Oklahoma, Norman, Oklahoma

EMILY LENHARDT

School of Meteorology, University of Oklahoma, Norman, Oklahoma

CAROL L. SILVA

University of Oklahoma National Institute for Risk and Resilience, Norman, Oklahoma

HANK C. JENKINS-SMITH

University of Oklahoma National Institute for Risk and Resilience, Norman, Oklahoma

(Manuscript received 15 March 2021; review completed 27 August 2021)

ABSTRACT

Forecasters are responsible for predicting the weather and communicating risk with stakeholders and members of the public. This study investigates the statements that forecasters use to communicate probability information in hurricane forecasts and the impact these statements may have on how members of the public evaluate forecast reliability. We use messages on Twitter to descriptively analyze probability statements in forecasts leading up to Hurricanes Harvey, Irma, Maria, and Florence from forecasters in three different groups: the National Hurricane Center, local Weather Forecast Offices, and in the television broadcast community. We then use data from a representative survey of United States adults to assess how members of the public wish to receive probability information and the impact of information format on assessments of forecast reliability. Results from the descriptive analysis indicate forecasters overwhelmingly use words and phrases in place of numbers to communicate probability information. In addition, the words and phrases forecasters use are generally vague in nature -- they seldom include rank adjectives (e.g., “low” or “high”) to qualify blanket expressions of uncertainty (e.g., “there is a chance of flooding”). Results from the survey show members of the public generally prefer both words/phrases and numbers when receiving forecast information. They also show information format affects public judgments of forecast reliability; on average, people believe forecasts are more reliable when they include numeric probability information.

1. Introduction

Driven by scientific advances in meteorology, risk communication, and the needs of the emergency management community, the National Weather Service (NWS) is rapidly expanding use of probability information in forecasts, products, and decision support services. Although this is true across the NWS, one might argue the National Hurricane Center (NHC) has the most experience in this domain. Beginning with early work on tropical cyclone (TC) strike probabilities (Jarrell and Brand 1983; Sheets 1985), NHC meteorologists, along with those at Weather Forecast Offices (WFOs) and television stations along the coast, have been communicating probability information for many years. How do they do it? Do they use numbers, words and phrases, or some combination of both? If they use words and phrases to express uncertainty, how much information do the words and phrases communicate? Do these choices matter? Do they affect public comprehension and response to forecast information?

Building on recent research in severe thunderstorm and tornado risk communication (Lenhardt et al. 2020), this study continues to address these important questions by analyzing patterns of probability communication in the TC domain. In addition to expanding the domain to tropical cyclones, the study addresses the following questions researchers have yet to address:

1. Do NHC, WFOs, and broadcast meteorologists (BMs) communicate probability information in the same way? Or, are there significant differences across the groups that we might relate to jurisdiction and audience?
2. What “subjects” (forecast features) do NHC, WFOs, and BMs focus on when communicating probability information? Are they more likely to focus on uncertainty about timing, for example, than the location or intensity of a TC?
3. What type of probability information (words, numbers, both) do members of the public want? How do public preferences align with current practices?
4. Do choices that meteorologists make about how to communicate probability information affect public comprehension and response to forecast information? If so, how?

We address the first two questions by examining forecast communication on Twitter during Hurricanes Harvey, Irma, Maria, and Florence. We address the next two questions by examining public responses to a new, yearly survey of the United States public that focuses on TC forecast and warning communication. To contextualize the study results, we begin with a short review of past literature on probability communication in the weather domain. Next, we describe the data, methods, and results. We conclude by discussing the implications of the results for improving communication of probability information in the weather enterprise.

2. Past research

Uncertainty is inherent in nearly all scientific prediction; risk communicators (such as emergency managers or public information officers) often use probability information to convey uncertainty to stakeholders and members of the public. Given the prevalence of this practice, there is considerable research on how to effectively communicate probability information in risk messages. Recent reviews of this literature (e.g., Morss et al. 2008; Joslyn and LeClerc 2012; Grounds and Joslyn 2018) indicate a few core findings that cut across hundreds of studies. These reviews generally agree on two key points: (1) average members of the public are able to understand and use probability information when making decisions about risk, as long as consideration is given to information presentation; and (2) assuming appropriate presentation, probability information generally improves decision satisfaction and quality.

These findings obviously beg questions about how to “appropriately” present probability information. Answers to these questions vary, but most studies and best practice reports indicate that risk communicators should use numeric probability information when possible (Trevena et al. 2006; Trevena et al. 2013; Peters 2017; Jenkins et al. 2019; Bonner et al. 2021; Mandel and Irwin 2021). Whereas this may seem like relatively simple advice, it can be very difficult to implement in practice because some types of numeric probability information can be confusing. Probability of Precipitation (PoP) forecasts provide a perfect example. People interpret phrases like “there is a 30% chance of rain” in multiple, mutually contradictory ways that are often inconsistent with expert intentions (Murphy et al. 1980; Gigerenzer et al. 2005; Morss et al. 2008; Joslyn et al. 2009; Juanchich and Sirota 2016; Juanchich and

Sirota 2018). This confusion stems from the fact that precipitation on a given day in a given area is a single event probability and that risk communicators generally give insufficient attention to the reference class when presenting PoP forecasts. In addition to some types of information, some groups of people are less able than others to understand and use complex numeric information when making decision (Peters et al. 2006; Reyna et al. 2009; Cokely et al. 2018; Peters 2020). Likewise, a variety of cognitive biases and heuristics affect the way people interpret numeric probability information when making decisions (Keller et al. 2006; Slovic 2010; Kahneman 2011).

Driven by some of these challenges, many risk communicators are uncomfortable providing numeric probability information to the public, so they translate numeric information into words and phrases when issuing public statements. This practice can lead to inconsistent interpretations of forecast information among members of the public and interpretations of probability information inconsistent with expert intentions (Budescu et al. 2014). Likewise, it can reduce trust in the information (Gurmankin et al. 2004) and exacerbate risk perceptions and biases (Berry et al. 2002; Berry and Hochhauser 2006). Because of this, researchers broadly agree it is best to use numeric translations when using words and phrases to communicate probability information (e.g., Harris et al. 2017). For example, the Intergovernmental Panel on Climate Change (IPCC) uses a combination of phrases and numeric intervals (for example: very likely 90–100%) to indicate certainty in scientific statements (Budescu et al. 2012). If numeric translations are not available, research suggests that communicators strive to include rank adjectives (also known as “qualifying expressions”) like “low,” “medium,” and “high” to convey a rough sense of magnitude when expressing the possibility of an outcome (Lenhardt et al. 2020). For instance, sentences like “there is a low chance of severe thunderstorms” generate more consistent interpretations than “there is a chance of severe thunderstorms.”

In a recent study, Lenhardt et al. (2020) examined the extent to which meteorologists adhere to past guidance when communicating information about the probability of severe thunderstorm and tornado events. Their analysis of 8900 tweets indicates WFO meteorologists routinely include probability information in forecasts, but very rarely include numbers. Less than 1% of the probability messages they analyzed included numeric information. Instead, 99%+ of the messages used words and phrases

to convey the possibility of an event. Furthermore, 95% of the messages used words and phrases did not include rank adjectives to “qualify” the possibility of the event. For example, 40% of the messages simply said storms/tornadoes are “possible,” giving no information about the magnitude of the possibility. These results are specific to the severe weather domain, and it is worth noting that TC probabilities are generally point probabilities and generally larger than severe weather probabilities. The use of words like “possible” and “chance” have been identified in other studies on communicating probability, and are valuable to study in more domains because of the ramifications on how people interpret and interact with forecast information (Morss et al. 2008; Budescu et al. 2014).

3. Background and expectations

Overall, the study by Lenhardt et al. (2020) indicates there is considerable room for improvement in how WFO meteorologists convey probability information about severe thunderstorms and tornadoes. We advance this line of research in multiple ways. First, we assess the generalizability of the findings to other weather domains, in this case, the TC domain. This is an important advancement because meteorologists have access to relatively little probabilistic guidance about when and where severe convective storms and tornadoes will materialize. It is possible this lack of information flows downstream to affect the way meteorologists communicate about the probability of these events to members of the public. If this is true, then we should observe a different pattern of communication in the TC domain, where probabilistic guidance is more readily available to meteorologists. When probability information is available, do meteorologists adhere to guidance from past research? Do they use numbers more often than words and phrases? When using words and phrases, do they use rank adjectives to indicate the relative magnitude of a probability?

Next, we assess the generalizability of the findings to different types of meteorologists. WFO meteorologists are an important link in the TC risk communication chain, but the chain also includes NHC and BMs. Although we have no specific expectations about how patterns of communication will vary across this set of actors, it is probable there will be differences from each group. NHC, who specialize in TCs, may have more expertise in the TC domain and may be more conversant with probabilistic TC guidance than

WFOs and BMs, who work across many domains. This specialization, expertise, and familiarity may result in more frequent use of numeric information and more informative words/phrases. Alternatively, BMs specialize in public risk communication, so they may be more likely to adhere to best practices than NHC and WFOs, who have less training in risk communication.

Finally, we assess the generalizability of the findings to different features (i.e., subjects of uncertainty) in a TC forecast. There are many subjects of uncertainty in a forecast, some relate to timing (when is an event going to happen?), location (where is an event going to happen?), and intensity (how bad is an event going to be?). We want to know if communication strategies vary depending on the feature meteorologists are highlighting in a given message. Do meteorologists use more precise language (e.g., numbers, rank adjectives) when discussing the timing and location of an event, and less precise language when discussing the possible intensity of an event? Again, we do not have specific expectations for this portion of the analysis, but it is plausible that forecast models and NWS products affect communication choices.

In addition to assessing the language that meteorologists use to convey probability information, Lenhardt et al. (2020) indicates public interpretation of the language meteorologists most commonly use is highly variable. Although this is an important finding, the study provides little information about the types of probability people want and, more importantly, the effect of information format on public comprehension and response to forecast information. We address both points in this research. On the first point, we hypothesize that most members of the public want meteorologists to use both words, phrases, and numbers when communicating forecast information because the combination covers multiple bases — it provides a true sense of the probability (the numbers) and a subjective sense (“gist”) of how experts interpret the probability (e.g., do the words and phrases suggest 15% is high or low?).

More important than preferences, we examine the impact of information format on public perceptions about the reliability of a forecast. As we note above, previous research indicates that probability information can enhance and improve decision making (Jenkins et al. 2019). A reliable, trustworthy source is important in decision making; the higher the levels of trust in a source, the more likely the audience will follow the recommendations in the message, like hazard forecasts

(Morss et al. 2018). People generally believe forecasts are more reliable when meteorologists explicitly convey uncertainty information (Savelli and Joslyn 2013). We conjecture this is true, depending on the language meteorologists use to convey probability information. Specifically, we hypothesize public assessments of reliability will be highest when forecasters use precise language (e.g., numbers) to convey probability information and lowest when forecasters use vague words and phrases (e.g., there is a “chance” of flooding). We are less certain of how members of the public will view forecasts including both words and phrases, and numeric translations. On the one hand, we expect most people want words and numbers when viewing forecasts, so seeing words and phrases in addition to numbers may enhance public confidence. On the other hand, some people may view words and phrases as “hedgies” that forecasters use because they do not trust the numbers. If this is the case, then including words and phrases with numbers may harm public assessments of forecast reliability. Resolving this question is critically important because it may provide guidance about how to communicate probability information when relatively precise numbers are available — should meteorologists stick to the numbers, or include words and phrases to help people interpret the numbers?

4. Data and methods

Following Lenhardt et al. (2020), we used text messages on Twitter (“Tweets”) to assess the language meteorologists use to convey information about the probability of TC hazards to members of the public. The messages were collected using the University of Oklahoma (OU) Center for Risk and Crisis Management’s Severe Weather and Social Media data collection platform, which taps into Twitter’s streaming application program interface to continuously collect and archive tweets that contain certain keywords that reference weather threats (such as “tornado” or “hurricane”). The program collects the text of each tweet and relevant metadata, including the author’s username and description (see Ripberger et al. 2014 for more information).

TC events, especially significant hurricanes, generate a significant volume of data on Twitter. In some cases, we have seen 5+ million tweets per day about a given hurricane. We limit this volume of data by focusing on four significant hurricanes — Harvey, Irma, Maria, and Florence. We chose these events

because they affected different portions of the United States, and therefore attracted tweets by a wide variety of WFOs and BMs. Table 1 provides information about the data collection stream for each hurricane, including data collection time windows, keywords, and users. For example, on 23 August 2017, we began collecting every tweet around that contained one or more of these words: hurricane, harvey, #hurricane, #Harvey, #HurricaneHarvey, HurricaneHarvey. We concluded this data collection on 1 September 2017. Though we collected 20+ million tweets using these keywords during the Harvey timeframe, we focused on tweets from three specific groups: the NHC, WFOs, and BMs in areas hit by Harvey.

Following data collection, we evaluated every tweet to ensure it met two basic conditions: (1) it included some type of forecast information (not simply a report on where the storm was or what it was doing); and (2) the forecast used some type of probability information (not only deterministic information). In all, 2138 tweets met these conditions — 94 tweets from the NHC, 424 from WFOs, and 1620 from BMs. After limiting the tweets to this set, analysts read and classified each forecast message along two binary dimensions: (1) use of numbers to convey probability information; and (2) use of words to convey probability information. Note these dimensions were not mutually exclusive, some forecasts used both numbers and words. Next, all forecasts using words to convey probability information were given an additional binary code for use (or nonuse) of rank adjectives to “qualify” the magnitude of otherwise vague probability statements. For example, a use of a rank adjective would include “high” or “low” connecting to the word “chance”. While making this determination, we also documented the specific words meteorologists used to convey the probability information.

Table 2 displays the coding categories and examples. Intercoder reliability was assessed using Cohen’s kappa for four independent coders, calculated via the utility ReCal3 (Freelon 2010; Hallgren 2012). Values above 0.80 are considered near perfect agreement, which is met for all of the categories used in this study (Landis and Koch 1977).

In addition to the categorical codes described above, the subjects connected to each probability expression were analyzed to see if any patterns emerged. We focused on three subjects: timing ($\kappa = 0.76$), location ($\kappa = 0.79$), and intensity ($\kappa = 0.68$). Timing statements indicate when a hazard may occur; location statements

indicate where it may occur; and intensity statements indicate the potential strength and magnitude of the forecasted hazard. These subject matter codes were more latent in nature, so the reliability scores among four coders was understandably lower than desired. However, we feel these scores were high enough to discuss briefly in context and as a possible avenue for future analyses.

We used data from the Tropical Cyclone and Society Survey to assess public preferences about the format of probability information and the effect of format on judgements of forecast reliability. The TC Survey is an annual survey of United States adults (age 18+) designed and fielded by the OU Center for Risk and Crisis Management. We used the 2020 iteration of the survey (TC20) for this analysis. TC20 was administered online using two different samples that were provided by Qualtrics: (1) a demographically representative sample of 1000 adults who live across the United States, and a demographically representative oversample of 2000 adults who live in County Warning Areas that historically have been affected by TC events. Inasmuch as TC20 included many questions, we analyzed data from two parts of the survey. First, we assessed responses to this question:

Some forecasters use words and phrases in place of numbers when describing the probability that an event will happen. For example, they might use the phrase “slight chance” in place of “15% chance” when describing the probability of rain at a location.

When you get information from forecasts, do you want:

- 1 – Numbers*
- 2 – Words*
- 3 – Both*

Next, we analyzed data from a survey experiment that was designed to isolate the effect of information format on reliability judgements. The experiment began by telling respondents:

Now we are going to show you some example forecast messages from past hurricanes and ask you how reliable the messages seem to you. We know this can be difficult without more information. Just give us your first impression.

Table 1. Description of Twitter Data.

Hurricane	Users (and Tweet Count) by Source Group
Harvey Date: Aug 23 - Sep 1, 2017 Searchterms: hurricane, harvey, #hurricane, #Harvey, #Hurricane-Harvey, HurricaneHarvey	NHC (170): NHC_Atlantic WFOs (1238): NWS_LCH, NWSHouston, NWSCorpus, NWS-Brownsville, NWSSanAntonio, NWSShreveport, NWSMobile Top 5 BMs (9151): Fox26Houston, KHOU, abc13houston, KPRC2, ksatnews (+51 more)
Irma Date: Sep 2 - Sep 11, 2017 Searchterms: hurricane, irma, #hurricane, #Irma, #HurricaneIrma, HurricaneIrma	NHC (179): NHC_Atlantic WFOs (870): NWSKeyWest, NWSSanJuan, NWSMiami, NWS Tampa Bay, NWS Jacksonville, NWSCharlestonSC, NWS-Melbourne Top 5 BMs (12463): ActionNewsJax, Fox35News, CBSMiami, WPBF25News, abcactionnews (+50 more)
Maria Date: Sep 18 - Sep 26, 2017 Searchterms: hurricane, #hurricane, maria, #Maria, #Hurricane-Maria, HurricaneMaria	NHC (143): NHC_Atlantic WFOs (153): NWSSanJuan, NWSMoreheadCity, NWSCharleston-SC, NWSWakefieldVA, NWSWilmingtonNC (+7 more) Top 5 BMs (381): wcti12, WTKR3, 13NewsNow, WAVY_News, wncn9
Florence Date: Sep 4 - Sep 18, 2018 Searchterms: hurricane, #hurricane, #florence, #hurricaneflorence	NHC (184): NHC_Atlantic WFOs (739): NWSRaleigh, NWSMoreheadCity, NWS Columbia, NWSWilmingtonNC, NWSWilmingtonNC Top 5 BMs (6623): WLTX, wpdeabc15, WNCN, ABC11_WTVD, WRAL (+24 more)

Table 2. Description of Coding Categories and Intercoder Reliability. Information format codes are bolded in the example column for clarity. Tweets could contain multiple information formats, so the categories are not mutually exclusive.

Information Format	Example	Cohen's kappa (κ)
Numbers	Probability of tropical storm force winds 10% or less for most of SE Louisiana through Monday AM. #LAWx #Harvey	0.90
Words (w/no rank adjectives)	#Harvey's heavy rainfall isn't over yet - an additional 10-25 inches expected over the next 5 days for much of the area #houwx #glswx #txwx	0.88
Words (w/rank adjectives)	Here are the latest estimates of when tropical-storm-force winds could arrive & are most likely to arrive with #Irma	0.83

After seeing this text, respondents were shown a hypothetical wind and flooding forecast that was posted by the NHC on Twitter (Fig. 1).

As the examples in Fig. 1 illustrate, the first condition in the experiment used words only to convey probability information; the remaining conditions used numbers only, words and numbers (percentage point estimate), or words and numbers (percentage range). Table 3 shows the text used for each condition; all other features of the message were the same (e.g., time, date, retweets, likes). Each respondent was randomly assigned to one condition in each forecast area (wind and flooding). After viewing the forecasts, respondents were asked to answer this question:

Using this information alone, how reliable does this forecast message seem to you?

- 1 – Not at all reliable
- 2 – Slightly reliable
- 3 – Moderately reliable
- 4 – Very reliable
- 5 – Extremely reliable

5. Findings

What formats do meteorologists use to convey probability information in TC forecasts? As shown in Fig. 2, our results clearly demonstrate that

Table 3. Description of Survey Experiment.

Condition	Wind Forecast	Flooding Forecast
Words Only	There is a chance of hurricane force winds in Key West, FL beginning tomorrow afternoon at 4:00.	We expect significant flooding in Houston tomorrow evening as the hurricane slows and continues into Texas.
Numbers Only	There is a 50% chance of hurricane force winds in Key West, FL beginning tomorrow afternoon at 4:00.	There is an 80% chance of significant flooding in Houston tomorrow evening as the hurricane slows and continues into Texas.
Words and Numbers (Point Estimate)	There is a moderate (50%) chance of hurricane force winds in Key West, FL beginning tomorrow afternoon at 4:00.	We expect (80% chance) to see significant flooding in Houston, TX tomorrow evening as the hurricane slows and continues into Texas.
Words and Numbers (Range)	There is a moderate (40% - 60%) chance of hurricane force winds in Key West, FL beginning tomorrow afternoon at 4:00.	We expect (70% - 90% chance) to see significant flooding in Houston, TX tomorrow evening as the hurricane slows and continues into Texas.



Figure 1. Example wind and flooding forecast messages used in the survey experiment; note that these forecasts use words alone to convey probability information. *Click image for an external version; this applies to all figures hereafter.*

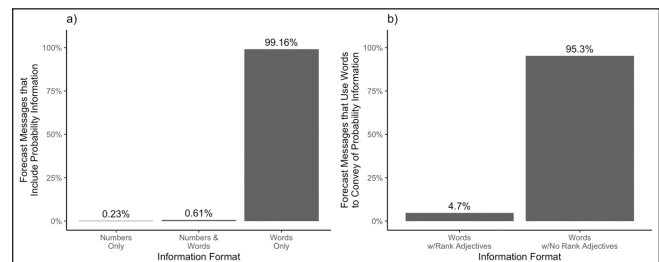


Figure 2. a) the distribution of forecast messages that include probability information (n = 2138 tweets) by information format; b) the distribution of forecast messages that use words to convey probability information by use of rank adjectives (n = 2132 tweets).

meteorologists almost always used words in place of numbers to communicate probability information about TCs in the data set. Of the 2138 tweets analyzed, 2120 (99.16%) used words alone, five (0.23%) used numbers alone, and 13 (0.61%) used a combination of words and numbers. Furthermore, when using words, meteorologists rarely use rank adjectives (“qualifying expressions”) like “low,” “medium,” and “high” to convey a rough sense of magnitude when expressing the possibility of a TC threat. Only 100 (4.7%) of the 2132 tweets that used words to convey probability information included a rank adjective.

To contextualize these findings, Fig. 3 displays the most common base words that meteorologists used to convey probability information. “Expect” was by far the most common base word; 932 (45.9%) of the 2032 messages we analyzed used some variant of the word to convey information about the probability of an event. The next most common base words were “could,” “likely,” and “possible.”

Here are a few examples using these words to express probability information:

- Hurricane force winds are **expected** to extend well inland over portions of the Florida Panhandle and portions of southeast Alabama and southwest Georgia as Michael moves inland, where hurricane warnings are in effect.
- If we were to see tropical storm force winds this week, they **could** arrive sometime Monday or Tuesday.
- Harvey is far from over. East Texas is **likely** to receive more rain this week.
- **Possibility** of tornadoes in Collier County from Hurricane Irma.

Although the information in these examples is clearly important, all of the forecasts use words in place of numbers to convey probability information and do not include rank adjectives to orient readers to the probability’s extent. For example, statements like “If we

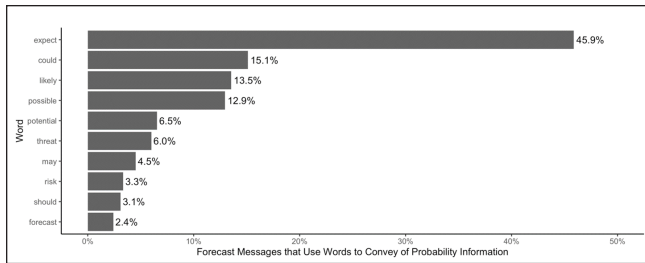


Figure 3. The base words that meteorologists most commonly use to convey information about the probability of TC hazards.

were to see tropical storm force winds this week, they could arrive sometime Monday or Tuesday” provide little information about how likely the winds are. Readers do not know if the probability is low, medium, or high, which would enhance the tweet’s message. This is true of most the messages in the analysis.

The results above indicate a strong pattern in how meteorologists generally convey probability information in social media messages. We now turn to possible differences across groups of meteorologists and subjects of uncertainty. As the NHC, WFOs, and BMs cover different scales of forecasts owing to their location and expected audiences, we were curious to see if or where the use of numerical information varied. Beginning with the former research question, do NHC, WFOs, and broadcast meteorologists communicate probability information in the same way? Or are there significant differences across the groups that we might relate to expertise, jurisdiction, and audience? As shown in Fig. 4, we find very modest differences across the groups. Meteorologists at the NHC were responsible for 94 forecast messages in the dataset; all of them (100%) used words alone to convey probability information. WFO meteorologists sent 424 messages; 1 (0.2%) used numbers alone, 1 (0.2%) used numbers and words, and 422 (99.5%) used words alone. Broadcast meteorologists were responsible for 1620 of the messages; 4 (0.2%) used numbers alone, 12 (0.7%) used words and numbers, and 1604 used words alone (99.0%). In addition to basic information format, our findings indicate rough comparability in the proportion of messages that use rank adjectives to qualify basic words phrases like “we expect *x*” or “*y* could happen.” Only five (5.3%), 30 (7.1%), and 65 (4.0%) messages from NHC, WFOs, and BMs (respectively), used some type of rank adjective to clarify the underlying probability of “*x*” or “*y*.” In short, there is little evidence that NHC, WFOs, and broadcast meteorologists communicate probability information in different ways.

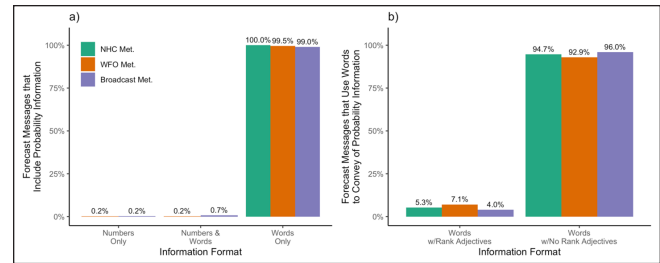


Figure 4. a) the distribution of forecast messages that include probability information by information format and type of meteorologist (NHC Met. $n = 94$; WFO Met. $n = 424$; Broadcast Met. $n = 1620$); b) the distribution of forecast messages that use words to convey probability information by use of rank adjectives and type of meteorologist (NHC Met. $n = 94$; WFO Met. $n = 423$; Broadcast Met. $n = 1615$).

What subjects of uncertainty (timing, location, intensity) do meteorologists focus on when communicating probability information? Do information formats vary, depending on these subjects? Our findings with respect to these questions are shown in Fig. 5. It is important to note a few things about this portion of the analysis while reviewing these findings. First and foremost, it was often difficult and somewhat subjective to discern the subject of uncertainty in a given message, so we urge readers to view these results as exploratory. Next, many messages referenced multiple subjects when conveying uncertainty, so these categories were not mutually exclusive. Finally, it is interesting to note that intensity/magnitude was the most common subject of the uncertainty statements, but location and timing were not far behind. Of the 1920 messages with a discernable subject, 1290 (67.2%) focused on intensity/magnitude; 1104 (57.5%) focused on location; and 961 (50.0%) focused on uncertainty with respect to timing. Despite these slight differences, there were no underlying patterns in the relationship between information format and subject. As shown in Fig. 5, the overwhelming majority of messages with discernable subjects used words alone to express uncertainty. Moreover, most of the messages did not include rank adjectives. If anything, messages focused on uncertainty with respect to location were more likely to use numbers and rank adjectives, but the difference between these messages and the others was minute.

The results above indicate meteorologists of all types almost always use words in place of numbers when conveying probability information about TC threats in social media messages. This is true regardless of the type of probability information they are

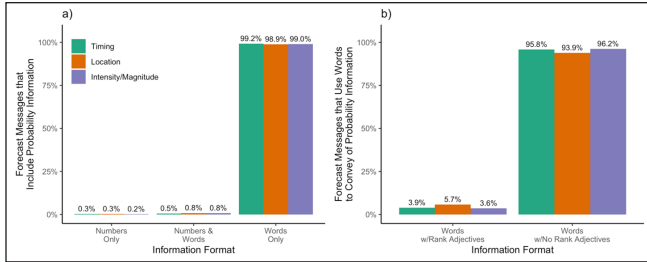


Figure 5. a) the distribution of forecast messages that include probability information by information format and subject (timing n = 961; location n = 1104; intensity/magnitude n = 1290); b) the distribution of forecast messages that use words to convey probability information by use of rank adjectives and subject (timing n = 958; location n = 1100; intensity/magnitude n = 1287).

highlighting, be it uncertainty about timing, location, or intensity/magnitude. The starkness of this finding begs an important question — is this the type of information members of the public want in TC forecasts? Are members of the public looking for words, numbers, or both in TC forecasts? Our findings are shown in Fig. 6. When we asked survey respondents across the United States and hurricane prone portions of the country, relatively few said they are looking for words alone (13.5% and 14.3% respectively). Likewise, relatively few said they were looking for numbers alone (13.2% and 10.3%). Rather, a strong majority (73.3% and 75.4%) said they want meteorologists to use words and numbers when conveying probability information.

Public preferences obviously contrast with current practices, but do the choices meteorologists make about how to communicate probability information affect public comprehension and response to forecast information? If so, how? As we explain above, we used a survey experiment to address this question. Survey respondents were randomly assigned to a forecast message that used one of four information formats (words only, numbers only, words and numbers with point estimates, and words and numbers with probability ranges) to communicate the probability of flooding and hurricane force winds. They were then asked to assess the reliability of the forecast. The results from this experiment are shown in Fig. 7, that plots the mean forecast reliability score by information format condition. On average, the patterns in the plot indicate survey respondents who saw messages with words alone judged the forecast to be less reliable than respondents who saw messages with numbers

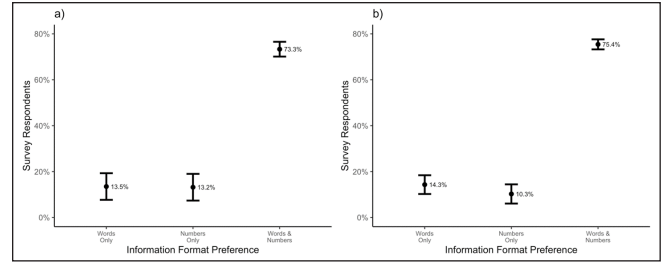


Figure 6. a) the percentage of survey respondents in the national sample who chose each format (n = 986); b) the percentage and of survey respondents in the TC sample who chose each format (n = 1970); error bars indicate 95% confidence intervals.

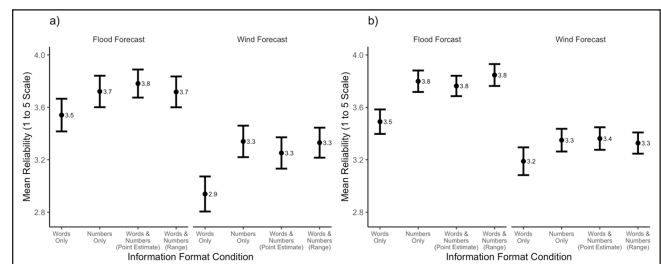


Figure 7. a) the mean response on the reliability scale from survey respondents in the national sample who were randomly assigned to each information format condition (n = 992); b) the mean response on the reliability scale from survey respondents in the TC sample who were randomly assigned to each information format condition (n = 1986); error bars indicate 95% confidence intervals.

alone. This pattern is evident for both hazards in both samples, though it is more subtle in the flood forecast for the national sample and the wind forecast for the TC sample. In these cases, the mean difference between the words only condition and numbers only condition is 0.18 and 0.16, respectively. In the wind forecast for the national sample and the flood forecast for the TC sample, the mean difference is roughly twice as large — 0.40 and 0.31, respectively. Just as interesting, the findings indicate little or no difference between forecasts that use numbers alone and those that use words in addition to numbers. This finding suggests judgements of reliability seem to hinge on the words versus numbers dichotomy — words are satisfactory, as long as meteorologists include numeric translations.

6. Discussion

This study was motivated by the rapid expansion of probability information in NWS forecasts, products, and decision support services. Although many experts believe this is a positive development, probability information can be complex and difficult to convey to members of the public. Past research on risk communication suggests two overarching points on this topic: (1) average members of the public are able to make sense of and use probability information when making decisions about risk, as long as consideration is given to information presentation; and (2) assuming appropriate presentation, probability information generally improves decision satisfaction and quality.

What does appropriate presentation look like? Research in this area is evolving, but most studies agree risk communicators ought to use numeric probability information as often as possible (e.g., Peters 2017; Cokely et al. 2018; Peters 2020). Despite this consensus, many communicators are uncomfortable providing numeric probability information to average members of the public, so they translate numeric information into words and phrases when issuing public statements. This can cause inconsistent interpretations among members of the public and interpretations inconsistent with expert intentions. To prevent this, researchers broadly agree it is best use numeric translations when using words and phrases to communicate probability information. In other words, when communicating forecast information, it is best to use both words and numbers whenever possible. As a last resort, if numeric translations are not available, research suggests communicators strive to include rank adjectives like “low,” “medium,” and “high” to convey a rough sense of magnitude when expressing the possibility of an outcome.

Do meteorologists adhere to this guidance when conveying probability information to members of the public? A recent study of social media forecast messages from WFOs during severe thunderstorm and tornado events indicates they do not (Lenhardt et al. 2020). In this domain, meteorologists almost exclusively rely on words alone to convey probability information. Furthermore, they rarely use rank adjectives to convey a rough sense of probability magnitude when expressing the possibility of a severe thunderstorm or tornado threat. Although these findings are important, they focus on a single group of meteorologists (those at WFOs) and a single threat domain (severe thunderstorms or tornados) where models provide less explicit probability guidance

than other domains.

In the present study, we sought to test the generalizability of the findings above by assessing the extent to which meteorologists use words, phrases, and numbers when conveying probability information in the TC domain, where probability information is more prevalent, both in models and products. Additionally, we sought to expand upon previous research by exploring patterns of communication across three groups (NHC, WFOs, and BMs) and in reference to different subjects of uncertainty (timing, location, and intensity).

Our findings were striking. They align almost exactly with previous research in the severe thunderstorm and tornado domains. Nearly all of the forecast messages we assessed used words alone to convey probability information and very few made use of rank adjectives to help readers make sense of the probabilities that underly the words. Considering this similarity, we reject the idea meteorologists are using words to convey probability information because numbers are not available. In many cases, explicit numeric information is available in the TC domain. This seems to be true of all meteorologists, regardless of specialization, expertise, and familiarity with probability information and/or risk communication. Likewise, it seems to be true across the variety of events meteorologists focus on when issuing public statements. Although there is numeric probabilistic guidance available to meteorologists, we recognize those probabilities are generally point probabilities, meaning they are more difficult to communicate when discussing a large area. Furthermore, a lack of training on when and how to use numbers to communicate uncertainty also likely plays a role in meteorologists choosing to use words instead of numbers.

Next, we sought to understand if current communication practices are also inconsistent with public preferences. How do the choices that meteorologists make about how to communicate probability information affect public comprehension and response to forecast information? Our findings indicate roughly 75% of people want forecasters to use words and numbers when conveying probability information. If we add this to the 15% of people who are looking for numbers alone, we conclude that nearly 90% of people are looking for information meteorologists rarely include in messages on social media.

Perhaps more important than preferences, the results from our survey experiment suggest the format meteorologists use to convey probability information

can affect public judgments about the reliability of a forecast. On average, people believe forecast messages with words alone are less reliable than messages that use numbers alone or a combination of words and numbers. Although we do not demonstrate a direct connection between assessments of reliability and responsiveness to forecasts and warnings, there is considerable literature on the relationship between trust and protective action (e.g., Ripberger et al. 2015). When people do not trust forecasts and warnings, they are less likely to take protective action in response to them. Although we do not believe information formats alone significantly affect levels of trust that members of the public place in meteorologists, we worry that using words alone to communicate probability information may cause some people to question the reliability of a specific message and hesitate to act upon it.

7. Conclusions

Overall, our results suggest meteorologists can vastly improve risk communication by including numeric information in messages to convey probability information. We suggest meteorologists use a combination of words and numbers to convey the probability of an outcome; numbers to give people an underlying sense of the probability and words to help them contextualize the probability — is the probability large or small in comparison to average, or past events? If numeric information is not available, we urge meteorologists to use rank adjectives or qualifying like “low,” “medium,” and “high” to convey a rough sense of magnitude when expressing the possibility of an outcome. For instance, statements like “there is a moderate chance of flooding” convey significantly more information than statements like “there is a chance of flooding.”

This research, along with Lenhardt et al. 2020, highlights the complexity that comes with communicating probabilistic information from NHC, WFOs, and BMs. We recommend forecasters (where possible) include numeric information when communicating probabilities. Where not possible, we recommend groups standardize the use of incorporating rank adjectives (e.g., high or low) with probabilistic phrases (e.g., chance or likelihood) for public-facing forecasts. This can be done via social media toolkits so training is consistent among meteorologists. Additionally, writing the recommendation for using numerical information and/or rank adjectives and

probabilistic phrases into the toolkits would aid in the overall adoption of these findings. Even though these findings pertain to the severe and tropical forecasting domains, the suggestion for incorporating numerical information and more descriptive words should be considered in the training for all hazard types, as decisions pertaining to the weather are made across all types of threats.

We realize these suggestions do not address many of the difficulties that come with the inclusion of probability information in forecast messages. For example, past research provides a variety of recommendations about how to effectively display probability information in graphics such as plots and maps. This project analyzed text alone, but meteorologists often use graphics to communicate forecast information in social media messages. Analysis of these graphics will be an important next step in assessing the extent to which meteorologists adhere to guidance from past research when conveying probability information and identifying opportunities for improvement. The same is true of other mechanisms and venues meteorologists use to disseminate forecasts, such as live broadcasts and in decision support presentations to partners in the emergency management community. We have yet to examine patterns of communication in these settings, but strongly believe such analyses will generate important insight about how to convey probability information more effectively in the future.

Acknowledgments. The data for this project were collected with support from the University of Oklahoma. The analysis was conducted with support from NOAA Award No. NA19OAR0220120.

REFERENCES

- Berry, D. C., P. R. Knapp, and T. Raynor, 2002: Is 15 percent very common? Informing people about the risks of medication side effects. *Int. J. Pharm. Pract.*, **10** (3), 145–151, [CrossRef](#).
- _____, and M. Hochhauser, 2006: Verbal labels can triple perceived risk in clinical trials. *Drug Inf. J.*, **40** (3), 249–258, [CrossRef](#).
- Bonner, C., and Coauthors, 2021: Current best practice for presenting probabilities in patient decision aids: fundamental principles. *Med. Decision Making*, 1–13, [CrossRef](#).

- Budescu, D. V., H. H. Por, and S. B. Broomell, 2012: Effective communication of uncertainty in the IPCC reports. *Climatic Change*, **113**, 181–200 (2012). [CrossRef](#).
- _____, _____, _____, and M. Smithson, 2014: The interpretation of IPCC probabilistic statements around the world. *Nat. Clim. Change*, **4**, 508–512. [CrossRef](#).
- Cokely, E. T., A. Feltz, S. Ghazal, J. N. Allan, D. Petrova, and R. Garcia-Retamero, 2018: Skilled decision theory: from intelligence to numeracy and expertise. *The Cambridge Handbook of Expertise and Expert Performance*, Cambridge University Press, pp. 476–505. [CrossRef](#).
- Freelon, D. G., 2010: ReCal: intercoder reliability calculation as a web service. *Int. J. Internet Sci.*, **5** (1), 20–33. [Available online at www.ijis.net/ijis5_1/ijis5_1_freelon.pdf.]
- Gigerenzer, G., R. Hertwig, E. Van 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** (3), 623–629. [CrossRef](#).
- Grounds, M. A., and S. L. Joslyn, 2018: Communicating weather forecast uncertainty: do individual differences matter? *J. Exper. Psychol. Appl.*, **24** (1), 18–33. [CrossRef](#).
- Gurman, A. D., J. Baron, and K. Armstrong, 2004: The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication. *Med. Decis. Making*, **24** (3), 265–271. [CrossRef](#).
- Hallgren, K. A., 2012: Computing inter-rater reliability for observational data: An overview and tutorial. *The Quant. Methods for Psychol.*, **8**, 1, 23–34. [CrossRef](#).
- Harris, A. J. L., H. H. Por, and S. B. Broomell, 2017: Anchoring climate change communications. *Climatic Change*, **140**, 387–398. [CrossRef](#).
- Jarrell, J., and S. Brand, 1983: Tropical cyclone strike and wind probability applications. *Bull. Amer. Meteor. Soc.*, **64** (9), 1050–1056. [CrossRef](#).
- Jenkins, S. C., A. J. Harris, and R. M. Lark, 2019: When unlikely outcomes occur: the role of communication format in maintaining communicator credibility. *J. Risk Res.*, **22** (5), 537–554. [CrossRef](#).
- Joslyn, S., L. Nadav-Greenberg, L., and R. M. Nichols, 2009: Probability of precipitation: Assessment and Enhancement of end-user understanding. *Bull. Amer. Meteor. Soc.*, **90** (2), 185–194. [CrossRef](#).
- Joslyn, S. L., and J. E. LeClerc, 2012: Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *J. Exper. Psychol. Appl.*, **18** (1), 126–140. [CrossRef](#).
- Juanchich, M., and M. Sirota, 2016: How to improve people’s interpretation of probabilities of precipitation. *J. Risk Res.*, **19** (3), 388–404. [CrossRef](#).
- _____, and _____, 2018: Not as gloomy as we thought: reassessing how the public understands probability of precipitation forecasts. *J. Cog. Psychol.*, **31** (1), 116–129. [CrossRef](#).
- Kahneman, D., 2011. *Thinking, Fast and Slow*. Macmillan. 512 pp.
- Keller, C., M. Siegrist, and H. Gutscher, 2006: The role of the affect and availability heuristics in risk communication. *Risk Anal.*, **26** (3), 631–639. [CrossRef](#).
- Landis J. R. and G. G. Koch, 1977: The measurement of observer agreement for categorical data. *Biometrics*, **33** (1), 159–174. [CrossRef](#).
- 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** (5), 64–78. [CrossRef](#).
- Mandel, D. R., and D. Irwin, 2021: On measuring agreement with numerically bounded linguistic probability schemes: A re-analysis of data from Wintle, Fraser, Wills, Nicholson, and Fidler (2019). *PLoS One*, **16** (3), e0248424, 1–13. [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** (5), 974–991. [CrossRef](#).
- _____, C. L. Cuite, J. L. Demuth, W. K. Hallman, and R. L. Shwom, 2018: Is storm surge scary? The influence of hazard, impact, and fear-based messages and individual differences on responses to hurricane risks in the USA. *Int. J. Disaster Risk Reduct*, **30**, 44–58. [CrossRef](#).
- Murphy, A. H., S. Lichtenstein, B. Fischhoff, and R. L. Winkler, 1980: Misinterpretations of precipitation probability forecasts. *Bull. Amer. Meteor. Soc.*, **61** (7), 695–701. [CrossRef](#).
- Peters, E., D. Västfjäll, P. Slovic, C. K. Mertz, K. Mazzocco, and S. Dickert, 2006: Numeracy and decision making. *Psychol. Sci.*, **17** (5), 407–413. [CrossRef](#).
- _____, 2017: Overcoming innumeracy and the use of heuristics when communicating science. *The Oxford Handbook of the Science of Science Communication*, Oxford University Press, 389–398. [CrossRef](#).
- _____, 2020: *Innumeracy in the Wild: Misunderstanding and Misusing Numbers*. Oxford University Press. 314 pp.
- Reyna, V. F., W. L. Nelson, P. K. Han, and N. F. Dieckmann, 2009: How numeracy influences risk comprehension and medical decision making. *Psychol. Bull.*, **135** (6), 943–973. [CrossRef](#).
- Ripberger, J. T., H. C. Jenkins-Smith, C. L. Silva, D. E. Carlson, and M. Henderson, 2014: Social media and severe weather: do tweets provide a valid indicator of public attention to severe weather risk communication? *Wea., Climate Soc.*, **6** (4), 520–530. [CrossRef](#).

- _____, C. L. Silva, H. C. Jenkins-Smith, D. E. Carlson, M. James, and K. G. Herron, 2015: False alarms and missed events: the impact and origins of perceived inaccuracy in tornado warning systems. *Risk Anal.* **35** (1), 44–56. [CrossRef](#).
- Savelli, S. and S. Joslyn, 2013: The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied Cognitive Psychology*, **27** (4), 527–541. [CrossRef](#).
- Sheets, R. C., 1985: The National Weather Service Hurricane Probability Program. *Bull. Amer. Meteor. Soc.*, **66** (1), 4–13. [CrossRef](#).
- Slovic, P., 2010: *The Feeling of Risk: New Perspectives on Risk Perception*. Routledge. 456 pp.
- Trevena, L. J., A. Barratt, P. Butow, and P. Caldwell, 2006: A systematic review on communicating with patients about evidence. *Jour. Eval. Clinical Practice*, **12** (1), 13–23. [CrossRef](#).
- _____, and Coauthors, 2013: Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. *BMC Med. Infor. Decis. Making*, **13** (2), 1–15. [CrossRef](#).