ABSTRACT: Probabilistic forecast information is rapidly spreading in the weather enterprise. Many scientists agree that this is a positive development, but incorporating probability information into risk communication can be challenging because communicators have little guidance about the most effective way to present it. This project endeavors to create such guidance by initiating a “living systematic review” of research studies that empirically examine the impact of risk messages that use probability information on protective action decision-making, intentions, and behaviors. In this article, we explain how we began the review, map the current state of the literature, synthesize core findings, provide actionable recommendations to assist forecasters in risk communication, and introduce an online platform that scholars and forecasters can use to interact with the data from the review. We conclude with two key points from the review that necessitate emphasis: the research literature strongly suggests that 1) average people can make sense of and use probability information if consideration is given to information presentation and 2) assuming appropriate presentation, probability information generally improves decision quality.

SIGNIFICANCE STATEMENT: Probability information is increasingly common in weather forecasts, but forecasters have relatively little guidance on the most effective way to communicate this information to members of the public. This project synthesizes the research literature to provide actionable recommendations to assist forecasters who are working to include probability information in risk communication messages.

KEYWORDS: Probability forecasts/models/distribution; Communications/decision making; Decision making

1. Motivation

Driven by advances in computation, ensemble modeling, and data assimilation, probabilistic forecast information is rapidly spreading in the weather enterprise. Many scientists agree that this is a positive development but incorporating probability information into risk communication can be challenging, as probabilities are notoriously difficult to communicate effectively to lay audiences (e.g., National Research Council 2006; AMS 2008; National Research Council 2012). What does the research literature say about the best way to include probability information in risk communication? What is the evidence base for different communication practices? This project endeavors to address these questions by initiating a “living systematic review” of relevant research from past studies and new studies as they become available.

2. Methodology

A systematic review is a literature review that uses a transparent and repeatable methodology to identify relevant research from past studies, evaluate results from those studies, and synthesize findings. Historically, systematic reviews have been static; they synthesize the literature at a point in time and become out of date almost as soon as they are complete. To prevent this, living systematic reviews are beginning to replace static reviews. Living systematic reviews follow the same steps but are updated as new research becomes available. Most systematic reviews, living or static, include the following steps:

1) Define the study domain.
2) Search for and identify relevant studies.
3) Extract key topics, questions, methods, and findings from relevant studies.
4) Evaluate the quality of relevant studies.
5) Analyze and combine the studies to identify common topics, questions, methods, and findings. This review includes two additional steps:
6) Assess common findings to develop recommendations to assist forecasters in communicating uncertainty and probabilities.
7) Develop a living platform that incorporates new studies and relevant findings into the review as they become available.

We use these steps in the sections that follow to describe our living review of research literature on the use of probability information in risk communication.

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

DOI: 10.1175/WCAS-D-21-0034.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
a. Study domain

The review focuses on research studies that directly examine the impact of probability information on protective action decision-making, intentions, and behaviors. Most of the studies in the review focus on the “best” or most effective way to communicate probability information. They address questions like: are people more likely to take protective action when probability information is given verbally or numerically? We do not include studies that indirectly examine these relationships by way of implication or suggestion. For example, we do not include studies that explore the relationship between numeracy and risk perceptions, which may have important implications for how people use probability information when making decisions.

b. Identification of relevant studies

We use a combination of three methods to search for and identify relevant studies: 1) electronic search databases, 2) past literature reviews, and 3) citation chains. Between all three sources, we were able to identify 327 unique studies that were relevant to the study domain. This list will continue to grow as the review continues.

1) ELECTRONIC SEARCH DATABASES

In phase 1 of the search, we use ProQuest, Web of Science, and EBSCO Academic Search Elite to identify potentially relevant studies that focus on communicating probability information in the weather and climate domains.1 We restrict the domains at this point in the process to ensure that we are identifying the studies that are most relevant to the audience of this review (see Table 1 for a list of search terms by domain). We rely on the next two phases (past reviews and citation chains) to identify potentially relevant studies in adjacent domains (health, insurance, etc.). The first database searches were conducted on 29 July 2019; additional searches were undertaken on 1 September 2020 to update the list of potentially relevant studies. Through these searches, we were able to identify 1559 potentially relevant studies; 725 of the studies were unique across the three databases.

After identifying potentially relevant studies, two researchers independently screen the title and abstract of each study to ensure that three inclusion criteria are met: 1) the study fits within the study domain (see above); 2) the study reports on new findings from a new research project (e.g., it is not a literature review, essay, or workshop report); and 3) the study uses a generalizable, transparent, and repeatable methodology to conduct the research. In many (but not all) cases, this leads to the exclusion of qualitative studies because most of them do not meet the generalizability criterion. If reviewers do not agree during the screening phase, they review the entire contents of the study to see if it meets the criteria above. If they disagree at this phase, they discuss and come to an agreement. Of the 725 unique studies that we were able to identify at the start of the review, 93 met the inclusion criteria in the first screening. Of these, 29 met the criteria following in-depth review of the study.

2) PAST LITERATURE REVIEWS

In phase 2, we use past literature reviews to identify potentially relevant studies. To begin this review, we were able to identify 12 past literature reviews that provide valuable information about potentially relevant studies (see Table 2 for a list of these past reviews). Using the two-stage screening methodology we describe above, we identified 37 new studies that met the inclusion criteria above, bringing the set of relevant studies to 66.

---

1 Although it is relatively comprehensive, we do not use Google Scholar because the search results it provides are not repeatable.
In phase 3, we use citation chains to identify potentially relevant studies. First, we use “backwards” citation chains that include all references IN the relevant studies that we identify in phases 1 and 2. Next, we use “forwards” citation chains that include all references TO the relevant studies that we identify in the first two phases. In all, there were 1523 unique references IN and 2279 unique references TO the 66 relevant studies that we were able to identify in the first two phases. Using the two-stage methodology we describe above, we came up with 255 new studies that met the inclusion criteria, bringing the cumulative set of relevant studies to 327.

Because this is a living systematic review, it is important to reiterate that this is only the beginning of the review. We plan to repeat these phases every few months to make sure that we are including the most up-to-date research.

c. Extracting key information

Following identification, we review all relevant studies, extract key information, and store it in a spreadsheet. In addition to basic bibliographic information, we note relevant research questions, variables, research methodologies, information about research subjects, domains of study, and primary findings.

d. Evaluating study quality

While extracting key information about each study, we assess the quality of each study. There are many ways to assess quality; we use three indicators of validity: 1) external validity (are the results generalizable to the population of interest?); 2) internal validity (are we sure that variation in x causes variation in y?); and 3) domain validity (how relevant is the study domain to weather hazards and forecasting?). Each dimension is independently given a score by two researchers on a 3-point scale (1 = low; 2 = medium; 3 = high). We use the mean value of these scores to measure validity along each dimension and the overall validity of each study.

3. Results

a. Common topics: Quantity of evidence

We were able to identify 5 primary research topics and 13 secondary topics in the 327 studies we are using to begin this review. While this list of topics may evolve as the review continues, an examination of the relative frequency of each topic provides a snapshot of the literature at this point in time. It provides valuable information about the quantity of evidence we were able to identify in each topic area. Figure 1 provides this information by showing the percentage of studies that address the primary topics. Note that the topics are not mutually exclusive; many studies address more than one topic. As Fig. 1 indicates, the most common topics in the current set of relevant studies are public understanding and use of probability information in decision-making (label a), communicating probability information using words and phrases (label b), and communicating probability information using visualizations (label c). The least common topic is communicating probability information to a heterogeneous population (label e). The discrepancy between these topics suggests that we know a lot about how people use probability information to make decisions and the extent to which words and phrases facilitate this process, whereas we know relatively little about how different types of people use and interpret different types of probability information.

Figure 2 provides more information by showing the percentage of studies within each primary topic that address

![Figure 1](image-url)
secondary topics of relevance. For example, the figure shows that studies on the communication of probability information using words and phrases (Fig. 2b) frequently address 1) numeric translations of words and phrases, but rarely address 3) severity and probability conflation. Likewise, studies on communication of probability information using numbers (Fig. 2c) are more likely to focus on 1) probabilities as percentages than 2) probabilities as frequencies. Again, these discrepancies provide valuable data on the amount of evidence we have on each of the secondary topics.

b. Common topics: Quality of evidence

Information about quantity provides one metric for assessing the strength of evidence in each of the topic areas. Information about quality of evidence provides a second metric. We assess the quality of evidence from each study in the review by evaluating the external, internal, and domain validity of the study. When we average these across topics, we can discriminate between topics with high-quality evidence and topics with low-quality evidence. These averages will change as the review continues, but they provide valuable information about the current state of research.

Figure 3 provides this information by showing the mean validity of studies that address the primary topics. Note that the sizes of the points in the plot reflect mean domain validity; large points indicate high domain validity, small points indicate low domain validity. The lines with dashes indicate overall mean validity scores; on average, studies on topics to the right of the line have more internal validity than studies to the left of the line and studies above the line have more external validity than studies on topics below the line. Given this orientation, the relatively large point in the top-right quadrant indicates that, on average, studies on public understanding and use of probability information in decision-making (label a) have high validity. By comparison, the relatively small point in the bottom-left quadrant indicates that studies on communicating probability information using words and phrases (label b) generally have low validity. Interestingly, studies on communicating...
probability information using numbers (label c) typically have more internal but less external validity than studies on the use of words in phrases; this is because these studies often use experiments to identify causality, but the subjects of the experiments are rarely representative of the U.S. population.

Figure 4 provides more information by plotting the average validity of studies by secondary topic within each of the primary topics. The plots illustrate, for example, that studies on the value of probability information (Fig. 4a, label 1) and understanding probability information (Fig. 4a, label 2) have high levels of validity whereas studies on verbal directionality (Fig. 4b, label 2) and reference class ambiguity (Fig. 4c, label 3) typically have low levels of validity. If quality is the metric, these findings suggest that we know a lot about the first two areas of study, whereas more research is necessary on the latter two.

c. Common findings

We were able to identify more than 100 unique findings in the 327 studies we are using to begin this review. It is not possible to describe all of these findings in a single article. We therefore attempt to synthesize and summarize as many findings as possible in the sections below.

1) PUBLIC UNDERSTANDING AND USE OF PROBABILITY INFORMATION IN DECISION-MAKING

(ii) Value of probability information

Do members of the public benefit from probability information, or are they better off with deterministic statements? Some forecasters express a desire to “boil down” complex probability information to a deterministic point forecast for fear of confusing members of the public (Pappenberger et al. 2013). Strong evidence in the research literature indicates that these fears are unfounded. Nearly all of the studies we review indicate that people make better decisions, have more trust in information, and/or display more understanding of forecast information when forecasters use probability information in place of deterministic statements (Ash et al. 2014; Bolton and Katok 2018; Grounds and Joslyn 2018; Grounds et al. 2017; Joslyn and LeClerc 2012, 2016; Joslyn and Demnitz 2019; Joslyn et al. 2007; LeClerc and Joslyn 2012; Marimo et al. 2015; Miran et al. 2019; Nadav-Greenberg and Joslyn 2009; Roulston and Kaplan 2009; Roulston et al. 2006; Joslyn and Grounds 2015). However, it is important to note that both experts and the public sometimes have difficulty interpreting probability information, and different communication formats can affect understanding (Bramwell et al. 2006). Along these lines, many studies emphasize the importance of making probabilistic forecasts as straightforward and easy to understand as possible in order to avoid “information overload” (Durbach and Stewart 2011). It is also important to note that these findings refer to probability information beyond a general acknowledgment of epistemological uncertainty (e.g., “this forecast is based on estimates; it is impossible to ever know for sure what will happen”), as these types of overly broad statements can undermine trust in forecasts (Howe et al. 2019).

(ii) Understanding probability information

Many studies also examine how people understand and interpret probability information. Often, these studies simply
give subjects a probability of precipitation forecast (PoP) and ask them to interpret it in order to assess their level of understanding. Gigerenzer et al. (2005), Morss et al. (2008), Sink (1995), Zabini et al. (2015), and Abraham et al. (2015) all show, to various degrees, that a majority or a substantial proportion of the public is unable to give the correct interpretation of a PoP forecast, generally considered to be something like “it will rain somewhere in the forecast area on X% of days like today” (Gigerenzer et al. 2005; Morss et al. 2008; Sink 1995; Zabini et al. 2015; Abraham et al. 2015). However, Juanchich and Sirota (2019) argue that previous studies use a cumbersome “correct” answer, and that “X% of simulations predict rain in the forecast area” is a more “fluid” and more easily understood response category; when using this “correct” answer, the vast majority are able to give the correct PoP interpretation (Juanchich and Sirota 2019). Another prior study, Murphy et al. (1980), argues something similar—that “people do not have trouble understanding what ‘30% chance’ means, but . . . they do have trouble understanding exactly what the probability refers to in this kind of forecast” (Murphy et al. 1980). These findings should reassure forecasters that the public can correctly understand probabilities but should underscore the need to explain the events the forecast refers to in an intuitive and clear way.

Multiple studies also offer findings about the process by which members of the public think through probabilistic risk information. For instance, a group of studies indicate that most people intuitively infer uncertainty even when given a deterministic forecast (Savelli and Joslyn 2012; Joslyn and Savelli 2010; Morss et al. 2008, 2010). Moreover, inclusion of uncertainty can lead to increased worry but can also be mitigated through the use of textual and visual formats (Han et al. 2011). These findings suggest that people think about forecast events in probabilistic terms even when not explicitly told to do so. Several more studies find that higher probabilities (regardless of context or direction) may lead people to view a forecast as more accurate; for instance, the same forecast would likely be taken to be more accurate if it reported a 70% chance of sun rather than a 30% chance of rain (Bagchi and Ince 2016; Løhre et al. 2019; Juanchich and Sirota 2017). In a similar vein, some studies suggest that people consistently misinterpret confidence intervals and forecast periods. For instance, two studies find that both experts and nonexperts implicitly interpret forecast events as being more likely toward the end of the forecast period (e.g., if there were an X% chance that a given event would occur sometime in a given week, people will, on average, perceive that the event is more likely to happen on Friday than on Monday) (Doyle et al. 2014; McClure et al. 2015), and one study finds that a significant proportion of the public is unsure of how to understand the distribution of possible outcomes denoted by confidence intervals (Dieckmann et al. 2015). These findings underscore the need to clarify what forecast periods and confidence intervals mean in the context of a given forecast rather than assuming that will be clear to the audience.

A few studies also identify some common ways that people misinterpret or dismiss probability information. Perhaps the most significant is motivated reasoning, the tendency to interpret new information in a way that supports preexisting beliefs. Four studies in the review explicitly focus on the effect of motivated reasoning on how people interpret probability information; all indicate that probability information, especially information about politically sensitive topics like climate change, is susceptible to misinterpretation when it contradicts preexisting beliefs (Budescu et al. 2012; Dieckmann et al. 2017; Piercey 2009; Nurse and Grant 2020). Interestingly, more numerate people seem to be more susceptible to these effects (Nurse and Grant 2020) and using verbal probability expressions seems to encourage more motivated reasoning than using visual or numeric expressions (Budescu et al. 2012; Dieckmann et al. 2017; Piercey 2009). Another common way that probability information can become “distorted” is what Hohle and Teigen (2015) call the “trend effect.” In short, people often interpret recent forecasts in light of past forecasts. A “moderate” risk, for instance, will cause more worry if it has been upgraded from a “low” risk than if it has been downgraded from a “high” risk (Hohle and Teigen 2015; Løhre 2018).

2) Communicating probability information using words and phrases

(i) Numeric translations of words and phrases

Experts and nonexperts routinely use verbal probability expressions like “unlikely” or “a good chance” to indicate uncertainty; this practice is particularly common in the weather domain. The first core finding in this area of the review is very simple: there is strong evidence that risk communicators should always include a numeric “translation” for any verbal probability expressions used, and that translation should appear directly in or next to the verbal expression itself (Carey et al. 2018; Connelly and Knuth 1998; Dorval et al. 2013; Fortin et al. 2001; Hill et al. 2010; Zabini et al. 2015; Wintle et al. 2019). For example, a verbal expression like “severe thunderstorms are possible this evening” would be more effectively rephrased as “severe thunderstorms are possible (20% chance) this evening” (Budescu et al. 2014). Explicit statements of the upper and lower bound (e.g., 0%-33%) implied by an expression (e.g., “likely” or “unlikely”) improve accuracy of interpretation versus a statement alone (Harris et al. 2017). This is important not only because it helps people to correctly interpret the meaning of a forecast, but also because people generally prefer mixed formats (e.g., a numeric probability and a verbal probability expression together, or a number and a visualization) to singular ones (Carey et al. 2018; Connelly and Knuth 1998; Dorval et al. 2013; Fortin et al. 2001; Hill et al. 2010; Sink 1995; Zabini et al. 2015). Members of the public demonstrate a basic understanding of probabilistic forecasts; however, uncertainty is best communicated through combined use of numeric and verbal expressions to meet the needs of heterogeneous audiences (Kox et al. 2015). Translations are important because less numerate people tend to focus on narrative evidence when evaluating risk communications (the context, their perceptions about the likelihood of comparable events, etc.), while more numerate people tend to focus on the

(ii) Verbal directionality

The next core finding in this area addresses the importance of “directionality” in verbal probability statements. “Directionality” can be positive or negative (Teigen and Brun 1995). Positive statements focus the probability that an event will happen (e.g., “it is possible that the hurricane will affect town x”) and negative statements focus on the probability that it will not happen (e.g., “it is likely that the hurricane will miss town x”). Generally, research in this area suggests that positive statements can cause people to overestimate the probability of an event and, consequently, engage in behaviors that are in line with the target outcome even if that outcome is very unlikely (e.g., take protective action even if there is a very small chance that a hurricane will affect town x). Negative statements can have the opposite effect; they can cause people to underestimate the probability of an event and decide not to engage in protective actions (Honda and Yamagishi 2006, 2009, 2017; Teigen and Brun 1995, 1999, 2000, 2003; Budescu et al. 2003). Researchers are still exploring the communicative function of these statements, but some evidence suggests that the direction of a statement conveys implicit information about a speaker’s reference point (McKenzie and Nelson 2003; Honda and Yamagishi 2017). Positive statements may indicate that a probability is increasing or higher than a speaker’s reference point. Negative statements indicate the opposite; that it is decreasing or low in comparison with a speaker’s point of reference.

(iii) Severity and probability conflation

Another core finding on verbal probability expressions pertains to the “severity effect,” which is the tendency of people to implicitly interpret verbal probability expressions as more likely when they describe more severe or undesirable outcomes (Bonnefon and Villejoubert 2006; Fischer and Jungerman 1996; Harris and Corner 2011; Weber and Hilton 1990). For example, someone who interprets a “slight chance” of rain showers to mean a 1%–5% chance will likely interpret a “slight chance” of a hurricane to mean something closer to a 10%–15% chance. This is important for forecasters to consider when using verbal probability statements, as it may suggest different interpretations of the same words and phrases, depending on the situation.

(iv) Choosing words and phrases

The studies in this area of the review provide a few core findings on word choice. For instance, when deciding whether to use a word like “can” or “will,” be aware of the “extremity effect”: when shown a probability and asked what “can” happen, people tend to focus on the most extreme possible values, and when asked what “will” happen, they tend to focus on the more likely scenarios (Teigen and Filipuková 2013; Teigen et al. 2018, 2014). In a similar vein, Teigen et al. (2013) find that people often use and interpret words like “improbable” to refer to events that are not just unlikely (something like a 10%–20% chance, for instance), but nearly impossible (closer to a ~1% chance, for instance). Often, “improbable” is implicitly understood to refer to events that have not happened yet but have a small chance of happening in the future, even when experts have another definition in mind. Clarifying such terms and providing explicit numeric “translations” helps to reduce these misunderstandings (Teigen et al. 2013). Last, research strongly indicates that forecasters should avoid vague verbal probability terms (such as “it is possible” or “there is a chance”), as they can be particularly problematic in communication due to variable interpretation (Fillenbaum et al. 1991; Reyna 1981; Lenhardt et al. 2020). In summary, words and phrases play an important role in the communication of probability information. As a result, forecasters and audiences would benefit from careful consideration of translations, directionality, severity, and word choice to ensure clear communication.

3) Communicating Probability Information using Numbers

(i) Probabilities as percentages

Experts and nonexperts also use a variety of numeric formats (e.g., percentages, frequencies, odds) to communicate probability information. Numeric probabilities are most commonly expressed as a percentage (e.g., “a 30% chance of rain”) or as a frequency (e.g., “a 3 in 10 chance of rain”). Some evidence suggests that percentage formats promote higher levels of risk comprehension than frequency or fraction formats (Fuller et al. 2002; Cuite et al. 2008; Joslyn and Nichols 2009), but there are also a few studies that suggests the opposite, that frequency formats promote more accurate judgements than percentage formats (Knapp et al. 2013). Along these lines, multiple studies conclude that there are no consistent benefits to percentage or frequency formats (Hendrickx et al. 1989; Joslyn et al. 2009; Neace et al. 2008; Knapp et al. 2016; Evans et al. 2000; Ruiz et al. 2013; Strathie et al. 2017). In place of a consistent benefit, the literature suggests that different circumstances may call for different formats, depending on the task at hand and the context of the risk (Cuite et al. 2008; Knapp et al. 2009; Sinayev et al. 2015; Wallsten et al. 1986).

(ii) Probabilities as frequencies

Variation in frequency formats, such as the use of “1-in-x” ratios (e.g., “1 in 33 people”) versus “N-in-X × N” ratios (e.g., “3 in 1000 people”) can also influence risk judgements and decisions (Grimes and Snively 1999; Pighin et al. 2015; Oudhoff and Timmermans 2015; Denes-Raj et al. 1995; Bell and Tobin 2007; Carey et al. 2018; Pighin et al. 2011; Sirota et al. 2018). For example, Pighin et al. (2015) suggests that 1-in-x formats can inflate probability and risk perceptions in comparison with numerically equivalent N-in-X×N formats. A study by Oudhoff and Timmermans (2015) finds the same thing but adds that 1-in-x formats may be easier for people to interpret. Because of these findings, researchers often urge caution in the use of 1-in-x formats; while they may be more
intuitive for some segments of the population, they can elevate probability perceptions and bias risk perceptions.

(iii) Reference class ambiguity

Communication of single-event probabilities can be challenging because common probability messages like “there is a 30% chance of rain” do not specify the reference class, which can cause confusion and misinterpretation (Gigerenzer et al. 2005; Joslyn et al. 2009; Neace et al. 2008; Knapp et al. 2016; Ruiz et al. 2013; Strathie et al. 2017). Gigerenzer et al. (2005) argues that one way to improve this is to explicitly include a reference class when providing single-event probabilities. Rather than simply stating, “There is a 30% chance of rain tomorrow,” a message that includes the reference class might say something like: “There is a 30% chance of rain on days like tomorrow. This means that on 3 of 10 days like tomorrow, it will rain” (Gigerenzer et al. 2005). Putting this argument to the test, studies by Neace et al. (2008) and Juanchich and Sirota (2019) show that specifying a reference class generally improves interpretation of single-event probability information.

(iv) Gain and loss frames

Somewhat analogous to the positive versus negative frames we describe above, communicators can use probability information to convey equivalent information about the outcome of an event or action. “Gain” frames emphasize positive results and the benefits of protective action. “Loss” frames emphasize negative results and the risks of inaction. Most studies agree that gain/loss frames affect perceptions and actions, but there is little consensus about the direction of these effects. Some studies suggest that gain frames are more effective at promoting preventive behaviors than loss frames (Garcia-Retamero and Cokely 2011). Others show that messages with negative or loss frames can increase risk perceptions and encourage precautionary actions (Banks et al. 1995; Chua et al. 2006; Garcia-Retamero and Cokely 2011; Teigen and Brun 2003). More in-depth work on the subject indicates that the impact of loss or gain frames can interact with forecast uncertainty; when losses loom large and forecast uncertainty is high, individuals can engage in a type of fatalism or “wishful thinking” that decreases protective action intentions in response to psychologically distant hazards like climate change (Morton et al. 2011). Along these lines, a study by Kuhn (1997) suggests that negative frames encourage people to gravitate toward vague or uncertain forecasts even if more precise information is available. Given these findings, many researchers agree that using negative frames to prompt action can have deleterious effects on risk comprehension and judgement, even if they can increase risk perceptions (Smithson et al. 2012; Armstrong et al. 2002).

4) COMMUNICATING PROBABILITY INFORMATION USING VISUALIZATIONS

(i) Visualizations in non-weather domains

Visualizations often provide an effective way to communicate probability information to people who have a difficult time with words and numbers (Johnson and Slovic 1995, 1998; Ulph et al. 2009; Okan et al. 2015). Some probability visualizations are common in the weather domain, but many are not; icon arrays (also known as pictographs), risk ladders, and survival curves, for example, are common in medicine and epidemiology, but rare in weather risk communication. Icon arrays, graphics that use icons in grids to represent “at risk” populations (usually in proportion to the whole population), are among the most common visualizations in modern risk communication practice and scholarship. Many studies show that they can increase risk comprehension and avoidance actions (Ancker et al. 2011; Dowen et al. 2017; Tubau et al. 2019; Galesic et al. 2009; Garcia-Retamero et al. 2010; Garcia-Retamero and Galesic 2009; Keller and Siegrist 2009; Schirillo and Stone 2005; Garcia-Retamero and Cokely 2014; Garcia-Retamero and Galesic 2010; Leonhardt and Robin Keller 2018; Stone et al. 1997; Taylor et al. 2018; Witteman et al. 2014; Zikmund-Fisher et al. 2014, 2011). While there are many mechanisms that relate icon arrays to risk comprehension and action, many scholars theorize that they are effective because they communicate risk information in ways that show exact percentages while simultaneously conveying “gist” impressions (Hawley et al. 2008; Hess et al. 2011). These gist impressions can be especially important because they help people with relatively low numeracy evaluate risks and make informed decisions (Garcia-Retamero et al. 2010; Garcia-Retamero and Galesic 2009; Garcia-Retamero and Cokely 2014; Garcia-Retamero and Galesic 2010). While there may be other ways to communicate risk information, say, with bar plots, most studies agree that icon arrays provide more transparent representations of risk that generally promote higher comprehension (Correll and Gleicher 2014; Dieckmann et al. 2015; Schapira et al. 2006; Newman and Scholl 2012).

Risk ladders, which typically compare a given risk (e.g., dying from a lightning strike) with others (e.g., dying from a heart attack), are also relatively common. They can help to increase comprehension by relating an abstract or new risk to one that people might be more familiar with, but some studies suggest that they are not especially effective among people with low levels of numeracy (e.g., Keller et al. 2009). Survival curves are similar; they are very common in medicine and epidemiology, but many people, especially people with low levels of numeracy misinterpret them if they are not given adequate instruction (Armstrong et al. 2001; Mazur and Hickam 1990).

Although the studies in this area focus on visualizations that are rarely used in weather domains, many of them provide broad advice about how to increase the effectiveness of visualizations more broadly. For example, nearly all of them emphasize the important point that there is not a single “best way” to visually communicate risk; different situations and audiences require different visualization choices (Barnes et al. 2016; Bisantz et al. 2005; Etnel et al. 2020; Garcia-Retamero and Dhami 2013; Kreye et al. 2012; Lorenz et al. 2015; Sanyal et al. 2009). Because of this, it is important that communicators pretest visualizations with relevant audiences before disseminating them en masse. Additionally, studies in this area frequently stress the importance of explanatory labels and numbers in graphics to provide adequate context and
information to people with low numeracy and little experience interpreting graphic information (Okan et al. 2015).

(ii) Visualizations in weather domains

There are many noteworthy studies on the use of probability information in weather visualizations; however, most of them focus on specific products across disparate domains that are difficult to compare. For example, several studies focus on ways to improve hurricane track visualizations, whereas others focus on ways to improve severe thunderstorm and tornado warnings. Given the different nature of these risks and the information that these visualizations attempt to convey, we urge caution in generalizing from one study to the next. Nevertheless, there are a few common findings in this area of the literature that warrant note. First and foremost, the studies generally agree that including probability information in forecast graphics improves risk comprehension and increases protective action intentions in high-risk areas (Nadav-Greenberg et al. 2008; Joslyn et al. 2013; Ash et al. 2014; Miran et al. 2017, 2019; Jon et al. 2018; Klockow-McClain et al. 2020).

There is less agreement on the most effective way to include probability information in graphics. When it comes to basic representations of probability, some studies focus on how to plot uncertainty, using relatively common plots in statistics that highlight most likely outcomes (boxplots, line plots, density plots, etc.) or ensemble/simulation representations that highlight alternative possibilities. These studies generally agree that ensemble/simulation representations promote risk comprehension and awareness of unlikely (but possible) outcomes, but they may distract some people from scenarios that forecasters believe are most likely (Padilla et al. 2017; Toet et al. 2019).

A separate line of research focuses on the use of colors when conveying probability information. Some studies on this topic indicate that colors have relatively little impact on risk perceptions, comprehension, and avoidance behaviors (Ash et al. 2014; Miran et al. 2017, 2019). Others suggest that colors may provide important information, especially at low ends of a probability scale that people may otherwise ignore (Klockow-McClain et al. 2020). While this literature provides limited guidance on the exact colors to use, a few studies suggest that polychromatic schemes may be preferable to monochromatic schemes (Sherman-Morris et al. 2015; Miran et al. 2017, 2019) and warm colors may indicate more risk than cool colors (Klockow-McClain et al. 2020).

Last, and perhaps most important, scholars who focus on weather visualizations join with scholars who focus on non-weather visualizations in emphasizing the importance of including text explanations and descriptive labels in probability graphics (Boone et al. 2018) and the mindfulness of context and communication goals when developing visualizations. Visualization choices, like forecasts, often come down to tradeoffs and visualizations are likely to be more effective if communicators act with intention and purpose when making choices about how to visually convey probability information (Ash et al. 2014).

5) Communicating probability information to a heterogeneous population

Many of the studies in the review include findings that, while important, are overly specific or difficult to generalize and use for forecasters. The consensus of these findings, though, is clear: efforts to communicate probability information must consider the heterogeneous nature of audiences and the different contexts and biases they entail. For instance, when it comes to both members of the public and experts, individuals sometimes have difficulty accurately interpreting forecasts that include probability information due to a number of different factors, including context (Kim et al. 2014; Morss et al. 2010; Hohle and Teigen 2015; Lohre 2018; Windschittl and Weber 1999), motivated reasoning (Dieckmann et al. 2017), and numeracy (Kong et al. 1986; Bramwell et al. 2006; Harris et al. 2013; Rinne and Mazzocco 2013; Juanchich and Sirota 2016). Forecasters play a critical role in setting the context for communicating probability information through their use of language (Connelly and Knuth 1998; Franic and Pathak 2000; Kox et al. 2015; Pappenberger et al. 2013). The way in which forecasters frame messages can influence how audiences interpret risks (Harris et al. 2009; Keller et al. 2006; Wilson et al. 2019). For example, there is a real risk of inducing audiences to overestimate risk when describing small or long-term risk using a “once in X years” expression (Grounds et al. 2018; Bell and Tobin 2007), time uncertain (e.g., “X will happen within the next 10–30 years”) (Ballard and Lewandowsky 2015), and longer time frames (Keller et al. 2006; Doyle et al. 2014; McClure et al. 2015; Morss et al. 2016). Audiences are also prone to motivated reasoning when presented with probability information and uncertainty (Savelli and Joslyn 2012; Joslyn and Savelli 2010; Morss et al. 2008, 2010; Dieckmann et al. 2009; Rabinovich and Morton 2012).

Although uncertainty often elicits more risk-averse behavior (Ramos et al. 2013), individuals have a tendency to engage in additivity neglect (e.g., responses totaling more than 100%) (Riege and Teigen 2013), anchoring (Losee et al. 2017), and favoring higher absolute numbers (Denes-Raj and Epstein 1994; Maglio and Polman 2016). Last, less numerate people tend to focus on narrative evidence when evaluating risk communications (the context, their perceptions about the likelihood of comparable events, etc.), while people with higher levels of numeracy generally focus on the stated probability of the risk with greater accuracy (Dieckmann et al. 2009; Gardner et al. 2011).

More than anything, these findings indicate that forecasters cannot overlook the context of a specific forecast (e.g., is the threat common in an area) and differences across populations (e.g., biases, experiences, and numeracy) when developing strategies to communicate probability information.

4. Recommendations for forecasters

We believe that these recommendations are tangible, but there are many policies and practices within the NWS that may make them difficult to implement. For example, decades of training and policy dictate that forecasters communicate
the probability of precipitation using specific expressions of uncertainty like “slight chance,” “chance,” and “likely” that relate to underlying probabilities. In many cases, forecasters use these expressions alone, which violates the recommendation that communicators provide numeric translations when discussing the probability of an event. This is somewhat easy issue to address because there are specific guidelines that relate the words to numbers. A “slight chance” of rain means that there is a 15%–24% (20%) chance of rain. When possible, we encourage forecasters to include both forms of information in the messages they provide. While this practice may address some issues, translations alone may not be enough to overcome the confusion that especially vague words like “chance” generate. Such words clearly violate the recommendation that communicators use rank adjectives like low, medium, or high chance to indicate the magnitude of a probability. This is a more difficult issue to address because these terms are relatively uncommon in the NWS, and forecasters have no guidance about how to consistently differentiate between the terms. New guidance and training may be necessary if we expect NWS forecasters to formally adopt this recommendation.

In addition to policies and practices, communication technology may limit the immediate practicality of some recommendations in the NWS. Some communication systems and products preprogram and limit the words and phrases that forecasters can use to indicate the probability of an event. Others limit the length of messages or encourage brevity (e.g., bullet points) in the information that forecasters provide. These limits may complicate a forecaster’s ability to provide important information about how to interpret the probability of an event, such as the reference class. They also make it difficult for forecasters to provide probability information in different formats to different audiences. Accordingly, we recognize that new technology may be necessary to assist NWS forecasters in following these recommendations.

5. Interactive platform

As we noted above, it is not possible to list and meaningfully summarize all of the findings from this review within the confines of a single report. Moreover, this is a living systematic review that will grow and improve as scholars continue to study the inclusion of probability information in risk communication messages. Because of this, we believe that the results of this review are best shown in an interactive format on a platform that we update as new research becomes available. The Probability Communication (ProbCom) platform serves this purpose. A snapshot of the interactive platform is shown in Fig. 5, and it is accessible online (https://crcm.shinyapps.io/probcom).

6. Conclusions and next steps

Many forecasters are struggling to figure out the best or most effective way to include probability information in risk communication. The living systematic review we introduce in this article is meant to assist in this struggle by documenting the evidence base for different communication practices. The review is ongoing, and it is difficult to meaningfully synthesize everything we are learning, but a few overarching points...
<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Example</th>
<th>Exemplary studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use probability information in place of deterministic statements in forecasts</td>
<td>Replace “these storms will cause heavy downpours and flooding” with “there is an extremely high (90%) chance that these storms will cause heavy downpours and flooding”</td>
<td>Joslyn and LeClerc (2012); Grounds and Joslyn (2018)</td>
</tr>
<tr>
<td>Use numeric point estimates if they are available and appropriate; if they are not, use numeric probability ranges or predictive intervals to emphasize uncertainty</td>
<td>Replace “the forecast is rapidly evolving, but we may see more than 10 inches of snow in the metro area tomorrow morning” with (point estimate) “the forecast is rapidly evolving, but there is a 15% chance that we will see more than 10 inches of snow in the metro area tomorrow morning” or (predictive interval) “the forecast is rapidly evolving, but there is a 50% chance that we will see 6 to 14 inches of snow in the metro area tomorrow morning”</td>
<td>Grounds et al. (2017); Løhre et al. (2019)</td>
</tr>
<tr>
<td>Include numeric translations next to words/phrases that indicate probability information</td>
<td>Replace “thunderstorms are possible this evening” with “thunderstorms are possible (30% chance) this evening”</td>
<td>Budescu et al. (2014)</td>
</tr>
<tr>
<td>If comprehension of probability information is especially important, use numeric probabilities alone or first (before words/phrases)</td>
<td>Replace “thunderstorms are possible (30% chance) this evening” with “There is a 30% chance of thunderstorms this evening”</td>
<td>Jenkins et al. (2019)</td>
</tr>
<tr>
<td>When using words and phrases to communicate probability information, include rank adjectives (like low, medium, and high) to indicate the magnitude of the probability; this is especially important if numeric translations are not available</td>
<td>Replace “there is a chance of snow and ice this morning along I-75” with “there is a low/medium/high chance of snow and ice this morning along I-75”</td>
<td>Lenhardt et al. (2020)</td>
</tr>
<tr>
<td>Use probability (percentage) formats when possible; frequency (fraction) formats can be effective, but they can also generate confusion</td>
<td>Replace: “there is a 1 in 10 chance that this storm will produce a tornado” with “there is a 10% chance that this storm will produce a tornado”</td>
<td>Cuite et al. (2008); Joslyn and Nichols (2009)</td>
</tr>
<tr>
<td>Include information about the reference class when using probability information</td>
<td>Replace “there is a low chance of tornadoes in the Oklahoma City metro area tomorrow afternoon and evening” with “there is a low (2%) chance of tornadoes in the Oklahoma City metro area tomorrow afternoon and evening: on 1 in 50 days like today, there will be a tornado within 25 miles of your location”</td>
<td>Gigerenzer et al. (2005); Juanchich and Sirota (2019)</td>
</tr>
<tr>
<td>Be aware of directionality; positive statements, which focus on the probability that an event will happen, can promote overestimation and unnecessary protective actions; negative statements, which focus on the probability that an event will not happen can promote underestimation and underreaction</td>
<td>Positive statement: “there is a small chance of drizzle this morning” vs negative statement: “drizzle is unlikely this morning”</td>
<td>Honda and Yamagishi (2017); Teigen and Brun (2003)</td>
</tr>
<tr>
<td>When possible, include probability information in forecast visualizations</td>
<td>Supplement maps showing deterministic warning boxes/polygons with maps showing probability information; e.g., probability grids [“Forecasting a Continuum of Environmental Threats” (FACETs)] and Potential Storm Surge Flooding map (NHC)</td>
<td>Gerst et al. (2020)</td>
</tr>
<tr>
<td>Use visualizations to increase comprehension of probability information; icon arrays and ensemble plots can be especially effective when teaching people how to interpret probability information; note that this might not be appropriate immediately before or during high-impact events</td>
<td>Supplement “models indicate that there is an 80%–90% chance that hurricane force winds will affect Miami, Florida, in the next 5 days” with a probabilistic wind speed graphic with an icon array that explains the probability information in the product</td>
<td>Toet et al. (2019); Garcia-Retamero and Cokely (2014)</td>
</tr>
<tr>
<td>Pay attention to the audience when using probability information; most people can grasp basic probability information, but depth of comprehension depends on numeracy and experience</td>
<td></td>
<td>Dieckmann et al. (2009)</td>
</tr>
</tbody>
</table>
warrant emphasis. First and foremost, this is a considerable body of research that addresses many of the questions that forecasters are grappling with. Next, in nearly all domains, the research strongly suggests that 1) average people can make sense of and use probability information if consideration is given to information presentation and 2) assuming appropriate presentation, probability information generally improves decision quality. These findings obviously beg questions about how to present probability information. We hope the recommendations and examples we provide will assist in answering some of these questions but recognize that we are not able to speak to every question forecasters might have. In some cases, this is because we cannot summarize all the literature in a single article. In these cases, we hope that forecasters and others will use ProbCom to find potentially relevant answers to specific questions. In other cases, we cannot speak to questions because there are relatively few (if any) quality studies that are applicable to the weather domain. These areas require future research and while there are many different ways to prioritize research needs, the findings in this review suggest that top priorities ought to include 1) more research on the use of specific words and phrases (perhaps in combination with numbers and ranges) in probability communication and 2) more research on the use of basic visualizations such as icon arrays to assist in communicating probability information to portions of the population that struggle with words and numbers; and 3) more research on how members of the public integrate (and possibly conflate) probability and severity information when judging and making decisions about low probability, high-impact events.

Before we close, it is important to mention a few limitations of this review and notes of caution. The first and most obvious limitation is that we necessarily simplify complex arguments and findings when aggregating and summarizing results across the studies. While necessary for synthesis, this process results in an important loss of nuance that may have implications for the recommendations we derive. We therefore strongly encourage forecasters who wish to implement the recommendations we provide to carefully review relevant studies and take note of nuance when developing implementation plans. This is one reason we identify the exemplary studies in Table 3. We believe that these studies warrant especially close attention. Next, we caution that research on probability communication using visualizations and probability communication to heterogeneous populations is particularly difficult to synthesize because most studies in these areas test specific communication strategies for specific risks, purposes, and audiences. It is therefore extremely difficult to identify systematic similarities and abstract common findings from the studies. We do the best we can, but we recognize that the common (core) findings and recommendations we make from these two areas of the reviewed literature are neither as clear nor prescriptive as are the others. With time and feedback, we hope to clarify some of these vagaries. In the meantime, we suggest that forecasters use the ProbCom site to review the studies that look most interesting and relevant to the problems they are trying to solve—there are many high-quality studies on these topics, and we are confident that some of them will help forecasters address some of the communication challenges they are facing.

Acknowledgments. This research was supported by the National Oceanic and Atmospheric Administration under Award NA20OAR4590345. The authors thank Jennifer Sprague-Hilderbrand for her role in making this project possible.

Data availability statement. Data and more results from this research are available online (https://crcm.shinyapps.io/probcom/).

REFERENCES


Jenkins, S. C., A. J. L. Harris, and R. M. Lark, 2019: When unlikely outcomes occur: The role of communication format


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, https://doi.org/10.1371/journal.pone.0213522.


