Risk and Uncertainty Communication Using Probabilistic Information: A Systematic Review and Assessment of Existing Research

Joe Ripberger
Andrew Bell
Carol Silva
Hank Jenkins-Smith

NATIONAL INSTITUTE
for RISK and RESILIENCE
The UNIVERSITY of OKLAHOMA

Motivation

- Probabilistic forecast information is rapidly proliferating, injecting in a new wave of uncertainty into the forecast and warning process.

- Most scientists agree that this is a positive development **but** incorporating probability information into risk communication can be challenging because probabilities are notoriously difficult to communicate effectively to lay audiences.

- What does the research literature say about the “best” way to include probability information in risk communication?

    - What is the evidence base for different practices?

Project

- **Systematic review** of research literature on the inclusion of probability uncertainty information in risk messages

- Timeline: August 1, 2019 – September 30, 2020

- Deliverables:
  - Bibliographic archive of relevant research with topic tags and summary notes
  - Summary report that highlights existing knowledge, gaps, and priorities for future research
  - Summary report with recommendations to assist in the practice of communicating uncertainty and probabilities
  - Presentation of results to NWS/OWAQ partners

Systematic reviews

- Type of literature review that uses a transparent and replicable methodology to identify relevant research from past studies, evaluate results from those studies, and synthesize findings both qualitatively and quantitatively

- Steps in a systematic review:
    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
    6. Define certainty levels for common findings

Systematic reviews

- Type of literature review that uses a transparent and replicable methodology to identify relevant research from past studies, evaluate results from those studies, and synthesize findings both qualitatively and quantitatively

- Steps in a systematic review:
  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
  6. Define certainty levels for common findings

We are here →

Systematic reviews

- Type of literature review that uses a transparent and replicable methodology to identify relevant research from past studies, evaluate results from those studies, and synthesize findings both qualitatively and quantitatively

- Steps in a systematic review:
  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
  6. Define certainty levels for common findings

Search for and identify relevant studies

- Search Methodology:
  1. Electronic search databases
     - ProQuest, Web of Science, and EBSCO Academic Search Elite
  2. Previous literature reviews
  3. Citation chains
     - References IN articles
     - References TO articles

- Inclusion Criteria:
  - Original research (not a literature review, essay, or workshop report)
  - Directly study the communication of a specific uncertainty or probability (not perception of risk, uncertainty, or probability alone)
  - Replicable quantitative methodologies (not interviews, observations, or focus groups)

Search for and identify relevant studies

Electronic Search Methodology

ProQuest Search

Set #	Search Terms	# of Results
1	ti(communicat* OR perception OR inform* OR messag* OR understand*) AND PEER(yes)	679,783
2	ti(risk OR probabil* OR uncertain*) AND PEER(yes)	707,081
3	ab(weather OR climat* OR meterolog* OR "global warming" OR forecast*) AND PEER(yes)	860,375
4	(ab(experiment* OR survey* OR data* OR statistic*) OR ti(experiment* OR survey* OR data* OR statistic*)) AND PEER(yes)	13,223,159
5	S1 AND S2 AND S4	13,442
6	S1 AND S2 AND S3 AND S4	**788**
7	S1 AND S2 AND S3	1,350

Web of Science Search

Set #	Search Terms	# of Results
1	ti=(communicat* OR perception OR inform* OR messag* OR understand*)	733,747
2	ti=(risk OR probabil* OR uncertain*)	814,479
3	ts=(weather OR climat* OR meterolog* OR "global warming" OR forecast*)	716,340
4	ts=(experiment* OR survey* OR data* OR statistic*) OR ti=(experiment* OR survey* OR data* OR statistic*))	10,657,241
5	S1 AND S2 AND S4	**7,007**
6	S1 AND S2 AND S3 AND S4	**565**

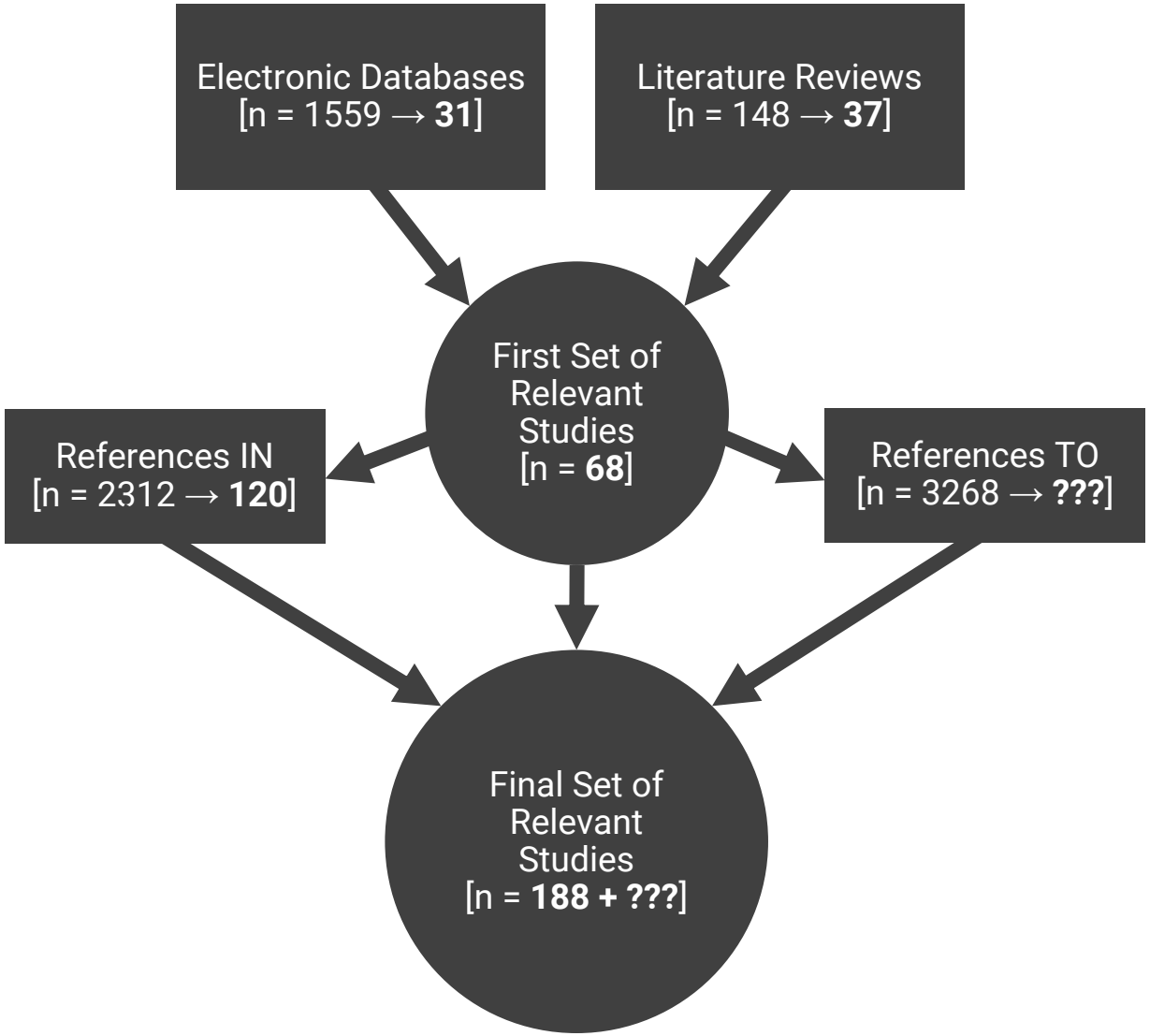
EBSCO (Academic Search Elite) Search

Set #	Search Terms	# of Results
1	ti(communicat* OR perception OR inform* OR messag* OR understand*)	364,341
2	ti(risk OR probabil* OR uncertain*)	315,672
3	ab(weather OR climat* OR meterolog* OR "global warming" OR forecast*)	322,317
4	S1 AND S2 AND S3	378
5	ab(experiment* OR survey* OR data* OR statistic*) OR ti(experiment* OR survey* OR data* OR statistic*))	5,060,185
6	S1 AND S2 AND S5	**3,493**
7	S1 AND S2 AND S3 AND S5	**206**

Search target: articles about probabilistic risk/uncertainty communication in the weather and climate domain that use quantitative methodologies

Search results: 1559 possibly relevant articles

Search for and identify relevant studies



LINK to list of relevant studies

1. Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B. & Katsikopoulos, K. V. 'A 30% Chance of Rain Tomorrow': How Does the Public Understand Probabilistic Weather Forecasts? *Risk Analysis* **25**, 623–629 (2005).
2. Roulston, Bolton, G., Kleit, A. & Sears-Collins, A. A Laboratory Study of the Benefits of Including Uncertainty Information in Weather Forecasts. *Weather and Forecasting* **21**, 116–122 (2006).
3. Roulston, M. S. & Kaplan, T. R. A laboratory-based study of understanding of uncertainty in 5-day site-specific temperature forecasts. *Meteorological Applications* **16**, 237–244 (2009).
4. Hoekstra, S. & Brooks, H. A Preliminary Look at the Social Perspective of Warn-on-Forecast: Preferred Tornado Warning Lead Time and the General Public's Perceptions of Weather Risks. *Weather, Climate, and Society* **3**, 128–140 (2011).
5. McClure, J., H. Doyle, E. E. & Velluppillai, J. M. A tale of two cities: Judgments about earthquake and aftershock probabilities across time windows. *International Journal of Disaster Risk Reduction* **14**, 15–26 (2015).
6. Cuite, C. L., Weinstein, N. D., Emmons, K. & Colditz, G. A Test of Numeric Formats for Communicating Risk Probabilities: *Medical Decision Making* (2008) doi:10.1177/0272989X08315246.
7. Sanyal, J., Zhang, S., Bhattacharya, G., Amburn, P. & Moorhead, R. J. A User Study to Compare Four Uncertainty Visualization Methods for 1D and 2D Datasets. *IEEE Trans. Vis. Comput. Graph.* **15**, 1209–1218 (2009).
8. Highhouse, S. A verbal protocol analysis of choice under ambiguity. *Journal of Economic Psychology* **15**, 621–635 (1994).
9. Zikmund-Fisher, B. J., Fagerlin, A., Roberts, T. R., Derry, H. A. & Ubel, P. A. Alternate Methods of Framing Information About Medication Side Effects: Incremental Risk Versus Total Risk of Occurrence. *Journal of Health Communication* **13**, 107–124 (2008).
10. Teigen, K. H. & Brun, W. Ambiguous probabilities: when does p=0.3 reflect a possibility, and when does it express a doubt? *Journal of Behavioral Decision Making* **13**, 345–362 (2000).
11. Durbach, I. N. & Stewart, T. J. An experimental study of the effect of uncertainty representation on decision making. *European Journal of Operational Research* **214**, 380–392 (2011).
12. Kreye, M., Goh, Y., Newnes, L. & Goodwin, P. Approaches to displaying information to assist decisions under uncertainty. *Omega* **40**, 682 (2012).
13. Dieckmann, N. F., Peters, E. & Gregory, R. At Home on the Range? Lay Interpretations of Numerical Uncertainty Ranges. *Risk Analysis* **35**, 1281–1295 (2015).
14. Newman, G. E. & Scholl, B. J. Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychon*

NOTE THAT THIS LIST WILL CHANGE AS THE REVIEW CONTINUES

Systematic reviews

- Type of literature review that uses a transparent and replicable methodology to identify relevant research from past studies, evaluate results from those studies, and synthesize findings both qualitatively and quantitatively

- Steps in a systematic review:
  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
  6. Define certainty levels for common findings

Extract key topics, questions, methods, and findings

Study	Source	Study Type	n	Population	Location	Exp. Treatment	Outcome Measure(s)	Outcome Result	Summary of Findings	F
Abraham et al (2015)	ES	Survey	274	Public	UK	None	Understanding of PoP forecasts (30% chan	27% - correct interpretation 34% - other 57% of resp. 40 y/o or below	This study examines how correc Most people cannot correctly int	W
Ancker et al (2011 i)	LR	Quasi-Experiment	165	Public	Online and a hospital waiti	Each participant show Only able to look at an	Accuracy of estimates	Random arrangements: Mean ina Sequential: Mean inaccuracy diff	"Although average estimates we	
Ancker et al (2011 ii)	LR	Quasi-Experiment	165	Public	Online and a hospital waiti	Static graphics vs. inte	Risk feelings, risk estimates, intent to take	No main effects for risk estimate	"A game-like graphic that allow	
Armstrong et al (2001)	LR	Experiment	246	Public	Philadelphia	Control group: Answe Treatment group: Did t	Understanding of survival curves	% Correctly identifying number % Correctly identifying change i	Most people can understand a su	
Armstrong et al (2002)	LR	Experiment	451	Public	Philadelphia	Participants given eith	Accuracy of understanding, willingness to	"Participants who received the in Statistically sig. differences in ef	"Framing graphic risk informati Possibly due to task matching ef	
Ash et al (2014)	LR	Experiment	501	Undergrads	South Carolina (USA)	Random dot somewhe 1 of 3 warning visualiz	1-5 scale: If you were located at this dot, h 1-5 scale: If you were located at this dot, h	Original, deterministic design ha However, the two probabilistic o	The researchers conclude that fe	
Brun & Teigen (1988) Study 1	LR	Quasi-Experiment	64 3 group	Psychology undergrads	Bergen, Norway	None	Participants were given lists of probability Also asked to give an estimate of how muc Asked to pick the "best" expressions	High standard deviations for each Estimated ambiguity was a lot lo Estimates in news context (group	There is a great deal of ambiguit The amount of ambiguity is subs	
Brun & Teigen (1988) Study 2	LR	Quasi-Experiment	66 phys 64 pare 24 psyc	Medical professionals, parents, and psych undergrads	Norway	None	Participants were given lists of probability Also asked to give an estimate of how muc Asked to pick the most "emotionally charg	Physician group had lowest / mos Substantial variation in all groups Ambiguity still higher than peopl	Again, there is a great deal of an The amount of ambiguity is subs	
Brun & Teigen (1988) Study 3	LR	Quasi-Experiment	23 stud	Students	Bergen, Norway	None	"Participants were given lists of probability Also asked to give an estimate of how muc	A lot of variability copmared to s Still within-group variability too	Context is really important in hc	
Budescu et al (2009)	ES	Experiment	223	60% Student Volunteers	America (U of Illinois)	Control, translation, w	Probability judgements based on informatic	When looking at all 5 terms: Control - 12.9% consistent, 62.5% Translation - 19.1%, 57.3%, 23.6 Narrow - 39% Wide - 22%	The public consistenty misinterp	
Budescu et al (2012) Study 1	ES	Experiment	556	Public	America	Control, translation (gi	Probability judgements based on informatic	Low consistency with IPPC guid 20.76% consistent - Control 18.81% - translation 30.12 - VN	The public consistenty misinterp Ideology and views on CC chan Alternative presentations (VN)	
Budescu et al (2012) Study 2	ES	Analysis of Experiment	556	Public	America	None (Analyzing expe	Probability judgements based on informatic	Correlation with consistency wit Belief in GCC - 0.15 Perception of causes - 0.14 Perception of consequences - 0.1 Numeracy score - 0.14	Believers in GCC gave higher es Democrats gave higher estimate Belief in GCC, numeracy, educa	

Systematic reviews

- Type of literature review that uses a transparent and replicable methodology to identify relevant research from past studies, evaluate results from those studies, and synthesize findings both qualitatively and quantitatively

- Steps in a systematic review:
  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
  6. Define certainty levels for common findings

Evaluate the quality of relevant studies

- Indicators of quality (validity):
  1. External validity (EV): sample size and generalizability (national survey vs. survey of college students in OK)
  2. Internal validity (IV): confidence in causality (experiment vs. correlation)
  3. Domain validity (DV): weather, climate, health, etc.
- Scoring system (3 points)
  - 1 = low; 2 = medium; 3 = high
  - Each study can range in validity from 3 (low on all thee indicators) to 9 (high on all three indicators)

Evaluate the quality of relevant studies

Journal of Experimental Psychology: Applied
2012, Vol. 18, No. 1, 126–140

1076-898X/11/$12.00 DOI: 10.1037/a0025185

Uncertainty Forecasts Improve Weather-Related Decisions and Attenuate the Effects of Forecast Error

Susan L. Joslyn and Jared E. LeClerc
University of Washington

Although uncertainty is inherent in weather forecasts, explicit numeric uncertainty estimates are rarely included in public forecasts for fear that they will be misunderstood. Of particular concern are situations in which precautionary action is required at low probabilities, often the case with severe events. At present, a categorical weather warning system is used. The work reported here tested the relative benefits of several forecast formats, comparing decisions made with and without uncertainty forecasts. In three experiments, participants assumed the role of a manager of a road maintenance company in charge of deciding whether to pay to salt the roads and avoid a potential penalty associated with icy conditions. Participants used overnight low temperature forecasts accompanied in some conditions by uncertainty estimates and in others by decision advice comparable to categorical warnings. Results suggested that uncertainty information improved decision quality overall and increased trust in the forecast. Participants with uncertainty forecasts took appropriate precautionary action and withheld unnecessary action more often than did participants using deterministic forecasts. When error in the forecast increased, participants with conventional forecasts were reluctant to act. However, this effect was attenuated by uncertainty forecasts. Providing categorical decision advice alone did not improve decisions. However, combining decision advice with uncertainty estimates resulted in the best performance overall. The results reported here have important implications for the development of forecast formats to increase compliance with severe weather warnings as well as other domains in which one must act in the face of uncertainty.

Keywords: risk-seeking, decision making, uncertainty

Many important decisions are made under uncertainty, such as choosing medical treatment, choosing retirement investments, and deciding whether to take precautionary action against severe weather. It is now possible in several domains to quantify uncertainty, providing accurate numeric assessments to end-users. This is especially true of weather forecasts. Recent technical advancements have resulted in methods for calculating forecast probabilities that are both calibrated and precise (Gneiting & Raftery, 2005; Sloughter, Raftery, Gneiting, & Fraley, 2007). However, very little of this information reaches the general public. At present, most public weather forecasts, like information provided to nonexperts in a broad range of domains, remain deterministic. They provide a single value, such as a nighttime low temperature of 32 °F, implying unrealistic certainty.

Although there are strong theoretical arguments for the economic benefit of uncertainty forecasts (Murphy, 1977; Thompson, 1952) and although psychological research suggests that numeric estimates lead to a more precise understanding than do verbally described categories (Budescu, Broomel & Por, 2009), there is still widespread reluctance to provide numeric uncertainty to the general public for fear it will be misused. Indeed, many of the crucial decisions based on weather forecasts are similar to those that have been shown in laboratory studies to lead to a decision error referred to as risk-seeking (Kahneman & Tversky, 1979). Weather-related decisions often concern costly precautionary action and must be made early, when the probability of adverse weather is well below 50%. Experimental participants asked to choose between options that involve losses tend to prefer a gamble with a moderately low probability, even when the probability-weighted value, called the expected value (Bernoulli, 1954), is less than the certain alternative (Kahneman & Tversky, 1984). In laboratory settings in which all of the relevant information is provided to the participant, this error, referred to as "risk seeking," and the opposite "risk averse" error, preferring a sure option to a risky option with equal or greater expected value, can be explained by cumulative prospect theory (Tversky & Kahneman, 1992). According to cumulative prospect theory, both errors result from a combination of the effects of a nonlinear utility function, which is steeper for losses than for gains, and a nonlinear subjective probability weighting function. We focus here on the risk-seeking error, which has been demonstrated in diverse domains including medicine (Nightingale, 1987), used car purchasing (Betts & Taran, 2005), and international relations (Haas, 2001). Such evidence suggests that, in a variety of situations involving loss, people assume more risk than is economically rational. This is true of both professionals as well as nonexperts, suggesting that it is not merely a matter of lack of education or experience.

This article was published Online First August 29, 2011.
Susan L. Joslyn and Jared E. LeClerc, Department of Psychology, University of Washington.
This research was supported by the National Science Foundation, Under Grant ATM 0724721.
Correspondence concerning this article should be addressed to Susan L. Joslyn, Department of Psychology, Box 351525, University of Washington, Seattle, WA 98195. E-mail: susanj@u.washington.edu

- External validity: 1 (low)
  - Survey of 304 University of Washington psychology students
- Internal validity: 3 (high)
  - Multiple high-quality survey experiments
- Domain validity: 3 (high)
  - Winter weather decisions (road salt experiments)
- **Total validity score: 7 (medium to high)**

Systematic reviews

- Type of literature review that uses a transparent and replicable methodology to identify relevant research from past studies, evaluate results from those studies, and synthesize findings both qualitatively and quantitatively

- Steps in a systematic review:
  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**
  6. Define certainty levels for common findings

Analyze and combine studies

- Common topics:
    1. General understanding of probability information
    2. Opinions and perceptions about probability information
    3. Probabilistic vs. deterministic information
    4. Verbal expressions of uncertainty
    5. Numeric formats and framing
    6. Visualizations, graphs, and icons
    7. Affect and emotion
    8. Communicating small or long-term risk
    9. Other/misc.
- Common questions (study outcome measures):
    - How does probability information impact *risk comprehension*?
    - How does probability information impact *protective action decisions/intentions/behaviors*?
- Methods:
    - Surveys that vary in size and scope; research designs that vary in validity
- Findings/recommendations:
    - Outcome result (statistics)
    - 2-3 sentence plain text summary of primary findings
    - Identify/infer a recommendation from findings

Analyze and combine studies

Study: ES98	
Reference	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, 623–629.
Abstract	The weather forecast says that there is a "30% chance of rain," and we think we understand what it means. This quantitative statement is assumed to be unambiguous and to convey more information than does a qualitative statement like "It might rain tomorrow." Because the forecast is expressed as a single-event probability, however, it does not specify the class of events it refers to. Therefore, even numerical probabilities can be interpreted by members of the public in multiple, mutually contradictory ways. To find out whether the same statement about rain probability evokes various interpretations, we randomly surveyed pedestrians in five metropolises located in countries that have had different degrees of exposure to probabilistic forecasts––Amsterdam, Athens, Berlin, Milan, and New York. They were asked what a "30% chance of rain tomorrow" means both in a multiple-choice and a free-response format. Only in New York did a majority of them supply the standard meteorological interpretation, namely, that when the weather conditions are like today, in 3 out of 10 cases there will be (at least a trace of) rain the next day. In each of the European cities, this alternative was judged as the least appropriate. The preferred interpretation in Europe was that it will rain tomorrow "30% of the time," followed by "in 30% of the area." To improve risk communication with the public, experts need to specify the reference class, that is, the class of events to which a single-event probability refers.
Topic	General understanding of probability information
Outcome Measure	Risk comprehension
Findings	Single event probabilities can be interpreted by members of the public in multiple, mutually contradictory ways
Validity	EV = 2 (med); IV = 1 (low); DV = 3 (high); Total validity score = 6
Recommendation	Specify the reference class when communicating a single-event probability

- Note: when complete, these "study cards" will populate a searchable database of studies

Systematic reviews

- Type of literature review that uses a transparent and replicable methodology to identify relevant research from past studies, evaluate results from those studies, and synthesize findings both qualitatively and quantitatively

- Steps in a systematic review:
  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
  6. **Define certainty levels for common findings**

Define certainty levels for common findings

- Indicators of certainty:
  1. Consistency of evidence (do all studies say the same thing?)
  2. Quantity of evidence (how many studies are there?)
  3. Quality of evidence (on average, how much validity do studies have?)

- Scoring system (3 points)
  - 1 = low; 2 = medium; 3 = high
  - Each recommendation can range in certainty from 3 (low on all thee indicators) to 9 (high on all three indicators)

Define certainty levels for common findings

Recommendation 1: THIS INFORMATION WILL CHANGE AS THE REVIEW CONTINUES	
Recommendation	Present both numerical and verbal probability information in a risk message. People prefer numerical information for its accuracy but use verbal statements to express probabilities to others. Presenting both makes sure that people have the right information no matter the purpose for which it is used.
Relevant Studies	Brun & Teigen (1988); Shaw & Dear (1990); Weber & Hilton (1990); Wogalter et al. (1999)
Consistency of Evidence	High
Quantity of Evidence	High
Validity of Evidence	High
Certainty	High

Define certainty levels for common findings

Recommendation 2: THIS INFORMATION WILL CHANGE AS THE REVIEW CONTINUES	
Recommendation	For single unique events, express proportions as percentages if possible.
Relevant Studies	Peters et al. (2011)
Consistency of Evidence	Medium
Quantity of Evidence	Low
Validity of Evidence	Medium
Certainty	Low to Medium

Define certainty levels for common findings

Recommendation 3: THIS INFORMATION WILL CHANGE AS THE REVIEW CONTINUES	
Recommendation	When possible, use positive frames (i.e., chance of survival) in place of negative frames (i.e., chance of death) when communicating probabilities; in some cases, both may be necessary.
Relevant Studies	Peters et al. (2011); Gigerenzer (2014); Pidgeon & Fischhoff (2011)
Consistency of Evidence	Medium
Quantity of Evidence	Medium
Validity of Evidence	Medium
Certainty	Medium

Next Steps

1. Complete list of relevant studies
2. Compile database with core information about each study; assess validity
3. Produce list of recommendations; assess certainty

- Deliverables:
  - Bibliographic archive of relevant research with topic tags and summary notes
  - Summary report that highlights existing knowledge, gaps, and priorities for future research
  - Summary report with recommendations to assist in the practice of communicating uncertainty and probabilities
  - Presentation of results to NWS/OWAQ partners

Questions?

Contact:

Joe Ripberger (jtr@ou.edu)
Andrew Bell (andrewjbell@ou.edu)
Carol Silva (clsilva@ou.edu)
Hank Jenkins-Smith (hjsmith@ou.edu)