Tornadoes, Social Science, and the False Alarm Effect

JOSEPH E. TRAINOR

Disaster Research Center, School of Public Policy and Administration, University of Delaware, Newark, Delaware

DANIELLE NAGELE

NOAA/National Ocean Service/Center for Operational Oceanographic Products and Services, Silver Spring, Maryland, and Disaster Research Center, School of Public Policy and Administration, University of Delaware, Newark, Delaware

BRENDA PHILIPS

Resource Economics, Center for Collaborative Adaptive Sensing of the Atmosphere, Electrical and Computer Engineering, University of Massachusetts Amherst, Amherst, Massachusetts

BRITTANY SCOTT

Disaster Research Center, Department of Sociology, University of Delaware, Newark, Delaware

ABSTRACT

Despite considerable interest in the weather enterprise, there is little focused research on the "false alarm effect." Within the body of research that does exist, findings are mixed. Some studies suggest that the false alarm effect is overstated, while several recent efforts have provided evidence that FAR may be a significant determinate of behavior. This effort contributes to the understanding of FAR through a sociological analysis of public perceptions and behavioral responses to tornadoes. This analysis begins by addressing public definitions of FAR and then provides two statistical models, one focused on perception of FAR and one focused on behavioral response to tornado warnings. The authors' approach incorporates a number of sociological and other social science concepts as predictors in both of these models. Findings provide a number of important insights. Most notably, it is found that 1) there is a wide degree of variation in public definitions of false alarm, 2) actual county FAR rates *do not* predict perception of FAR, 3) actual county FAR rates *do* predict behavioral response, and 4) planning and family characteristics are also influential. Another major contribution is to illustrate the significant complexity associated with analysis of false alarms. Conclusions discuss the limits of this analysis and future direction for this type of research.

1. Background

National Weather Service (NWS) forecasters issue tornado warnings based on a variety of factors, including environmental conditions, scientific knowledge of storm evolution, ground truth information, radar data, characteristics of the population, characteristics of the built environment, and even "gut or instinct." When the evidence suggests that tornado formation is imminent or a tornado is present, forecasters issue a deterministic tornado warning (Andra et al. 2002). The tornado warning includes several key pieces of information, including a geographic area, a valid time, the duration, and text describing the potential hazard, its path, and advice for protective action. In current practice, local NWS warning forecast offices (WFOs) issue the initial tornado warning, and then the warning information is disseminated to the public through broadcast media, outdoor sirens, social media, call-out systems, word of mouth, texting, or phone calls. NWS, Red Cross, and Federal Emergency Management Agency (FEMA) policy recommend that members of the public should shelter in place immediately after receiving a tornado warning,

Corresponding author address: Joseph E. Trainor, School of Public Policy and Administration, University of Delaware, 166 Graham Hall, Newark, DE 19716.
E-mail: jtrainor@udel.edu

Severe Weather Forecasted

FIG. 1. Warning performance typology.

except for those in cars or in mobile homes who are advised to seek other shelter.

Because of limitations of our scientific knowledge, observational technology, models, and forecaster interpretation tornado warnings are not always accurate. To measure warning system performance, the NWS tracks and reports a number of performance measures, including warning lead time, missed events, probability of detection (POD), and false alarm ratio (FAR; National Weather Service 2011; see also http://www.nws.noaa.gov/oh/rfcdev/docs/Glossary_Forecast_Verification_Metrics.pdf). The most relevant for the purpose of this analysis are derived from a four-square typology (see Fig. 1) based on the intersection of 1) whether or not a warning was issued and 2) whether or not an event was observed. The four categories that result are hit, miss, false alarm, and all clear. Over time, forecasters and offices are evaluated based on the proportions between these categories of prediction.[1] These metrics are used to measure how well a forecaster or office predicts the touchdown of a tornado within a specific geographic area and within a specific time period. Many weather safety professionals give significant attention to these measures and in particular believe that issuing too many false alarms (FAs) is the modern-day version of the well-known fable "The Boy Who Cried Wolf." They believe that that if one "cries warning" too often, the public will eventually become complacent and not take protective action measures that can prevent injuries, death, and property damage (Breznitz 1984). If true, this cry-wolf effect, or

"false alarm effect," is especially relevant for tornadoes, where 74% of tornado warnings have been false alarms over the past 5 years (NOAA/NWS Office of Climate, Water, and Weather Services 1986a). Even so, it is important to recognize that NWS performance metrics, while important, do not directly evaluate how warnings are linked to public perception and behavioral response. It is important to develop that linkage scientifically.

In looking more closely at the NWS definition of tornado false alarm, the agency says it is an "unverified" warning (i.e., one where the tornado funnel did not touch down in the warned area during the effective time of the warning). Unlike warnings, however, which are broadly disseminated, there are no official or informal mechanisms to notify the public that a false alarm occurred or why. Nor does the NWS actively communicate annual average FARs to the public. As a result, the public is left to gain an impression of false alarms through experiencing severe weather and severe weather communications around them and determining if they think warnings are accurate or not. With this in mind, it is empirically important to explore the connections and gaps between individual and organizational views on system performance. Answers to these questions could provide important insight into how to improve the current approach to warning policy and procedures and could significantly enhance our understanding of how perceptions of tornado risks develop more generally. Such an understanding is important if we are to move from simplistic performance measures to those more calibrated to the limits of forecast, human perception, and behavioral response [see Barnes et al. (2006) for an example].

Our research begins to address some of these questions by examining the impacts of tornado false alarms on public perception and behavior. The current analysis adopts a sociological perspective on risk perception. The approach suggests that it is "a basic sociological task to explain how social agents create and use boundaries to demarcate that which is (or is not) dangerous" (Clarke and Short 1993, p. 379). Further, the view calls on social scientists to explore how social circumstances impact how we develop these perceptions and calls for a clearer understanding of how institutional views on what is/was dangerous reconcile with the ways that normal people develop their views (Tierney 1999). Building on this sociological tradition, we focus on how social demographics, contextual conditions, and institutional warnings influence the perception of false alarms and behavioral response to tornadoes. In addition, we integrate variables from social psychology, economics, and NWS false alarm ratios (a measure of the institutions view on accuracy of warnings). In the first model, we present predictors of households' perceptions of false alarm. In the second model, we explore the relationship between these measures and protective action responses

---

[1] For example, the NWS FAR = FAs/(hits + FAs); POD = hits/(misses + hits).

during significant severe weather events, many of which produced actual tornados.

2. Literature review

Despite expectations to the contrary, much of the hazards literature suggests no cry-wolf effect or negative influence of false alarm on household perceptions and behavior. This finding holds true for hurricanes (Dow and Cutter 1998), tornadoes (Schultz et al. 2010), and floods (Barnes 2006). However, several recent analyses have begun to provide alternative results, particularly for tornado warnings. Notably, in their statistical analysis of historical tornado casualty data, Simmons and Sutter (2006) have shown that high NWS FARs within a region increases the likelihood of fatalities and casualties by 29%–40%. Similarly, Ripberger et al. (2015) have also noted a FAR effect by modeling the link between false alarms and missed events with trust in NWS and by then linking trust to reductions in hypothetical future response to tornadoes.

Our research contributes to this evolving interdisciplinary conversation on FAR in several ways, some of which have been explicitly called for by Ripberger et al. (2015) and Barnes (2006). First, it provides new insights into how members of the public define a false alarm. Second, it explores household perceptions of false alarm frequency at the univariate level. Third, it explores the degree to which demographics, other social factors, and NWS measures of false alarm are related to public perceptions of false alarm. Finally, we explore how perceptions of FAR, NWS FAR, and other social variables influence protective action behavior by the public. Despite some sampling limits, the approach includes several methodological advances as well. One important distinction is that we measure people's *actual* protective action behavior to a tornado warning, an approach that addresses some of the biases inherent in research that uses hypothetical examples, stated intentions, or events that occurred in the distant past as a way to elicit perceptions and future behavior. Further, our work is also unique in that our data collection efforts for each event were completed within one month of the event's occurrence. While this did come with some tradeoffs, as discussed in the limitations section, the speed of data collection helped to reduce retrospective bias, helping to ensure that the warned events could be meaningfully discussed. Finally, the approach also provides some value in that it highlights the complexity of the warning response process when viewed across the system. In what follows, we present a conceptual introduction to the three analyses included in this paper, followed by a discussion of the methodology used to collect the data.

Next, we present findings for each analysis. Finally, we present a discussion of the conclusions, limitations, and future directions for research on false alarms.

3. What is a false alarm?

As discussed above, our first analysis below focuses on responses to an open-ended survey question that asked respondents to define the concept of false alarm. In recognition that gaps often exist between official and lay definitions of terminology, the primary analytical goal was to explore what the public thinks of when the term "false alarm" is used. This part of the data analysis focused on responses to the open survey question "Could you please describe what a false alarm is?" For this question, interviewers typed the verbatim responses of interviewees. Using an inductive open coding process, coders were asked to read all the responses, take notes, and inductively code important dimensions of the responses using thematic analysis (Boyatzis 1998) to capture variations in respondents' ideas. These multiple coders then met to compare notes and to develop a set of focused codes (Corbin and Strauss 2008) that would be uniformly applied across all of the responses. Three individual coders were then asked to code all responses into these categories. Coders met to reconcile any differences in their application of the codes for each case and were instructed to consider all views in coming to a final determination. The resulting data are discussed in the analysis section and included in both models.

4. Perception of false alarm

In the extant warning literature, risk perceptions are typically treated as independent variables and as such are used to explain behavior. It has been uncommon for analysts to consider the drivers of perception (Tierney 1999; Ripberger et al. 2015). The perception of the FAR model presented below was developed to improve our understandings of how risk perceptions generally, and perception of false alarm specifically, are developed. Our hypotheses for this model are presented in the empirical review section below. While a few variables retest patterns observed in previous studies, given the lack of existing research on how perceptions of false alarm develop, we also extended previous findings on "perception of risk" in order to create hypotheses. We developed these hypotheses based on the assumption that those who have in prior research been found to respond to warnings will likely have lower or reduced perceptions of the frequency of false alarm occurrence. The logic behind this assumption is that, given their concern, these groups will also prefer more information

TABLE 1. Summary of hypotheses.

Perception of false alarm hypotheses	Behavioral model hypotheses
PH1: Females will have a lower perception of false alarms.	BH1: Females will be more likely to take protective actions.
PH2: Elderly will have a lower perception of false alarms	BH2: Elderly will be more likely to take protective actions
PH3: Nonwhites will have a higher perception of false alarm	BH3: Nonwhites will be less likely to take protective actions.
PH4: Tornado experience will reduce perception of false alarms.	BH4: Tornado experience will increase the likelihood of taking protective action.
PH5: Families with a disaster plan will perceive fewer false alarms.	BH5: Families with a disaster plan will be more likely to take protective action.
PH6: Families with children will perceive fewer false alarms.	BH6: Families with children will be more likely to take protective action.
PH7: Increased FARs will result in increased perceptions of false alarms.	BH7: Increased FAR will reduce the likelihood of taking protective action.
PH8: Residents of counties with a higher average number of tornadoes will have an increased perception of false alarms.	BH8: Residents of counties with a higher average number of tornadoes will be less likely to take protective actions.
PH9: Greater trust in local weather providers will result in decreased perceptions of false alarms.	BH9: Greater trust in local weather providers will result in increased likelihood of taking.
	BH10: As perception of FAR increases, the likelihood of taking protective actions decreases.

and as a result will be less likely to perceive warnings as false alarms regardless of other factors.

5. Approach to the protective action analysis

Our second model focuses on behavioral responses to tornado warning. It draws on much of the same theoretical and empirical material as the perception model in order to formulate hypotheses about patterns of protective action behavior. In particular, it explores sociological context, past experiences, planning, the influence of FAR perception, and the NWS FAR measure. Given the high level of attention on FAR by the NWS, some might be surprised to know there has been little direct empirical evidence to support the claim that reducing false alarms will have a significant influence on human behavioral response to warnings. Further, many have found that false alarms do not significantly impact an individual's decision to take protective action (e.g., evacuate) in future disasters (Dow and Cutter 1998; Benight et al. 2004), and most assert that FAR has little impact on an individual's perception of future warnings or future behavior (Atwood and Major 1998; Dow and Cutter 1998; Riad and Norris 1998; Benight et al. 2004). It is only recently that analyses have emerged that contradict these assertions. Most notably, Simmons and Sutter's economic analyses (Simmons and Sutter 2009, 2011) have found a strong connection between false alarm rate with tornado injury and morbidity. However, as the authors explicitly note, this is an indirect measure of the link between FAR and action. While their analyses are important and insightful, the limited availability of meaningful protective measures and the inability to capture actual behaviors are significant and

important intervening factors. Work by Ripberger et al. (2015) further establishes the possibility of a false alarm effect by modeling the link between false alarms and missed events with trust in NWS and by then linking trust to hypothetical future response to tornadoes. These recent works provide new insights and new controversy around how false alarms influence behavior. In light of their work, both have called for more detailed and novel analyses of these patterns. Our second model takes up this call by providing a statistical analysis that includes factors others have found important for predicting protective actions. The specific details are presented below.

6. Empirical foundations and specific hypotheses

In this section we provide the empirical basis for our hypotheses in our perception model (listed as PHx) and behavioral models (BHx). Table 1 provides a summary of the hypotheses.

a. Gender

In regards to gender, research shows that females are more tolerant of warnings and false alarms than males and are more likely to evacuate and have a stronger influence on others to do the same (Barnes 2006; Riad and Norris 1998). Similarly, they are also more likely to respond to a warning message (Bateman and Edwards 2002; O'Brien and Atchison 1998) and to shelter in safe locations (Comstock and Mallonee 2005). It has also been found that women perceive hazards and threats as more serious and risky than men; in other words, they generally have a higher risk perception (Fothergill 1996; Lerner et al. 2003). While it is worth noting that

Nagele and Trainor (2012) were unable to support these prior results, significant evidence exists of this relationship. Based on these findings, we propose the following:

PH1: Females will have a lower perception of false alarms.

BH1: Females will be more likely to take protective actions.

b. Age

There is still much disagreement in the literature with regards to the influence of age on warning response (Perry and Lindell 1997). Some suggest that the elderly tend to be more tolerant of increased warnings and consequently false alarms than younger respondents. Some prior research has suggested that the elderly are more likely to respond to messages (Aguirre et al. 1998; Baker 1987; Cutter and Barnes 1982) and more likely to understand a warning (Blanchard-Boehm 1998). Other studies have found no effect (Edwards 1993; Baker 1979; Mileti and Darlington 1997). Others still have found that the elderly are less likely to believe warnings (Mack and Baker 1961; Hodge et al. 1981) or respond (Gruntfest 1977, 1997). Despite a mixed empirical record, we propose the following:

PH2: Elderly will have a lower perception of false alarms.

BH2: Elderly will be more likely to take protective actions.

c. Race

In addition to gender and age, we will also test for a racial effect. Flynn et al. (1994) advanced the concept of a "white male effect" when they found that white, male conservatives with a higher social economic status were less concerned with common U.S. risks than other demographics. Similarly, other researchers have found that membership in a minority group reduces the likelihood of believing (Turner 1976; Cuthbertson and Nigg 1987) or responding to a warning message (Drabek and Boggs 1968; Perry et al. 1981; Lindell and Perry 1987; Mileti and O'Brien 1992; Edwards 1993). As with other factors, there are, however, mixed results where race was found to correlate with increased likelihood of responding (Riad and Norris 1998). We propose the following hypotheses:

PH3: Nonwhites will have a higher perception of false alarm.

BH3: Nonwhites will be less likely to take protective actions.

d. Experience

With respect to experiencing prior events, researchers have found evidence that prior hazard experiences influence decision-making (Dow and Cutter 1998) and can make people more likely to believe and to respond to tornado warnings (Hodler 1982). Prior experience with disasters or hazards leads to greater response to warnings (Mileti and O'Brien 1992; Landry and Rogers 1982; Lachman et al. 1961). Similarly, others have found that experience also increases the likeliness of people to prepare (Blanchard-Boehm and Cook 2004) as well as their desire to react more proactively in future events (Simmons and Sutter 2007). According to Mileti and Sorrenson (1990), our experiences influence our perception and response to warnings; thus, the more experience we have with severe weather, the lower our perception of false alarms will be.

PH4: Tornado experience will reduce perception of false alarms.

BH4: Tornado experience will increase the likelihood of taking protective action.

e. Emergency plan

Past research has shown an increased likelihood of taking protective action when a family has some sort of emergency plan in place for their household (Balluz et al. 2000, Blanchard-Boehm and Cook 2004). Those families that take the time to develop a plan value increased preparedness and are likely to respond to a warning (Nagele and Trainor 2012). Continuing to follow the same line of reasoning in previous sections, we propose that those with a family plan have a higher risk perception and thus a lower perception of false alarms.

PH5: Families with a disaster plan will perceive fewer false alarms.

BH5: Families with a disaster plan will be more likely to take protective action.

f. Children

Finally, the presence of children has often been correlated with a higher risk perception and intention to evacuate (Riad and Norris 1998; Houts et al. 1984; Drabek and Stephenson 1971). Similarly, those with children are more likely to respond to warning messages (Edwards 1993; Carter et al. 1983; Turner et al. 1981; Flynn 1979; Wilkinson and Ross 1970). Thus, the presence of children (under 18) will influence one's perception of risk, and consequently one's perception of false alarms.

PH6: Families with children will perceive fewer false alarms.

BH6: Families with children will be more likely to take protective action.

g. County FAR

Most discussions of false alarms begin with the assumption that increases in false alarm rates are recognized by the public and lead to increased perception of false alarms. We suggest that it is vital to empirically test that assumption. Therefore, in addition to sociodemographic characteristics, we propose that the NWS false alarm will have an impact on a person's perception of false alarms. In the behavioral model, we include the NWS FAR as an independent variable in order to explore its impact on an individual's perception of false alarms.

PH7: Increased FARs will result in increased perceptions of false alarms.

BH7: Increased FAR will reduce the likelihood of taking protective action.

h. County tornadoes

Along with the NWS FAR, the average number of tornadoes a county experiences may also impact perception, even if those tornadoes have never affected a family. While the NWS FAR may not increase, a higher number of tornado events in the area may impact awareness and perceptions in a number of ways. In addition, areas with frequent tornadoes may be desensitized to weather alerts and thus more critical of warning accuracy. For instance, they may consider anything outside of their community a "miss," when in actuality the tornado had hit within the warning area. We propose that the number of tornadoes an area gets will have an influence on people's perception of false alarms.

PH8: Residents of counties with a higher average number of tornadoes will have an increased perception of false alarms.

BH8: Residents of counties with a higher average number of tornadoes will be less likely to take protective actions.

i. Trust

The relationship between trust and false alarm is complicated (White and Eiser 2006). It is likely endogenous. What we mean is that people who believe the warning system is prone to false alarms may be less trusting of weather providers [see Ripberger et al. (2015) for an extended discussion of trust]; at the same time, those who do trust providers are probably less likely to have a higher perception of false alarms. This complexity makes it difficult to specify a linear relationship. That being said, for the purposes of this analysis, we are focused on FAR as an independent variable and as such will explore the effect of trust on FAR. Building on the extant evidence, we propose that the more credible or trustworthy the public finds weather sources, the more likely they are to respond to a warning from them (Perry 1982; Cutter 1987; Gruntfest 1997). If they are receiving information from a trustworthy, official source, they are more likely to believe the warning (Baker 1984, 1987; Cola 1996). We propose that those who already have a high opinion of their weather providers will perceive false alarm frequencies in their area as lower and will be more likely to take protective actions.

PH9: Greater trust in local weather providers will result in decreased perceptions of false alarms.

BH9: Greater trust in local weather providers will result in increased likelihood of taking protective actions.

j. False alarm and behavior

To further explore the effect of false alarm on behavior, our analysis will focus on the relationships between FAR and perceptions of FAR with protective action. A logically derived hypothesis captures the essence of the false alarm effect we wish to test:

BH10: As perception of FAR increases, the likelihood of taking protective actions decreases.

7. Methodology

This analysis was developed using a quantitative dataset created at the University of Delaware Disaster Research Center (DRC) as part of the National Science Foundation (NSF)-funded Collaborative Adaptive Sensing of the Atmosphere (CASA) Engineering Research Center. The data were collected by telephone interviews with an instrument that aimed to better understand public response to tornado and severe storm warnings by bringing together knowledge from social science disciplines that focus on weather warnings. The major topics the survey addresses include 1) receipt of warnings and alerts; 2) severe storm/tornado impacts; 3) confirmation/verification behavior; 4) access, use, and familiarity with specific sources of information; 5) multiple types of protective actions; 6) damage to property; 7) insurance coverage; 8) lead time, watch, warnings, and false alarms; 9) experience with previous hazards; 10) preparedness activities; and 11) demographic; and 12) socioeconomic variables. Our survey combines the respondents' attitudinal and perceptual questions about false alarms and tornado warnings with actual behavioral

response from a recent tornado warning that occurred in the respondents' county. By having respondents recount very recent behavior, we tried to lessen the impact of retrospective bias. The final instrument included 120 questions and took respondents between 15 and 45 min depending on their path through the skip patterns. The mean time to completion was 34 min. The survey was administered as a telephone interview using a computer assisted telephone interviewing (CATI) system. In evaluating these data, it should be noted that we did not focus on a single severe weather event. Instead, we developed a method to collect data from multiple events using a set of fixed methodological conventions. Over the course of the data collection, we put in place systematic procedures that improved the reliability and validity of the information we collected, each of which will be discussed in detail below.

a. Sampling

Data collection occurred during 2008, 2009, and 2010. The first year we started data collection in June as soon as the survey system was readied and continued until August. For subsequent years, we began searching for storms mid-February and ended around August. The only exception to this pattern was one major event in February 2009 where the system was deployed early because of an unusual tornado that was relevant to the larger CASA project. This data collection period covers the months when most tornado events occur in the United States.

Our sampling approach can be described as a two-stage process. Stage one involved the selection of geographic areas where households were likely to have encountered the need to make protective action decisions. For this phase we chose to select counties where a tornado warning was issued or a tornado was confirmed to have occurred. In the second phase, households were randomly selected from the county identified in phase one. More details are provided below. This first stage of our sampling process identified significant weather or warning events that occurred in the United States in the two weeks prior to the initiation of data collection. Because there is no sampling frame for the "population" of people who make protective action decisions, we collected data from selected counties in which tornado events or warnings occurred. Since it was not possible to survey every county that had a tornado in these years, we employed theoretical replication logic in order to select counties that provided variation in the population demographics. This sampling approach involves the selection of case counties in a deliberate and theoretically informed way. In our case, this involved varying the selection of events by including areas with varied demographic mixes.

Theoretical replication is common in multiple-case-study research [see Yin (2009) for a more detailed description]. We also focused on selecting counties where there was a high probability of a tornado event based on initial news and NWS reporting. In total, we collected data from households for 17 weather events over the 3-yr period as summarized in Table 2.

During stage two of our sampling process, landline telephone numbers from each of the counties above were obtained by purchasing a sample from Genesys, a third party sample provider. In each case, 1000 random digit dialing (RDD) numbers were requested. After obtaining the initial 1000 numbers, we had Genesys purge business and disconnected numbers from the initial sample. To purge the business numbers, a database comprised of nonresidential yellow page businesses is utilized. The distinction of *nonresidential* is important because over one million households nationwide use their residential phone number for business purposes as well. The generated sample was compared to this database, and any matching telephone numbers were purged from the sample. The remaining numbers that were not purged from the sample were then examined to determine if they were disconnected. Finally, a list of telephone numbers was provided for each county; it served as the sample for each event on average each sample contained about 500 potential households after the purge process.

b. Calling process and response rates

Given our focus on choices made during multiple severe weather events, it was necessary to adopt a process for identifying, selecting, and transitioning between severe weather events in order to build a sample. Our data collection process relied on a three-week cycle. During week 1, a "storm searcher" developed a list of candidate events by reviewing TV news and NWS warning products. The primary focus was to select events that were significant enough to produce tornados. During weeks 2 and 3, our call center was activated at different times of the day with a focus on calling between 1800 and 2100 local time (LT) Monday through Thursday and on Saturdays from 1200 to 1600 LT. All interviewers were trained on the instrument, concepts, and soft refusal call conversion techniques. The survey did not offer an incentive. After three weeks of calling we terminated data collection for that event and would start the cycle again as soon as a suitable next case was identified. This process optimized the timeliness of our calling by ensuring that we did not call more than a few weeks after an event occurred, greatly reducing retrospective bias. It did so, however, at the expense of not fully exhausting our samples. Each phone number was called up to four times

TABLE 2. Sample data and response rates.

Date	Event county	State	Event type	N	Response rate	Cooperation rate	Refusal rate	Contact rate	2010 Census		
									% female	% white	% over 65
5 Jun 2008	Tulsa	OK	False alarm	21	6%	22%	6%	28%	51%	69%	12%
11 Jun 2008	Riley	KS	EF4	91	14%	46%	14%	31%	48%	83%	7%
11 Jul 2008	Kandiyohi	MN	EF3	54	11%	43%	11%	26%	50%	93%	16%
4 Aug 2008	DuPage	IL	EF1	39	6%	41%	5%	14%	51%	78%	12%
10 Feb 2009	Carter	OK	EF4	58	10%	45%	9%	21%	51%	74%	15%
10 Feb 2009	Oklahoma	OK	Two tornadoes (EF1 and EF2)	53	7%	39%	7%	19%	51%	65%	12%
26 Mar 2009	Jasper	MS	EF1	65	18%	37%	15%	47%	52%	46%	16%
3 May 2009	Jefferson	AL	EF1	52	12%	44%	12%	27%	53%	53%	13%
13 May 2009	Caddo	OK	EF2	46	10%	61%	10%	16%	48%	63%	15%
12 Jun 2009	Shelby	TN	EF1	37	12%	42%	12%	30%	52%	52%	10%
20 Jul 2009	Denver	CO	False alarm	51	14%	42%	14%	33%	50%	69%	10%
2 Apr 2010	Rush Springs	OK	Conflicting assessments	47	14%	26%	14%	53%	51%	88%	18%
24 Apr 2010	Yazoo	MS	Multiple tornadoes up to EF4	44	10%	23%	10%	42%	46%	40%	12%
10 May 2010	Pottawatomie	OK	Three tornadoes (EF1, EF3, EF3)	44	9%	20%	9%	47%	52%	76%	14%
6 Jun 2010	Monroe	MI	Two tornadoes (EF1 and EF2)	42	13%	22%	13%	58%	51%	94%	13%
12 Jul 2010	Lancaster	SC	Two tornadoes (EF0 and EF0)	43	11%	15%	11%	78%	51%	72%	15%
25 Jul 2010	Bronx	NY	EF1 (missed event)	17	4%	8%	4%	53%	53%	28%	11%
			Average	47	11%	34%	10%	37%	51%	67%	13%
			Total	804				Sample values	66%	78%	31%

to make contact with the residence and attempt the interview. The number of completed interviews per event ranged from 17 to 91. The final N for the survey was 804. Our cooperation rates and refusal rates are within typical ranges but the responses rate of 11% is lower than others have typically reported despite being average in comparison to those used reported by many professional data collection services (Pew Research Center 2012). Contact rates (average 37%)[2] to some degree clarify the discrepancy between good cooperation rates (average 34%)[3] and refusal rates (average 10%)[4] but low response rates. We believe that the patterns illustrated reflect the consequences of only calling each number four times over the two-week field period rather than calling each number 10–12 times, as is more typical for phone survey. This choice was made in order to reduce retrospective bias, but did come at the cost of increased

noncontact and a subsequent reduction in response rate. The comparison of our sample to census data for the counties shows that our sample does vary from the census in that we overrepresent females, whites, and those over 65. These results are similar to other surveys completed around the same time before the American Association of Public Opinion Research officially recommended the adoption of cell phone RDD samples (AAPOR 2008). While the sampling approach leaves room for improvement, even with these limitations, these data still represent the best available data on perception and actual human behavioral response to date. In the limitation section, we discuss possible ways to improve on this approach.

c. Dependent variables

To examine potential effects on protective action, we used a dependent variable that asked respondents "Was protective action taken in response to this event?" It was coded as 0 (did nothing) and 1 (took some sort of protective action). Because protective action can include many different types of activities, we also used a question that asked what actions the respondent took. Protective actions were coded into four categories: do nothing, seek

---

[2] Contact rates measure the number of phone numbers that we were able to determine were eligible or ineligible for participation.

[3] Cooperation rates measure the percentage of eligible households that agreed to participate in the survey.

[4] Refusal rates measure the percentage of eligible households that refused to participate.
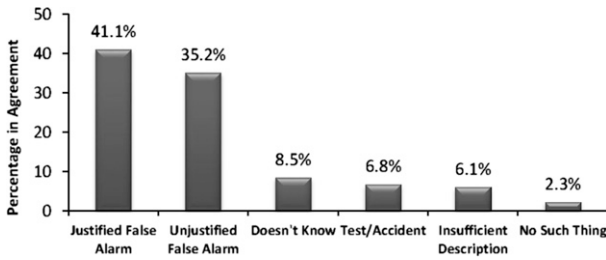
FIG. 2. Respondent definitions of false alarm.

more information, protect property, and shelter. Because of low counts within protecting property, we combined "information seeking" and "protecting property" into one category. The final variable was measured as 0 (did nothing), 1 (seek information and/or protect property), and 2 (shelter).

To further explore what may influence perception of false alarm, we used a dependent variable representing this perception. Respondents were asked to rate the frequency of false alarms in their area on a scale of 1 to 10, with 1 being "never" and 10 being "all the time." To accurately represent the variable while taking into account the low frequencies, particularly at the mid and higher end of the scale, the variable was simplified into 0 (ratings 1–3), 1 (ratings 4–7), and 2 (ratings 8–10).

d. Independent variables

To test hypotheses 1–6, several sociodemographic variables were included in the regression. Age was used as a continuous variable in the regression. Race was simplified into 0 (white) and 1 (nonwhite). Gender was treated as a dichotomy of 0 (male) and 1 (female). To determine the presence of children in the home, we asked if anyone under 18 was living with the respondent. This variable was coded as 0 (no) and 1 (yes). A variable addressing past experiences with tornadoes was included as well. This question asked "How many tornadoes have you experienced in all?" It was simplified into a binary variable of 0 (no experience) and 1 (at least one prior experience). The family emergency plan variable was also included. It was coded as 0 (no family emergency plan present) and 1 (family emergency plan present). Trust in local weather providers was measured on a scale coded 1–5, with 1 being the lowest trust and 5 being the highest trust. In addition to sociodemographic variables, a variable representing false alarm ratio for the sampled county was also used. Data for this variable were downloaded from the NWS performance management website (NOAA/NWS Office of Climate, Water, and Weather Services 1986b). This rate is calculated by dividing the number of false alarms by the total number of warned events by county. A tornado warning would be considered a false alarm if a tornado was not observed. The average FAR for the three years leading up to the event, including the year of the event, was determined for each county within the survey. This was a numeric variable representing the rate as a percentage for each county. We also added a variable for the total average number of tornadoes impacting each county for the three years leading up to the event.

After observing a possible correlation between understanding of the term FA and perception of FAR, we also added dummy variables representing the categorized definitions of FA given by respondents. These dummy variables (unjustified false, justified false alarm, test, insufficient description, do not know, no such thing) are described in more detail in the next section. Justified false alarm was left out as the reference variable for both regressions.

All variables were coded using Statistical Package for the Social Sciences (SPSS). SPSS was also used for the data analysis. Given the sample limits, we chose to also generate correlation matrices in addition to regression models in order to provide a more complete view of the data relationships. Results and discussion of
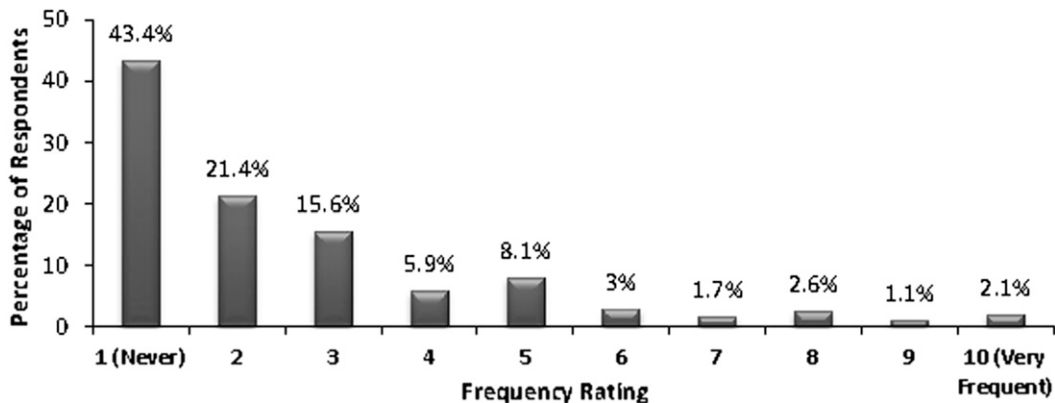

FIG. 3. Respondent rating of false alarm frequency.

TABLE 3. FA perception correlation matrix. Single asterisk indicates correlation is significant at the 0.05 level (two tailed); double asterisk indicates correlation is significant at the 0.01 level (two tailed).

	FA perception	Age	No. of tornadoes	Actual FAR	Trust	Gender	Race	Children
FA Perception	1	−0.045	0.087*	0.075*	−0.168**	−0.017	0.033	0.085
Age	−0.045	1	−0.028	−0.048	0.162**	0.034	−0.166**	−0.475**
No. tornadoes	0.087*	−0.028	1	0.183**	0.057	0.057	0.207**	0.003
Actual FAR	0.075*	−0.048	0.183**	1	−0.122**	−0.089*	0.142**	0.005
Trust	−0.168**	0.162**	0.057	−0.122**	1	0.092*	0.015	−0.119**
Gender	−0.017	0.034	0.057	−0.089*	0.092*	1	0.043	0.079
Race	0.033	−0.166**	0.207**	0.142**	0.015	0.043	1	0.203**
Children	0.085	−0.475**	0.003	0.005	−0.119**	0.079	0.203**	1
Family plan	−0.021	−0.059	−0.034	−0.036	0.034	−0.052	−0.072*	0.073
Past experience	−0.031	−0.053	−0.044	−0.111**	0.002	−0.006	−0.005	0.043
Define: Justified	−0.156**	0.025	0.026	−0.041	−0.026	−0.031	−0.033	−0.005
Define: Not justified	0.152**	0.002	0.014	0.023	0.024	0.058	0.079*	0.036
Define: Insufficient	−0.063	0.040	−0.011	−0.011	0.028	−0.022	0.021	−0.046
Define: Test	0.075	−0.013	−0.007	0.052	−0.042	−0.041	−0.064	−0.037
Define: Does not exist	0.079*	0.016	−0.019	0.011	−0.066	0.018	−0.075*	−0.038
Define: Does not know	−0.043	−0.079*	−0.031	−0.006	0.050	0.005	0.004	0.046

these matrices are included to provide a full picture of the potential relationships between the variables. Some correlations, though unconfirmed by the regressions, may warrant further analysis through refinement and retesting of the variables. Ordinal logistic regressions and a multinomial regression were then performed to explore the relationship between the dependent variables and the independent variables further. Our measured approach to the presentation of survey results and our extended discussion of conclusions, connections to extant literature, limits, and future directions are intended to inform those who would improve on this work.

8. Findings

a. Definition of false alarms

Figure 2 illustrates the patterns we observed in responses to an open-ended question that asked respondents to describe what a false alarm is. Six categories captured the essence of most replies.

The categories that emerged from our analysis are quite meaningful for several reasons. First, the results show that the public does not agree on what this term means. In other words, when we ask people about false alarm, they think of different things. If we are to use FAR as a measure of success within the weather community, then the definition of false alarm must consider how the concept relates to different factors that might impact the public's views on what constitutes a successful or unsuccessful warning. In looking more specifically at the breakdown of frequencies, we also see several other important insights.

The largest majority of respondents (almost half of the respondents at 41.1%) defined false alarms as an event that did not occur as predicted, meaning that the alert, warning, or prediction did not materialize in the way it was described. This could mean a situation in which the "weatherman's" predictions were not perfect and the storm did not produce a tornado. Similarly, it could also mean that a tornado was on the ground, but the storm weakened before it hit the warned area. This category of response falls closest to the actual definition of a false alarm. Respondents recognize there was a reason to alert, but because of changing weather (e.g., "a funnel cloud that never touched down") or imperfect forecasting ("weathermen aren't perfect and [it] goes another way"; "they thought it would form into a tornado but it didn't"; "when Doppler radar says there may [be one] and they call it but it never happens"), the event did not occur in the way it was originally thought. In other words, respondents implied that the alarm was justified or within reason, but turned out to be false.

The next most popular definition we have labeled as unjustified warnings and this category makes up 35.2%. This group believes a false alarm is when the alert, warning, or prediction is made without just cause. Responses tended to vary in the degree of blame but ranged from deliberate falsified information to more accidental and mistaken instances such as someone thinking they saw a tornado when they did not. For example, some respondents defined false alarms as "someone says something is happening and it does not really exist" or "alarm that comes when there's really nothing out there to harm you or your property." Some respondents believe they are being provided false information (e.g., "the

TABLE 3. *Extended*

Family plan	Past experience	Define: Justified	Define: Not justified	Define: Insufficient	Define: Test	Define: Does not exist	Define: Does not Know
−0.021	−0.031	−0.156**	0.152**	−0.063	0.075	0.079*	−0.043
−0.059	−0.053	0.025	0.002	0.040	−0.013	0.016	−0.079*
−0.034	−0.044	0.026	0.014	−0.011	−0.007	−0.019	−0.031
−0.036	−0.111**	−0.041	0.023	−0.011	0.052	0.011	−0.006
0.034	0.002	−0.026	0.024	0.028	−0.042	−0.066	0.050
−0.052	−0.006	−0.031	0.058	−0.022	−0.041	0.018	0.005
−0.072*	−0.005	−0.033	0.079*	0.021	−0.064	−0.075*	0.004
0.073	0.043	−0.005	0.036	−0.046	−0.037	−0.038	0.046
1	0.089*	−0.002	0.034	0.006	−0.041	−0.021	−0.015
0.089*	1	−0.028	0.034	0.032	−0.041	−0.017	0.015
−0.002	−0.028	1	−0.616**	−0.213**	−0.226**	−0.123**	−0.255**
0.034	0.034	−0.616**	1	−0.188**	−0.199**	−0.109**	−0.225**
0.006	0.032	−0.213**	−0.188**	1	−0.069	−0.038	−0.078*
−0.041	−0.041	−0.226**	−0.199**	−0.069	1	−0.040	−0.083*
−0.021	−0.017	−0.123**	−0.109**	−0.038	−0.040	1	−0.045
−0.015	0.015	−0.255**	−0.225**	−0.078*	−0.083*	−0.045	1

evidence that has been given to us is false'') or even being lied to (e.g., ''they lied to us and they don't know what they're talking about''). Some also described false alarm as a situation in which a spotter or someone from the general public thought they saw a tornado or funnel and were actually mistaken (e.g., ''tornado spotter thought they saw a tornado and was wrong'' and ''when someone says they saw a tornado but didn't see one'').

Surprisingly, 8.5% of respondents, the third most common response, did not seem to know what false alarm meant. Most of the time they simply responded ''don't know'' or ''couldn't say,'' but others provided single-word answers and refused to elaborate. Examples in this category included ''no'' or ''depends.'' Other times more elaborate answers were given, such as ''in the next town over they have a whistle that goes off,'' ''when it's over they sound the all clear,'' or ''when you have to go outside.'' While this is a relatively small portion of the sample, it is still important to note that some of the public could not articulate any meaningful response or at the very least did not understand the question.

The fourth most common (6.8%) false alarm definition provided by our respondents was a malfunction of the warning system, such as setting off a warning siren by accident or sounding a siren to test a warning system. For instance, this could mean a weekly test of a siren system. It could also refer to a situation in which a forecaster ''hit a wrong button.'' This definition seems to be most strongly related to scenario in which warnings go out for no reason other than a broken or faulty system. Therefore, this definition is not related to the forecasting process, but it refers to the overall condition and function of the mechanical warning system in place.

The fifth most popular definition (6.1%) simply stated that no threat occurred; the respondent gave no further elaboration or explanation beyond this. In this case, the description of false alarm was insufficient or incomplete. Consequently, these responses were not able to fit into any of the above categories because the respondent gave no indication of cause or reason for the nonoccurring threat (e.g., ''nothing materializes'' or ''didn't happen''). In addition, they do not mention an alert, warning, or prediction in association with the event. Respondents in this category appear to agree that a false alarm means something did not occur, but they were either unable or unwilling to explain why or how this happened.

The final group of respondents is those that do not believe in false alarms, or deny their existence. While they are a small group (2.3%), it is worth noting that a portion of the sample does not believe false alarms exist. For example, one respondent claimed, ''there really is no such thing as a false alarm with tornadoes.'' Therefore, a small segment of our sample believes that all warnings matter regardless of any other elements because they must always have some truth or cause behind them. They may also believe that one should not question the possible threat of a tornado because they are one of the most unpredictable severe weather events.

In thinking across the categories, over 80% of respondents recognized false alarm as a predicted event that did not happen. In other words, four-fifths of our sample had a mostly accurate understanding of the most central element of the concept of false alarm. It is also critical to note that within that group over 75% offered unprompted comments regarding blame and responsibility. One group

made it clear that the warnings were unjustified, while the other group felt that despite a nonevent, an alert was warranted. In addition, many responses in the "no such thing" category imply that all issued warnings are necessary and justified regardless of the outcome; consequently false alarm does not exist. In other words, the public wanted the concept to capture responsibility. This gives credence to the idea that the public understanding of false alarm should be considered further if it is to be used as a metric for forecaster success. It also suggests that the public to some degree wants to distinguish along the continuum of warning Barnes et al. (2006) suggest by distinguishing between true errors/mistakes and missed predictions linked to uncertainty.

Based on these results, we added the following hypotheses to the quantitative analyses:

PH11: Those that define false alarm as a justified mistake will have a lower perception of false alarm occurrence.
PH12: Those that define false alarm as an unjustified mistake will have a higher perception of false alarm occurrence.
PH13: Those that define false alarm as a test will have a higher perception of false alarm occurrence.
PH14: Those that suggest that there is no such thing as a false alarm will have a lower perception of false alarm occurrence.

b. Perception of false alarms analysis

As discussed above, it is important to understand public perceptions of how frequent false alarms happen. In our survey, when respondents were asked "On a scale of one to ten (10 being "very frequent" and 1 being "never"), could you tell us how frequent false alarms are in your area?," the responses were striking and far from the pattern we expected. Figure 3 shows the distribution of respondent ratings. A total of 64.8% rate the frequency as only a 1 or 2. This means that a substantial portion of the sample believes there are very few to no false alarms in its area. Though counterintuitive to many in the weather enterprise, this general pattern has been repeated in two separate studies conducted by our team that measured perceptions of tornado FA occurrence.[5] Despite the large number of low FA frequency ratings, most areas have an NWS false alarm ratio over 50% in actuality. In other words, when asked in a survey, people tend to perceive FAR as much lower than it actually is.

c. Predicting perceptions of false alarm

While the false alarm frequency ratings themselves provide meaningful results, it is also important to understand what factors drive this perception. The literature review above presented a number of independent and control variables, including demographics, past experience, and trust in weather providers. To explore these possible correlations, we generated a correlation matrix shown in Table 3.

The correlation matrix shows that NWS FAR (PH7) and average number of tornadoes (PH8) were positively correlated with perception of FA. In addition, results supported a negative correlation between trust in weather providers and FA perception (PH9). Matrix results suggest a possible link between understanding of the term FA and perception of FA. People who defined FA as a justified mistake were less likely to perceive a high frequency of false alarms (PH11). On the other hand, those who viewed it as an unjustified mistake (PH12) were more likely to have a higher perception. It should be noted for all of these correlations that the strength of association is very weak even for those variables with significance. These findings illustrate the complexity of the process of developing FAR perceptions. The full results are shown in Table 4.

To explore the hypotheses above in more depth, we also ran an ordinal regression using perception as the dependent variable and age, race, gender, past experience, the presence of children under 18, existence of a family emergency plan, trust in local weather providers, county FAR, number of tornadoes, and FA definition categories as independent variables. The results from the regression show that while the NWS FAR for the county did not significantly influence FA perception, trust did. In agreement with our hypothesis (PH9), those who had greater trust in their local weather providers had a lower perception of false alarms ($B = -0.295$, Sig. $= 0.030$). This finding corresponds with other prior work that suggest if people are familiar with their information sources and find them credible and trustworthy, they will probably also see them in an overall positive light. Some significance was found between understanding of the term FA and FA perception. Those defining a false alarm as an unjustified mistake were actually more likely to rate the FA frequency as low, thus forcing us to reject hypothesis PH11. This result is counterintuitive and difficult to interpret because our questions were not initially designed to explore why people may view FAs as misinformation. These results reinforce the idea that perceptions are complex and further targeted research is needed to fully grasp respondent understanding of FA and its impact. None of the other variables were significant in this model, so

---

[5] It should be noted that both studies were post event quick response surveys and a recent significant event may be impacting perceptions.

TABLE 4. False alarm perception regression parameters.

		Estimate	Std. Error	Sig.
Threshold	Medium (4–7)[a,b]	−0.096	0.894	0.915
	High (8–10)	1.827	0.908	0.048
Location	Age	−0.004	0.008	0.635
	County tornado no.	0.012	0.009	0.221
	County tornado FAR	0.006	0.007	0.416
	Trust	−0.295[c]	0.136	0.030
	Male	0.275	0.248	0.271
	White	0.112	0.307	0.714
	No children	−0.361	0.268	0.177
	No family plan	0.257	0.237	0.279
	No past experience	−0.240	0.245	0.326
	Define: Unjustified[d]	−0.843[e]	0.284	0.003
	Define: Insufficient	0.472	0.417	0.257
	Define: Test	−0.902	0.519	0.082
	Define: Does not know	−0.069	0.436	0.875

[a] Chi-square value of 30.839; sig. of 0.004.
[b] Psuedo-R-square—Cox and Snell, 0.071; McFadden, 0.054; Nagelkerke, 0.095.
[c] Coefficient is significant at the 0.05 level.
[d] Reference category is the "justified FA" definition.
[e] Coefficient is significant at the 0.01 level.

hypotheses PH1–PH7 were rejected. Table 5 below provides a review of our findings relative to our hypotheses and prior research findings.

d. Protective action analysis

As discussed above, the ultimate goal of warning systems is to make people safe. As a result, it is important to improve our knowledge of protective action behavior associated with tornadoes. As discussed in section 4, it is intuitive to believe that higher false alarm rates would lead to a decrease in protective action taken, much like the story about the boy who cried wolf. On the other hand, there is lack of empirical support for this assertion. To explore this possible connection, we created a correlation matrix shown in Table 6.

Table 6 offers support for several hypotheses. Most notable, NWS tornado FAR (BH7) has a negative impact on sheltering behavior. The correlation matrix also generated several other important results. As shown in previous research, women (BH1) and those with children (BH6) are more likely to shelter and those with a family plan (BH5) are less likely to do nothing. On the other hand, those with more tornadoes in their area (BH8) are less likely to shelter and those that define false alarm as a test are more likely to do nothing. As with the prior correlation matrix, it should again be noted that all of these associations are very weak.

The logistic regression shown in Table 7 explores these results further. Because the perception of FARs does not mirror the actual FAR in the analysis above, we also added a variable representing FA perception. All other variables are the same as those used in the previous regression. Three of the variables were significant in the above model; county FAR, presence of children, and existence of a family emergency plan. In agreement with BH10, the likelihood of taking some sort of

TABLE 5. False alarm perception model.

	Prior findings	Empirical foundation	Current study	
			Correlations	Regression
PH1: Females will have a lower perception of false alarms.	Supported	Barnes (2006)	Rejected	Rejected
PH2: Elderly will have a lower perception of false alarms.	Supported	Barnes (2006)	Rejected	Rejected
PH3: Nonwhites will have a higher perception of false alarm.	—	—	Rejected	Rejected
PH4: Tornado experience will reduce perception of false alarms.	—	—	Rejected	Rejected
PH5: Families with a disaster plan will perceive fewer false alarms.	—	—	Rejected	Rejected
PH6: Families with children will perceive fewer false alarms.	—	—	Rejected	Rejected
PH7: As FAR Increases perception of FAR will increase.	—	—	Supported	Rejected
PH8: Residents of counties with a higher average number of tornadoes will have an increased perception of false alarms.	—	—	Supported	Rejected
PH9: Greater trust in local weather providers will result in decreased perceptions of false alarms.	Supported	Baker (1984, 1987); Cola (1996); Lindell and Perry (1987); Ripberger et al. (2015)	Supported	Supported
PH11: Those that define false alarm as an justified mistake will have a lower perception of false alarm.	—	—	Supported	Rejected
PH12: Those that define false alarm as an unjustified mistake will have a higher perception of false alarm.	—	—	Supported	Rejected
PH13: Those that define false alarm as a test will have a higher perception of false alarm.	—	—	Rejected	Supported
PH14: Those that suggest that there is no such thing as a false alarm will have a lower perception of false alarm.	—	—	Rejected	N/A

TABLE 6. Protective action correlation matrix. Single asterisk indicates correlation is significant at the 0.05 level (two tailed); double asterisk indicates correlation is significant at the 0.01 level (two tailed).

	Action: Do nothing	Protect prop/seek	Action: Shelter	FA perception	Age	No. of tornadoes	Actual FAR	Trust	Gender
Action: Do nothing	1	−0.407**	−0.331**	0.041	0.062	0.069	0.171**	−0.057	−0.027
Action: Protect property/seek info	−0.407**	1	−0.727**	−0.005	−0.037	0.038	0.065	0.022	−0.087*
Action: Shelter	−0.331**	−0.727**	1	−0.026	−0.008	−0.091*	−0.196**	0.021	0.110**
FA perception	0.041	−0.005	−0.026	1	−0.045	0.087*	0.075*	−0.168**	−0.017
Age	0.062	−0.037	−0.008	−0.045	1	−0.028	−0.048	0.162**	0.034
No. of Tornadoes	0.069	0.038	−0.091*	0.087*	−0.028	1	0.183**	0.057	0.057
Actual FAR	0.171**	0.065	−0.196**	0.075*	−0.048	0.183**	1	−0.122**	−0.089*
Trust	−0.057	0.022	0.021	−0.168**	0.162**	0.057	−0.122**	1	0.092*
Gender	−0.027	−0.087*	0.110**	−0.017	0.034	0.057	−0.089*	0.092*	1
Race	0.012	−0.029	0.021	0.033	−0.166**	0.207**	0.142**	0.015	0.043
Children	−0.069	−0.074	0.125*	0.085	−0.475**	0.003	0.005	−0.119**	0.079
Family plan	−0.145**	0.069	0.038	−0.021	−0.059	−0.034	−0.036	0.034	−0.052
Past experience	−0.028	−0.035	0.057	−0.031	−0.053	−0.044	−0.111**	0.002	−0.006
Define: Justified	−0.044	0.039	−0.007	−0.156**	0.025	0.026	−0.041	−0.026	−0.031
Define: Unjustified	−0.043	0.059	−0.029	0.152**	0.002	0.014	0.023	0.024	0.058
Define: Insufficient	0.043	−0.018	−0.013	−0.063	0.040	−0.011	−0.011	0.028	−0.022
Define: Test	0.088**	−0.085	0.021	0.075	−0.013	−0.007	0.052	−0.042	−0.041
Define: Does not exist	0.032	−0.006	−0.018	0.079*	0.016	−0.019	0.011	−0.066	0.018
Define: Does not know	0.017	−0.068	0.058	−0.043	−0.079*	−0.031	−0.006	0.050	0.005

protective action decreases as the county tornado FAR increases ($B = -0.020$, Sig. $= 0.003$). This suggests that people are statistically less likely to take action in areas with more false alarms, when controlling for relevant sociodemographics, false alarm perception, and number of tornadoes. To our knowledge, this is the first quantitative empirical evidence of a direct link between FAR and protective behaviors. However, we encourage significant caution in the interpretation of this result. There is still a need for further corroboration and investigation. In agreement with past literature, lack of children (BH6; $B = -0.882$, Sig. $= 0.001$) and lack of a family emergency plan (BH5; $B = -0.770$, Sig. $= 0.001$) also lead to a decrease in protective action behavior. No other variables were significant in this model. It is also important to recognize that protective action can mean many different things. For this reason, we also used a more complex version of the protective action variable: do nothing, information seeking/protecting property, and sheltering. Table 8 shows the results of a second logistic regression with the complex protective action variable as the dependent variable.

This version of the analysis shows that respondents in counties with higher FARs are less likely to seek more information and/or protect their property ($B = -0.027$, Sig. $= 0.017$). Similarly, they are also less likely to take shelter ($B = -0.049$, Sig. $= 0.000$), findings that support the hypotheses that high FAR reduces protective action (BH7). The only other significant variable within this model was existence of a family plan (BH5). Again, our results agreed with past literature; lack of a family emergency plan decreases the likelihood of information seeking/protecting property (BH5; $B = -0.797$, Sig. $= 0.043$) and sheltering ($B = -1.033$, Sig. $= 0.012$). Table 9 below provides a review of our findings relative to our hypotheses and prior research.

9. Discussion and conclusions

a. Understanding of false alarm

This study has provided a more in-depth look at public understanding, perceptions, and actions associated with the concept of false alarm. After categorizing respondent definitions of false alarm, significant policy implications become apparent. First, over 80% of the respondents accurately understood the concept of false alarm. While they may not have used the exact wording or terminology, they showed an understanding of the idea that an alert was given, but did not occur as anticipated. On the other hand, there was also a substantial portion of the sample that believed false alarm meant the threat never existed at all. This definition is quite alarming given they believe the alerts were unjustified, misinformed, or even outright lies. This understanding of false alarm could breed animosity toward those providing weather information. It could also lead to mistrust and skepticism in future events. These findings give support to the notion that the distinction between mistakes and chance make a difference to people, thus

TABLE 6. *Extended*

Race	Children	Family plan	Past experience	Define: Justified	Define: Unjustified	Define: Insufficient	Define: Test	Does not exist
0.012	−0.069	−0.145**	−0.028	−0.044	−0.043	0.043	0.088*	0.032
−0.029	−0.074	0.069	−0.035	0.039	0.059	−0.018	−0.085	−0.006
0.021	0.125*	0.038	0.057	−0.007	−0.029	−0.013	0.021	−0.018
0.033	0.085	−0.021	−0.031	−0.156**	0.152**	−0.063	0.075	0.079*
−0.166**	−0.475**	−0.059	−0.053	0.025	0.002	0.040	−0.013	0.016
0.207**	0.003	−0.034	−0.044	0.026	0.014	−0.011	−0.007	−0.019
0.142**	0.005	−0.036	−0.111***	−0.041	0.023	−0.011	0.052	0.011
0.015	−0.119**	0.034	0.002	−0.026	0.024	0.028	−0.042	−0.066
0.043	0.079	−0.052	−0.006	−0.031	0.058	−0.022	−0.041	0.018
1	0.203**	−0.072*	−0.005	−0.033	0.079*	0.021	−0.064	−0.075*
0.203**	1	0.073	0.043	−0.005	0.036	−0.046	−0.037	−0.038
−0.072*	0.073	1	0.089*	−0.002	0.034	0.006	−0.041	−0.021
−0.005	0.043	0.089*	1	−0.028	0.034	0.032	−0.041	−0.017
−0.033	−0.005	−0.002	−0.028	1	−0.616**	−0.213**	−0.226**	−0.123**
0.079*	0.036	0.034	0.034	−0.616**	1	−0.188**	−0.199**	−0.109**
0.021	−0.046	0.006	0.032	−0.213**	−0.188**	1	−0.069	−0.038
−0.064	−0.037	−0.041	−0.041	−0.226**	−0.199**	−0.069	1	−0.040
−0.075*	−0.038	−0.021	−0.017	−0.123**	−0.109**	−0.038	−0.040	1
0.004	0.046	−0.015	0.015	−0.255**	−0.225**	−0.078*	−0.083*	−0.045

providing support for the idea of a continuum of warning (Barnes et al. 2006). Further, our analyses of perception and protective action suggest that these distinctions have effects on what people think and do. Our results demonstrate the need for further investigation into the complexities of public understanding of false alarm. If the public does not see false alarm as a simple ''hit'' or ''miss,'' perhaps the terminology is too constraining to accurately represent the public's understanding of such events. With less rigid definitions of the performance measures of false alarm, hit, and miss, the public may feel less of a need to place blame and instead a greater understanding of the real uncertainty associated with forecasting.

b. Perception of false alarm

It is important to note that over 60% of respondents in our sample believe that there are very few to no false alarms in their area. Contrary to anecdotal evidence, many of the people in our sample felt that there were not many false alarms in their area. It is equally as important that the actual FARs have no statistically significant influence on people's perception of false alarms. In other words, actual FAR numbers alone cannot accurately represent how the public judges forecasting success. This is supported by the significance of trust as a variable. The more faith people have in their local weather providers, the lower they tend to rate false alarm frequencies. In other words, if they like them and believe them as providing a good service, they are less inclined to judge them harshly. In addition, the significance

of trust has already been supported by several past studies (Baker 1984, 1987; Cola 1996; Lindell and Perry 1987). It is just as important to note that the sociodemographic variables were not significant. Past research has shown that women and elderly have a lower perception of false alarms (Barnes 2006; Riad and Norris 1998), while nonwhites had a higher perception (Flynn et al. 1994; Lindell et al. 1980). We were not able to support these conclusions. This suggests that future work must be done in order to determine the true effect these variables have on FA perception.

Our results are also limited by the fact that we do not know our respondent's sensitivity to false alarms; as discussed in the previous section, we do not know what situation they see as a hit and what they see as a miss. This can drastically influence how they may rate false alarms in their areas. Other relevant questions may be, ''How long after a false alarm is there an influence on perception and action?''; ''Do multiple subsequent false alarms have complex effects on perception?''; and ''What does a hit between misses do to perception of false alarms?''

c. Protective action

In general, our findings suggest that people do not really feel there are many false alarms in their area, regardless of the NWS FAR. In addition, their perception of false alarm frequency was not a significant predictor of their protective action. On the other hand, the NWS FAR does indeed appear to impact their decision

TABLE 7. Simple protective action regression parameters.

Did you take any actions to protect yourself, your family, or your property from this event?[a,b,c]		B	Std. error	Sig.
Yes	Intercept	2.141	1.008	0.034
	Age	−0.009	0.008	0.254
	County tornado no.	−0.002	0.010	0.799
	County tornado FAR	−0.020[d]	0.007	0.003
	FA frequency perception	0.165	0.215	0.441
	Trust	0.266	0.145	0.066
	Male	−0.279	0.247	0.258
	White	−0.348	0.329	0.291
	No children	−0.882[e]	0.275	0.001
	No family plan	−0.770[e]	0.237	0.001
	No past experience	−0.221	0.237	0.351
	Define: Unjustified[f]	0.368	0.269	0.172
	Define: Insufficient	−0.484	0.493	0.327
	Define: Test	0.901	0.492	0.067
	Define: Does not exist	−0.112	0.719	0.876
	Define: Does not know	−0.145	0.455	0.751

[a] Chi-square value of 56.156; sig. of 0.000.
[b] Psuedo-R-square—Cox and Snell, 0.142; McFadden, 0.113; Nagelkerke, 0.192.
[c] Reference category is "No."
[d] Coefficient is significant at the 0.05 level.
[e] Coefficient is significant at the 0.001 level.
[f] Reference category is the "justified FA" definition.

to take action. Not only are people in areas with higher FARs more likely to take no action upon hearing a warning, but in particular less likely to shelter. This result provides some evidence that even though they may not accurately perceive the false alarm rate, the actual incidence of false alarms may be influencing their behaviors. Thus, while there is merit to working toward lower false alarm rates, there are still parts of the causal puzzle missing.

Considering that people's definition of false alarm made an impact on their perception of FAR, it is intuitive to assume that these variables may also factor into protective action decision-making. Strangely enough, these were not significant predictors of protective action. We did find a negative correlation between believing a false alarm was a test of the system and taking action, but this was not confirmed in the regression. There is some merit to this idea though. These people are not seeing false alarm as a forecaster inaccuracy or mistake, but as a legitimate test or malfunction of the system, so FAR should not impact their actions. While none of the other variations on false alarm definition were significant, it is not conclusive proof that public conceptualization of false alarm is not an important factor. Future work is needed to determine just how people conceptualize this term. While asking for a definition provides some insight into the public's understanding of false alarm, it cannot fully convey the complexities involved in this understanding. Other relevant questions could include, "What do you

consider a well forecasted event?"; "Would you prefer less missed events even if that meant more false alarms?"; and "Would you consider a tornado that does not impact your community a false alarm?"

It has long been known that a variety of factors go into protective action decision-making. Our findings in regards to other relevant variables also merit discussion. In contrast to previous research (Perry and Lindell 1997; Lindell et al. 1980), we did not find age or race to be significant predictors of protective action. On the other hand, our correlation matrix did show that women were more likely to take shelter, but interestingly less likely to protect property/seek more information (Comstock and Mallonee 2005; Bateman and Edwards 2002; O'Brien and Atchison 1998; Fothergill 1996; Lerner et al. 2003). Perhaps they feel the best course of action is to shelter immediately rather than preparation or information-seeking activities. Our findings also agreed with past research in that those with a plan were indeed more likely to take shelter and protect property/seek more information (Balluz et al. 2000; Blanchard-Boehm and Cook 2004; Nagele and Trainor 2012). Additionally, families with children have been shown to be more likely to take action (Edwards 1993; Carter et al. 1983; Turner et al. 1981; Flynn 1979; Wilkinson and Ross 1970). We could not conclude that they are more likely to take shelter or protect property/ seek more information in particular, but we could agree that they are less likely to do nothing at all. Trust in weather providers is perhaps the most intuitive factor, and

TABLE 8. Complex protective action regression parameters.

What type of protective action did you take?[a,b,c]		*B*	Std. error	Sig.
Increase awareness/protect property	Intercept	4.323	1.744	0.013
	Age	−0.004	0.014	0.773
	County tornado no.	−0.001	0.014	0.947
	County tornado FAR	−0.027[d]	0.011	0.017
	FA frequency perception	−0.011	0.336	0.974
	Trust	−0.037	0.232	0.875
	Male	0.094	0.396	0.811
	White	−0.504	0.578	0.383
	No children	0.102	0.444	0.818
	No family plan	−0.797[d]	0.394	0.043
	No past experience	0.260	0.387	0.501
	Define: Unjustified	−0.119	0.415	0.774
	Define: Insufficient	0.499	0.847	0.556
	Define: Test	1.391	1.084	0.199
	Define: Does not exist	−0.413	1.218	0.735
	Define: Does not know	0.950	0.821	0.247
Shelter	Intercept	5.910	1.825	0.001
	Age	0.006	0.015	0.670
	County tornado no.	−0.003	0.016	0.852
	County tornado FAR	−0.049[e]	0.012	0.000
	FA frequency perception	0.134	0.353	0.703
	Trust	−0.073	0.245	0.765
	Male	−0.436	0.424	0.304
	White	−0.986	0.593	0.096
	No children	−0.371	0.465	0.426
	No family plan	−1.033[d]	0.413	0.012
	No past experience	0.010	0.410	0.981
	Define: Unjustified[f]	0.212	0.438	0.628
	Define: Insufficient	0.267	0.894	0.766
	Define: Test	0.846	1.133	0.455
	Define: Does not exist	0.056	1.211	0.963
	Define: Does not know	0.411	0.891	0.645

[a] Chi-square value of 47.343; sig. of 0.023.
[b] Psuedo-*R*-square—Cox and Snell, 0.138; McFadden, 0.076; Nagelkerke, 0.161.
[c] Reference category is "do nothing."
[d] Coefficient is significant at the 0.05 level.
[e] Coefficient is significant at the 0.001 level.
[f] Reference category is the "justified FA" definition.

indeed it has been shown to influence protective action (Perry 1982; Cutter 1987; Gruntfest 1997).

In conclusion, our findings provide a greater depth of understanding of how false alarm interacts with protective actions. Most importantly, we have reported some evidence that 1) to some degree, people associate the term false alarm with blame suggesting a need to revisit the continuum of warning approach; 2) perceptions of false alarm are not driven by false alarm rates, but can to some degree be explained by other variables; and 3) in addition to confirming some prior findings, our work provides the first evidence of a false alarm effect. Finally, this study has shown that even simple concepts like false alarm are significantly more complex than they appear, and good policy needs extensive, detailed analysis to understand these phenomenon and in turn their implications.

10. Limitations and future directions

As discussed in the body of this analysis, while our approach provides important insights into the false alarm phenomenon, there are a number of important shortcomings that provide ample opportunity for future researchers to improve on the approach taken here. The first opportunity comes from the low response rates in our sample. While we made a deliberate decision to focus on a short window of time after events, it is possible that through larger staffing, use of incentives, and/or a reduced/focused instrument response rates could be improved and future researchers could test the patterns observed here. Second, a project focused exclusively on false alarms could also tackle additional theoretical explanations for perception and behavioral changes. As noted in the analyses,

TABLE 9. False alarm behavior model.

	Prior findings	Empirical foundation	Current study		
			Correlations	Simple regression	Complex regression
BH1: Females will be more likely to take protective actions.	Supported	Flynn and Chalmers (1980); O'Brien and Atchison (1998); Riad and Norris (1998); Bateman and Edwards (2002); Barnes (2006); Comstock and Mallonee (2005)	Supported	Rejected	Rejected
	Rejected	Nagele and Trainor (2012)			
BH2: Elderly will be more likely to take protective actions.	Supported	Cutter and Barnes (1982); Baker (1987); Aguirre et al. (1998); Barnes (2006); Riad and Norris (1998);	Rejected	Rejected	Rejected
	Rejected	Edwards (1993); Baker (1979); Mileti and Darlington (1997); Mack and Baker (1961); Hodge et al. (1981); Gruntfest (1977, 1997)			
BH3: Nonwhites will be less likely to take protective action.	Supported	Flynn et al. (1994); Lindell et al. (1980); Drabek and Boggs (1968); Perry et al. (1981);	Rejected	Rejected	Rejected
	Rejected	Riad and Norris (1998)			
BH4: Increases in tornado experience will increase the likelihood of taking protective action.	Supported	Hodler (1982); Mileti and O'Brien (1992); Landry and Rogers (1982); Lachman et al. (1961)	Rejected	Rejected	Rejected
BH5: Existence of a family plan will increase the likelihood of taking protective action.	Supported	Balluz et al. (2000); Blanchard-Boehm and Cook (2004); Nagele and Trainor (2012)	Supported	Supported	More likely to shelter, more likely to protect/seek information
BH6: Families with children will be more likely to take protective action.	Supported	Riad and Norris (1998); Houts et al. (1984); Drabek and Stephenson (1971); Edwards (1993); Carter et al. (1983) Turner et al. (1981); Flynn (1979); Wilkinson and Ross (1970)	Supported	Supported	Rejected
BH7: Increased FARs will result in decrease protective action.	—	—	Supported	Supported	Less likely to shelter/seek information, less likely to shelter
BH8: Residents of counties with a higher average number of tornadoes will have an increased perception of false alarms.	—	—	Supported	Rejected	Rejected
BH9: Greater trust in local weather providers will result in decreased perceptions of false alarms.	Supported	Baker (1984, 1987); Cola (1996); Lindell and Perry (1987)	Rejected	Rejected	Rejected
BH10: As the perception of false alarm increases, the likelihood of taking protective action decreases.	Supported	Cola (1996); Simmons and Sutter (2009); Ripberger et al. (2015)	Rejected	Rejected	Rejected
	Rejected	Dow and Cutter (1998); Benight et al. (2004); Atwood and Major (1998)			
BH11: Those that define false alarm as a justified mistake will have a lower perception of false alarm.	—	—	Rejected	Rejected	Rejected
H12: Those that define false alarm as an unjustified mistake will have a higher perception of false alarm.	—	—	Rejected	Rejected	Rejected
H13: Those that define false alarm as a test will have a higher perception of false alarm.	—	—	Supported	Rejected	Rejected
H14: Those that suggest that there is no such thing as a false alarm will have a lower perception of false alarm.	—	—	Rejected	Rejected	Rejected

many of the factors we expected to predict these dependent variables were not significant, and those that were showed weak correlations. This is important because these measures are the ones that are most often cited as explaining this phenomenon. It is clear that the link between NWS FAR, perception, and behavior is quite complicated. Among other important variables one might consider the impact of spatial specificity and personalization of warnings. Third, research should consider the temporal sequencing of false alarms, hits, and misses. The data we collected for this project came after major events as described in the analysis respondents reported low perceptions of false alarm. It is important to explore the degree to which this is connected to the events. A longitudinal analysis of warnings and perception would be uniquely situated to accomplish this and would allow for pre-event and postevent comparisons. A final possibility would be to include a suite of risk perception measures, in particular affective measures, which the current dataset did not include. There is much to be done to understand the false alarm effect.

REFERENCES

AAPOR, 2008: Cell phone task force report. American Association of Public Opinion Research, 60 pp.

Aguirre, B. E., D. Wenger, and G. Vigo, 1998: A test of emergent norm theory of collective behavior. *Sociol. Forum*, **13**, 301–320, doi:10.1023/A:1022145900928.

Andra, D. L., E. M. Quoetone, and W. F. Bunting, 2002: Warning decision making: The relative roles of conceptual models, technology, strategy, and forecaster expertise on 3 May 1999. *Wea. Forecasting*, **17**, 559–566, doi:10.1175/1520-0434(2002)017<0559:WDMTRR>2.0.CO;2.

Atwood, L. E., and A. M. Major, 1998: Exploring the "cry wolf" hypothesis. *Int. J. Mass Emerg. Disasters*, **16**, 279–302.

Baker, E. J., 1979: Predicting response to hurricane warnings: A reanalysis of data from four studies. *Mass Emerg.*, **4**, 9–24.

——, 1984: Public response to hurricane probability forecasts. NWS Tech. Rep., 35 pp.

——, 1987: Warning and evacuation in Hurricanes Elena and Kate. Tech. Rep., Dept. of Geography, Florida State University.

Balluz, L., L. Schieve, T. Holmes, S. Kiezak, and J. Malilay, 2000: Predictors for people's response to a tornado warning: Arkansas, 1 March 1997. *Disasters*, **24**, 71–77, doi:10.1111/1467-7717.00132.

Barnes, L. R., 2006: False alarms: Warning project research findings and warning accuracy conceptual model. WAS*IS Presentation. [Available online at http://www.sip.ucar.edu/wasis/boulder06/ppt/False%20alarms%20-%20Barnes.ppt.]

——, C. C. Benight, E. C. Gruntfest, M. H. Hayden, and D. M. Schultz, 2006: False alarms and close calls: A conceptual

model of warning accuracy. *Wea. Forecasting*, **22**, 1140–1147, doi:10.1175/WAF1031.1.

Bateman, J. M., and B. Edwards, 2002: Gender and evacuation: A closer look at why women are more likely to evacuate for hurricanes. *Nat. Hazards Rev.*, **3**, 107–117, doi:10.1061/(ASCE)1527-6988(2002)3:3(107).

Benight, C., E. Gruntfest, and K. Sparks, 2004: Colorado Wildfires 2002. Quick Response Rep. 167, Natural Hazards Center, University of Colorado Boulder, 9 pp. [Available online at http://www.colorado.edu/hazards/research/qr/qr167/qr167.html.]

Blanchard-Boehm, R. D., 1998: Understanding public response to increased risk from natural hazards: Application of the hazards risk communication framework. *Int. J. Mass Emerg. Disasters*, **16**, 247–278.

——, and M. J. Cook, 2004: Risk communication and public education in Edmonton, Alberta, Canada on the 10th anniversary of the "Black Friday" tornado. *Int. Res. Geogr. Environ. Educ.*, **13**, 38–54, doi:10.1080/10382040408668791.

Boyatzis, R. E., 1998: *Transforming Qualitative Information: Thematic Analysis and Code Development.* SAGE Publications, 200 pp.

Breznitz, S., 1984: *Cry Wolf: The Psychology of False Alarms.* Lawrence Erlbaum Associates, 280 pp.

Carter, M. T., S. Kendall, and J. P. Clark, 1983: Household response to warnings. *Int. J. Mass Emerg. Disasters*, **9**, 94–104.

Clarke, L., and J. F. Short Jr., 1993: Social organization and risk: Some current controversies. *Annu. Rev. Sociol.*, **19**, 375–399, doi:10.1146/annurev.so.19.080193.002111.

Cola, R. M., 1996: Responses of Pampanga households to lahar warnings: Lessons from two villages in the Pasig-Potrero River watershed. *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines*, C. G. Newhall and R. S. Punongbayan, Eds., University of Washington Press, 141–149.

Comstock, R. D., and S. Mallonee, 2005: Comparing reactions to two severe tornadoes in one Oklahoma community. *Disasters*, **29**, 277–287, doi:10.1111/j.0361-3666.2005.00291.x.

Corbin, J., and A. Strauss, 2008: *Basics of Qualitative Research.* 3rd ed. Sage Publications, 379 pp.

Cuthbertson, B. H., and J. M. Nigg, 1987: Technological disaster and the nontherapeutic community: A question of true victimization. *Environ. Behavior*, **19**, 462–483, doi:10.1177/0013916587194004.

Cutter, S. L., 1987: Airborne toxic releases: Are communities prepared? *Environment*, **29**, 12–31, doi:10.1080/00139157.1987.9931330.

——, and K. Barnes, 1982: Evacuation behavior and Three Mile Island. *Disasters*, **6**, 116–124, doi:10.1111/j.1467-7717.1982.tb00765.x.

Dow, K., and S. L. Cutter, 1998: Crying wolf: Repeat responses to hurricane evacuation orders. *Coast. Manage.*, **26**, 237–252, doi:10.1080/08920759809362356.

Drabek, T. E., and K. S. Boggs, 1968: Families in disaster: Reactions and relatives. *J. Marriage Fam.*, **30**, 443–451, doi:10.2307/349914.

——, and J. S. Stephenson III, 1971: When disaster strikes. *J. Appl. Soc. Psychol.*, **1**, 187–203, doi:10.1111/j.1559-1816.1971.tb00362.x.

Edwards, M. L., 1993: Social location and self protective behavior: Implications for earthquake preparedness. *Int. J. Mass Emerg. Disasters*, **11**, 293–303.

Flynn, C. B., 1979: *Three Mile Island Telephone Survey: Preliminary Report on Procedures and Findings.* University of Michigan, 100 pp.

——, and J. A. Chalmers, 1980: The social and economic effects of the accident at Three Mile Island. U.S. Nuclear Regulatory Commission Rep. NUREG/CR-1215, 99 pp.

Flynn, J., P. Slovic, and C. K. Mertz, 1994: Gender, race, and perception of environmental health risks. *Risk Anal.*, **14**, 1101–1108, doi:10.1111/j.1539-6924.1994.tb00082.x.

Fothergill, A., 1996: Gender, risk, and disaster. *Int. J. Mass Emerg. Disasters*, **14**, 33–56.

Gruntfest, E., 1977: What people did during the Big Thompson Flood. Working Paper 32, University of Colorado, 62 pp.

——, 1997: Warning dissemination and response with short lead times. *Flood Hazard Management: British and International Perspectives*, J. Handmer, Ed., GEO Books, 191–202.

Hodge, D., V. Sharp, and M. Marts, 1981: Contemporary responses to volcanisim: Case studies from the Cascades and Hawaii. *Volcanic Activity and Human Ecology*, P. D. Sheets and D. K. Grayson, Eds., Academic New York, 221–248.

Hodler, T. W., 1982: Residents' preparedness and response to the Kalamazoo tornado. *Disasters*, **6**, 44–49, doi:10.1111/j.1467-7717.1982.tb00743.x.

Houts, P. S., M. K. Lindell, T. W. Hu, P. D. Cleary, G. Tokuhata, and C. B. Flynn, 1984: The protective action decision model applied to evacuation during the Three Mile Island crisis. *Int. J. Mass Emerg. Disasters*, **2**, 27–39.

Lachman, R., M. Tatsuoka, and W. Bonk, 1961: Human behavior during the tsunami of 1960. *Science*, **133**, 1405–1409, doi:10.1126/science.133.3462.1405.

Landry, T., and G. Rogers, 1982: Warning confirmation and dissemination. Rep., Center for Social and Urban Research, University of Pittsburgh.

Lerner, J. S., R. M. Gonzalez, D. A. Small, and B. Fischhoff, 2003: Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychol. Sci.*, **14**, 144–150, doi:10.1111/1467-9280.01433.

Lindell, M. K., and R. W. Perry, 1987: Warning mechanisms in emergency response systems. *Int. J. Mass Emerg. Disasters*, **5**, 137–153.

——, ——, and M. R. Greene, 1980: Race and disaster warning response. Research Rep., Battelle Human Affairs Research Centers, 13 pp.

Mack, R. W., and G. W. Baker, 1961: *The Occasion Instant: The Structure of Social Response to Unanticipated Air Raid Warnings*. National Research Council, 69 pp.

Mileti, D. S., and J. H. Sorenson, 1990: Communication of emergency public warnings: A social science perspective and state of-the-art assessment. Oak Ridge National Rep. ORNL-6609, 145 pp.

——, and P. W. O'Brien, 1992: Warnings during disaster: Normalizing communicated risk. *Soc. Probl.*, **39**, 40–57, doi:10.2307/3096912.

——, and J. Darlington, 1997: The role of searching in shaping reactions to earthquake risk information. *Soc. Probl.*, **44**, 89–103, doi:10.2307/3096875.

Nagele, D., and J. E. Trainor, 2012: Geographic specificity, tornadoes, and protective action. *Wea. Climate Soc.*, **4**, 145–155, doi:10.1175/WCAS-D-11-00047.1.

National Weather Service, 2011: National Weather Service Instruction 10-1601. Accessed 17 August 2015. [Available online at http://www.nws.noaa.gov/directives/sym/pd01016001curr.pdf.]

NOAA/NWS Office of Climate, Water, and Weather Services, 1986a: Verification: Severe weather (updated daily). NWS Performance Branch Verification Program. Subset used: January 2009–January 2014, accessed 1 September 2014. [Available online at https://verification.nws.noaa.gov/content/pm/verif/index.aspx.]

——, 1986b: Verification: Severe weather (updated daily) and interactive product database (updated daily). NWS Performance Branch Verification Program. Subset used: October 2007–December 2010, accessed 12 December 2012. [Available online at https://verification.nws.noaa.gov/content/pm/verif/index.aspx.]

O'Brien, P. W., and P. Atchison, 1998: Gender differentiation and aftershock warning response. *The Gendered Terrain of Disaster: Through Women's Eyes*, E. Enarson and B. H. Morrow, Ed., Greenwood Publishing Group, 173–180.

Perry, R. W., 1982: *The Social Psychology of Civil Defense.* Lexington Books, 127 pp.

——, and M. K. Lindell, 1997: Aged citizens in the warning phase of disasters: Re-examining the evidence. *Int. J. Aging Hum. Dev.*, **44**, 257–267, doi:10.2190/RT3X-6MEJ-24AQ-03PT.

——, ——, and M. R. Green, 1981: *Evacuation Planning in Emergency Management*. D. C. Heath, 201 pp.

Pew Research Center, 2012: Assessing the representativeness of public opinion surveys. Accessed 17 August 2015. [Available online at http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/.]

Riad, J. K., and F. H. Norris, 1998: Hurricane threat and evacuation intentions: An analysis of risk perception, preparedness, social influence and resources. Preliminary Paper 271, Disaster Research Center, University of Delaware, 34 pp.

Ripberger, J. T., 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**, 44–56, doi:10.1111/risa.12262.

Schultz, D. M., E. C. Gruntfest, M. H. Hayden, C. C. Benight, S. Drobot, and L. R. Barnes, 2010: Decision making by Austin, Texas, residents in hypothetical tornado scenarios. *Wea. Climate Soc.*, **2**, 249–254, doi:10.1175/2010WCAS1067.1.

Simmons, K. M., and D. Sutter, 2006: Improvements in tornado warnings and tornado casualties. *Int. J. Mass Emerg. Disasters*, **24**, 351–369.

——, and ——, 2007: The Groundhog Day Florida tornadoes: A case study of high-vulnerability tornadoes. Quick Response Rep. 193, Natural Hazards Center, 9 pp. [Available online at http://www.colorado.edu/hazards/research/qr/qr193/qr193.html.]

——, and ——, 2009: False alarms tornado warnings, and tornado casualties. *Wea. Climate Soc.*, **1**, 38–53, doi:10.1175/2009WCAS1005.1.

——, and ——, 2011: *Economic and Societal Impacts of Tornadoes.* Amer. Meteor. Soc., 282 pp.

Tierney, K. J., 1999: Towards a critical sociology of risk. *Sociol. Forum*, **14**, 215–242.

Turner, B. A., 1976: The organizational and interorganizational development of disasters. *Adm. Sci. Quart.*, **21**, 378–397, doi:10.2307/2391850.

Turner, R. H., J. M. Nigg, D. H. Paz, and B. S. Young, 1981: *Community Response to Earthquake Threat in Southern California: Part 10, Summary and Recommendations*. Institute for Social Science Research, University of California, Los Angeles, 133 pp.

White, M. P., and J. R. Eiser, 2006: Marginal trust in risk managers: Building and losing trust following decisions under uncertainty. *Risk Anal.*, **26**, 1187–1203, doi:10.1111/j.1539-6924.2006.00807.x.

Wilkinson, K. P., and P. J. Ross, 1970: Citizen response to warnings of Hurricane Camille. Social Science Research Center Rep., Mississippi State University, 60 pp.

Yin, R. K., 2009: *Case Study Research: Design and Methods*. 4th ed. Sage Publications, 240 pp.