5-26-2020

What matters the most? Understanding individual tornado preparedness using machine learning

Junghwa Choi
University of Nebraska at Omaha, junghwachoi@unomaha.edu

Scott Robinson

Romit Maulik

Wesley Wehde
East Tennessee State University

Recommended Citation

Junghwa Choi, Scott Robinson, Romit Maulik and Wesley Wehde (2020). What matters the most? Understanding individual tornado preparedness using machine learning. Natural Hazards 103, 1183-1200. https://doi.org/10.1007/s11069-020-04029-1

What matters the most? Understanding individual tornado preparedness using machine learning

https://doi.org/10.1007/s11069-020-04029-1

Junghwa Choi1, Scott Robinson2, Romit Maulik3, & Wesley Wehde4

1 The University of Oklahoma, 455 West Lindsey, 205 DAHT, Norman, OK 73019, USA

2 Henry Bellmon Chair of Public Service, The University of Oklahoma, 455 West Lindsey, 304E DAHT, Norman, OK 73019, USA

3 Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont, IL 60439, USA

4 East Tennessee State University, 301C Rogers-Stout Hall, P.O. Box 70651, Johnson City, TN 37614, USA

Abstract

Scholars from various disciplines have long attempted to identify the variables most closely associated with individual preparedness. Therefore, we now have much more knowledge regarding these factors and their association with individual preparedness behaviors. However, it has not been sufficiently discussed how decisive many of these factors are in encouraging preparedness. In this article, we seek to examine what factors, among the many examined in previous studies, are most central to engendering emergency prepared- ness in individuals particularly for tornadoes by utilizing a relatively uncommon machine learning technique in disaster management literature. Using unique survey data, we find that in the case of tornado preparedness the most decisive variables are related to personal experiences and economic circumstances rather than basic demographics. Our findings contribute to scholarly endeavors to understand and promote individual tornado prepared- ness behaviors by highlighting the variables most likely to shape tornado preparedness at an individual level.

Keywords:

Disaster management; Emergency preparedness; Machine learning; Random forest regression; Tornado preparedness

1 Introduction

While disasters cause tremendous damage to individual human life and property, the risks associated with them can be significantly reduced if the society and its members are appropriately prepared. Previous disasters have shown that the level of emergency preparedness, particularly at an individual level, can reduce negative outcomes of these events such as mortality and property damage (Diekman et al. 2007; Keim 2008; Paton 2008). Therefore, in addition to the development of emergency response plan and training first-responders for emergency preparedness, it is also important to promote emergency preparedness at individual level.

Given the importance of individual emergency preparedness, scholars from various disciplines have attempted to identify the variables most closely associated with individual emergency preparedness. While the associated factors vary depending on the specific type of hazards such as earthquakes, hurricanes and tornadoes (Perry and Lindell 1991; Murphy et al. 2009, 2018), numerous studies have found that basic demographic characteristics and disaster-specific variables such as disaster experience and risk perception may explain variations in individual emergency preparedness for various disastrous situations. Furthermore, there is evidence that political disposition variables may also influence emergency preparedness and other general risk-mitigating behaviors of individuals. This line of study has primarily been conducted using statistical modeling techniques such as multiple and multivariate linear regression.

Given these scholarly contributions, we now have much more knowledge regarding factors associated with individual preparedness for various hazards. However, the dominant statistical approach (relying on variations of multiple regression techniques) raises concerns about the confounding influence of multiple variables—especially when calculating the size of various effects. Because of this, it is not clear how important and decisive many of these factors are in encouraging preparedness. Is preparedness driven by relatively immutable demographic characteristics? Does disaster experience shape individual preparedness behaviors? We know that there are many statistically significant factors—but which are the most often decisive? Therefore, we seek to identify what factors, among the many tested in the previous studies, are most central to engendering emergency preparedness.

To that end, we look at levels of individual preparedness for tornadoes specifically. There are several reasons why we choose to look at tornado preparedness. First, tornadoes are one of the most salient hazards; tornadoes pose immediate harm to the public and their properties in more than 30% of states in the USA. Given the scale and scope of this issue, it is important to understand individual preparedness for tornadoes and seek to reveal the ways in promoting it. Furthermore, scholars have previously heavily focused on individual preparedness for seismic hazards (see review in Choi and Wehde 2020). Given that hazard type may affect individual behaviors during emergencies differently, it is useful to increase efforts for understanding individual preparedness beyond seismic hazards. By answering this question, we expect to have more accurate predictions regarding individual emergency preparedness and eventually provide better practical implications for practitioners.

To examine the relative decisiveness of factors, we use a relatively uncommon technique in this literature: random forest regression (RFR). RFR is a machine learning technique that assesses the effects of all explanatory variables simultaneously and allows us to understand what factors are more decisive than others to determine the value of dependent variables (Breiman 2001). This machine learning technique allows us to assess the decisiveness of variables without making strong parametric assumptions and avoiding the common problems related to high correlations between the explanatory variables (i.e., multicollinearity). This technique answers questions related to how often a variable is an influential part of classifying respondents as undertaking preparedness activities rather than estimating an effect on a conditional mean of preparedness activities (as regression approaches do).

In the following article, we first review previous studies of individual emergency preparedness to determine the variables we will analyze. Given that there are insufficient studies on emergency preparedness for tornadoes specifically (Choi and Wehde 2020), this paper reviews the emergency preparedness literature for various hazards including tornadoes as well as general risk mitigation behaviors. Next, we provide a brief introduction to RFR and how this technique helps us to assess the decisiveness of variables. We then utilize data from the 2013 Severe Weather and Society Survey and ft models using RFR to ascertain independent variable decisiveness. While previous studies have heavily focused on gender and race to understand individual preparedness for emergencies, we find that the most decisive variables are related to personal experience, economic circumstances and trust in government rather than basic demographics. Our analysis highlights the interventions to accentuate disaster experience and trust in government to promote individual preparedness for tornadoes. Finally, we end with a discussion of the methodological implications of our research for future studies in disaster management.

2 Literature review

Scholars in various disciplines have sought to understand variations in the level of individual emergency preparedness for various hazards and other risk-mitigation behaviors. Because individual decisions and behaviors are rather complex, many categories of potential explanatory factors have been examined.

First, scholars have found that basic individual demographic characteristics explain variations in individual emergency preparedness. While the effects of demographic characteristics may vary depending on locations and specific disasters, scholars have found the contingent effect of a plethora of demographic variables reviewed here. There is the greatest degree of consensus that as income and education increase, generally, so do preparedness levels (Fothergill and Peek 2004; Russell et al. 1995; Ablah et al. 2009; Edwards 1993). Most studies find that females are more likely to be prepared for disasters and emergencies than their male counterparts (Murphy et al. 2009; Robinson et al. 2019; Eisenman et al. 2006; Mulilis et al. 2000). Home-ownership is often associated with increased preparation for emergencies as well (Perry et al. 2001; Murphy et al. 2009). Furthermore, those who have more children at home are more likely to be prepared for disasters and emergencies (Edwards 1993; Russell et al. 1995; Baker and Cormier 2013; Olympia et al. 2010).

In some areas, the evidence on the influence of demographic characteristics is contradictory or mixed. Ablah et al. (2009) and Lindell and Whitney (2000) find that age is positively associated with emergency preparedness while Heller and coauthors (2005) find a negative relationship between age and preparedness for emergencies. Many studies have found that racial and ethnic minorities are more likely to be prepared for emergencies (Eisenman et al. 2006; Torabi and Seo 2004) while others find no differences between white respondents and minority counterparts (Murphy et al. 2009).

Additionally, there is evidence that disaster-specific variables are associated with individual emergency preparedness. The set of disaster-specific variables generally include individual risk perception, previous disaster experience, and the level of knowledge on a specific disaster of individuals, to name a few (Lindell and Whitney 2000). Individual risk perception of specific disasters is a significant determinant of individual emergency preparedness as well as other risk mitigating behaviors (Miceli et al. 2008; Funk et al. 2010; Lai et al. 2018; Murphy et al. 2009; Paton 2008; Palm et al. 1990; Pennings and Grossman 2008). While a majority of these studies have shown that those who perceive higher risks associated with disasters tend to prepare for these events more, other studies have also found that risk perception is not significantly associated with individual risk mitigating behaviors (Russell et al. 1995; Jackson 1981; Mileti and Darlington 1997; Lindell and Whitney 2000).[1] Previous studies have also found that individuals are more likely to take risk mitigating behaviors or prepare for emergencies when they are knowledgeable regarding specific disasters and risks associated with them (Bord et al. 1998; Jaeger et al. 1993; Leiserowitz 2006). Furthermore, some scholars have also found that people who have previously experienced disasters and emergencies are more likely to prepare for potential emergencies (Norris et al. 1999; Mulilis et al. 2003; Kapucu 2008).

Political dispositions are also often investigated to explain individual emergency preparedness and other risk mitigating behaviors—though this research is far more limited than the study of demographics and disaster-specific characteristics. Political dispositions are generally measured as political ideology, party identification or attitudes toward government. These concepts work to represent the central belief systems through which individuals process their decisions, such as being prepared for an emergency, that is related to policy (Taber and Lodge 2006; Rudolph and Evans 2005). It is reasonable to expect that individual decisions related to policy filtered through political lenses may eventually affect their actual behaviors. Scholars have found that political dispositions, mainly trust in federal and local government, may affect individual emergency preparedness although this effect is contingent on locations and specific type of disasters (Ablah et al. 2009; Perry and Lindell 1991; Murphy et al. 2009; Basolo et al. 2009; Arlikatti et al. 2007; van der Weerd et al. 2011; Murphy et al. 2018).

1 Some scholars argue that the mixed results may be the results of different measurement strategies of risk perception.

3 Data and methods

While previous studies show variables associated with individual preparedness for various hazards including tornadoes, scholars have not investigated which variables, among many, are the most or least decisive factors explaining individual emergency preparedness. Therefore, we seek to examine the relative decisiveness of these variables.

To answer this question, we draw on data from the 2013 Severe Weather and Society survey which was fielded in eight weekly waves between May 8th and June 27th with each wave consisting of approximately 500 randomly selected members of the same SurveySpot Internet panel.[2] Because we are interested in emergency preparedness in the context of tornadoes, we geographically restricted our sample so that only people living in a tornado-prone region of the USA commonly known as Tornado Alley were asked to participate. Members of the panel qualified as living in a tornado-prone region if the address they registered with SSI is located in one of the high-vulnerability regions listed by Ashley (2007) in his seminal study of tornado climatology. A total of 3976 participants living in tornado prone areas from Alabama, Arkansas, Georgia, Illinois, Indiana, Kansas, Kentucky, Louisiana, Michigan, Mississippi, Missouri, North Carolina, Ohio, Oklahoma, South Carolina, Tennessee, and Texas were recruited for this survey.[3] We also oversampled members of the sample that reside in rural settings so as to maximize geographic coverage and combat the urban bias typically associated with Internet access and participation in web-based surveys (Couper 2000). Our survey measures the perceptions, opinions and preferences of Americans regarding natural disaster issues, particularly tornadoes, perceived risks, trust in various levels of government and agencies, knowledge about tornadoes, political dispositions and other basic demographic characteristics.[4] By utilizing this data, we seek to investigate the decisiveness of variables which may influence individual emergency preparedness, with a specific focus on tornadoes.

The dependent variable in this study is the level of individual emergency preparedness for tornadoes based on specific protective actions. Each respondent was asked to select items they currently have available at their residence in case of emergencies. These items include (1) a disaster response plan for them and their family, (2) an emergency preparedness kit containing first-aid supplies, flashlights, batteries, etc., (3) supplies of water and food, (4) generators to provide electricity, (5) designated place to

2 At the time, the SurveySpot panel consisted of approximately two million households with about five million household members. In addition to this large panel, SSI maintains a subpanel of approximately 400,000 members whose demographics (e.g., race, gender, and education) are roughly proportionate to national census characteristics. Our sample was randomly drawn from the 400,000 census-balanced subpanel.

3 See Table 2 in “Appendix”

4 See Tables 3 and 4 in “Appendix” for survey questions

provide the most shelter from tornadoes within their house, and (6) specially constructed room or other facilities on your property designed to provide shelter from tornadoes and they are recommended by FEMA. This article creates a measure of the level of individual emergency preparedness using this survey question (0 = “not prepared at all’ to 6 = “fully prepared”). If a respondent who answered that he or she does not have any of those items, the respondent was coded as 0 (“not prepared at all”). In contrast, the respondent who answered that he or she has all of these items was coded as 6 (“fully prepared”).

This study includes seven important demographic variables based on the findings of previous studies (Murphy et al. 2009; Fothergill and Peek 2004; Russell et al. 1995; Ablah et al. 2009; Edwards 1993; Robinson et al. 2019; Eisenman et al. 2006; Mulilis et al. 2000). We include gender (0 = “female”, 1 = “male”), age (respondent’s reported age in years), education (dummy variables for a high school education or less, some college or a bachelors degree, and graduate education), race (dummy variable for White), and income (variable ranging from 1 to 4 where 1 represents “less than $50,000” and 4 indicates “$150,000 or more”). We also include home-ownership of individuals (0 = “ do not live in their own property”, 1 = “ live in their own property instead of renting a house”) and the number of children individuals live with (0 = “None” to 4 = “four or more”).

This study includes variables to measure disaster-specific factors such as perceived risks (Risk perception) and previous tornado experience (Disaster experience). These factors have been previously determined to explain the variations in individual preparedness for various hazards (Paton 2008; Murphy et al. 2009; Lindell and Whitney 2000; Funk et al. 2010; Lai et al. 2018; Palm et al. 1990; Pennings and Grossman 2008; Bord et al. 1998; Jaeger et al. 1993; Leiserowitz 2006; Norris et al. 1999; Mulilis et al. 2003; Kapucu 2008; Perry et al. 2001). To measure risk perception, each respondent was asked to rate how much risk they think tornadoes impose to them and their family (From 0 = “no risk” to 10 = “extreme risk”). In order to measure disaster experience, we included two ordinal variables: Damage experience and Experience of active tornado. To measure Damage experience, the respondents were asked to answer how many tornadoes they have personally seen while they were active (1= “None” to 4 = “More than fve”). Additionally, the respondent were asked if they or their members of family, neighbors, friends or associates ever experienced property damage, personal injury, or loss of life from a tornado (0 = “No” to 4= “Yes for them personally, for family, for neighbors, for close friends or associates”). By including these two ordinal variables to measure individual experience with tornadoes, we expect to have a more accurate understanding of the role of disaster experience. This article also includes individual knowledge regarding tornadoes (Disaster knowledge). The respondents were asked to answer six questions regarding common myths about tornadoes. These statements were either true or false; each question was recoded where 1 = correct and 0 = incorrect. Based on these recoded questions, we create a scale of individual knowledge

(0 = “not knowledgeable at all” to 6 = “fully knowledgeable”).[5]

This study, finally, includes several measures of political dispositions based of research from political science (Murphy et al. 2018; Choi and Wehde 2020). We consider individuals’ confidence in federal and local government as political disposition variables (Trust in fed, Trust in local). The respondents were asked to rate how much of the time they trust the federal and local government to do what is right for the American people (0 = “none of the time” to 10 = “all of the time”). Additionally, respondents were asked to report their political ideology (1 = “strongly liberal” to 7 = “strongly conservative”) and party identification (Republican, Independent, or Democrats). Based on this we created dummy variables for Republican. Summary statistics are in Table 1.

4 Random forest regression

The existing literature focuses on regression-based models that seek to identify the difference in expected levels of preparedness (the dependent variable) based on ensemble lists of suspected factors related to preparedness (all of the independent variables). This approach provides information about the statistical significance of proposed factors and, sometimes, an effect size related to a variable under specific parametric assumptions. The parametric assumptions become quite strong as one includes interrelated independent variables—as is almost universally the case. We sought an alternative model to, instead, assess the decisiveness of the variable rather than an expected difference. This search led us to a class of models relatively new to emergency management studies—machine learning models, specifically random forest regression.

Random forest regression (RFR) is an ensemble machine learning technique which may be utilized for efficient regression and classification tasks on large data sets (Hastie et al. 2005). It is built around the utilization of many decision trees (thus leading to the “forest” terminology). In simple terms, the technique imagines that one wants to most efficiently predict the level of some variable (for this study, preparedness) with as little data as possible. It proposes simple decision rules that help split the sample to explain the observed dependent variable.

A decision tree is an efficient algorithm for splitting a data set according to its various independent variables leading to a branch-like structure with each “node” corresponding to a splitting based on one variable. The decision to split according to one independent variable as against the others is taken by measuring the effectiveness of the split. This effectiveness is generally measured through different metrics such as impurity or variance reduction in the split data sets which is explained in the following. We may define the impurity for a regression task as

5 Question wordings for knowledge questions and correct answers are in the Appendix. For development of this scale, see Allan et al. (2017).

$$H\left(Q_m\right)=\frac{1}{N_m}\sum_{i=0}^{N_m}\left(y_i-\bar{y}_m\right)^2,$$

With $\bar{Y}_m$ is the mean of the dependent variable Y_i given by

$$\bar{y}_m=\frac{1}{N_m}\sum_{i=0}^{N_m} y_i,$$

and where N_m is the number of data points at a particular node *m*. The set Q_m represents the data that resides at a node. The total impurity at this node may then be expressed as

$$G(Q_m,\theta)=\frac{n_{\text{left}}}{N_m}H\left(Q_{\text{left}}(\theta)\right)+\frac{n_{\text{ right}}}{N_m}H\left(Q_{\text{ right }}(\theta)\right),$$

and where n_{left} and n_{right} correspond to the number of children data points in the left and right branches arising from the node *m* and

$$\begin{aligned}Q_{\text{left}}(\theta) &=(x,y)|x_j\leq t_m,\\ Q_{\text{right}}(\theta) &=(x,y)|x_j>t_m\end{aligned}$$

are the left and right split data sets, respectively. Note that the splitting of Q_m depends on a independent variable x_j and a threshold t_m at each node. The choice of t_m is generally given by the median value of the attribute x_j, i.e., the left data set is comprised of all samples with independent variable x_j less than t_m. The decision tree splits the data Qm in a manner that minimizes $G(Q_m,\Theta)$ by choosing an optimal dimension *j* for the independent variable. This branching is performed recursively for every node (corresponding to a different *j*) until certain user-defined criteria are met such as $N_m \leq N_{tol}$ at which point the node is denoted a leaf. A branching may also be terminated if a certain maximum depth is reached. A new data point (i.e., an unseen sample) can then follow the branching trajectory (by tracking the order of splits by *j* and *Eq. 3* and placement into left or right branches by t_m) and reach a leaf. In case N_{tol} criteria are set, the prediction of the decision tree for a sample is the average prediction value of the dependent variables at this leaf. We note that there exists multiple impurity metrics for splitting trees such as Gini impurity and entropy. From a high-level perspective, node impurity is a measure of the homogeneity of the data at a node and user defined homogeneity metrics may be used for the purpose of splitting the data at a node.

Table 1: Summary statistics

Variables	Mean	Standard Dev.	Min	Max
Dependent variable				
Emergency preparedness	1.71	1.45	0	6
Demographic characteristics				
Income	1.52	0.74	1	4
Race	0.78	0.41	0	1
Children	0.51	0.92	0	4
Education				
High school	0.97	0.17	0	1
College	0.37	0.48	0	1
Graduate	0.1	0.3	0	1
Age	45.51	16.51	18	99
Gender	0.47	0.5	0	1
Homeownership	0.53	0.5	0	1
Rural	0.28	0.45	0	1
Political dispositions				
Political ideology	4.17	1.59	1	7
Party identification	0.28	0.45	0	1
Trust in Fed Gov.	3.46	2.58	0	10
Trust in local Gov.	4.89	2.47	0	10
Issue-specific variables				
Risk perception	5.34	2.26	0	10
Disaster experience	0.73	0.45	0	1
Disaster knowledge	3.62	1.25	0	6

Figure 1 shows an example of one of several decision trees in the random forest regression which explains the level of individual emergency preparedness with this study's data. This decision tree splits a random subset of our total data set into several branches terminating at leaves (or when a maximum depth of 4 branches is reached). As shown here, the first variable used to split the data is the high school degree (*high school*, dummy variable). We remind the reader that the decision to split is made through the arguments of impurity reduction. The tree continues to branch according to the same rule. The MSE field in the tree indicates the value given in Eq. 1 and the value field indicates the mean of the dependent variable at each node.

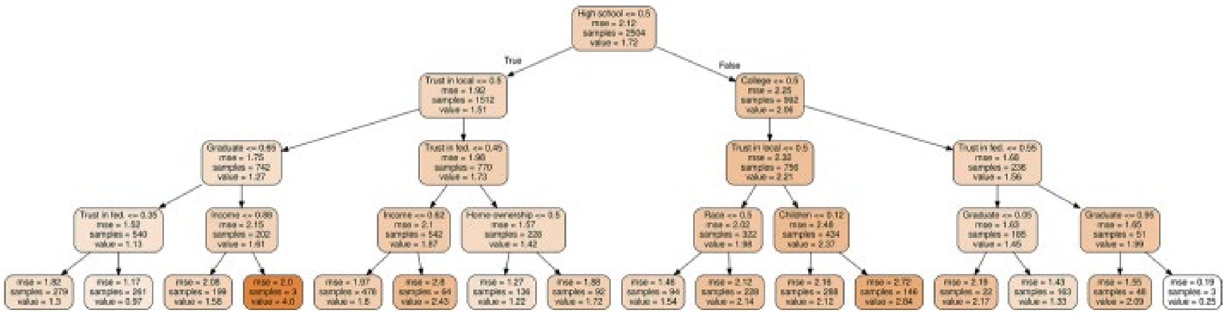

Fig. 1 An example of a decision tree

RFRs utilize ensembles of decision trees (one example of which is shown in Fig. 1) by selecting random subsets of the total data for each tree to obtain a consensus on regression predictions. This leads to greater accuracy than the utilization of one decision tree solely which is prone to the phenomenon of 'overfitting', implying a lack of generalization of the model. Each tree utilizes a random subset of the data through sampling with replacement and may therefore obtain a completely different branching structure depending on the distribution of that particular data set. The generation of multiple decision trees as estimators also encourages generalization. RFRs are well suited to the modeling of nonlinear interactions between independent variables for which they are considered more robust than linear regression methods. In addition to greater robustness in model building, RFRs also provide their users a metric of independent variable importance. Each decision tree therefore provides different estimates for impurity reduction for each variable for each tree. Average impurity reductions may then be obtained for the entire forest and used to rank the relative importance of variables. In other words, variables with the most average impurity reduction are classified as more important. The RFR is a good fit for our endeavor to assess the relative importance of features affecting individual preparedness in the event of a natural disaster. A greater discussion of the mathematics of the RFRs utilized here can be found in Breiman (2001).

We note, however, that unlike linear regression methods, RFRs do not lead to a parameterized model and do not provide correlation coefficients or the results one would need to generate effect sizes and related statistics. In that sense, the RFR results are decidedly 'black-box' in comparison with more conventional regression methods. For each tree, N samples are drawn randomly with replacement from the training set (where N is the total number of samples) and the tree is built on this new version of the training data. This introduces randomness in the training procedure since trees will each be trained on slightly different training sets. In expectation, drawing N samples with replacement from a data set of size N will select approximately $2N/3$ unique samples from the original set (Louppe 2014). This explains the 2513 'unique' samples at the head node in Fig. 1. Note that it is possible to implement individual trees with a random number of independent variables utilized for each tree. However, we chose to keep all decision variables available to all trees.

Due to the socio-economic nature of our dataset, we execute several validation strategies before concluding on the importance of factors. These are outlined below. A first crucial step to ensure that the feature importance aren't numerical artifact is through the use of a multi-step cross-validation. In this phase, 40 different random selections are made with each comprising 80% of the data which we denote the *training data.* Subsequently, for each selection, a hyperparameter optimization is performed to determine the optimal depth for branching. This optimization scans between a minimum branching depth of 4 to a maximum depth of 15 to determine when branching should be stopped to account for the lowest mean absolute validation error. Note that the validation error comes from the 20% of the total data set that is kept aside for each of the 40 assessments—commonly denoted *validation data* in machine learning parlance.

Our implementation of the RFR uses the well-known scikit-learn machine learning package for Python. We utilize 100 trees with each tree using all the independent variables and keep $N_{tol} = 1$. Feature rankings (obtained from the forest with the optimal depth) are then ranked according to their importance on this set. Finally, one is left with 40 sets of feature importances. We stack all the feature importances and display box plots which also show median, quartile, minimum, maximum and outlier values for the ranking a variable obtained as shown in Fig. 2. We can thus be more confident about our conclusions in this manner. One can clearly observe that the results indicate a clear trend.

For each of our 40 assessments, we record the validation mean absolute error and find the average over the entire experiment. We obtain a mean absolute error of 0.798. We remind the reader that the output variable ranges from 0 to 6 and is ordinal and within that context, this error is deemed acceptable for the purpose of discerning feature importance. We would like to clarify here that the work seeks not to use the RFR as a predictive model directly—although this can be an interesting follow up study. Some strategies to extend the current framework to a predictive model would require removal of unimportant features recursively and assessment of test error improvement, feeding the identified features into another machine learning framework such as a neural network, potentially. However, they are outside this manuscript's scope of study.

5 Results

In this paper, we calculated modal rankings of variable decisiveness with regard to individual preparedness for tornadoes. Figure 2 presents the results. We remind the readers that a higher modal value implies lower ranking and consequently a lower importance in our formulation. For example, first place is best while higher rankings are less decisive. Box-plots for each explanatory variable are presented to provide more detailed information about the variance of variable decisiveness. In addition, we also provide the percentages in relative decisiveness of each explanatory variable in Fig. 3. While random forest regressors do not provide quantitative interpretation with these percentages in relative decisiveness, it helps us identify the clusters of decisive factors among all explanatory variables.

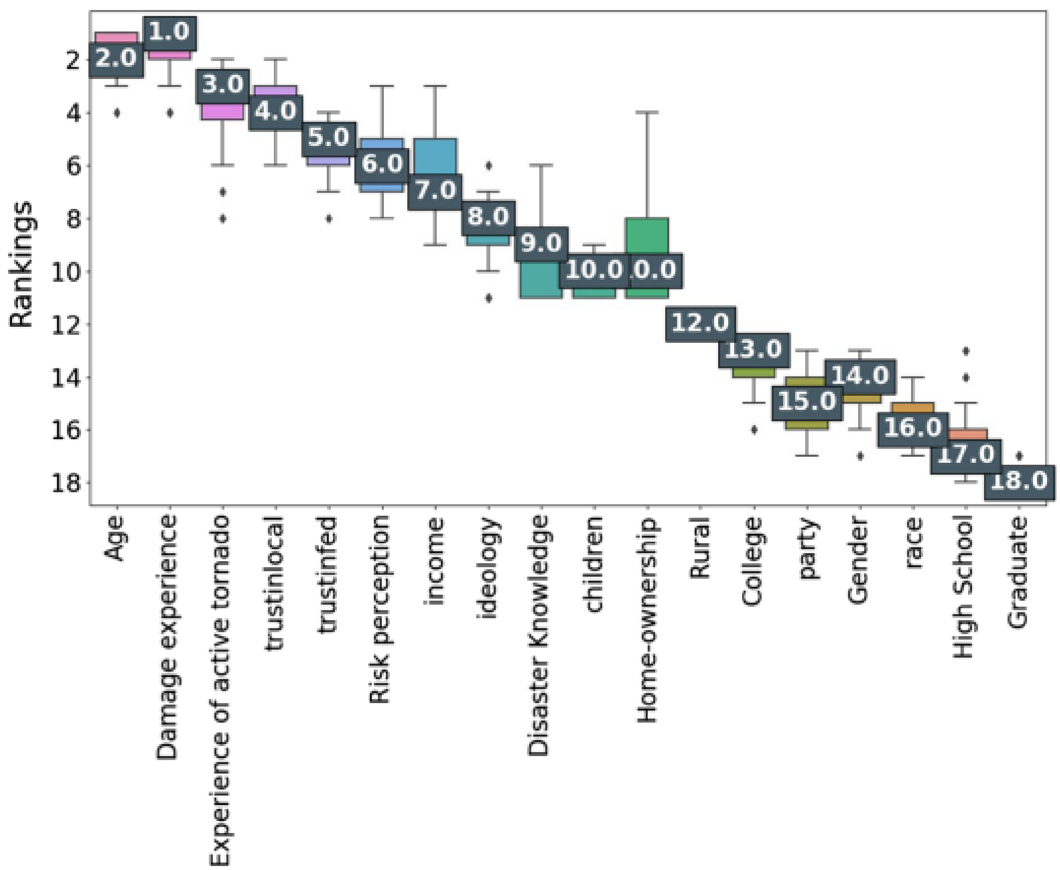

Fig. 2 Relative decisiveness of independent variables

When we consider both Figs. 2 and 3, age and damage experience seem to be the most decisive factors to structure individual preparedness for tornadoes. Figure 3 particularly shows that damage experience and age are the most commonly decisive factor by a notable margin (approximately 0.14 and 0.13 in relative decisiveness, respectively). In addition, experience of active tornado (how many times they have seen active tornadoes) also seems to be one of the most decisive factors to shape individual preparedness for tornadoes. This is not surprising, but it illustrates the importance of personal history, and the possibility of learning, in the decision to prepare for tornadoes. This is a variable that is common—though not universal—within the literature.

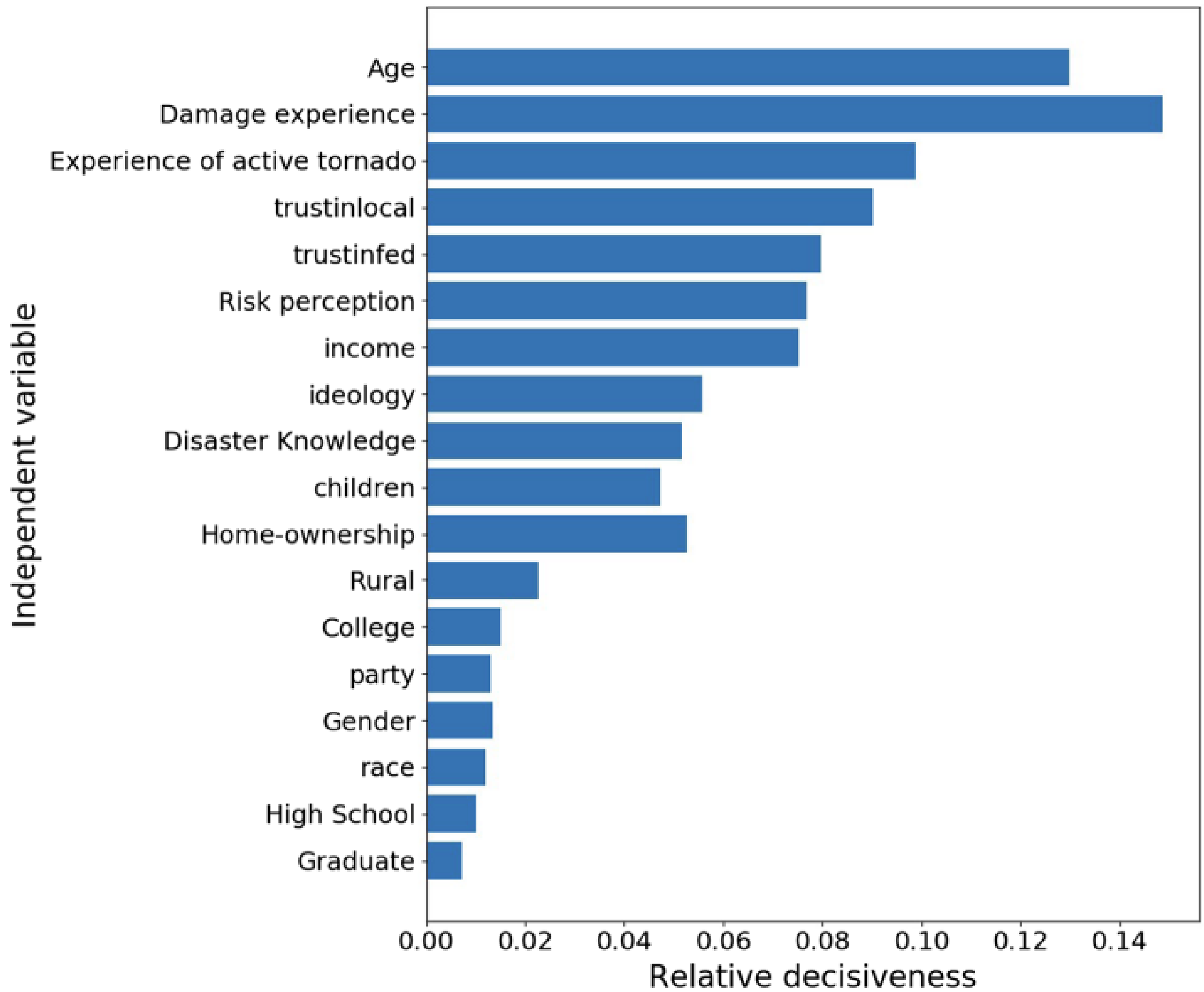

Fig. 3 Relative decisiveness of independent variables

Random forest regressor does not provide knowledge on how these relatively important variables affect individual preparedness for tornadoes, therefore, the direction of explanatory factors is beyond the scope of this paper. However, according to the previous studies and their findings (see Choi and Wehde (2020), Ablah et al. (2009)), it is expected that older people are more likely to prepare for tornadoes. Furthermore, people may be more likely to prepare for tornadoes when they have previous experience related to this hazard (Mulilis et al. 2003; Kapucu 2008).

Trust in local government as well as trust in federal government are also among the decisive factors in shaping individual preparedness for tornadoes. It is interesting that it is trust in local government that is decisive rather than trust in federal government, political ideology or party identification. This may suggest more attention to what it is about trust in local government that makes it more salient to preparedness decisions than more distant (but, often in other situations, more salient and familiar) political characteristics like party identification. Additionally, risk perception is considered as important as trust in government factors. The result also highlights economic motivation. Clearly income level defines a great deal of one’s economic experience and defines many options for preparedness. Higher income not only increases the options available, that is affordable, for respondents to use in preparation it is also associated with increased leisure time available to prepare for future events such as disasters.

The third set of variables, number of children, ideology, disaster knowledge and homeownership, seem to be similarly decisive at a low level. Finally, at the lowest end, our results suggest that a majority of basic demographic factors such as household location (rural), education, race and gender as well as party identification may be associated with differences in preparedness but is not a decisive factor in predicting levels of preparedness in our analysis.

6 Discussion

This article utilized a machine learning technique to examine the relative decisiveness of various factors tested in previous studies. While the analyses provide unique results, it should be noted that our conclusions are somewhat limited to tornado specific situations. Scholars have argued that the dynamics of preparedness for different types of disasters may either affected differently by or be affected by different individual attitudes and categories of variables (DeYoung and Peters 2016; Choi and Wehde 2020; Murphy et al. 2018). Therefore, we suggest future research to investigate and compare the relative decisiveness of factors for different types of natural and human-induced hazards to better understand individual emergency preparedness in the future.

The results show that the most decisive variables in preparedness for tornadoes are age, personal experiences, trust in government, risk perceptions, and economic motivations. The RFR results related to the decisiveness of the variables reveal that some of the most commonly discussed components of individual level models are not the most decisive in the decision trees. For instance, race has been emphasized in many previous studies. Our analysis shows that, while related, it is not a decisive factor in predicting levels of preparedness for tornadoes.[6] Rather, the most decisive variables are related to personal experiences, economic circumstances, and trust in government. These results should direct our effort to better understand the role that disaster experiences, economic situation and public trust in government have in determining household disaster preparedness. These results suggest emergency managers and policymakers concerned with disaster preparedness, at least for tornadoes, focus on interventions that will also increase trust. For example, clear and transparent communication strategies and deliberative processes may serve to increase preparedness both directly and indirectly through improvements in trust.

6 It is worth noting that the measurement of race in this study is simplistic. Future work on decisiveness should use more precise measures of race before the concept is relegated to secondary importance in preparedness research.

While RFR allows us to answer somewhat different and under-investigated questions, these results do not replace the continuing important work with traditional regression approaches. We acknowledge that the RFR method has some inherent limitations due to its black-box nature and have attempted to discuss these throughout to serve as warnings. Furthermore, it should be noted that RFR and its quantitative interpretation is somewhat limited. For instance, according to our analyses, we can confidently say that disaster experience is more important than gender since there is a large difference between their relative decisiveness. However, it does not tell us as much about individual characteristics that are much more closely ranked such as age and experience with tornadoes. Furthermore, it does not provide estimates of effect size such as the increase in the expected value of preparedness given a unit increase in an independent variable. Consequently, these results cannot provide results in the form of "having experienced a tornado increases preparedness levels by X%" . Those sorts of statements are possible from a parametric regression approach but not from our RFR approach. There is clearly still room for traditional parametric models to help generate these sorts of statements—statements often essential for benefit/cost analysis, for example. However, the RFR approach provides a useful check and corrective for a literature that relies on the parametric approach—and the misinterpretations may fall victim to based on strong parametric assumptions. For example, regression-based models can provide misleading results (in terms of both statistical significance and effective sizes) when models include closely related variables—even if correlations are low due to differences in level of measurement. It is easy in parametric, regression based approaches to misrepresent the influence of trust in local government when included simultaneously with party identification. The decisiveness approach of the RFR lets us see which variables are most likely to help us categorize respondents into levels of preparedness. In this way, it helps us see which variables are most important in the decision trees implied by our survey data. This can help us focus our research attention to the variables that are most able to shape individual preparedness behaviors.

Acknowledgements

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research, under Contract DE-AC0206CH11357. This research was funded in part and used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. DOE or the United States Government. Additionally, the data from this project were collected with support from the University of Oklahoma's Office of the Vice President of Research.

Appendix

See Tables 2, 3 and 4.

Table 2: Proportion of respondents by state

State	Observation (%, proportion)	Population proportion (%)
Alabama	233 (5.9%)	4
Arkansas	184 (4.6%)	2
Georgia	348 (8.8%)	8
Illinois	262 (6.6%)	9
Indiana	309 (7.8%)	5
Kansas	176 (4.4%)	2
Kentucky	241 (6.1%)	3
Louisiana	198 (5.0%)	3
Michigan	228 (5.7%)	7
Mississippi	174 (4.4%)	2
Missouri	225 (5.7%)	5
North Carolina	43 (1.1%)	8
Ohio	369 (9.3%)	9
Oklahoma	198 (5.5%)	3
South Carolina	41 (1.0%)	4
Tennessee	347 (8.7%)	5
Texas	400 (10.1%)	21
Total	3976 (100%)	100

Table 3: Survey Questions

Variables	Survey questions and response options
Individual preparedness	w74_prep: Which of the following do you currently have available at your residence? Please select all that apply. 1—A disaster response plan for you and your family 2—An emergency preparedness kit containing such things as first-aid supplies, flashlights, batteries, etc 3—Supplies of water and food specially packaged or designated for use in emergencies 4—A generator (either mobile or fixed) to provide electricity in emergencies 5—A designated place within your home that you consider to provide the most shelter from tornadoes 6—A specially constructed room or other facility on your property designed to provide shelter from tornadoes (may be above or below ground) 7—None of the above
Gender	w15_gend: Are you male or female? 0—Female 1—Male
Age	w13_age: How old are you? [verbatim]
Education	w14_edu: What is the highest level of education you have completed? 1—Elementary or some high school 2—High school graduate/GED 3—Some college/vocational school 4—College graduate 5—Some graduate work 6—Master's degree 7—Doctorate (of any type) 8—Other degree [verbatim]
Race	w124_race: Which of the following best describes your race? 1—White 2—Black, African American, or Negro 3—American Indian or Alaska Native 4—Asian 5—Native Hawaiian or Pacific Islander 6—Two or more races 7—Some other race (please specify) [verbatim]
Income	w125_inc: Was the estimated annual income for your household in 2012: 1—Less than 50,000 [go to w125a_inc50] 2—At least 50,000 but less than 100,000 [go to w125b_inc100] 3—At least 100,000 but less than 150,000 [go to w125c_inc150] 4—150,000 or more [go to w125d_inc200]
Home-ownership	w7_rent: Which of the following categories best describes your living arrangements at that residence? 1—Live with family or friends and do not pay rent 2—Pay to rent or lease your primary residence (includes college or other dormitory rooms) 3—Own your primary residence (includes making mortgage payments or outright ownership with no mortgage payments)
Children	w75_child: How many children that attend public schools (grades K through 12) reside with you? 0—None [skip to RESPONSE random table] 1—One [go to w76_pickup] 2—Two [go to w76_pickup] 3—Three [go to w76_pickup] 4—Four or more [go to w76_pickup]

Table 4: Survey questions (continued)

Variables	Survey questions and response options
Risk perception	w31_risk: Considering both the likelihood and potential consequences of tornadoes throughout the year, and using a scale from zero to ten, where zero means no risk and ten means extreme risk, how do you rate the overall risk to you and your family from tornadoes? 0—No risk ————10—Extreme risk
Damage experience	w73_damg: Have you or members of your family, neighbors, friends, or associates ever experienced property damage, personal injury, or loss of life from a tornado? Please select all that apply. 1—No 2—Yes, for you personally 3—Yes, for family 4—Yes, for neighbors
Experience of active	w35_funnel: Not including movies, television, videos, web shots, or other pictures how many actual tornadoes have you, personally, ever seen while they were active? 1—None 2—One 3—Two to five 4—More than five
Disaster knowledge	w48_tall: Tall buildings protect large cities from tornadoes. 0—False 1—True w49_mtns: Mountains, rivers, and lakes do not protect nearby areas from tornadoes. 0—False 1—True w50_SW: When sheltering from tornadoes in a basement, the southwest corner is safest. 0—False 1—True w51_brdg: When driving, you should not take shelter from tornadoes under a bridge or overpass. 0—False 1—True w52_open: When sheltering in a house from a tornado, you should open all the windows to equalize the pressure inside and outside to prevent the house from exploding. 0—False 1—True w53_size: The shape and size of a tornado determine how destructive it is. 0—False 1—True
Individuals' confidence in federal government	w117_doright: On a scale from zero to ten, where zero means none of the time and ten means all of the time,how much of the time do you trust the federal government in Washington to do what is right for the American people? 0—None of the time ————10—All of the time

Table 4 (continued)

Variables	Survey questions and response options
Inidividuals' confidence in local government	w119_locgov: Again, on the same scale, how much of the time do you trust your county and local governments to do what is right for you and your fellow citizens in your local area? 0—None of the time. ——————10—All of the time
Political ideology	w122_ideol: On a scale of political ideology, individuals can be arranged from strongly liberal to strongly conservative. Which of the following categories best describes your views? 1—Strongly liberal 2—Liberal 3—Slightly liberal 4—Middle of the road 5—Slightly conservative 6—Conservative 7—Strongly conservative
Party identification	w120_party: With which political party do you most identify? 1—Democratic party [go to w121_iden] 2—Republican party (or GOP) [go to w121_iden] 3—Independent [skip to w122_ideol] 4—Other party (please specify) [verbatim],[skip to w122_ideol]

References

Ablah E, Konda K, Kelley CL (2009) Factors predicting individual emergency preparedness: a multi-state analysis of 2006 brfss data. Biosecur Bioterrorism Biodefense Strategy Pract Sci 7(3):317–330

Allan JN, Ripberger JT, Ybarra VT, Cokely ET (2017) Tornado risk literacy: beliefs, biases, and vulnerability. Nat Decis Mak Uncertain 61:284–290

Arlikatti S, Lindell MK, Prater CS et al (2007) Perceived stakeholder role relationships and adoption of seismic hazard adjustments. Int J Mass Emerg Disasters 25(3):218

Ashley WS (2007) Spatial and temporal analysis of tornado fatalities in the United States: 1880–2005. Weather Forecast 22(6):1214–1228

Baker LR, Cormier LA (2013) Disaster preparedness and families of children with special needs: a geographic comparison. J Commun Health 38(1):106–112

Basolo V, Steinberg LJ, Burby RJ, Joyce L, Maria CA, Chihyen H (2009) The effects of confidence in government and information on perceived and actual preparedness for disasters. Environ Behav 41(3):338–364

Bord RJ, Ann F, O’Connor RE (1998) Public perceptions of global warming: United states and international perspectives. Clim Res 11(1):75–84

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Choi J, Wehde W (2020) Trust in emergency management authorities and individual emergency preparedness for tornadoes. Risk Hazards Crisis Public Policy 11(1):12–34

Couper MP (2000) Web surveys: a review of issues and approaches. Public Opin Q 64(4):464–494

DeYoung SE, Peters M (2016) My community, my preparedness: the role of sense of place, community, and confdence in government in disaster readiness. Int J Mass Emerg Disasters 34(2):250–282

Diekman ST, Kearney SP, O’Neil ME, Mack KA (2007) Qualitative study of homeowners’ emergency preparedness: experiences, perceptions, and practices. Prehospital Disaster Med 22(6):494–501

Edwards ML (1993) Social location and self-protective behavior: implications for earthquake preparedness.

Int J Mass Emerg Disasters 11(3):293–303

Eisenman DP, Cheryl W, Jonathan F, Anna L, Claude S, Scot H, Lillian G (2006) Differences in individual level terrorism preparedness in Los Angeles county. Am J Prev Med 30(1):1–6

Fothergill A, Peek LA (2004) Poverty and disasters in the united states: a review of recent sociological findings. Nat Hazards 32(1):89–110

Funk S, Marcel S, Jansen VAA (2010) Modelling the influence of human behaviour on the spread of infectious diseases: a review. J R Soc Interface 7(50):1247–1256

Hastie T, Tibshirani R, Friedman J, Franklin J (2005) The elements of statistical learning: data mining, inference and prediction. Math Intell 27(2):83–85

Heller K, Alexander DB, Gatz M, Knight BG, Rose T (2005) Social and personal factors as predictors of earthquake preparation: the role of support provision, network discussion, negative affect, age, and education. J Appl Soc Psychol 35(2):399–422

Jackson EL (1981) Response to earthquake hazard: the west coast of North America. Environ Behav 13(4):387–416

Jaeger C, Dürrenberger G, Kastenholz H, Trufer B (1993) Determinants of environmental action with regard to climatic change. Clim Change 23(3):193–211

Kapucu N (2008) Culture of preparedness: household disaster preparedness. Disaster Prev Manag Int J 17(4):526–535

Keim ME (2008) Building human resilience: the role of public health preparedness and response as an adaptation to climate change. Am J Prev Med 35(5):508–516

Lai C-H, Chib A, Ling R (2018) Digital disparities and vulnerability: mobile phone use, information behaviour, and disaster preparedness in southeast Asia. Disasters 42:734–760

Leiserowitz A (2006) Climate change risk perception and policy preferences: the role of affect, imagery, and values. Clim Change 77(1–2):45–72

Lindell MK, Whitney DJ (2000) Correlates of household seismic hazard adjustment adoption. Risk Anal 20(1):13–26

Louppe G (2014) Understanding random forests: from theory to practice. arXiv preprint arXiv:1407.7502

Miceli R, Sotgiu I, Settanni M (2008) Disaster preparedness and perception of food risk: a study in an alpine valley in Italy. J Environ Psychol 28(2):164–173

Mileti DS, DeRouen DJ (1997) The role of searching in shaping reactions to earthquake risk information. Soc Probl 44(1):89–103

Mulilis J-P, Duval ST, Bovalino K (2000) Tornado preparedness of students, nonstudent renters, and nonstudent owners: issues of pre theory 1. J Appl Soc Psychol 30(6):1310–1329

Mulilis J-P, Shelley DT, Rogers RY (2003) The effect of a swarm of local tornados on tornado preparedness: a quasi-comparable cohort investigation 1. J Appl Soc Psychol 33(8):1716–1725

Murphy ST, Michael C, Frank LB, Deborah G, Alfonso A (2009) Predictors of emergency preparedness and compliance. Disaster Med Public Health Prep 7:S1–S8

Murphy H, Greer A, Wu H-C (2018) Trusting government to mitigate a new hazard: the case of Oklahoma earthquakes. Risk Hazards Crisis Public Policy 9:357–380

Norris FH, Tenbroeck S, Krzysztof K (1999) Revisiting the experience-behavior hypothesis: the effects of hurricane Hugo on hazard preparedness and other self-protective acts. Basic Appl Soc Psychol 21(1):37–47

Olympia RP, Ruby R, Steven H, Uchechi A, Madeline G (2010) Natural disasters and mass-casualty events affecting children and families: a description of emergency preparedness and the role of the primary care physician. Clin Pediatr 49(7):686–698

Palm RI, Hodgson ME, Blanchard RD, Lyons DI (1990) Earthquake insurance in California: environmental policy and individual decision-making. Westview Press, Boulder

Paton D (2008) Risk communication and natural hazard mitigation: how trust influences its effectiveness. Int J Glob Environ Issues 8(1–2):2–16

Pennings JME, Grossman DB (2008) Responding to crises and disasters: the role of risk attitudes and risk perceptions. Disasters 32(3):434–448

Perry RW, Lindell MK (1991) The effects of ethnicity on evacuation decision-making. Int J Mass Emerg Disasters 9(1):47–68

Perry RW, Lindell MK, Tierney KJ (2001) Facing the unexpected: disaster preparedness and response in the United States. Joseph Henry Press, Washington, DC

Robinson SE, Pudlo JM, Wehde W (2019) The new ecology of tornado warning information: a natural experiment assessing threat intensity and citizen-to-citizen information sharing. Public Adm Rev 79:905–916

Rudolph TJ, Evans J (2005) Political trust, ideology, and public support for government spending. Am J Polit Sci 49(3):660–671

Russell LA, Goltz JD, Bourque LB (1995) Preparedness and hazard mitigation actions before and after two earthquakes. Environ Behav 27(6):744–770

Taber CS, Lodge M (2006) Motivated skepticism in the evaluation of political beliefs. Am J Polit Sci 50(3):755–769

Torabi MR, Seo D-C (2004) National study of behavioral and life changes since September 11. Health Educ Behav 31(2):179–192

van der Weerd W, Timmermans DRM, Beaujean DJMA, Oudhof J, van Steenbergen JE (2011) Monitoring the level of government trust, risk perception and intention of the general public to adopt protective measures during the infuenza a (H1N1) pandemic in the Netherlands. BMC Public Health 11(1):575

Publisher's Note